

THE WORLD BANK ECONOMIC REVIEW

Volume 28 • 2014 • Number 2

“Economic Development as Opportunity Equalization”

John E. Roemer

The Measurement of Educational Inequality: Achievement
and Opportunity

Francisco H. G. Ferreira and Jérémie Gignoux

Economic Growth and Equality of Opportunity

Vito Peragine, Flaviana Palmisano, and Paolo Brunori

Children’s Health Opportunities and Project Evaluation: Mexico’s
Oportunidades Program

Dirk Van de gaer, Joost Vandebossche, and José Luis Figueroa

SYMPOSIUM ON CONFLICT AND GENDER

Armed Conflict, Gender, and Schooling

Mayra Buvinić, Monica Das Gupta, and Olga N. Shemyakina

Short- and Long-Term Impact of Violence on Education:
The Case of Timor Leste

Patricia Justino, Marinella Leone, and Paola Salardi

Education and Civil Conflict in Nepal

Christine Valente

Schooling, Violent Conflict, and Gender in Burundi

Philip Verwimp and Jan Van Bavel

www.wber.oxfordjournals.org

THE WORLD BANK ECONOMIC REVIEW

EDITOR

Andrew Foster, *Brown University*

CO-EDITORS

Peter Lanjouw and Luis Servén, *World Bank*

The editorial team would like to thank former editors Elisabeth Sadoulet and Alain de Janvry for overseeing the review process of the articles in this issue.

ASSISTANT TO THE EDITOR

Marja Kuiper

EDITORIAL BOARD

Harold H. Alderman, *World Bank (retired)*
Chong-En Bai, *Tsinghua University, China*
Pranab K. Bardhan, *University of California,
Berkeley, USA*

Kaushik Basu, *World Bank*
Thorsten Beck, *Cass Business School,
City University London, UK*

Johannes van Biesebroeck, *K.U. Leuven,
Belgium*

Maureen Cropper, *University of Maryland,
USA*

Asli Demirgüç-Kunt, *World Bank*

Jean-Jacques Dethier, *World Bank*

Quy-Toan Do, *World Bank*

Frédéric Docquier, *Catholic University of
Louvain, Belgium*

Eliana La Ferrara, *Università Bocconi, Italy*

Francisco H. G. Ferreira, *World Bank*

Augustin Kwasi Fosu, *United Nations
University, WIDER, Finland*

Caroline Freund, *World Bank*

Paul Glewwe, *University of Minnesota, USA*

Philip E. Keefer, *World Bank*

Norman V. Loayza, *World Bank*

Jeremy Magruder, *University of California,
Berkeley, USA*

William F. Maloney, *World Bank*

David J. McKenzie, *World Bank*

Jaime de Melo, *University of Geneva,
Switzerland*

Ugo Panizza, *UNCTAD*

Nina Pavcnik, *Dartmouth College, USA*

Vijayendra Rao, *World Bank*

Martin Ravallion, *Georgetown University,
USA*

Jaime Saavedra-Chanduvi, *World Bank*

Claudia Sepúlveda, *World Bank*

Dominique Van De Walle, *World Bank*

Christopher M. Woodruff, *University of
California, San Diego, USA*

The *World Bank Economic Review* is a professional journal used for the dissemination of research in development economics broadly relevant to the development profession and to the World Bank in pursuing its development mandate. It is directed to an international readership among economists and social scientists in government, business, international agencies, universities, and development research institutions. The Review seeks to provide the most current and best research in the field of quantitative development policy analysis, emphasizing policy relevance and operational aspects of economics, rather than primarily theoretical and methodological issues. Consistency with World Bank policy plays no role in the selection of articles.

The Review is managed by one or two independent editors selected for their academic excellence in the field of development economics and policy. The editors are assisted by an editorial board composed in equal parts of scholars internal and external to the World Bank. World Bank staff and outside researchers are equally invited to submit their research papers to the Review.

For more information, please visit the Web sites of the Economic Review at Oxford University Press at www.wber.oxfordjournals.org and at the World Bank at www.worldbank.org/research/journals.

Instructions for authors wishing to submit articles are available online at www.wber.oxfordjournals.org. Please direct all editorial correspondence to the Editor at wber@worldbank.org.

SUBSCRIPTIONS: A subscription to *The World Bank Economic Review* (ISSN 0258-6770) comprises 3 issues. Prices include postage; for subscribers outside the Americas, issues are sent air freight.

Annual Subscription Rate (Volume 28, 3 Issues, 2014): *Institutions*—Print edition and site-wide online access: £197/\$297/€297, Print edition only: £181/\$272/€272, Site-wide online access only: £159/\$239/€238; *Corporate*—Print edition and site-wide online access: £296/\$444/€444, Print edition only: £272/\$407/€407, Site-wide online access only: £237/\$356/€357; *Personal*—Print edition and individual online access: £47/\$71/€71. US\$ rate applies to US & Canada, Euros€ applies to Europe, UK£ applies to UK and Rest of World. There may be other subscription rates available; for a complete listing, please visit www.wber.oxfordjournals.org/subscriptions. *Readers with mailing addresses in non-OECD countries and in socialist economies in transition are eligible to receive complimentary subscriptions on request by writing to the UK address below.*

Full prepayment in the correct currency is required for all orders. Orders are regarded as firm, and payments are not refundable. Subscriptions are accepted and entered on a complete volume basis. Claims cannot be considered more than four months after publication or date of order, whichever is later. All subscriptions in Canada are subject to GST. Subscriptions in the EU may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Personal rates are applicable only when a subscription is for individual use and are not available if delivery is made to a corporate address.

BACK ISSUES: The current year and two previous years' issues are available from Oxford University Press. Previous volumes can be obtained from the Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. E-mail: psc@periodicals.com. Tel: (518) 537-4700. Fax: (518) 537-5899.

CONTACT INFORMATION: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnls.cust.serv@oup.com. Tel: +44 (0)1865 353907. Fax: +44 (0)1865 353485. *In the Americas, please contact:* Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. E-mail: jnlorders@oup.com. Tel: (800) 852-7323 (toll-free in USA/Canada) or (919) 677-0977. Fax: (919) 677-1714. *In Japan, please contact:* Journals Customer Service Department, Oxford University Press, Tokyo, 4-5-10-8F Shiba, Minato-ku, Tokyo, 108-8386, Japan. E-mail: custserv.jp@oup.com. Tel: +81 3 5444 5858. Fax: +81 3 3454 2929.

POSTAL INFORMATION: *The World Bank Economic Review* (ISSN 0258-6770) is published three times a year, in February, June, and October, by Oxford University Press for the International Bank for Reconstruction and Development/THE WORLD BANK. Send address changes to *The World Bank Economic Review*, Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009. Periodicals postage paid at Cary, NC and at additional mailing offices. Communications regarding original articles and editorial management should be addressed to The Editor, *The World Bank Economic Review*, The World Bank, 3, Chemin Louis Dunant, CP66 1211 Geneva 20, Switzerland.

ENVIRONMENTAL AND ETHICAL POLICIES: Oxford Journals, a division of Oxford University Press, is committed to working with the global community to bring the highest quality research to the widest possible audience. Oxford Journals will protect the environment by implementing environmentally friendly policies and practices wherever possible. Please see <http://www.oxfordjournals.org/ethicalpolicies.html> for further information on environmental and ethical policies.

DIGITAL OBJECT IDENTIFIERS: For information on dois and to resolve them, please visit www.doi.org.

PERMISSIONS: For information on how to request permissions to reproduce articles or information from this journal, please visit www.oxfordjournals.org/jnls/permissions.

ADVERTISING: Advertising, inserts, and artwork enquiries should be addressed to Advertising and Special Sales, Oxford Journals, Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK. Tel: +44 (0)1865 354767; Fax: +44(0)1865 353774; E-mail: jnlsadvertising@oup.com.

DISCLAIMER: Statements of fact and opinion in the articles in *The World Bank Economic Review* are those of the respective authors and contributors and not of the International Bank for Reconstruction and Development/THE WORLD BANK or Oxford University Press. Neither Oxford University Press nor the International Bank for Reconstruction and Development/THE WORLD BANK make any representation, express or implied, in respect of the accuracy of the material in this journal and cannot accept any legal responsibility or liability for any errors or omissions that may be made. The reader should make her or his own evaluation as to the appropriateness or otherwise of any experimental technique described.

PAPER USED: *The World Bank Economic Review* is printed on acid-free paper that meets the minimum requirements of ANSI Standard Z39.48-1984 (Permanence of Paper).

INDEXING AND ABSTRACTING: *The World Bank Economic Review* is indexed and/or abstracted by *CAB Abstracts*, *Current Contents/Social and Behavioral Sciences*, *Journal of Economic Literature/EconLit*, *PAIS International*, *RePEc (Research in Economic Papers)*, and *Social Services Citation Index*.

COPYRIGHT © 2014 The International Bank for Reconstruction and Development/THE WORLD BANK

All rights reserved; no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission of the publisher or a license permitting restricted copying issued in the UK by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1P 9HE, or in the USA by the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Typeset by Techset Composition India Private Limited, Chennai, India; Printed by The Sheridan Press.

THE WORLD BANK ECONOMIC REVIEW

Volume 28 · 2014 · Number 2

“Economic Development as Opportunity Equalization” <i>John E. Roemer</i>	189
The Measurement of Educational Inequality: Achievement and Opportunity <i>Francisco H. G. Ferreira and Jérémie Gignoux</i>	210
Economic Growth and Equality of Opportunity <i>Vito Peragine, Flaviana Palmisano, and Paolo Brunori</i>	247
Children’s Health Opportunities and Project Evaluation: Mexico’s Oportunidades Program <i>Dirk Van de gaer, Joost Vandenbossche, and José Luis Figueroa</i>	282

SYMPOSIUM ON CONFLICT AND GENDER

Armed Conflict, Gender, and Schooling <i>Mayra Buvinić, Monica Das Gupta, and Olga N. Shemyakina</i>	311
Short- and Long-Term Impact of Violence on Education: The Case of Timor Leste <i>Patricia Justino, Marinella Leone, and Paola Salardi</i>	320
Education and Civil Conflict in Nepal <i>Christine Valente</i>	354
Schooling, Violent Conflict, and Gender in Burundi <i>Philip Verwimp and Jan Van Bavel</i>	384

“Economic Development as Opportunity Equalization”

*John E. Roemer**

Economic development should be conceived of as the degree to which an economy has implemented an efficient and just distribution of economic resources. The ubiquitous measure of GDP per capita reflects a utilitarian conception of justice, where individual utility is defined as personal income, and social welfare is the average of utilities in a population. A more attractive conception of justice is opportunity-equalization. Here, a two-dimensional measure of economic development is proposed, based upon viewing individuals' incomes as a consequence of circumstances, effort, and policy. The first dimension is the average income level of those in the society with the most disadvantaged circumstances, and the second dimension is the degree to which total income inequality is due to differential effort, as opposed to differential circumstances. This pair of numbers is computed for a set of 22 European countries. No country dominates all others on both dimensions. The two-dimensional measure induces a partial ordering of countries with respect to development. JEL codes: O1, D3, D63

INTRODUCTION¹

Suppose we are concerned with the inequalities that exist in a society with respect to the distribution of some desirable good or advantage – wealth, life expectancy, literacy, or wage-earning capacity. The causes of inequality in that distribution can be partitioned into two categories: those for which individuals should not be held responsible, and those for which they should be. We need not here be concerned with the problem of free will, and the possibility that people are not responsible for anything if they lack free will, because every society has a conception of responsibility, and we may take that as the politically salient conception. Thus, in many societies, it is thought wrong that an individual's income

*John Roemer is a professor of political science and economics at Yale University; his e-mail address is john.roemer@yale.edu I am grateful to referees of this journal for suggestions leading to important revisions, and to the many scholars with whom I have discussed these ideas. This article is based upon a lecture given at the ABCDE meeting in Paris, 2011. A supplemental appendix to this article is available at <http://wber.oxfordjournals.org/>.

1. This section reviews previous work of the author on the conceptualization of equality of opportunity (see Roemer [1993, 1998, 2002]).

THE WORLD BANK ECONOMIC REVIEW, VOL. 28, NO. 2, pp. 189–209
Advance Access Publication August 8, 2013

doi:10.1093/wber/lht023

© The Author 2013. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

be strongly correlated with her parent's education or social position, for, assuming that that correlation reflects causality, these family characteristics seem to be ones from which children should not differentially benefit or suffer. On the other hand, most societies believe that adults should be held responsible for various choices that they make, assuming that they possess adequate information about the alternatives. Let us call the social and biological aspects of a person's environment for which society believes he should not be responsible his *circumstances*, those choices and actions for which he should be held responsible, his *effort*, and the desirable good whose distribution we are concerned with the *objective*.

When we have a data set that permits us to measure the inequality in the distribution of the objective, and its correlation with circumstances and effort, it is usually necessary (because data sets are finite) to choose a fairly small number of circumstances, each of which can take on a fairly small number of values. Thus, one circumstance might be parental education, which one could partition into three values; another might be race, partitioned into three categories, and so on. Call a vector of circumstances a *type*. Thus, one may partition the population of the data set into a finite number of types, where a type is the set of individuals with (approximately) the same vector of circumstances. Denote the types by $t = 1, \dots, T$. Denote the level of the objective with which we are concerned (income, wage-earning capacity, or life expectancy) by u , which is a function of circumstances, policy, and effort. Thus, $u^t(e, \varphi)$ is the average level of the objective for individuals of type t whose effort choices are summarized by the vector e if the policy is φ . Denote the policy space by Φ . In this formulation, any characteristic of the individual is either a component of circumstance, or of effort.

Effort here is measured so that increasing effort produces an increasing value of the objective. In this way, effort's role in the functions u^t differs from its relationship to utility in economic theory. For example, if the objective is health status, then *refraining* from smoking constitutes positive effort, although that abstinence may lower 'utility' in the usual sense, where the utility function is a representation of subjective preferences.

If the population faces a policy $\varphi \in \Phi$, there will ensue a distribution of effort in each type; denote the distribution functions of these probability distributions by $G_\varphi^t(\cdot)$. These distribution functions will, of course, have characteristics that reflect type – that is, circumstances. For instance, we will find different distributions of smoking behavior in different socio-economic types. Because the goal of equal-opportunity policy is to compensate persons for their circumstances, we should compensate them as well for the effect of their circumstances on their effort. How can we decide when two persons, of different types, have expended *comparable degrees* of effort? I propose to measure the *degree* of a person's effort by his rank in the distribution G_φ^t . Rank sterilizes out of the distribution aspects of it that reflect circumstances. Thus, for example, if we view 'years of education chosen' as effort, and persons in two different types both rank at the 80th centile of the distributions of years of education in their respective types, we will

declare them to have expended equal *degrees* of effort (although their actual years of education may be quite different).

We may thus define the function

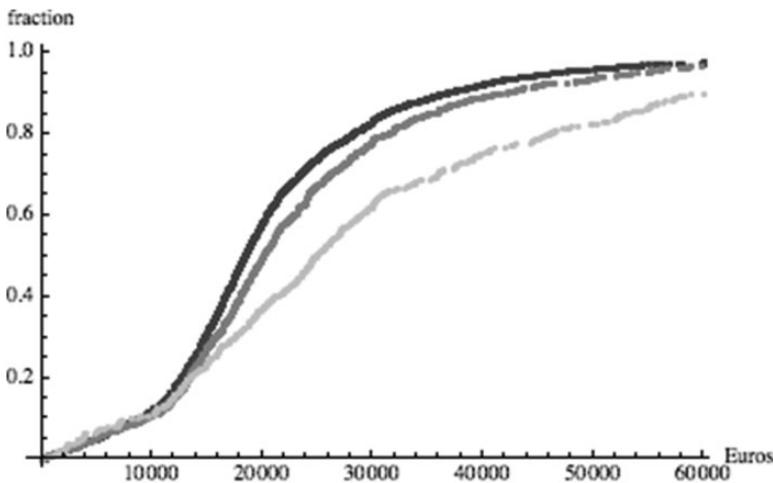
$$v^t(\pi, \varphi) = u^t\left(\left(G_\varphi^t\right)^{-1}(\pi), \varphi\right) \tag{1}$$

which is the (average) value of the objective, when the policy is φ , of the individuals at the π^{th} quantile of the distribution of effort of their type. If effort is uni-dimensional, the function v^t is well-defined. If e is multi-dimensional, then in general it is not, and we should then replace vectors of effort with, for example, the linear combination of its components that best explains the value of the objective. For practical purposes, however, in many applications, one need never measure effort: one can simply define the values $v^t(\pi, \varphi)$ directly as the level of the objective in type t at the π^{th} quantile of the distribution of the objective in that type. Implicitly, this approach assumes that effort is declared to be that constellation of choices that enhance the value of the objective, conditional upon type.

In figure 1, we see the distribution function of post-fisc income of three types of men in France, in 2005, where the unique circumstance defining type is the level of education of the individual's more educated parent. The yellow curve is the distribution function of those men whose parent had at least some tertiary education; the red curve, of those men whose parent had 12 years of education, and the blue curve, of those men whose parent had less than 12 years of education.

Taking the objective to be post-fisc income, the inverses of the functions in the graph are the functions $v^t(\pi)$. So to see the graphs of the three functions v^t ,

FIGURE 1. CDFs of Income for Three Types of French Male Worker



simply reflect figure 1 over the vertical axis and then rotate it 90 degrees clockwise.

Holding persons responsible for their effort means that if two individuals in the same type (who are exposed to identical policy treatments by hypothesis) sustain different values of the objective, there is no inequity, because by hypothesis, these different values are due to differential effort, something for which they are responsible. However, differences between the *functions* v^t are ethically undesirable – a reflection of unequal opportunities—because individuals are not responsible for their type/circumstances. Therefore, the goal of policy should be so render the functions $\{v^t(\cdot, \varphi)\}$ as similar as possible. Since we identify individuals at the same rank of their distributions as having expended equal degrees of effort, the goal is to choose the policy φ to render the distribution functions (in Figure 1) as close together as they can be (that is, to minimize the horizontal distance between the functions).

But we do not want equality of distribution functions at a low level: we therefore desire some kind of maxi-minimization. Suppose we fixed a particular value of π ; inequality in the T numbers $\{v^t(\pi, \varphi) | t \in T\}$ is due not to differential effort, by hypothesis, but to differential circumstances. Thus, we should choose the policy to

$$\max_{\varphi \in \Phi} \min_{t \in T} v^t(\pi, \varphi). \quad (2)$$

However, we are concerned with *every* level of effort: a reasonable way of addressing all effort levels is to take the average of the numbers being maximized in (2), that is, to choose policy to

$$\max_{\varphi \in \Phi} \int_0^1 \min_t v^t(\pi, \varphi) d\pi. \quad (3)$$

I call the solution to (3) the *equal opportunity policy*. It must be emphasized that this policy is conditional upon the definition of circumstances, and the choice of policy space.²

Define $W^{EO}(\varphi) = \int_0^1 \min_t v^t(\pi, \varphi) d\pi$. $W^{EO}(\cdot)$ defines an ordering on Φ ; that is to say that:

$$\varphi \succeq_{EOP} \varphi' \Leftrightarrow W^{EO}(\varphi) \geq W^{EO}(\varphi').$$

In words, $W^{EO}(\varphi)$ is the average value of the lower envelope of the objective functions, across types.

2. A generalization of program (3) is provided in Roemer (2012), where a large family of possible equality-of-opportunity measures is proposed.

There is a special case of interest. Typically, there are constraints that we impose upon policies, so that the policy space for the problem, Φ , is fairly small – for example, we may limit ourselves to affine income tax policies, a uni-dimensional (small) set in the large set of income-tax policies. In this case, it may well be that there is one type – denote it type 1 – that for all policies $\varphi \in \Phi$ is unambiguously the most disadvantaged one, in the sense that its distribution function is dominated (first-order stochastically) by the others at every policy. This is virtually the case in figure 1, where the distribution functions are stacked almost unambiguously – it is obvious that the most disadvantaged type is the one of men whose more educated parent had fewer than twelve years of education, as its income distribution function is virtually FOSD by the distributions of the other two types.³ In this case, the left-hand envelope of the distribution functions is simply the distribution function of a single type, and equation (3) reduces to:

$$W^{EO}(\varphi) = \int_0^1 v^1(\pi, \varphi) d\pi \equiv \bar{v}^1(\varphi) \tag{4}$$

where $\bar{v}^1(\varphi)$ is the average value of the objective⁴ in the most disadvantaged type under policy φ . In this case, the equal-opportunity ethic directs us to *choose the policy to maximize the average value of the objective in the most disadvantaged type* – assuming that this type is unambiguously the most disadvantaged, for any feasible policy.⁵

It is worthwhile contrasting the equal-opportunity ethic with one of its main competitors, the utilitarian ethic. Denote by f^t the fraction of the population in type t . The utilitarian policy maximizes the ordering given by:

$$W^U(\varphi) = \sum_{t=1}^T f^t \int_0^1 v^t(\pi, \varphi) d\pi \tag{5}$$

i.e., the average value of the objective in the population.

3. First-order stochastic dominance does not hold for very low incomes in figure 1; almost surely, this is the case because the children of highly educated parents are going to university, and earning very low incomes early in their careers, less than those who have entered the working-class after secondary school.

4. Recall that the area above a distribution function and bounded by the line at ordinate value one is the mean of the distribution.

5. Indeed, an alternative proposal, due to van de Gaer (1993), is to implement equality of opportunity by maximizing the function $W^*(\varphi) = \varphi \min \int v^t(\pi, \varphi) d\pi$. In $W^*(\cdot)$, the ‘min’ and ‘integral’ operators are commuted, with respect to $W^{EO}(\cdot)$. Fleurbaey and Peragine (2012) call W^{EO} an ‘ex post’ approach, and W^* an ‘ex ante’ approach to measuring equality of opportunity. In Roemer (2012), I offer reasons for my preference for the measure W^{EO} . However, what the text has just pointed out is that, in special cases, the two measures coincide.

A third ordering, associated with John Rawls, is the ordering which maximizes the minimum value of the objective in the population; I will write:

$$W^R(\varphi) = \min_{\pi, t} v^t(\pi, \varphi). \quad (6)$$

We see that the equal-opportunity ethic lies ‘between’ utilitarianism and the Rawlsian difference principle; it is less extreme than the Rawlsian formulation, in that it maximizes an *average* of minima across effort levels. Actually, my naming of (6) as the Rawlsian view is not quite fair, for Rawls wrote that the difference principle should apply to ‘social groups,’ not individuals. If we take the different types to be the relevant social groups, then, at least in the special case where (4) holds, the equal-opportunity ethic maximizes the minimum objective value over social groups, and hence possesses a Rawlsian ancestry.

In the general case, however, if the distribution functions cross, the solution of (3) does not entail maximizing the average value of the most disadvantaged type, but rather, maximizing the area above the left-hand envelope of the distribution functions of the types, and bounded by the horizontal line of height one.

To summarize, we have provided an ordering of policies with respect to the equal-opportunity ethic. That ordering takes as data a particular social view of personal responsibility, summarized in a set of circumstances and an implied typology – a partition of the population – and a policy space. The objective for which opportunities are to be equalized is typically some measurable and interpersonally comparable kind of advantage, the kind of thing a ministry in a government might be concerned with, such as income, health, life expectancy, or educational achievement.

ECONOMIC DEVELOPMENT

Economic development should be measured by the extent to which a society has achieved a desirable distribution of advantage. Desirability should include considerations of both efficiency and justice or fairness. Indeed, the most common measure of economic development, GDP per capita, is based upon the utilitarian ethic, which computes the level of social welfare as the average of the utilities in the population, where utility is taken to be proportional to income.

The human development index (HDI) is not utilitarian, because it is not an average of a value of some kind of advantage over a population. But it is a convex combination of the average of three kinds of advantage over a population: the individual’s income (or consumption), his degree of literacy (which could be coded as 0 or 1), and his life expectancy.⁶ The human development index is an average of the averages of these three dimensions over the whole

6. An individual’s life expectancy can be defined as the average age of death of a cohort of persons with the individual’s characteristics. I am assuming that the life expectancy in a country is the average of life expectancies of individuals, so defined.

society. So the HDI does not essentially depart from utilitarian practice in that it looks only at population averages, although it looks at three averages instead of only one. To be precise, this description is valid for the HDI as defined up to the 2009 Human Development Report of the UNDP. But in 2010, the Human Development Report introduced the ‘inequality adjusted human development index,’ which, although still consequentialist, is not utilitarian.

Neither of these measures of development is sensitive to the distinction between circumstances and effort. They are *consequentialist* measures of how well an economic system is doing, in that the data required to assess the system’s desirability are the values of various outcomes for members of the population. The equal-opportunity view, however, focuses upon the distinction between circumstances and effort. Thus, to assess the desirability of a system, it requires not only the data just mentioned, but also knowledge of the *type* of each individual. It is a non-consequentialist measure, for it will assess differently the same outcome for two individuals, if they have different circumstances. Utilitarianism condemns inequalities if their elimination would increase average or total welfare (however it is measured); opportunity egalitarianism condemns them to the extent they are due to circumstances beyond the control of the individuals concerned. The views are quite different.

Opportunity egalitarianism is not only a superior ethic to utilitarianism, it is the one implicitly endorsed by members of many societies. Suppose one asks the proverbial man on the street, “Do you think that the inequality between the rich Mr. *A* and poor Ms. *B* is unjustified?” it is unlikely that he will answer, “Only if a redistribution from *A* to *B* would increase their total welfare.” But he might well answer, “It depends upon how hard they each worked.” In other words, the popular views of justice are not consequentialist, they are based upon notions of *desert*, and desert is based upon measurements of effort. Our man on the street must know more than the aggregate distribution of the objective to assess whether that distribution is fair – he must know the (disaggregated) distributions of the objective by type. The source of the inequality matters, ethically speaking, but these sources are ignored by looking only at outcomes.⁷

We cannot maintain that the most common measures of economic development are value free: they are derived from a utilitarian ethic. To this claim one might object that the measure of GDP per capita has nothing to do with utilitarianism, it is simply a proxy for technological development. But this cannot be right, because economists are not interested in technological prowess per se: we are interested in human welfare. We would not consider a society highly developed which possessed a fine technology run by slaves, whose product all went, but for the slaves’ subsistence, to the prince. So an attempt to justify the GDP per capita measure of development as a value-free measure of technological

7. There is substantial survey and experimental work which examines the views that people in various societies have concerning distributive justice. An extensive discussion of this literature is found in Gaertner and Schokkaert (2012).

accomplishment has the undesirable consequence of obliterating the distinction between economics and engineering – namely, that economics must always focus upon human welfare.⁸

Therefore, we should use the best conception of justice or social welfare to derive a measure of economic development. Perhaps the extent to which opportunities have been equalized is not the best such conception, but it dominates, so I believe, utilitarian measures. It is better to measure the level of economic development by some statistic that reflects equality of opportunity rather than by a utilitarian measure.

But what version of equality of opportunity should we use to evaluate economic development in a panel of countries? What should be the circumstances and the objective? To be most similar to GDP per capita, the objective should be income – let us say, post-fisc income including the per capita value of public goods. Begin with circumstances that include the educational level and occupations of the parents of the individuals, and the ethnicity and gender of the individual.⁹ Then calculate the number $\hat{W}^{EO} = W^{EO}(\varphi^*)$, where φ^* is the status quo policy. This is the first component of the two-dimensional measure of economic development that I propose. The value \hat{W}^{EO} is the average income of those in society who are most disadvantaged by circumstances. Choosing it as a measure of economic development is inspired by the view, attributed to Mohandes Gandhi, among many others, that “a nation’s greatness is measured by how it treats its weakest members”.¹⁰ This proposal is not new. Indeed Bourguignon, Ferreira, and Walton (2007) propose a dynamic and closely related version of \hat{W}^{EO} : namely they say that the present discounted value of the average income of the most disadvantaged type should be maximized for development policy to be equitable.

THE DEGREE OF OPPORTUNITY EQUALITY

I have proposed to measure the *level of development* of a society as the value of the equal-opportunity social welfare function. Of course, we are highly restricted in our ability to measure economic development when we must use a single number to capture it. In particular, applying the measure defined in (4) to income does not allow us to distinguish the *wealth* of the society from the *degree* to which the society has succeeded in eliminating injustice – that being the influence of circumstances upon inequality. To do this, I propose a second measure, which I call the *degree of opportunity equality*.¹¹

8. I do not wish to denigrate the goals of engineers. However, there is a sense in which engineers are interested in technological efficiency as a goal, while economists are interested in it only insofar as technological efficiency is necessary for Pareto efficiency (in which the goal is human welfare).

9. How one treats gender depends upon whether one uses households or the individual as the unit.

10. The quotation is ubiquitously attributed to Gandhi, although the original statement of it is obscure.

11. Again, this proposal is not new. It is a special case of the ‘inequality of opportunity ratio (IOR)’ defined in Ferreira and Gignoux (2011). Ferreira and Gignoux’s preferred measure of inequality is not CV^2 but the ‘mean logarithmic deviation.’ The same idea for measuring the degree of inequality due to circumstances is proposed in Checchi and Peragine (2010) as well.

A society will have achieved equality of opportunity to the extent that the contribution of differential circumstances to total inequality in the distribution of the objective is small. Let the distribution function of the objective in a given society be H , and the distribution functions of the objectives in the types be H^t ; then

$$H = \sum f^t H^t. \tag{7}$$

Suppose we measure inequality in a distribution by the coefficient of variation squared (CV^2), that is:

$$C(H) = \frac{\text{var } H}{\mu^2} \tag{8}$$

where the mean of H is μ . Denote the mean of H^t by μ^t . Without loss of generality, suppose that we have enumerated types so that $\{\mu_t\}$ is a monotone increasing sequence. Define the distribution:

$$\Phi^T(x) = \begin{cases} 0, & \text{if } 0 \leq x \leq \mu_1 \\ \sum_{t=1}^k f_t, & \text{if } \mu_k < x \leq \mu_{k+1} \text{ for } k < n. \\ 1, & \text{if } x > \mu_n \end{cases} \tag{9}$$

Clearly the mean of Φ^T is μ . If Φ^T were the actual distribution of the objective in society, then everybody in a given type would have exactly the same value of the objective, equal to the mean of the objective in that type. Were this the case, then the contribution of effort to inequality would be nil, as no variation of accomplishment would exist within any type. Now it is well-known that we can decompose $C(H)$ as follows:

$$C(H) = C(\Phi^T) + \sum f^t (\rho^t)^2 C(H^t) \tag{10}$$

where $\rho^t = \mu^t / \mu$. Since both contributions in this decomposition are positive, it is natural to interpret $C(\Phi^T)$ as the amount of inequality due to circumstances, and $\sum f^t (\rho^t)^2 C(H^t)$ the amount of inequality due to effort. I therefore propose, as a measure of the *degree* of opportunity equalization, the index:

$$\eta = 1 - \frac{C(\Phi^T)}{C(H)}. \tag{11}$$

We may want to think of η as an upper bound on the fraction of inequality due to effort because surely some circumstances have not been taken into account, whose effect is measured, residually, as ‘effort.’

My suggestion is that we measure economic development by the ordered pair $d = (\hat{W}^{EO}, \eta)$. Note that neither component of d is a consequentialist measure. One cannot recover either \hat{W}^{EO} or η from knowledge of the distribution of income (more generally, the objective) alone. One must know, as well, the circumstances of individuals, which capture the concept of responsibility salient for the society in question.

COUNTRY CALCULATIONS OF THE LEVEL AND DEGREE OF DEVELOPMENT

In this section, I calculate the value of d for a set of OECD countries. The data upon which these calculations are based are taken from EU-SILC 2005. The sample consists of male workers, who are partitioned into three types, based upon the maximum of the worker's parents' educational levels:

Type 1: the worker's more educated parent had at most lower secondary education

Type 2: the worker's more educated parent had at least upper secondary education but not tertiary education

Type 3: the worker's more educated parent had at least some tertiary education.

The net income for each respondent is recorded, which includes earnings, self-employment income, after taxes and transfers. The *single characteristic* of type in these calculations is parental education which takes on three values.¹²

The fact that income does not include the value of public goods is a weakness of the measure. If a country has a high rate of taxation, and a substantial fraction of tax revenues finance public goods (as opposed to transfer payments), this will not be reflected in the income data. Transfer payments are included in the definition of income.

Figure 1 presents the income-distribution functions for France, by type, which is in many ways typical. Since the left-hand envelope of the three CDFs is, for all practical purposes, the CDF of type 1, the level of development is simply the mean of type 1's income. For France, the level and degree of development, as defined in the previous section, are:

$$(\hat{W}^{EO}, \eta) = (21684, 0.970).$$

(Incomes are measured in Euros.) It may surprise the reader that only about 3% of income inequality is attributed to circumstances, but this is quite typical for

12. I am grateful to Daniele Checchi and Francesco Scervini for providing me with the data set. For an exact description of the data set, see Checchi, Peragine, and Serlenga (2010). The computation of the degrees of development and the type-distributions of income were performed by the author using *Mathematica*; I will supply the code upon request.

advanced European countries, given that only one circumstance is specified. For Latin American countries, this number will be considerably larger. Indeed, [Ferreira and Gignoux \(2011\)](#), in their table 8, present their comparable measure to η , using the mean logarithmic deviation of income as the inequality measure, for a set of six Latin American countries. Their set of circumstances is denser than mine. For some countries (Guatemala), the degree of inequality attributable to circumstances is over 50%.

Figure 2 presents the income CDFs of the three types for one of the least developed countries in sample, Hungary, and for one of the most developed, Denmark.

We see that the inter-type dispersion is considerably more dramatic in Hungary than in Austria, while the CDFs in Denmark are very close together. The graphs of the three CDFs for the other countries in the sample are presented in an online appendix.

Figure 3 plots the ordered pairs (\hat{W}_j^{EO}, η) for all 22 countries in the sample.¹³

Some comments:

1. The eastern European countries are the worst off with respect to the index \hat{W}^{EO} : these comprise Lithuania (LT), Estonia (EE), the Czech Republic (CZ), Poland (PL), Latvia (LV), and Hungary (HU). (Slovenia (SI) does somewhat better.) But Spain (ES) is also very low on this measure. With respect to the degree of development, η , the eastern European countries span a range from about 92% to 99.5%.

2. Greece appears to do very well on the degree of development: I question whether the data are reliable.

3. We may define a partial order with respect to development; a country j *dominates* a country k if

$$(\hat{W}_j^{EO}, \eta_j) > (\hat{W}_k^{EO}, \eta_k).$$

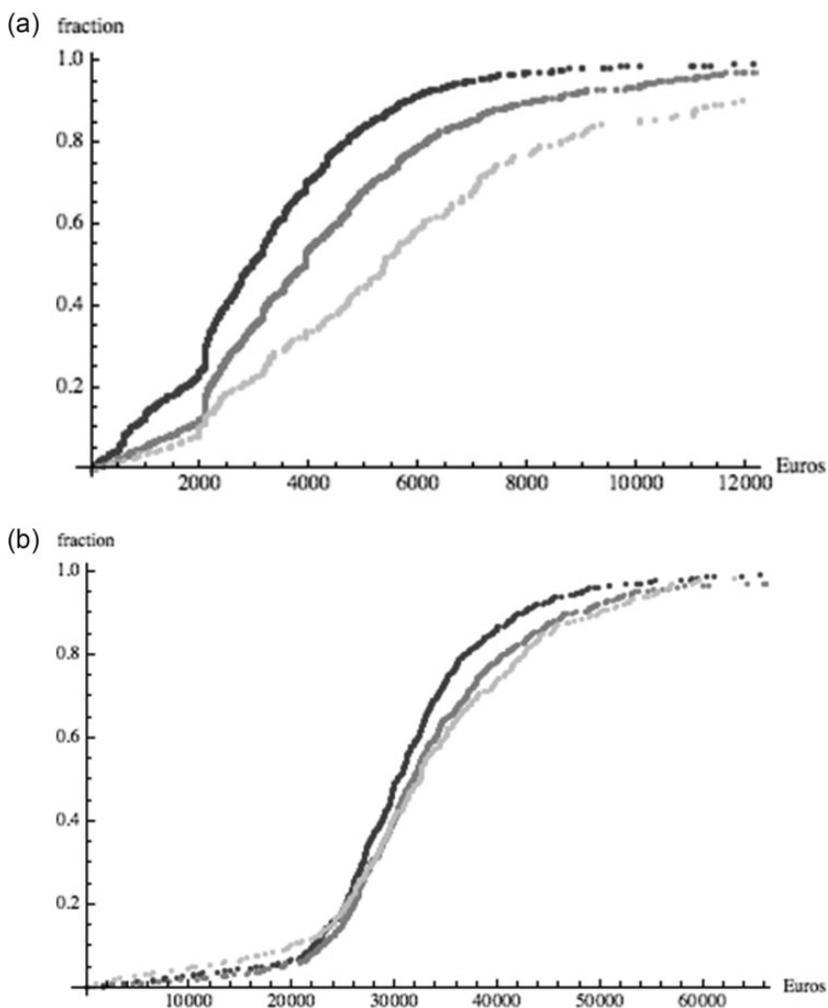
With regard to this partial order, no country in the sample dominates all others. Thus, we can say there exists no most developed European country. Conversely, however, there are five countries that are undominated by any other: Denmark (DK), Iceland (IS), Germany (DE), the UK, and the Netherlands (NL). These data are from 2005, and doubtless Iceland, post-crash, no longer enjoys this status.

Table 1 presents the same data as figure 1.

As noted, [Ferreira and Gignoux \(2011\)](#) calculate a similar statistic to η for six Latin American countries. Their calculation differs from the one presented here using the SILC data in two ways: they have a different set of circumstances, and they use a different measure of inequality. I have calculated an index (W^{EO}, η) for

13. EU-SILC also contains data for Cyprus, but there are so few observations that I do not consider the CDFs to be meaningful. I excluded as well Ireland from the sample, because I believe the data have been miscoded: according to the data, the middle type in Ireland is worse off than the most disadvantaged type.

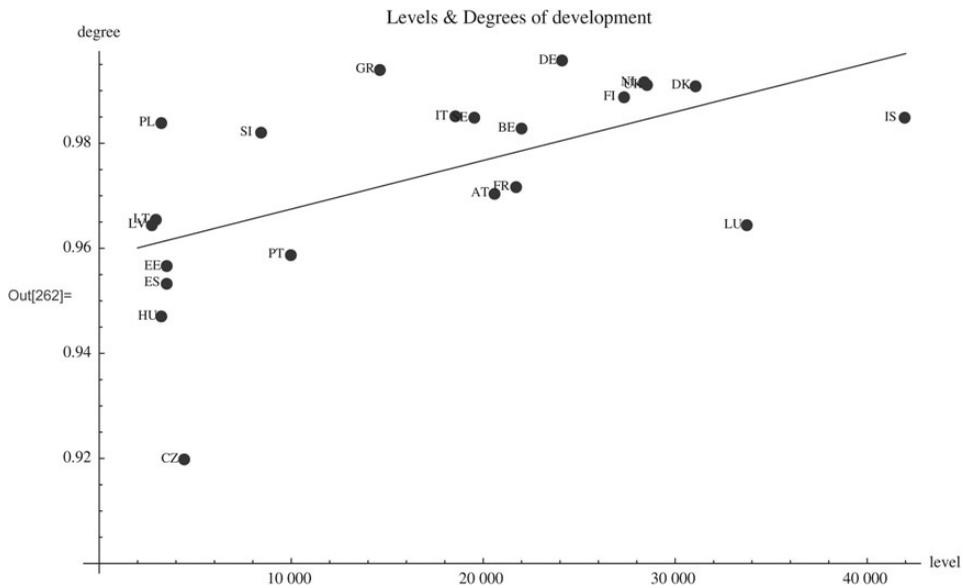
FIGURE 2. (a) CDFs of Income for Three Types of Hungarian Worker (b) Cdfs of Income for Three Types of Danish Worker



Brazil, using a data set which reports income of workers for a typology whose circumstances are race, gender of the head of household in which the worker was raised, and urban-rural.¹⁴ There are four races: white, mixed, black, and ‘other’ – thus 16 types. I limited my analysis to nine types, which comprise 94.5% of the sample, not including the four types of ‘other’ race, or the three rural types with female head-of-household parent. For this population, we compute $\eta = 0.984$, which is surprisingly high – only 1.6% of income inequality is attributed to these

14. I thank Sean Higgins for providing me with the Brazilian data, which were collected as part of the Commitment to Equity project. See Higgins and Pereira (in press) for details.

FIGURE 3. Ordered Pairs (W_i^{EO}, η_j) for European Countries with Best-Fitting (Least-Squares) Regression Line



circumstances. This contrasts with the IOR computation of [Ferreira and Gignoux \(2011\)](#), in which, in Brazil, about 32% of inequality is due to (their) circumstances, which are {ethnicity, father in agriculture, father’s education, mother’s education, birth region}. Surely the inclusion of parental education in the Ferreira-Gignoux data set increases the role of circumstances in generating inequality.

Figure 4a presents the distribution functions of disposable income for the Brazilian types (white, male household head, urban), (white, female household head, urban), (white, male household head, rural) in order of stochastic dominance (using the Higgens-Pereira data set). Thus, it appears that one has better opportunities if one is raised by a woman in the city than by a man in the countryside. Figure 4b presents the analogous three distribution functions for the mixed race. The order of stochastic dominance is the same as in figure 4a. Figure 4c places these two plots together: we see that even the *most disadvantaged* white type first-order-stochastically-dominates the *most advantaged* mixed type.

It turns out that the three black types also have distribution functions ordered in the same way. Figure 5 presents the distribution functions of the black types (in violet) superimposed upon figure 4c. We observe that the first two black types ((male household head, urban) and (female household head, urban)) have distribution functions that are virtually coincident with the analogous mixed types, and the distribution function of (black, male household head, rural) appears to FOSD the comparable ‘mixed’ type. Indeed, the distribution function of the (black, male household head, urban) type is virtually invisible in figure 5, as it coincides so closely with that of the (mixed, male household head, urban) type. The conclusion

TABLE 1. Values of W^{EO} and $W^*(\varphi) = \max_{\varphi} \min_t \int v^t(\pi, \varphi) d\pi$ for 22 OECD Countries

Country	Level	Degree
AT	20 574.6	0.970434
BE	21 971.4	0.982783
CZ	4438.18	0.919772
DK	31 040.6	0.990947
DE	24 099.6	0.995659
EE	3511.13	0.956504
ES	3511.13	0.953383
FI	27 328.7	0.988751
FR	21 683.9	0.971633
GR	14 609.	0.994
HU	3257.44	0.946974
IS	41 968.6	0.984799
IT	18 564.2	0.985046
LT	2942.47	0.965407
LU	33 755.2	0.964373
LV	2713.6	0.964469
NL	28 377.6	0.991502
PL	3246.67	0.983872
PT	9974.57	0.95876
SE	19 544.8	0.984744
SI	8404.	0.982
UK	28 535.6	0.991047

appears to be that there is racial discrimination in Brazil which favors whites over non-whites, but there is no special discrimination against black workers, who appear, if anything, to have somewhat better opportunities than ‘mixed’ workers.

Finally, Björklund, Jäntti and Roemer (2012) study income in Sweden, using a large data set that permits the partition of the population into over 1100 types, based on six circumstances, each partitioned into several levels. Using methods different from the ones discussed here, they conclude that at least 25% of income inequality is due to circumstances.¹⁵ Contrast this to the 1.5% figure for Sweden from table 1.

EQUITY ‘VERSUS’ DEVELOPMENT¹⁶

It is often said that equity and efficiency are competing goals – that equity is purchased at the expense of efficiency. There are two senses in which this phrase may be uttered. The first is that redistributive taxation may be purchased only at the cost of Pareto inefficiency, due to workers’ and firms’ facing different effective wages. The second sense is that redistribution may lower total output. These

15. Perhaps, most critically, IQ in adolescence is taken as a circumstance.

16. The point in this section is discussed more extensively in Roemer (2006).

FIGURE 4. (a) Distribution Functions of Three White Types in Brazil (b) Distribution Functions of Three 'mixed' Types in Brazil (c) Plotting Figures 4a and 4b in The Same Plane

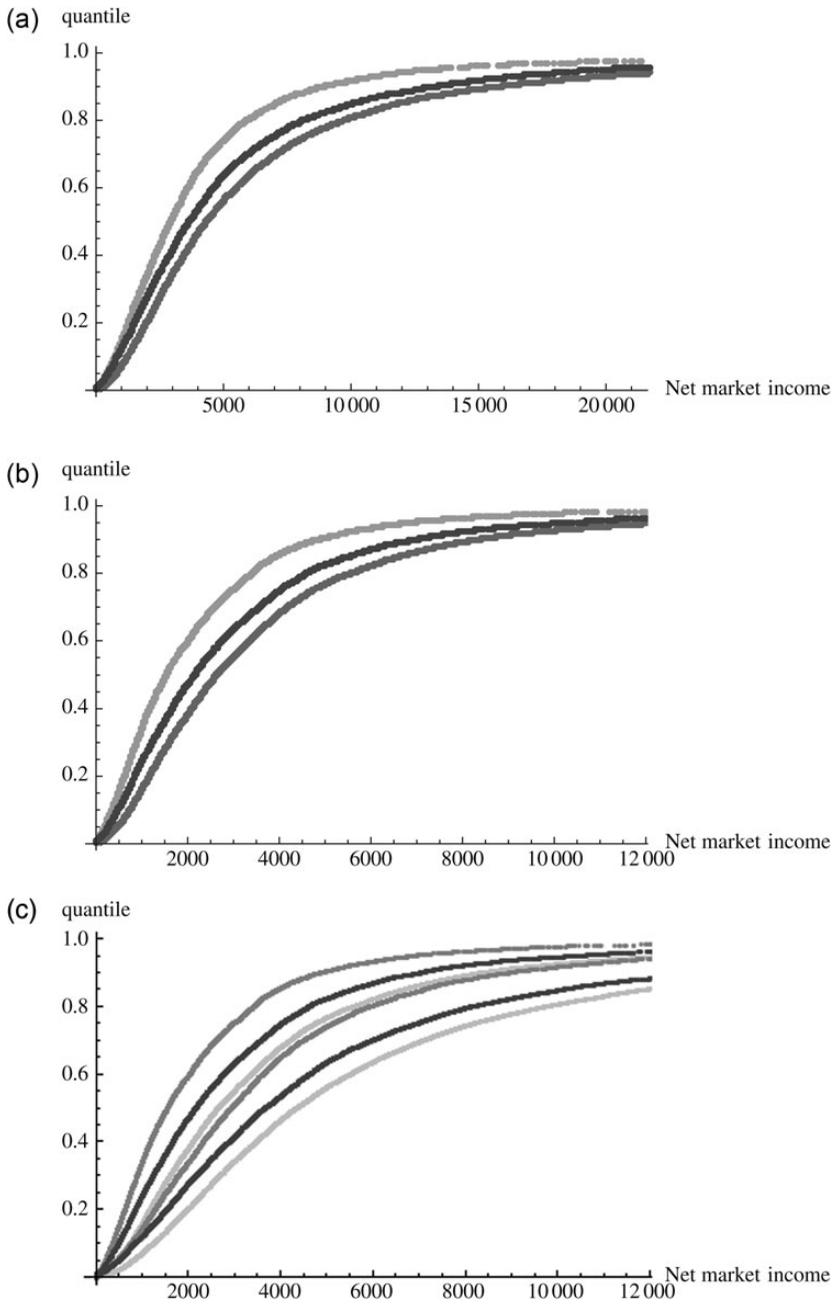
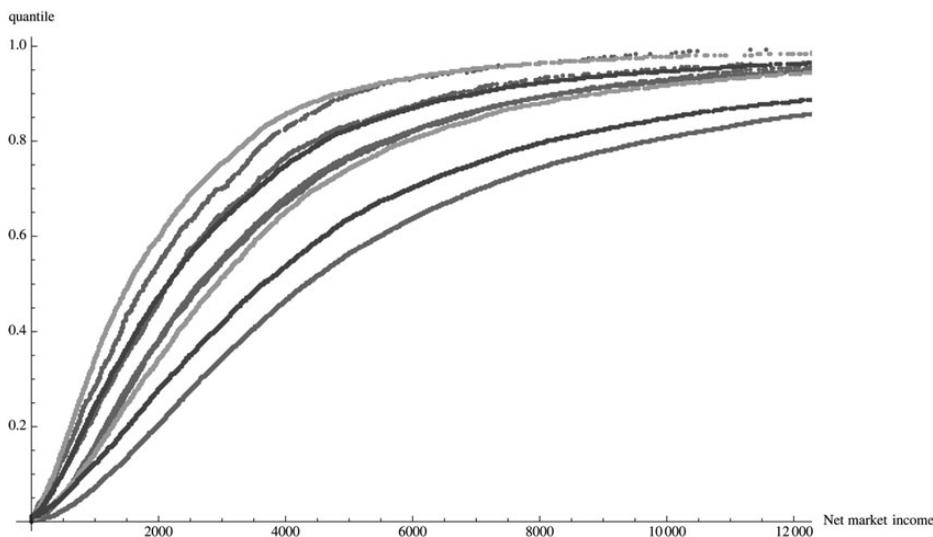


FIGURE 5. Superimposition of the Three Black Types on figure 4c, Brazil



two claims are in principle independent. There may be policies which re-allocate income in a more equitable manner, lower total output, but are not Pareto inefficient. (Think, for example, of re-allocating educational funds from tertiary education to secondary education in a poor country. This might have a purely redistributive effect, without significant consequences for Pareto efficiency.)

I wish to criticize the second usage of the phrase. Saying that there may be a trade-off between equity and efficiency *where efficiency is measured as total output* is equivalent to saying there is a trade-off between equity and the utilitarian measure of development, which (in its simplest form) is given by output per person. In fact, both the measures of equity that I have proposed in the ordered pair d , and output per capita, are measures of equity according to different normative criteria, as discussed in section 2. Indeed, because utilitarianism was the reigning conception of distributive justice until at least the 1970s, it is unsurprising that GDP per capita was the corollary measure of development in economics.

There is an increasing number of economists who argue that ‘improving equity improves efficiency.’ ([The World Development Report \[2006\]](#) presses this point, but the argument goes back many years.) My objection is not to the substantive claim, that equalizing opportunities often increases productivity and national income, but only to the tradition of assigning utilitarianism *primus inter pares* as the normative view which *defines* efficiency.

If the view of economic development I here advocate is adopted, there may be a significant change in policy evaluation. One would not have to justify investment in very disadvantaged social groups by showing that such investment increased total output. In the long run, such a conflict might not exist:

but often, policy makers must evaluate the consequences of their policy choices in the short run. If a country is evaluated on the basis of its ordered-pair statistic (\hat{W}^{EO}, η) rather than on GDP per capita, policies could be quite different.

A WORLD-BANK PROPOSAL FOR MEASURING EQUAL OPPORTUNITY

The World Bank has been an important innovator in bringing considerations of equal opportunity into economic development. Its two important publications, to date, have been the 2006 World Development Report, *Equity and Development*, and a monograph, *Measuring inequality of opportunities in Latin America and the Caribbean* (Paes de Barros et al., 2009). The more recent publication contains a wealth of information on the effects of social circumstances on various measures of achievement and output.

Paes de Barros et al. (2009) propose a measure of equality of opportunity. Consider a particular kind of opportunity, such as ‘attaining the sixth grade in elementary school.’ Let the total sixth-grade attendance in a country be H , and the total number of children of sixth-grade age be N , and define $\bar{p} = H/N$ to be the *access* on average of children to the opportunity of a sixth-grade education. \bar{p} measures the level of this opportunity in the country, but not the extent to which access is unequal to different children, based upon their social circumstances. Now using a logit model, estimate the probability that each child, j , in the country has of attending the sixth grade, where that probability is a function of a vector of circumstances; denote this estimated probability by \hat{p}_j . Define $D = 1/2\bar{p}N \sum |\hat{p}_j - \bar{p}|$. D measures the variation in access to the opportunity in question across children in the country. The normalization guarantees that $0 \leq D \leq 1$. Now define the *human opportunity index* as

$$O = \bar{p}(1 - D);$$

note that $0 \leq O \leq \bar{p}$.

The human opportunity index is a non-consequentialist measure of development, because the probabilities \hat{p}_j can only be computed knowing the circumstances of the children. The measure combines a concern with the level of provision of opportunities and the inequality of the distribution of them. This is to be contrasted with my ordered pair (\hat{W}^{EO}, η) , which separates these two concerns into two measures. Obviously, some information is lost in using a single measure rather than two measures.

The concern of the 2009 report is in large part with children. In my view, where children are concerned, all inequality should be counted as due to circumstances, and none to effort, and so the fact that the human opportunity index

does not explicitly make the distinction between effort and circumstances is unobjectionable.¹⁷ However, if the measure is used for addressing inequality of opportunity for adults, this may be a defect.

To study this, let us take an opportunity for adults – earning an income above M , measured in PPP exchange rates. Suppose there are three types of worker, according to the level of education of their more educated parent. Denote the distribution of income in type t as F^t ; let the fraction of type t be f^t and let F be the distribution of income in the society as a whole. Then $\bar{p} = 1 - F(M)$ is the average access to the opportunity in question in the country. Now for all members j of a given type, t , compute that $\hat{p}_j = 1 - F^t(M)$: this is because the probabilities \hat{p}_j are computed by taking the independent variables in the logit regression as the circumstances. Hence, the human opportunity measure is:

$$\begin{aligned} O &= \bar{p} \left(1 - \frac{1}{2\bar{p}} \sum f^t |1 - F^t(M) - (1 - F(M))| \right) \\ &= (1 - F(M)) - \frac{1}{2} \sum f^t |F(M) - F^t(M)|. \end{aligned} \quad (12)$$

Despite the fact that effort is not explicitly mentioned in defining the index, effort is reflected in measure, because the distributions F^t appear in the calculation. Indeed, the first term $1 - F(M)$ measures the level of opportunity in the country, while the second term is a penalty for the degree to which this opportunity is mal-distributed with respect to circumstances (e.g., if there were no inequality of opportunity, then $F^t(M) = F(M)$ for all t , and the penalty is zero).

In expression (12), the first term on the right-hand side, $1 - F(M)$, plays the role that \hat{W}^{EO} plays in my measure: it measures the level of development. But while \hat{W}^{EO} focuses upon how well off the most disadvantaged type is doing, $1 - F(M)$ is a level for the society at large. The second component of my measure, η , is explicitly derived to show the degree to which inequality is due to circumstances, while the second term on the right-hand side of (12) is a form of a variance. Certainly these two measures are getting at the same phenomenon. I have a slight preference for my proposal, as it is more carefully justified as measuring what we are concerned with. But these are minor criticisms; certainly, the measure O is in the spirit of thinking of economic development as opportunity equalization.

CONCLUSION

Inequality has become an important focus in development economics in recent years, and this is a step forward from the days when only GDP per capita was considered to be salient. But an important weakness in the entry of inequality into the field has been treating all inequality as having the same ethical status.

17. Children should only become responsible for their actions after an ‘age of consent’ is reached, which may vary across societies. Both nature and nurture fall within the ambit of circumstances for the child.

This is seen in the very large literature on the measurement of inequality, where the concern has been upon whether the statistical properties of various inequality measures conform to our intuitions concerning when equality is *large or small*. These discussions ignore the issue of whether inequality is *innocuous or undesirable* – that is, the ethical status of the inequality. The equal-opportunity literature introduced the latter distinction into economic theory, and it built on the introduction of the issue of responsibility into egalitarian political philosophy, through the writings of R. Dworkin (1981a, b), G.A. Cohen (1989) and R. Arneson (1989). For discussions of this literature, from economists' viewpoints, see Roemer (1996, 2009) and the treatment of Fleurbaey (2008).

It is useful to further compare the equal-opportunity approach to inequality to the approach represented by the human development index, based upon the work of Amartya Sen on functionings and capability. As is well known, Sen's (1980) major point was that there are objective measures of human functioning that are important for any conception of welfare, and the set of vectors of functionings, available to a person, which Sen defined as her *capability*, is a measure of the opportunities that she has. Sen's intervention was post-Rawls and pre-Dworkin: his main foil was Rawls's choice of primary goods as the equal-sandum, which he proposed replacing with capabilities; and his conception of responsibility was implicit in the idea that, if capabilities, so defined, were 'equal' (whatever that should mean) across persons, then if individuals chose different vectors of functioning from these sets, the result was of no ethical consequence. The treatment of responsibility, in Dworkin (1981, 1982), was significantly more explicit, and led to the equal-opportunity literature.

The proposal I have stated here, and the human development index (HDI), are complementary. The HDI broadens the objective of concern from income (GDP) to a set of functionings, but continues to average over the population as a whole, and ignores the source of inequality.¹⁸ The equal-opportunity approach – as I have advocated applying it to a set of countries – retains income as the objective, but disaggregates the population into types based upon circumstances that are beyond the control of individuals. The HDI approach says that human accomplishment along dimensions other than income is important, and the equal-opportunity approach says that inequality is bad only if it is of a certain kind.

Of course, it is possible to unite the two approaches. Instead of using income as the measure in my proposal, one could measure human development *disaggregated by types*, where type continues to be defined according to a set of circumstances, and then the two-dimensional index d would allow us to assess levels and degrees of development with regard to the various Sen-inspired functionings. It would be ideal to have data sets that permitted us to do this. The reason I have here proposed using only income is that I think, at this point, we do not have the data to compute the distribution of levels of human development by type for a

18. In the 2011 Human Development Report, the human development index is calculated by taking a geometric mean of national income, literacy, and longevity, rather than a convex combination of them.

large set of countries. However, the recent publication of the results of the Global Burden of Disease project (see the entire issue of *The Lancet*, December 13, 2012) indicates that this lacuna may be filled, as we may soon have available distributions of longevity by country and by type.¹⁹

Note that the issue of the ethical status of inequality is quite different from another way that inequality can be good or bad, and that is, with regard to its effect on incentives. Bad inequality in this sense – inequality that is bad for incentives – will be condemned by the utilitarian measure of GDP per capita, because its elimination will increase social output. This is to be distinguished from inequality that is bad because it reflects disadvantage due to circumstances: as I have emphasized, eliminating this kind of inequality is not – at least in the short run – synonymous with increasing total output or welfare.

The equal-opportunity approach, which focuses upon eliminating inequalities that are due to circumstances for which persons should not be held responsible, is both good ethics and also good policy – by which I mean it is policy supported by the majority of people in many countries. For we know from survey data that, globally, people believe injustice occurs when low incomes are due to bad luck as opposed to low effort. What differs across countries is the extent to which citizens attribute low incomes to bad luck as opposed to low effort: in Brazil, a much larger fraction believe poverty is due to bad luck than in the United States (and perhaps this reflects reality). Indeed, the popular moniker associated with equality of opportunity – it levels the playing field – can be interpreted as a way of saying that disadvantages that some face due to circumstances beyond their control should be eliminated before the competition for economic goods begins.

REFERENCES

- Arneson, R. 1989. "Equality and Equality of Opportunity for Welfare." *Philosophical Studies* 56: 77–93.
- Björklund, A., M. Jäntti, and J. Roemer. 2012. "Equality of Opportunity and the Distribution of Long-Run Income in Sweden." *Social Choice and Welfare* 39: 675–96.
- Bourguignon, F., F. Ferreira, and M. Walton. 2007. "Equity, Efficiency and Inequality Traps: A Research agenda." *J. Econ. Inequality* 5: 235–56.
- Checchi, D., and V. Peragine. 2010. "Inequality of opportunity in Italy." *J. Econ. Inequality* 8: 429–50.
- Checchi, D., V. Peragine, and L. Serlenga. 2010. "Fair and Unfair Inequalities in Europe." Working Paper No. 174. ECINEQ, Society for the study of economic inequality Palma de Mallorca, Balearic Islands. <http://www.ecineq.org/milano/WP/ECINEQ2010-174.pdf>
- Cohen, G.A. 1989. "On the Currency of Egalitarian Justice." *Ethics* 99: 906–44.
- Dworkin, R. 1981a. "What is Equality? Part 1: Equality of Welfare." *Philosophy & Public Affairs* 10: 185–246.
- . 1981b. "What is Equality? Part 2: Equality of Resources." *Philosophy & Public Affairs* 10: 283–345.

19. In this issue of *Lancet*, longevity and morbidity figures are given for almost all countries in the world. Information on these measures of welfare by type is not, however, reported.

- Ferreira, F., and J. Gignoux. 2011. "The Measurement of Inequality of Opportunity: Theory and Application to Latin America." *Rev. Income & Wealth* 57: 622–57.
- Fleurbaey, M. 2008. *Fairness, Responsibility and Welfare*. Oxford: Oxford University Press.
- Fleurbaey, M., and V. Peragine. 2013. "Ex Ante Versus Ex Post Equality of Opportunity." *Economica* 80: 118–30.
- Gaertner, W., and E. Schokkaert. 2012. *Empirical Social Choice*. Cambridge, UK: Cambridge University Press.
- Higgins, S., and C. Pereira. Forthcoming. "The Effects of Brazil's High Taxation and Social Spending on the Distribution of Household Income." *Public Finance Review*.
- Paes de Barros, R., F. Ferreira, J. Vega Molinas, and J Chanduvi Saavedra. 2009. *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: The World Bank.
- Rawls, J. 1971. *Theory of Justice*. Cambridge, MA: Harvard University Press.
- Roemer, J. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian Planner." *Philosophy & Public Affairs* 22: 146–66.
- . 1996. *Theories of Distributive Justice*. Cambridge, MA: Harvard University Press.
- . 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
- . 2002. "Equality of Opportunity: A Progress Report." *Social Choice & Welfare* 19: 455–72.
- . 2006. "Review Essay: The 2006 World Development Report 'Equity and Development.'" *Journal of Economic Inequality* 4: 233–44.
- . 2012. "On Several Approaches to Equality of Opportunity." *Econ. & Phil.* 28: 165–200.
- Sen, A. 1980. "Equality of What?" In S. McMurrin, ed., *The Tanner Lectures on Human Values*, Vol. 1. Salt Lake City: University of Utah Press.
- Van de gaer, D. 1993. "Equality of Opportunity and Investment in Human Capital." Ph.D. dissertation, K.U. Leuven, Flanders, Belgium.
- World Bank. 2006. *World Development Report: Equity and Development*. Washington DC: The World Bank.

The Measurement of Educational Inequality: Achievement and Opportunity¹

Francisco H. G. Ferreira and Jérémie Gignoux

Two related measures of educational inequality are proposed: one for educational achievement and another for educational opportunity. The former is the simple variance (or standard deviation) of test scores. Its selection is informed by consideration of two measurement issues that have typically been overlooked in the literature: the implications of the standardization of test scores for inequality indices, and the possible sample selection biases arising from the Program of International Student Assessment (PISA) sampling frame. The measure of inequality of educational opportunity is given by the share of the variance in test scores that is explained by pre-determined circumstances. Both measures are computed for the 57 countries in which PISA surveys were conducted in 2006. Inequality of opportunity accounts for up to 35 percent of all disparities in educational achievement. It is greater in (most of) continental Europe and Latin America than in Asia, Scandinavia, and North America. It is uncorrelated with average educational achievement and only weakly negatively correlated with per capita gross domestic product. It correlates negatively with the share of spending in primary schooling, and positively with tracking in secondary schools. JEL codes: D39, D63, I29, O54

Educational inequalities have long been a matter of significant policy concern, in both developed and developing countries. Some view educational achievement as a dimension of well-being in its own right, or at least as a fundamental input into a person's functionings and capacity to flourish (Sen, 1985). Education is also a powerful predictor of earnings, as we have known since the early days of work on human capital (for a review see, e.g., Psacharopoulos, 1994). More recent research has also found that inequality in educational

1. Francisco Ferreira (corresponding author) is a Lead Economist with the Development Research Group at the World Bank and a Research Fellow at the Institute for the Study of Labor (IZA); email: fferreira@worldbank.org. Jérémie Gignoux is an Assistant Professor at the Paris School of Economics (PSE) – French National Institute for Agricultural Research (INRA); email: gignoux@pse.ens.fr. We are grateful to the Editors, three anonymous referees and to Gordon Anderson, Markus Jäntti, Maria Ana Lugo, John Micklewright, Alain Trannoy and participants at conferences and seminars in Barcelona, Buenos Aires, Chicago, Oxford and St. Gallen for helpful comments on earlier drafts. We are solely responsible for any remaining errors. The views expressed in this paper are those of the authors, and should not be attributed to the World Bank, its Executive Directors, or the countries they represent.

THE WORLD BANK ECONOMIC REVIEW, VOL. 28, NO. 2, pp. 210–246
Advance Access Publication February 20, 2013

doi:10.1093/wber/lht004

© The Author 2013. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

achievement and earnings inequality are correlated, both over time within the United States and across countries (see, e.g., [Blau and Kahn, 2005](#); and [Bedard and Ferrall, 2003](#)). Education is also correlated with health status, and in some cases with political participation in the democratic process, so that inequalities in the former may translate into undesirable gaps and gradients in other dimensions as well.

For all of these reasons, people care about the distribution of education. Those concerned about fairness and social justice care also about the distribution of opportunities for acquiring a good education and, in particular, about the degree to which family background and other pre-determined personal characteristics determine a person's educational outcomes. Nevertheless, there is much less agreement on how those concepts—inequality in educational outcomes, and inequality of opportunity to a good education—should be measured. Constrained by data availability, early work comparing inequality in education across countries focused on educational attainment: the number of years of schooling a person had completed or, in some cases, broader 'levels' of education, such as primary, secondary, or higher. [Thomas, Wang and Fan \(2001\)](#) compiled a set of Gini coefficients for years of schooling for 85 countries, over the period from 1960 to 1990. [Castelló and Doménech \(2002\)](#) and [Morrisson and Murtin \(2007\)](#) also examine inequality in years of schooling across a large number of countries.

Interesting though those comparisons were, there is widespread agreement that a year of schooling is a problematic unit with which to measure "education." Does a student learn the same amount in 6th grade in Zambia as in Finland? Is the value of one year of schooling the same even across different schools in a single country or city? The growing availability of data on student performance in comparable tests has confirmed what one already suspected: that the answer to these questions is generally 'no'. The quality—and hence the ultimate value—of education varies considerably, both within and across countries.

Over the last decade, different projects have compiled school-based surveys that administer cognitive achievement tests to samples of students across a number of countries, as well as collecting (reasonably) comparable information about the students' families and the schools they attend. The OECD's *Program of International Student Assessment* (PISA) and the International Association for the Evaluation of Educational Achievement's *Trends in International Mathematics and Science Study* (TIMSS) are perhaps the best known, but the *Progress in International Reading Literacy Study* (PIRLS), which is applied to younger students, shares a number of common features.¹

But performance in a test, while probably preferable to a simple indicator of enrollment or attendance, is not a perfect measure of learning either. For one

1. There is also an International Adult Literacy Survey (IALS), which is applied to adults long after they have left school.

thing, tests and test items (i.e. questions) vary in difficulty. The final result is known to measure scholastic ability or learning achievement only imperfectly. For this reason, all of the aforementioned surveys present scores constructed from the raw results by means of Item Response Theory (IRT) models, which attempt to account for “test parameters,” so as to better infer true learning. This process generates an arbitrary metric for test scores, which are then typically standardized to some arbitrary mean and standard deviation.

Using these standardized test scores, a number of studies have attempted to provide international comparisons of educational inequality on the basis of *achievement*, rather than *attainment*. Micklewright and Schnepf (2007) and Brown et al. (2007) examine the robustness of measures of central tendency and dispersion in the distribution of student achievement obtained using different surveys, by comparing the measures and country rankings across them. Marks (2005), Schultz, Ursprung and Wossmann (2008), and Macdonald et al. (2010) examine the question of intergenerational persistence in educational achievement, which is closely related to that of inequality of opportunity, and present cross-country comparisons of measures of the association between student achievement and certain family characteristics.

This paper seeks to contribute to that literature by proposing two simple and closely-related measures of inequality—one for educational achievement and another for opportunity to acquire education—and reporting them for all countries that participated in the 2006 wave of PISA surveys. To measure inequality in achievement, we propose simply using the variance or the standard deviation of test scores. But we arrive at this simple proposal by considering the implications of two issues specific to the distribution of test scores for the measurement of inequality. These two issues are: (i) the fact that many common inequality indices are not ordinally invariant in the standardization to which IRT-adjusted test scores are generally subjected; and (ii) the fact that PISA student samples are likely to suffer from non-trivial selection biases in a number of countries. The choice of the variance (or the standard deviation) addresses the first issue. We also propose two alternative two-sample non-parametric procedures to assess the robustness of the inequality measure to the sample selection biases, and implement them in the four countries for which PISA sample coverage (as a share of the total population of 15 year-olds) is smallest.

The proposed measure of inequality of educational opportunity draws on the recent literature on inequality of opportunity in the income space, but is also adapted to the specificities of educational data and the resulting choice of measure for inequality in achievement. It also utilizes information on student background more comprehensively than all previous studies we are aware of, and is additively decomposable both across circumstances and population sub-groups. The measure is also isomorphic to (inverse) measures of educational mobility.

We report our measures of inequality in educational achievement and opportunity for the 57 countries that took part in the PISA 2006 exercise. Each measure was computed separately for each of the three tests applied by PISA: mathematics, reading and science. But there was a good measure of agreement between their rankings, and we often refer only to the math results in the text. We find considerable variation in the standard deviation of test scores, from lows of around 80 (for Indonesia, Estonia and Finland) to highs near 110 (in Belgium and Israel).² Similarly stark variation exists in our measure of inequality of opportunity, from 0.10 – 0.15 for Macau (China), Australia, and Hong Kong SAR, China, up to 0.33 – 0.35 in Bulgaria, France and Germany.

The paper is organized as follows. Section I describes the data sets we use. Section II considers the implications of test score standardization and of the PISA sampling frame for the measurement of inequality in educational achievement, and reports the standard deviation in test scores for our sample of countries. Section III proposes our measure of inequality of educational opportunity (IOp), discusses some of its properties, and presents results. Section IV applies the proposed measures by examining how they correlate with two educational policy indicators across countries. Section V concludes.

I. DATA

Two broad kinds of data are used for the analysis in this paper. The first is the complete set of PISA surveys, for all 57 countries that participated in the 2006 round. The second is a group of four household surveys, for Brazil, Indonesia, Mexico and Turkey, which are used as ancillary surveys in the two-sample non-parametric sample selection correction procedures described in Section 3. We briefly describe each of these in turn.

The PISA 2006 Data Sets

The third round of the Program of International Student Assessment surveys was conducted in 57 countries between March and November, 2006. Two earlier rounds were collected in 2000/2002 (in 43 countries), and in 2003 (in 41 countries). A fourth round has since been collected in 2009. Most OECD countries were surveyed, as were a number of developing countries in Asia, Latin America, North Africa and the Middle East. Table 1 lists all participating countries in the 2006 round, as well as their sample sizes.

In each country, fifteen year-olds enrolled in any educational institution, and attending grade 7 or higher, were sampled. All children surveyed took three tests: in reading, mathematics, and science.³ Their performance in these tests

2. But the low variance for Indonesia is a good example of the sensitivity of these measures to assumptions made about the nature of selection into the test-taking sample. Under our scenario of “extreme” selection on unobservables, the variance of math scores for Indonesia triples. See below.

3. The data for achievements in reading for the United States were not issued after a problem occurred during the field operations in that country.

TABLE 1. Sample Statistics, Mean Scores and the Standard Deviation in PISA Test Scores

	# Obs.	Coverage rate	Reading Mean	Reading SD (SE of SD)		Math Mean	Math SD (SE of SD)		Science Mean	Science SD (SE of SD)	
<i>Asia & North Africa</i>											
Azerbaijan	5184	0.88	355.0	70.26	2.12	476.8	47.96	1.64	385.3	55.68	1.92
Hong Kong SAR, China	4645	0.97	538.9	81.79	1.92	551.4	93.39	2.31	546.1	91.71	1.92
Indonesia	10647	0.53	383.9	74.79	2.39	380.7	80.01	3.18	384.8	70.06	3.26
Israel	4584	0.76	441.3	119.34	2.79	443.3	107.33	3.20	455.6	111.45	1.92
Japan	5952	0.89	409.5	102.38	2.34	389.2	91.01	2.06	427.1	100.12	2.01
Jordan	6509	0.65	500.2	94.09	2.24	525.6	83.71	1.95	533.7	89.86	1.89
Korea	5176	0.87	290.5	88.29	2.68	315.9	92.59	3.12	326.3	90.06	2.35
Kyrgyzstan	5904	0.63	556.1	102.10	2.51	547.2	86.98	2.03	521.9	83.86	2.03
Macao-China	4760	0.73	490.6	76.36	2.26	524.4	83.90	1.51	509.5	77.83	1.58
Qatar	6265	0.90	312.5	108.12	1.15	317.7	90.24	1.39	349.1	83.29	1.37
Russian Federation	5799	0.81	442.4	93.23	1.87	478.7	89.53	1.58	481.5	89.57	1.33
Chinese Taipei	8812	0.88	506.7	84.38	1.73	562.7	103.11	2.16	543.7	94.45	1.63
Thailand	6192	0.72	425.2	81.85	1.73	425.5	81.43	1.57	429.7	77.17	1.45
Tunisia	4640	0.90	379.0	97.30	2.49	363.9	91.95	2.34	384.2	82.38	2.05
Turkey	4942	0.47	452.9	92.90	2.75	428.2	93.24	4.32	427.6	83.20	3.14
<i>Latin America</i>											
Argentina	4339	0.79	383.9	124.22	3.63	388.1	101.14	3.48	398.3	101.24	2.62
Brazil	9295	0.55	389.2	102.46	3.34	365.6	92.02	2.65	385.3	89.28	1.93
Chile	5233	0.78	447.9	103.24	2.44	417.1	87.44	2.17	443.1	91.68	1.72
Colombia	4478	0.60	390.3	107.83	2.38	373.8	88.04	2.42	391.9	84.81	1.81
Mexico	30971	0.54	427.4	95.68	2.27	420.7	85.27	2.16	422.6	80.70	1.47
Uruguay	4839	0.69	424.7	121.22	2.03	435.5	99.30	1.77	437.7	94.44	1.73
<i>North America & Oceania</i>											
Australia	22646	0.87	508.7	96.25	1.43	516.3	85.79	1.03	523.1	94.19	1.14
Canada	14170	0.87	512.3	93.79	1.00	517.4	88.03	1.09	522.5	100.23	1.02
New Zealand	4823	0.84	522.7	105.21	1.58	523.8	93.27	1.20	532.7	107.30	1.36
United States	5610	0.85				474.7	89.75	1.90	488.3	106.07	1.68
<i>Eastern Europe</i>											
Bulgaria	4498	0.83	406.8	117.51	4.00	417.4	101.10	3.65	439.1	106.72	3.20
Czech Republic	5932	1.01	509.6	111.21	2.90	536.0	103.14	2.08	537.6	98.41	2.00
Estonia	4865	0.94	502.4	85.19	1.87	516.8	80.68	1.54	533.7	83.75	1.09
Croatia	5213	0.85	477.6	88.83	2.12	467.3	83.31	1.50	493.7	85.72	1.44

Hungary	4490	0.85	488.1	94.39	2.37	496.2	91.04	1.94	508.7	88.20	1.53
Lithuania	4744	0.93	469.3	95.54	1.51	485.6	89.80	1.73	486.5	89.99	1.52
Latvia	4719	0.85	484.9	90.70	1.69	491.2	82.81	1.51	493.8	84.38	1.30
Montenegro	4455	0.84	388.2	89.41	1.64	395.8	84.45	1.80	408.8	79.69	1.19
Poland	5547	0.94	512.6	100.22	1.48	500.9	86.52	1.13	503.3	89.87	1.11
Romania	5118	0.66	392.0	91.86	2.93	415.0	83.97	2.85	416.6	81.16	2.37
Serbia	4798	0.83	402.9	91.84	1.69	436.6	91.76	1.77	436.9	85.15	1.56
Slovak Republic	4731	0.95	470.6	105.08	2.51	495.1	94.53	2.47	491.2	93.15	1.79
Slovenia	6595	0.88	468.6	87.97	2.47	482.2	89.25	1.36	494.2	98.11	1.35
<i>Western Europe</i>											
Austria	4927	0.92	494.0	108.16	3.16	509.5	98.06	2.29	513.9	97.83	2.41
Belgium	8857	0.99	507.1	110.02	2.81	526.9	106.13	3.31	516.3	99.70	2.00
Switzerland	12192	1.02	496.6	94.07	1.71	528.3	97.44	1.60	508.0	99.31	1.61
Germany	4891	0.95	496.5	111.95	2.67	504.3	99.08	2.53	516.2	99.98	1.99
Denmark	4532	0.85	493.8	89.30	1.63	512.2	84.85	1.53	494.7	93.13	1.42
Spain	19604	0.87	479.5	88.84	1.14	501.7	88.92	1.09	504.5	90.54	0.97
Finland	4714	0.93	547.1	81.23	1.08	549.0	80.87	1.01	563.4	85.62	1.00
France	4716	0.91	488.7	103.95	2.75	496.4	95.58	1.96	496.1	101.57	2.09
United Kingdom	13152	0.94	495.6	101.92	1.69	497.3	88.92	1.31	514.3	106.79	1.50
Greece	4873	0.90	461.9	102.61	2.92	462.0	92.30	2.37	476.6	92.12	2.03
Ireland	4585	0.94	518.6	92.39	1.86	502.3	81.99	1.50	509.5	94.35	1.50
Iceland	3789	0.96	485.0	97.09	1.23	505.6	88.08	0.89	491.0	96.87	0.95
Italy	21773	0.90	477.0	108.76	1.74	473.6	95.82	1.66	487.2	95.56	1.31
Liechtenstein	339	0.84	510.7	95.14	2.93	524.9	93.05	2.17	522.3	96.96	2.10
Luxembourg	4567	1.03	480.1	99.85	0.72	490.5	93.15	0.73	486.8	96.53	0.67
Netherlands	4871	0.96	513.9	96.62	2.47	537.4	88.60	2.18	530.8	95.63	1.64
Norway	4692	0.97	484.4	105.15	1.92	489.8	91.58	1.38	486.9	96.12	1.98
Portugal	5109	0.78	476.8	98.82	2.28	470.9	90.65	1.97	479.0	88.56	1.71
Sweden	4443	0.97	509.0	98.21	1.77	503.2	89.66	1.37	504.2	94.21	1.40

Notes: The standard deviation (S.D.) of test scores is used as an ordinal measure of inequality in achievement, as discussed in the text. Standard errors reported in the columns next to the S.D. are bootstrapped.

Source: Authors' analysis based on data from PISA 2006.

forms the basis for the assessment of their learning or cognitive achievement. Yet, educationalists seem agreed that raw, unadjusted test scores are of little value. Test questions (or ‘items’) vary in their degree of difficulty, and simply adding up correct answers, or weighing them arbitrarily, does not correctly measure the latent variable of interest—cognitive achievement. Instead, the educational community in charge of international tests such as PISA, TIMSS, PIRLS and IALS processes raw scores through statistical techniques known as Item Response Theory (IRT). In essence, an item response model consists of an equation of the form:

$$p(s|\theta, \alpha). \tag{1}$$

Equation (1) gives the probability of scoring s in a given test, conditional on individual latent cognitive ability θ and test item parameters α (such as their difficulty). Given an additional assumption about the distribution of latent ability in the population (usually a normal law such as $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$) and an observed distribution of raw scores, $F(s)$, the IRT model is used to back out a distribution of the latent variable θ .⁴

Unfortunately, the inference of statistics summarizing distributions of unobserved latent variables such as θ is subject to a specific small-sample measurement error problem: Each pupil answers a limited number of items so that it is not possible to estimate individual abilities accurately. In this situation, the distribution of estimates for individual abilities obtained with traditional methods (such as maximum likelihood estimates) does *not* converge to the population distribution of these abilities as the number of examinees increases (Mislevy et al. 1992). These estimates of parameters of this distribution are thus inconsistent (although the asymptotic bias decreases with the number of items per examinee).

The standard solution to this measurement error problem in psychometrics is to draw a number of plausible values for the latent variable for each individual. The marginal distribution of ability for each pupil, conditional on his or her answers and a set of observables, is estimated, and a number M of draws from this distribution is obtained. These M draws are known as *plausible values* of a pupil’s score. To estimate a given statistic s , each of the M datasets containing one plausible value per pupil should be used separately to obtain a set of estimates \hat{s}_m . The final estimate \hat{s} of the statistic s is given by the average of the M estimates \hat{s}_m . For PISA, $M = 5$, which implies that five “data sets” are used separately to compute sample statistics (e.g. means and standard deviations). In conformance with the advice in the PISA Data Analysis Manual (OECD, 2009), all of the estimates presented in this paper are computed as averages of the summary statistics estimated separately for each of the five data

4. See Baker (2001) for a general introduction, and OECD (2006) for a description of how the IRT method is applied to PISA surveys.

sets (rather than as summary statistics of the distribution obtained by first averaging across plausible values).⁵

This use of Item Response Theory involves a number of functional form assumptions which are not innocuous. Brown et al. (2007) have shown, for instance, that the final distribution of test scores can be sensitive to differences in the specification of the model used to estimate equation (1).⁶ Here, however, we are concerned with the standardization that happens *after* the IRT adjustment (and the appropriate treatment of the distributions of plausible values generated in the process). Once that procedure is complete, and a new distribution of ‘adjusted’ test scores (which we denote by x) has been generated, this latter variable is standardized, according to a simple formula such as:

$$y_{ij} = \hat{\mu} + \frac{\hat{\sigma}}{\sigma} (x_{ij} - \mu) \quad (2)$$

In equation (2), x_{ij} denotes the (post-IRT, pre-standardized) test score for individual i in country j . μ and σ denote their original mean and standard deviation across all countries in the sample (the world, or the OECD, for example). $\hat{\mu}$ ($\hat{\sigma}$) is the new arbitrary mean (standard deviation) for the standardized distribution. In the PISA procedure, it has a value of 500 (100). It is the distributions of y_{ij} that are used in computing means and inequality indicators for each country j in the PISA data set. As we will see in the next section, the operation described by equation (2), even if the IRT procedure that precedes it is taken as given, poses serious issues for inequality measurement.

In addition to standardized test scores, the PISA data set contains information on a number of individual, family and school characteristics for each test-taker. The presence of these covariates accounts for a large part of the interest of the research community on the PISA data. For the analysis of inequality of opportunity in education, we focus on a subset of these covariates that are informative of the family background and other inherited circumstances of the child. Ten such variables are used: gender, father’s and mother’s education, father’s occupation, language spoken at home, migration status, access to books at home,

5. The sampling variance of population parameter estimates are computed using the Balanced Repeated Replication (BRR) weights provided within the data (PISA 2006). BRR is a replication method for multistage stratified sample designs similar to the Jackknife. The particular variant of the BRR known as Fay’s method was used. For PISA it consisted in forming pairs (called strata) of schools (the primary sampling units) and drawing a number of replicates of the sample (using a so-called Hadamard matrix). 80 replicates were performed. Each of these replicate attributes weight 1.5 to one of the school and weight 0.5 to the other in each strata, the selection being different for each replicate. The BRR weights are then computed as the product of students’ original sampling weights and the school weight (1.5 or 0.5) for each particular replication. See Mislevy (1991) and Mislevy et al. (1992) for a more detailed discussion.

6. Brown et al. (2007) investigated this question by applying the IRT model used in the 1999 TIMSS sample retrospectively to the 1995 sample, which had used a different specification for (1). Although changes were small for most developed countries, there were some non-trivial re-rankings among developing countries.

durables owned by the households, cultural items owned, and the location of the school attended (used as an indicator of a rural or urban upbringing).⁷

Parental education is measured by the highest level completed and is coded using ISCED codes into four categories: a) no education or unknown level; b) primary education (ISCED level 1); c) lower secondary education (ISCED level 2), upper secondary (ISCED level 3), or post-secondary non-tertiary education (ISCED level 4); and d) college education (ISCED level 5)). Father's occupation is classified using ISCO codes. We aggregate occupations into three broad categories: a) legislators, senior officials and professionals, technicians and clerks; b) service workers, craft and related trades workers, plant or machine operators and assemblers, and unoccupied individuals; and c) skilled agricultural and fishery workers, elementary occupations or unknown occupation. The variable for language spoken at home is a dummy identifying a language other than the language of the test. The migration status variable is a dummy identifying a first or second generation migrant as an individual who was, or whose parents were, born in a foreign country.⁸

The number of books at home variable, an indicator of parental human capital, is a categorical variable coded into four categories: a) 0 to 10 books; b) 11 to 25 books; c) 26 to 100 books; and d) more than 100 books. Ownership of durables, an indicator of family wealth, is captured by six dummy variables indicating the ownership of a) a dishwasher; b) a DVD or a VCR player; c) a cell phone; d) a television; e) a computer; f) a car. Ownership of cultural possessions is captured by three dummy variables indicating the ownership of a) books of literature; b) books of poetry; and c) works of arts (paintings are mentioned as an example of such works in the formulation of the question). School location is a proxy for the person's inherited spatial endowment and we recode it using three categories: a) villages or small towns (less than 15,000 inhabitants); b) towns (between 15,000 and 100,000 inhabitants); and c) cities (larger than 100,000 inhabitants). School location information was not collected in France; Hong Kong SAR, China; and Liechtenstein.

A final data issue worth highlighting is that of sample coverage and representativeness. PISA samples were designed to be representative of the population of 15 year-olds who are enrolled in grade 7 or higher in any educational institution. The samples are *not*, therefore, representative of the total population of

7. School-level variables are not used in this analysis deliberately, for reasons which should become clear in Section 4.

8. Naturally, non-random measurement error in these covariates would be undesirable. In particular, one might be concerned that information on family background elicited from children might be systematically misreported. To assess the seriousness of this problem, we use supplementary information on parental education asked directly of parents, which is available for sixteen countries in PISA 2006. For these countries, attainments reported in the child and parent questionnaires for both the mother and the father match exactly for 70 percent of children. Moreover there is no evidence that children tend to systematically report higher than actual attainments: The shares of children reporting higher and lower attainments than the parents are close to 15 percent for both fathers and mothers.

15 year-olds in each country: Children who dropped out of school before they turned fifteen, as well as those who are so delayed that they are in grade 6 or lower at age fifteen, are purposively excluded. In addition, sampling flaws induce an additional under-coverage of enrolled 15 year olds. PISA documentation suggests that this arises from the fact that their sampling frame (a listing of schools and sampling weights) is established in the year preceding the surveys, on the basis of current school enrollment on that year. But some schools close down between the two years, and new ones are not included in the sample. Changes in the enrollment of 15 year-olds arising from this process are not taken into account.

The PISA sample coverage rate, defined as the ratio of the covered student population (using PISA expansion factors) to the total population of 15 year-olds, varies considerably across countries, and is reported in column 2 of Table 1. Although coverage is typically high in OECD countries, it is low in many developing ones: Coverage rates are as low as 47% for Turkey, 53% for Indonesia, 54% for Mexico, and 55% for Brazil. Overall, coverage is less than 80% of the total population of 15 years-olds in fifteen countries. Table 2 provides a sense of the sources of exclusion for the four countries in our dataset with the lowest coverage rates, by decomposing those selected out of the sample into children no longer in school, children with excessive delays, and those missed due to PISA sampling issues. It should be obvious from these magnitudes that any international comparison of countries with vastly different coverage rates must seek to address the problem in some way, and we suggest two alternatives in Section 3.

Ancillary Household Survey Data Sets

Our proposed procedure to examine the sensitivity of inequality measures to sample selection, which is described below, relies on using information on fifteen year-olds from general-purpose household surveys. While these surveys may have their own sampling issues, these are not dictated by school enrollment or delay status, or by school closures, openings and reforms. We obtained such household surveys for the four countries with the lowest coverage rates in the 2006 PISA sample: those reported in Table 2. For Brazil, we used the Pesquisa Nacional por Amostra de Domicílios (PNAD) 2006. For Indonesia, we used the SUSENAS 2005. For Mexico, the Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) for 2006 was used. For Turkey, the Household Budget Survey (HBS) 2006 was used.

All four are large-sample household surveys with national coverage and representative down to the regional level, which are fielded on an annual basis by each country's national statistical authority. The PNAD 2006 collected information from a sample of about 119,000 households and 410,000 individuals; SUSENAS 2005 from 257,900 households and 1,052,100 individuals; the ENIGH 2006 from 20,900 households and 83,600 individuals; and the HBS 2006 from 8,600 households and 34,900 individuals. We restrict the samples

TABLE 2. PISA Sample Coverage: Analysis for Four Developing Countries

	Brazil	Indonesia	Mexico	Turkey
Expanded 15 year-old populations, using PISA data and weights				
Total population of 15-year-olds	3 390 471	4 238 600	2 200 916	1 423 514
Total enrolled population of 15-year-olds at grade 7 or above	2 374 044	3 119 393	1 383 364	800 968
Weighted number of students participating to the assessment	1 875 461	2 248 313	1 190 420	665 477
Coverage rate of the population of 15-year-olds, from PISA (%)	55.3	53.0	54.1	46.7
Total missed children (%)	44.7	47.0	45.9	53.3
Composition of those not covered by PISA samples				
Out-of-school children (%)	10.2	25.5	24.1	21.6
Delays of more than two years (%)	19.8	0.9	13.1	22.2
PISA sampling issues (%)	14.7	20.6	8.8	9.5

Source: PISA 2006 surveys; PNAD 2006 for Brazil, Susenas 2005 for Indonesia; ENIGH 2006 for Mexico, and HBS 2006 for Turkey. The share of fifteen year-olds who are not enrolled in school comes from the ancillary household surveys. Those delayed by more than two years come from household surveys, and are checked with PISA administrative records. The last row is derived as a residual.

to children aged 15, for which we have 7,626 observations in the PNAD 2006; 22,600 in the SUSENAS 2005; 1,921 in the ENIGH 2006; and 683 in the HBS 2006. Although some children in boarding schools and other institutions are likely to be out of the sample frame, those samples should otherwise be representative for the total population of 15 year-olds.

In these four countries, these are the staple surveys for assessing the distribution of household income and, in some cases, consumption expenditures. But they also collect information on other topics, including labor supply, education and migration. We use information on parents' characteristics for estimating the total population of 15 year-olds in groups defined by similar gender, mother's education and father's occupation. The classification of the family background variable can be made comparable with the ones in the PISA by appropriate aggregation of coding categories. Parental characteristics are missing for orphans, children who do not live with their parents, or whose parents did not report their education. For instance, the information on mother's education is missing for about 15.0% of 15 year-olds in the PNAD 2006, 8.7% in the SUSENAS 2005, 11.9% in the ENIGH 2006, and 3.8% in the HBS 2006. When comparing the two surveyed populations, children with missing parental background information in the household surveys are not dropped, but associated with those with the same information missing in the PISA survey.

II. MEASURING INEQUALITY IN EDUCATIONAL ACHIEVEMENT

Measures of inequality in educational achievement are based on distributions of standardized test scores (y_{ij}), constructed from the IRT-adjusted scores (x_{ij}) by means of a transformation such as equation (2). In the case of PISA, the transformation is given by (2) exactly, with $\hat{\mu} = 500$, and $\hat{\sigma} = 100$. That operation involves both a translation of the original distribution (by the difference between the new arbitrary mean and the original mean, re-scaled) and a rescaling (by the ratio of the new to the original standard deviations).

In the field of inequality measurement it is usual to impose axioms, or desirable properties, that individual indices should respect. Three common such axioms are:

- (i) *symmetry*: which requires that the measure be insensitive to any permutation of the y vector;
- (ii) *continuity* in any individual income;
- (iii) and the *transfer principle*: which requires that the measure should rise (strong axiom) or at least not fall (weak axiom) as a result of any sequence of mean-preserving spreads.

In addition, inequality indices often satisfy *either* one of two invariance axioms:

(iv-a): *scale invariance*: which requires that the index be insensitive to any re-scaling of the y vector: $I(y) = I(\lambda y)$, $\lambda > 0$, where y is the vector of interest, and λ is a positive scalar.

(iv-b): *translation invariance*: which requires that the index be insensitive to a translation of the y vector: $I(y) = I(y + a)$, $a \neq 0$, where a is a non-zero constant vector of the same dimension as y .

An important result, due to [Zheng \(1994\)](#), is that no inequality index that satisfies axioms (i)-(iii)—known as “meaningful” inequality measures—satisfies *both* (iv-a) and (iv-b). This impossibility result, in other words, states that no meaningful inequality index can be both scale- and translation invariant. A direct implication of Zheng’s result for the measurement of inequality of educational achievement using standardized data is stated below as our Remark 1:

Remark 1: No meaningful inequality index yields a cardinally identical measure for the pre- and post-standardization distributions of the same test scores.

Note that the remark derives from the standardization procedure (equation 2), rather than from the much more complex item response theory adjustments. It refers, therefore, to the measurement of inequality *in* IRT-adjusted test scores, and not to a comparison between adjusted and unadjusted scores.

How important is Remark 1? Clearly this depends on whether or not inequality indices applied to pre- and post-standardization distributions are ordinally equivalent—that is to say, whether they *rank* distributions in precisely the same way, regardless of cardinal differences in value. After all, standardization is just a change in metric. The (post-standardization) mean score in each country j , for example is simply:

$$\mu_j^y = \hat{\mu} + \frac{\hat{\sigma}}{\sigma} (\mu_j^x - \mu) \quad (3)$$

where μ_j^x is the pre-standardization mean in country j , and other notation is as in equation (2). Since every other term in (3) is a constant, μ_j^y and μ_j^x are ordinally equivalent. One is a monotonic (and in this case, affine) transformation of the other. Country ranks based on either would be identical. The only effect of standardization on country mean scores is a change in metric. Since this was the point of the process in the first place, there seems to be no cause for concern.

The same is true for percentile-based measures of dispersion, such as the inter-quartile ratio, or the absolute difference P95-P5 used by [Micklewright and Schnepf \(2007\)](#) to compare dispersion across 21 countries and three different surveys. Equation (2) is itself a monotonic, and therefore rank-preserving, transformation. Since each score y_i occupies precisely the same rank in its distribution as the original score x_i did in its distribution, rank- or percentile-

based measures—be they ratios or differences, will be cardinally different, but ordinally equivalent.

Yet this is not true of inequality measures in general. The post-standardization Gini coefficient in country j (G_j^y) for example, can be straightforwardly shown to relate to the pre-standardization Gini (G_j^x) as follows:

$$G_j^y = \frac{\mu_j^x \hat{\sigma}}{\mu_j^y \sigma} G_j^x. \tag{4}$$

Unlike in equation (3), the terms multiplying G_j^x are not all constants. In particular, the post-standardization Gini is a function of the ratio of pre- to post-standardization means, which varies with μ_j^x (see equation 3). The existence of a second argument in (4) implies that the post-standardization Gini coefficient is *not* ordinally equivalent to its pre-standardization analogue.

Most other common meaningful inequality measures do not share the linearity of the Gini, so their post- and pre-standardization formulae cannot be related as straightforwardly. Nevertheless, substitution of equations (2) and (3) into the formulae for the Generalized Entropy or the Kolm-Atkinson classes of inequality measures yield expressions that are functions of both the central distance indicators of the measure in question, and of the ratio of pre- to post-standardization means (μ_j^x / μ_j^y). For the Generalized Entropy (GE) class, for example:

$$GE_j^y = \frac{1}{\alpha^2 - \alpha} \left[\frac{1}{n_j} \sum_{i \in j} \left(\frac{\hat{\mu} + \hat{\sigma} / \sigma (x_{ij} - \mu)}{\mu_j^x} \right)^\alpha \left(\frac{\mu_j^x}{\mu_j^y} \right)^\alpha - 1 \right]. \tag{5}$$

These results give rise to our second remark:

Remark 2: A number of well-known inequality indices are not even ordinally equivalent when applied to pre- and post-standardization distributions.

Ordinal equivalence with respect to standardization is clearly a desirable property for an index used for measuring inequality in educational achievement. The standardization operation given by (2) is meant merely to adjust an arbitrary metric. It is not intended to fundamentally alter our judgment of how countries compare with one another in substantive terms. Yet, when indices such as the Gini or Theil index are applied to these standardized distributions, we cannot be confident that the original rank in post-IRT adjusted inequality is preserved.⁹

What then are the options for those interested in the distribution of educational achievement? One could, of course, rely on rank-based measures such as the inter-quartile range or percentile differences which, as noted above, are

9. Gamboa and Waltenberg (2012), for example, report Theil-L indices of post-standardized PISA test scores.

ordinally equivalent. However, these measures do not satisfy the transfer principle: a progressive transfer (from above) to the income recipient on the 95th percentile will, for example, cause the p95-p05 measure to indicate an increase in inequality. And of course, because such indices are insensitive by construction to any changes in incomes that do not affect those on the percentiles of reference, they also violate continuity.

A possible alternative would be to use an absolute measure of inequality—such as the variance, or the absolute Gini coefficient¹⁰—which are ordinally invariant in the standardization. The variance of a post-standardized distribution (V_j^y), for example, is a monotonic (linear) function of the pre-standardization variance (V_j^x), and does not depend on any other moment of the pre-standardization distribution:

$$V_j^y = \left(\frac{\hat{\sigma}}{\sigma}\right)^2 V_j^x. \quad (6)$$

The variance is seldom used as an inequality measure because it is scale-dependent: It increases with the mean. It also fails the transfer sensitivity axiom, by placing greater weight on transfers higher up the distribution than to those lower down. While these are not trivial concerns, it appears to us that in the context of distributions of educational achievement, they are less severe than violating either the transfer principle itself (like the percentile based measures) or ordinal invariance in the standardization, which allows an apparently innocuous operation to fundamentally alter distributional rankings. The variance (and the standard deviation, of course) is a meaningful measure of inequality in the precise sense that it satisfies axioms (i)-(iii) above. The variance is also additively decomposable, and shares of the variance obtained from some such decompositions can be shown to be cardinally invariant to standardization, as discussed in the next section.¹¹ These properties will prove instrumental in adapting an intuitive measure of inequality of opportunity to the context of education.

It should be noted that ordinal invariance is not a mere theoretical curiosity. For the countries listed in Table 1, country rankings for inequality in achievement in Mathematics differ considerably if one uses the variance (which preserves the original ordering) or the (relative) Gini coefficient (which does not). As an example, consider the positions of Mexico and Germany: Mexico is ranked the 15th most unequal country by the Gini, but only 44th most unequal by the variance. Germany has the 8th highest variance, but 22nd highest Gini.¹²

10. The absolute Gini coefficient, of course, is the standard (relative) Gini index scaled up by the mean.

11. An alternative ordinally invariant measure is the ratio of the within-country variance to the overall variance in the pooled sample of countries. This measure would also preserve cardinality. We are grateful to an anonymous referee for pointing this out.

12. The detailed comparison is available from the authors upon request.

For these reasons, we adopt the variance and the standard deviation as our basic measures of inequality of educational achievement. Because users of this kind of data are generally more comfortable with the standard deviation than its square, this is the variable we report. Columns 3-11 in Table 1 present the mean and standard deviation (S.D.) of the standardized test scores in reading, math and science, in that order, for all 57 countries in the 2006 PISA surveys. The column immediately to the right of each S.D. column reports its bootstrapped standard error. Among the countries with higher inequality in math scores are Western European countries such as Austria, Belgium, France, Germany, and Italy; East European ones such as Czech Republic and Bulgaria, Latin American countries such as Argentina and Uruguay, but also Israel and Taiwan, China. Among the ones with lower inequality in achievements are other European countries such as Croatia, Denmark, Estonia, Finland, Ireland, and Latvia, but also Asian countries such as Indonesia, Thailand and Jordan. Countries such as the UK, Japan, and the United States take intermediate rankings.¹³

Sample Selection Issues

Although we have established that the country ranking that can be derived from Table 1 is ordinally equivalent to the pre-standardization ranking, the issue of PISA sample selection remains a potential problem. As noted in Section 2, coverage rates range from a low of 0.47 in Turkey, to 1.02 in Switzerland.¹⁴ Selection would not be a problem if one were interested *exclusively* in the performance of 15 year-olds that are in school, and within a reasonable range of their expected grade of attendance. But this is likely to be an excessively narrow prism through which to assess a country's educational system and—even more so—to make international comparisons. Consider the example of two hypothetical “educational strategies,” illustrated by countries A and B, which have identical distributions of school and family characteristics, as well as of underlying ability in the population of 15 year-olds. Country A seeks to be inclusive, and allocates resources towards retaining as many students as possible in school, and towards promoting learning by those with the lowest demonstrated achievement. Country B, on the other hand, actively discourages enrollment by those with lower ability, and seeks to retain only the top half of performers in school by age 15. Looking only at the test scores for the samples of enrolled fifteen year-olds will naturally suggest that Country B has both a higher mean and a lower variance than country A, and thus a superior educational system altogether.

13. The inequality measures obtained for Azerbaijan seem particularly small and place the country as an outlier in all the analyses. It is unclear how much of this is due to the data collection procedures in this country, but such a different pattern is not likely due to real differences only.

14. One presumes that coverage rates in excess of 1.00 must be due either to statistical discrepancies in the estimates of 15 year-olds in the total population, or to errors of inclusion in the sample of test-takers.

This is not to suggest, of course, that Brazil, Indonesia, Mexico, Turkey, or any of the other countries with low coverage rates in Table 1 actively pursue an exclusionary strategy like that of hypothetical country B. But dropping out and lagging behind are, nevertheless, extremely likely to be selective processes, in the sense that they are correlated with family and student characteristics that also affect test scores. If one is interested in comparing the educational achievement of the *population* of fifteen year-olds across countries, therefore, the PISA samples suffer from selection bias.

Correcting for such biases is never simple, and even less so when non-participants are not observed at all in the sample (unlike, say, when seeking to correct for labor force participation on the basis of surveys that contain information on both earners and non-participants). While we do not offer a sample selection bias correction procedure for all countries in the PISA sample in this paper, we propose a simple two-sample non-parametric mechanism for assessing the sensitivity of our inequality measures to alternative assumptions about the sample selection process.

Denote the (density of the) distribution of test scores y in a particular country j by $f_j(y)$. Consider a vector of covariates X that is observed both in the PISA sample and in an ancillary household survey, which is representative of the full population of 15 year-olds. Note that the density of test scores in the PISA sample can be written as:

$$f_j(y) = \int \int \int \Phi_j(y, X) dX = \int \int \int g_j(y|X) \phi_j(X) dX. \quad (7)$$

In (7), Φ denotes the joint distribution of y and X , g denotes the conditional distribution of y on X , and ϕ denotes the joint density of the covariates in the vector X .¹⁵ If the joint density of the observable covariates X in a particular survey for country j is written $\phi_j(X|s = survey)$, then our first proposed estimate for a test-score distribution (density) corrected for sample selection on observables is given by:

$$f_j^{SO}(y) = \int \int \int g_j(y|X) \psi_j(X) \phi_j(X) dX \quad (8)$$

where

$$\psi_j(X) = \frac{\phi_j(X|s = HH)}{\phi_j(X|s = PISA)}. \quad (9)$$

15. The triple integral notation is short-hand for integrating out every element of X , so that there are as many integrals as there are elements in the vector of covariates common to both surveys. As it happens, in our application that dimension is three.

Equation (9) is simply the ratio of the density of fifteen year-olds whose observed characteristics X take certain values, in the ancillary household survey (HH), to the density of fifteen year-olds with the exact same observed characteristics in the PISA survey. $\psi_j(X)$ is a re-weighting function exactly analogous to that used by DiNardo, Fortin and Lemieux (1996) to construct counterfactual income densities in their study of inequality in the US. Whereas DiNardo et al. use the ratio of densities across different years (of the same survey), we use the ratio of densities across different surveys (for the same year). To the extent that test-taking (i.e. being in the PISA sample) is correlated with observed covariates in X , the counterfactual distribution in (8) should correct for the corresponding selection bias.¹⁶ In practice, this procedure was implemented by partitioning both the PISA and the ancillary household survey into cells with identical values for three observed covariates: gender, mother's education, and father's occupation, with the latter two variables classified as in Section 2.¹⁷ The ratios of densities in each cell in these partitions were used to construct the reweighting function (Equation 9), and both the S.D. and the IOP measures were computed over the counterfactual density of scores given by (8).

This procedure assumes that selection into the PISA sample is fully explained by observable variables, such as gender and family background. While such variables are likely to play a role in selection, it is also likely that other, unobserved variables do too. Within the set of girls with mothers with no formal education and fathers who work in agriculture, for example, it is possible that a higher proportion of high-ability students than low-ability students stay in school long enough to enter the PISA sample. This kind of selection would imply that equation (8) may overstate the achievement of those students who are counterfactually "brought back into" the sample: Simple re-weighting effectively assigns all those out-of-sample students the same scores obtained by students similar to them (in terms of the variables in X). If they are, in fact, likely to perform somewhat less well because of unobserved differences, the procedure overstates their true performance.

By its very nature, of course, selection on unobservables is harder to account for. The ancillary household surveys used to construct the reweighting function do not contain information on test scores. To provide another sensitivity test for the possible magnitude of sample selection bias driven by unobservables, we consider the (rather extreme) assumption that all those students who are counterfactually "re-introduced" into the PISA sample by the above procedure—a proportion given by $\psi_j(X) - 1$, for each X —do no better than those who are actually in the sample. In practice, we ascribe to them the lowest observed score for their cell in the partition.

16. The superscript SO stands for selection on observables.

17. Surveys were thus partitioned into 24 cells. Given the sample sizes reported earlier, particularly for Turkey's HBS and, to a lesser extent, Mexico's ENIGH, it was not possible to further refine the partition by using additional covariates.

In order to provide a sense of how sensitive our estimates of educational inequality (reported in Table 1) might be to sample selection, Table 3 reports the results of both of the above scenarios for the four countries with the lowest PISA coverage ratios in Table 1. To economize on space, Table 3 reports the effects of these ‘selection correction’ procedures both on the standard deviation of test scores and on our measure of inequality of educational opportunity, which is introduced in the next section. The first three columns report these measures (and standard errors) for the uncorrected, original PISA sample, for reading, math and science respectively. The next three report estimates for the correction that assumes selection on observables only (equation 8), and the final three for the correction that assumes selection on unobservables (with no common support).

The results in Table 3 provide a mixed message. Somewhat surprisingly, both inequality of achievement (measured by the standard deviation) and inequality of opportunity seem to be quite robust to selection on observables, despite very low coverage rates (of approximately 50% in these four countries). While this is encouraging, the same cannot be said for the estimates for selection on unobservables. Under these (admittedly extreme) assumptions, inequality in achievement increases by between 44% in Turkey and 92% in Mexico. Inequality of educational opportunity also rises in all countries, except Mexico.

It is possible to interpret these results as comforting, if one chooses to focus on the relative robustness of the measures to selection on observables, even in countries where PISA coverage is lowest. It seems likely that, if these observed variables account for most of the sample selection process, the estimates of educational inequality in Table 1 are robust for all countries. The fact that those estimates are sensitive to selection on unobservables can be minimized by the strength of the “no common support” assumption that assigns the very lowest grade in each cell to all those students counterfactually added to the sample.

Yet, it would probably be wiser to interpret the results from Table 3 as providing grounds for caution. We simply do not know how much selection into the PISA sample takes place on the basis of variables other than gender, mother’s education and father’s occupation. Until more is known about the composition of the group of fifteen year-olds that is excluded from the PISA sample, the possibility remains that inequality in countries with low coverage is underestimated. Investigation of that group of teenagers would seem like an important—but so far neglected—area of study for those interested in the distribution of educational achievement, particularly in developing countries.

III. A MEASURE OF INEQUALITY OF EDUCATIONAL OPPORTUNITY

At least as important as the total level of inequality in educational achievement is the question of how much of that inequality is explained by pre-determined circumstances, which individuals simply inherit, rather than controlling. While many may find some inequality in achievement—that might reflect differences in effort, or perhaps even differences in innate ability—quite acceptable, it is

TABLE 3. Inequality of Achievement and Opportunity in Low-Coverage Countries: Sensitivity to Different Assumptions on Selection into the PISA Sample

	PISA population without any correction			Correction assuming selection on observables			Correction assuming strong selection on unobservables		
	Reading	Math	Science	Reading	Math	Science	Reading	Math	Science
TURKEY									
Inequality (SD)	92.90 2.75	93.24 4.32	83.20 3.14	98.38	91.43	82.58	155.67	134.04	121.61
IOP	0.251 0.026	0.241 0.033	0.249 0.032	0.250	0.236	0.250	0.327	0.320	0.326
BRAZIL									
Inequality (SD)	102.46 3.34	92.02 2.65	89.28 1.93	102.86	90.44	86.75	179.82	146.68	146.17
IOP	0.268 0.020	0.318 0.005	0.286 0.021	0.265	0.309	0.262	0.404	0.404	0.385
MEXICO									
Inequality (SD)	95.68 2.27	85.27 2.16	80.70 1.47	95.63	85.02	79.18	196.85	162.79	136.99
IOP	0.278 0.024	0.261 0.002	0.271 0.024	0.267	0.242	0.255	0.256	0.250	0.228
INDONESIA									
Inequality (SD)	74.79 2.39	80.01 3.18	70.06 3.26	71.03	76.27	65.74	130.56	135.89	112.79
IOP	0.250 0.038	0.237 0.042	0.220 0.045	0.218	0.200	0.181	0.274	0.261	0.261

Notes: IOP denotes the measure of inequality of educational opportunity, defined in equation (13). It is the share of the total variance in test scores which is accounted for by the student's pre-determined circumstance variables. Source: Authors' analysis based on data described in the text.

common to come across arguments against unequal opportunities among students. These are differences in achievement that do not reflect the choices or actions of today's students, but only inherited circumstances beyond their control. That such inequalities are morally objectionable is today a dominant view among social justice theorists. See, for example, Cohen (1989), Dworkin (1981), Roemer (1998) and Fleurbaey (2008) for some of the classic references. There is also a positive argument against the inheritance of educational inequality, namely that if scarce opportunities for educational investment are allocated on some basis other than talent – such as inherited wealth, for example – this will lead to an inefficient allocation of resources.¹⁸

The applied literature on the measurement of inequality of opportunity has focused primarily on opportunities for the acquisition of income, but there is

18. See, e.g. Fernández and Galí (1999).

no reason it cannot be adapted to the space of educational achievement.¹⁹ Two main approaches characterize that empirical literature. Both approaches begin by seeking agreement on a set of individual characteristics which are beyond the individual's control, and for which he or she cannot be held responsible. These variables are known as 'circumstances'. Once a vector C of circumstances has been agreed upon, society can be partitioned into groups with identical circumstances. Formally, such a partition is given by a set of *types*: $\Pi = \{T_1, T_2, \dots, T_K\}$, such that $T_1 \cup T_2 \cup \dots \cup T_K = \{1, \dots, N\}$, $T_l \cap T_k = \emptyset, \forall l, k$, and the vectors $C_i = C_j, \forall i, j | i \in T_k, j \in T_k, \forall k$.

Given such a partition, the two approaches differ in how they define the benchmark of equality of opportunity. In the *ex-ante* approach, associated with [van de Gaer \(1993\)](#), the opportunity set faced by each type is evaluated, and equality of opportunity is attained when there is perfect equality in those values across all types. In practice, researchers have often used the mean income (or achievement) of the type as an estimate of the value of the opportunity set they face. Since equality of opportunity would imply equality in means across types, inequality of opportunity is then naturally seen as some measure of between-type inequality.

In the *ex-post approach*, associated with [Roemer \(1998\)](#), equality of opportunity obtains only when individuals exerting the same degree of effort, regardless of their circumstances, receive the same reward. Under certain assumptions, this amounts to requiring equality in the full conditional outcome distributions across all types. Inequality of opportunity would, in this case, best be captured by the (appropriately weighted) sum of inequality within groups characterized by the same degree of effort.²⁰ The two approaches are closely related but, for any society with a given joint distribution of achievement and circumstance variables, they yield different answers to the question "How much inequality of opportunity is there?" See [Fleurbaey and Peragine \(2012\)](#) for a formal discussion of the relationship between the two approaches.

In what follows, we adapt the *ex-ante* approach employed by [Ferreira and Gignoux \(2011a\)](#) to the distributions of test scores described earlier.²¹ These authors propose to measure inequality of opportunity (IOp) by between-type inequality. Specifically:

$$\theta_{IOp} = \frac{I(\{\mu_i^k\})}{I(y)} \quad (10)$$

19. Indeed [Cecchi and Peragine \(2005\)](#), the working paper version of their 2010 paper, do apply the concept to educational achievement measures. See also [Gamboa and Waltenberg \(2012\)](#) for a more recent treatment.

20. Under the standard Roemerian assumptions, these groups are [Cecchi and Peragine's \(2010\)](#) 'tranches'.

21. [Ferreira and Gignoux \(2011a\)](#), in turn, build on [Bourguignon et al. \(2007\)](#) and [Cecchi and Peragine \(2010\)](#).

where $\{\mu_i^k\}$ is the smoothed distribution corresponding to the distribution y and the partition Π .²²

Naturally, θ_{IOP} can be computed non-parametrically by means of a standard between-group inequality decomposition (provided the chosen inequality index $I()$ is properly decomposable). However, this procedure is data-intensive when the vector C is large. As the partition becomes finer, cells become small and sparsely populated, and the precision of the estimates of cell means declines, giving rise to an upwards bias in the estimation of θ_{IOP} . Following Bourguignon et al. (2007), Ferreira and Gignoux (2011a) then propose a parametric alternative for θ_{IOP} , based on an OLS regression of y on C :

$$\hat{\theta}_{IOP} = \frac{I(C'_i \hat{\beta})}{I(y)}. \tag{11}$$

$\hat{\beta}$ in (11) is the OLS estimate of the regression coefficients in a simple regression of y on C :

$$y_i = C'_i \beta + \eta_i. \tag{12}$$

In (11), $C'_i \hat{\beta}$ denotes the vector of predicted test scores from regression (12). Under the maintained assumption of a linear relationship between achievement and circumstances, this vector is equivalent to the smoothed distribution, since all individuals with identical circumstances are assigned their conditional mean scores.

Because of its unique path-independent decomposability properties, Checchi and Peragine (2010) and Ferreira and Gignoux (2011a) both use the mean logarithmic deviation as the inequality index $I()$. However, as shown above, the mean log deviation is not ordinally invariant in the standardization to which test scores are submitted, and it is therefore unsuitable for use in the present context. Following the discussion in Section 3, we use the simple variance as our inequality index $I()$. This choice yields our proposed measure of inequality of educational opportunity, as a special case of (11):

$$\hat{\theta}_{IOP} = \frac{Var(C'_i \hat{\beta})}{Var(y_i)}. \tag{13}$$

This index has a number of attractive features. First, it is extremely simple to calculate: It is simply the R^2 of an OLS regression of the child’s test score on a vector C of individual circumstances. In our application to the PISA data sets, C includes the following ten variables: gender, father’s and mother’s education, father’s occupation, language spoken at home, migration status, access to

22. A smoothed distribution is obtained from a vector y and a partition Π by replacing each element of y in a given cell T_k with the mean value of y in its cell, μ^k . See Foster and Shneyerov (2000).

books at home, durables owned by the households, cultural items owned, and the location of the school attended.

Second, despite its simplicity, it is a very meaningful summary statistic. It is a parametric approximation to the *lower bound* on the share of overall inequality in educational achievement that is explained by pre-determined circumstances. A formal proof is provided by Ferreira and Gignoux (2011). But the basic intuition is to note that (12) can be seen as the reduced form of a (linearized version of a) model such as:

$$y = f(C, E, u) \quad (14)$$

$$E = g(C, v). \quad (15)$$

In (14) and (15), y denotes achievement, and C denotes the vector of circumstances, as before. E denotes a vector of efforts: all variables that affect achievement and over which individuals do have some measure of control. u and v denote random shocks or innovations. Because 15 year-olds may conceivably affect the choice of school they attend, the class they are assigned to, and thus the teachers they interact with, all school characteristic variables, for example, are included in E . So are any direct measures of the student's own efforts in preparing for exams, for instance. Of course, efforts E can be influenced by circumstances C , but the reverse cannot happen. Variables can only be treated as circumstances if they are pre-determined and entirely exogenous to the individual.

Now return to (12) as a linearized reduced form of (14)-(15). We know that circumstances C are economically exogenous to y . We also know that all effort (E) variables (whether or not one could observe them in the data) are omitted deliberately: β is intended to capture the reduced-form effect of circumstances – both directly and through efforts. Since all relevant factors are classified into either circumstances or efforts, the only sources of bias to the estimates of β are omitted, unobserved circumstance variables. Although the observed vector C is economically exogenous, it may not be exogenous in the (econometric) sense that its components may be correlated with other (unobserved and thus omitted) circumstance variables. Individual elements of the vector $\hat{\beta}$ suffer from these omitted variable biases, and *cannot* be interpreted as causal estimates of the individual impact of a particular circumstance on test scores.

If one is interested, however, on the total *joint effect of all circumstances* on achievement and, more specifically, on the share of variation in y that is causally explained by the *overall* effect of circumstances (operating both directly and through efforts), then the R^2 of (12) – our $\hat{\theta}_{IOp}$ – yields a valid lower bound for the object of interest. By construction, the only missing variables in (12) are other circumstances. If any were added, $\hat{\theta}_{IOp}$ might rise, but it cannot fall. While individual coefficients in $\hat{\beta}$ may be biased, $\hat{\theta}_{IOp}$ is a lower bound estimate of the joint causal effect of all circumstances on achievement, and thus an appropriate

measure of inequality of opportunity. A formal proof is provided by [Ferreira and Gignoux \(2011a\)](#), for the perfectly analogous case of incomes.²³

A third attractive feature of (13) is that it allows for the use of more information on circumstances than previous studies, which typically rely on a smaller set of background variables, and thus capture a more limited share of heterogeneity in family resources. [Schultz, Ursprung and Wossmann \(2008\)](#), for example, focus on the number of books at home. [Macdonald et al. \(2010\)](#) look at the effect of gender and an index of household wealth but ignore, for example, information on parental education and occupation. [Gamboa and Waltenberg \(2012\)](#) see inequality of opportunity as determined by gender, parental education, and school type (public or private), which they treat as a circumstance. We consider the joint effect of all of these circumstances, and more.

A fourth attractive feature of $\hat{\theta}_{IOP}$ as a measure of inequality of educational opportunity is that, unlike any measure of the level of inequality (see Remark 1 above), it is a parametric estimator of a ratio (equation 10) that is cardinally invariant in the standardization of test scores. To see this, note that any subgroup mean is affected by standardization in a manner analogous to equation (3), so that:

$$Var\{\mu_i^k(y)\} = \left(\frac{\hat{\sigma}}{\sigma}\right)^2 Var\{\mu_i^k(x)\}. \tag{16}$$

Given (16) and equation (6), it follows that $\theta_{IOP} = Var\left(\left\{\mu_i^k(y)\right\}\right)/Var(y) = Var\left(\left\{\mu_i^k(x)\right\}\right)/Var(x)$.

A fifth attractive feature of this IOP measure is that it is neatly decomposable into components for each individual variable in the vector C. Equation (13) can be rewritten as:

$$\hat{\theta}_{IOP} = (var\ y)^{-1} \left[\sum_j \beta_j^2 var\ C_j + \frac{1}{2} \sum_k \sum_j \beta_k \beta_j cov(C_k, C_j) \right]. \tag{17}$$

This in turn can be written as the sum over all elements (denoted by j) of the C vector:

$$\hat{\theta}_{IOP} = \sum_j \hat{\theta}^j = \sum_j (var\ y)^{-1} \left[\beta_j^2 var\ C_j + \frac{1}{2} \sum_k \beta_k \beta_j cov(C_k, C_j) \right]. \tag{18}$$

23. Note, however, the implication that the cross-country comparisons reported in this paper are comparisons of that lower-bound measure. If additional circumstance variables were observed across all of these countries, those rankings might change.

This decomposition is an example of a Shapley-Shorrocks decomposition: it corresponds to the average between two alternative paths for estimating the contribution of a particular circumstance C_j to the overall variance. In the first (direct) path, all C_j , $j \neq J$ are held constant. In the second (residual) path, C_J is itself held constant, and its contribution is taken as the difference between the total variance and the ensuing variance. Either path is conceptually valid, and the Shapley-Shorrocks averaging procedure yields (18) as the path-independent additive decomposition.²⁴

Finally, $\hat{\theta}_{Iop}$ can be seen as isomorphic to a measure of intergenerational persistence of inequality, itself the converse of a measure of educational mobility.²⁵ In the canonical Galton regression of a child's outcome (y_{it}) on the parent's outcome ($y_{i,t-1}$):

$$y_{it} = \beta y_{i,t-1} + \varepsilon_{it} \quad (19)$$

the coefficient β is sometimes used as measure of persistence, and $1-\beta$ as a measure of mobility. An alternative that gives equal weight to the variance in both father's and son's distributions is the R^2 of (19) which is, of course, also the square of the correlation coefficient between the two outcomes in the population. If one were to replace the parent's outcome $y_{i,t-1}$ with a vector of parental or family background variables, (19) would transform into something very close to (12), and the R^2 measure of immobility into our measure of inequality of opportunity, $\hat{\theta}_{Iop}$. Indeed, the only pre-determined circumstance among the ten variables previously listed which is not a family background variable is the child's own gender. Apart from the child's own gender, one could see $\hat{\theta}_{Iop}$ as a measure of intergenerational persistence, or immobility, in which the missing value for the parent's own test scores, $y_{i,t-1}$, is replaced with a proxy vector of family background circumstances, C_i .

Having separately regressed test scores for each subject (in each country) on the vector C (equation 12), and computed the R^2 of each regression to obtain $\hat{\theta}_{Iop}$, we report them on Table 4. These are our estimates of the inequality of educational opportunity (IOp) given by equation (13). They range between 0 and 1, and can be interpreted straight-forwardly as a lower-bound on the share of the total variance in educational achievement that is accounted for by

24. See Shorrocks (1999) for the original application of the Shapley value to distributional decompositions. Ferreira et al. (2011) provide a formal proof that (18) is the Shapley-Shorrocks decomposition of the variance into the effects of individual circumstances.

25. Mobility is a multifaceted concept, and there are many distinct measures of it, often attempting to capture different aspects of "movement" across distributions. See Fields and Ok (1996) for a discussion. In the present context, we adopt a view of mobility as time- or origin-independence. See also Shorrocks (1978). Persistence would therefore correspond to the concept of origin-dependence, which is closely related to the notions of inequality of opportunity in both van de Gaer (1993) and Roemer (1998).

TABLE 4. Inequality of Educational Opportunity for Three PISA Subjects

	IOp Reading	Standard Error (Reading IOp)	IOp Mathematics	Standard Error (Math IOp)	IOp Science	Standard Error (Science IOp)
<i>Asia & North Africa</i>						
Azerbaijan	0.173	0.028	0.044	0.012	0.112	0.024
Hong Kong SAR, China	0.177	0.016	0.154	0.016	0.166	0.018
Indonesia	0.250	0.038	0.237	0.042	0.220	0.045
Israel	0.197	0.018	0.206	0.019	0.195	0.016
Japan	0.206	0.017	0.203	0.020	0.189	0.016
Jordan	0.346	0.024	0.272	0.024	0.271	0.019
Korea	0.214	0.022	0.209	0.021	0.173	0.019
Kyrgyzstan	0.314	0.023	0.306	0.027	0.269	0.023
Macao-China	0.127	0.012	0.102	0.009	0.111	0.008
Qatar	0.309	0.010	0.254	0.009	0.264	0.009
Russian Federation	0.238	0.021	0.165	0.020	0.183	0.020
Chinese Taipei	0.300	0.017	0.275	0.022	0.281	0.019
Thailand	0.325	0.023	0.230	0.021	0.265	0.022
Tunisia	0.215	0.024	0.273	0.031	0.191	0.026
Turkey	0.251	0.026	0.241	0.033	0.249	0.032
<i>Latin America</i>						
Argentina	0.289	0.024	0.315	0.007	0.312	0.026
Brazil	0.268	0.020	0.318	0.005	0.286	0.021
Chile	0.248	0.022	0.330	0.001	0.299	0.021
Colombia	0.181	0.018	0.216	0.007	0.193	0.018
Mexico	0.278	0.024	0.261	0.002	0.271	0.024
Uruguay	0.221	0.015	0.245	0.004	0.248	0.012
Australia	0.199	0.010	0.153	0.009	0.164	0.009
Canada	0.242	0.011	0.211	0.011	0.207	0.010
New Zealand	0.276	0.013	0.241	0.012	0.269	0.013
United States			0.279	0.020	0.282	0.019
<i>Eastern Europe</i>						
Bulgaria	0.377	0.028	0.331	0.030	0.364	0.030
Czech Republic	0.296	0.021	0.268	0.019	0.279	0.020
Estonia	0.271	0.013	0.206	0.013	0.208	0.012
Croatia	0.297	0.017	0.222	0.015	0.239	0.014
Hungary	0.345	0.023	0.326	0.022	0.326	0.019
Lithuania	0.318	0.017	0.279	0.017	0.262	0.016
Latvia	0.254	0.017	0.201	0.020	0.187	0.016
Montenegro	0.252	0.013	0.223	0.012	0.197	0.011
Poland	0.275	0.014	0.241	0.013	0.241	0.014
Romania	0.301	0.026	0.313	0.028	0.310	0.027
Serbia	0.311	0.018	0.276	0.017	0.255	0.016
Slovak Republic	0.292	0.026	0.317	0.030	0.297	0.024
Slovenia	0.336	0.018	0.263	0.016	0.268	0.014
<i>Western Europe</i>						
Austria	0.296	0.019	0.300	0.020	0.324	0.022
Belgium	0.335	0.015	0.329	0.018	0.338	0.015
Switzerland	0.313	0.013	0.282	0.013	0.322	0.012
Germany	0.368	0.021	0.351	0.018	0.352	0.019

(continued)

TABLE 4. Continued

	IOp Reading	Standard Error (Reading IOp)	IOp Mathematics	Standard Error (Math IOp)	IOp Science	Standard Error (Science IOp)
Denmark	0.229	0.015	0.219	0.014	0.249	0.017
Spain	0.243	0.013	0.239	0.012	0.258	0.013
Finland	0.247	0.014	0.179	0.010	0.167	0.011
France	0.305	0.019	0.335	0.019	0.345	0.018
United Kingdom	0.274	0.014	0.258	0.012	0.275	0.012
Greece	0.261	0.023	0.228	0.022	0.245	0.019
Ireland	0.259	0.018	0.235	0.017	0.240	0.016
Iceland	0.234	0.009	0.167	0.009	0.184	0.009
Italy	0.207	0.015	0.178	0.014	0.206	0.014
Liechtenstein	0.388	0.031	0.323	0.034	0.379	0.030
Luxembourg	0.344	0.008	0.291	0.008	0.328	0.009
Netherlands	0.247	0.022	0.271	0.023	0.283	0.023
Norway	0.271	0.016	0.195	0.014	0.220	0.018
Portugal	0.303	0.021	0.274	0.019	0.267	0.020
Sweden	0.265	0.014	0.233	0.012	0.250	0.013

Notes: IOp denotes the measure of inequality of educational opportunity, defined in equation (13). It is the share of the total variance in test scores which is accounted for by the student's pre-determined circumstance variables.

Source: Authors' analysis based on data from PISA 2006.

pre-determined circumstances (gender and family background) in each country. Bootstrapped standard errors are reported next to each IOp measure. The IOp estimates range between 12.7% and 38.8% of the total variance of test scores in reading; between 4.4% (10.2% excluding the outlier Azerbaijan) and 35.1% of the variance of test scores in math; and between 11.1% and 37.9% in Science.²⁶

No clear regional pattern emerges from the estimates presented in Table 4. Among the countries with the highest levels of inequality of opportunity, with shares above 30%, are Western European countries (such as Belgium, France, and Germany) but also Eastern European countries (such as Bulgaria and Hungary), and Latin American countries (such as Argentina, Brazil and Chile). Among the countries with the lowest IOp, with shares below 20%, are Asian countries (such as Azerbaijan, Macao (China), and Hong Kong SAR, China), Nordic countries (such as Finland, Iceland, and Norway), Russia, Australia and Italy. The United States, the UK, and Spain lie in an intermediate range, with shares close to 25%.

One can use these results to make specific comparisons. For example, the degree of inequality of educational opportunity seems to be significantly higher in a few large European countries, such as France and Germany, than in the

26. If one were interpreting these shares as proxies for the persistence measure given by the R^2 of (19), one should note that the numbers correspond to *squares* of the correlation coefficient. The square root of IOp for mathematics scores, for example, ranges from 0.21 to 0.59.

United States. However these inequalities are significantly lower in Nordic countries, such as Finland and Norway, or in Japan and Korea. Regarding developing economies, countries in Latin America tend to rank in the upper half of the distribution, while Asian countries, such as Indonesia and Thailand, rank in the lower half. Although the estimates are very imprecise for Indonesia, Thailand exhibits significantly lower inequalities than Latin American countries such as Brazil. The results for reading and science are not discussed in detail here, but IOP measures for the three subjects are highly correlated: The Spearman rank correlation coefficients for shares in Reading, Math and Science range from 0.75 to 0.92.

The absence of a clear geographical pattern in the cross-country distribution of inequality of educational opportunity is mirrored in the absence of a correlation between IOP and either the level of educational achievement, as measured by mean test scores, or the level of economic development, as measured by GDP per capita.²⁷ Simple regressions of IOP in mathematics against both mean achievement in mathematics and GDP per capita yield insignificant coefficients. Scatter plots and some additional robustness analysis are presented in the working paper version of this article, [Ferreira and Gignoux \(2011b\)](#).

[Ferreira and Gignoux \(2011b\)](#) also report the exact decomposition of inequality of opportunity into partial shares by individual circumstance, described in equation (18). These partial shares are functions of individual regression coefficients from (12) which, as noted earlier, are likely to be biased. The partial shares reported in our working paper version should therefore *not* be interpreted causally in any way. They are useful only as a description of the variables underpinning the overall (lower-bound) measure of inequality of opportunity.

IV. A DESCRIPTIVE APPLICATION: CORRELATIONS BETWEEN IOP AND EDUCATION POLICIES

As an illustration of potential applications, we now briefly investigate the cross-country correlation between the measure of inequality of educational opportunity presented in the previous section and two specific educational policy variables: the distribution of public spending across different levels of the education system, and the extent of early tracking of pupils between general and vocational schools or classes.

The incidence of public spending in education and the allocation of financial resources among the different segments of the education system have been examined by various studies (e.g. [Birdsall 1996](#); [Castro-Leal et al. 1999](#); and [Van de Walle and Nead 1995](#)). Given that children with disadvantaged backgrounds tend to drop out from school earlier than others, the allocation of

27. GDP per capita is measured at purchasing power parity exchange rates, in 2006 US prices; the data are from the World Development Indicators (WDI) database.

resources to the primary level of schooling is generally thought more likely to be progressive.

The impacts of tracking policies on the efficiency and equity of educational systems are another example of education policies that have received considerable attention in recent studies (e.g. Ariga et al. 2006; Bertocchi and Spagat 2004; Brunello and Checchi 2007; Brunello et al. 2006; Hanushek and Woessman 2006; Jabukowski et al. 2010; Manning and Pischke 2006; Pekkarinen et al. 2009). Theory does not provide clear-cut predictions for the effect of early tracking on educational achievements. On the one hand homogenous classrooms, and the associated specialization of teaching and curricula to the needs and abilities of specific students, could lead to efficiency gains. But on the other hand, disadvantaged groups might be harmed by unfavorable allocations of resources, including less well endowed schools, teacher sorting, peer effects, or differences in curricula²⁸. Moreover, since much of the early inequality in achievement—and thus the track placements themselves—are driven by differences in parental resources, a frequent concern has been that tracking might reinforce the effects of family background on educational achievements. I.e. that it might reduce inter-generational mobility, and exacerbate inequality of educational opportunity.

We briefly examine the correlation between our measure of IOp and these two policies, using data on the policy indicators from the UNESCO Institute for Statistics (UIS).²⁹ Our indicator of the distribution of educational expenditures is the share of spending in primary schools—defined as the first ISCED level, corresponding to grades 1 to 6—in total public educational expenditure. The indicator of tracking is the share of technical or vocational enrollment at the secondary level (including lower and upper secondary or the second and third ISCED levels, usually corresponding to grades 7 to 12) in total enrollment at that level.³⁰ The information on the distribution of education expenditure across levels is missing for six countries (Canada, Montenegro, Qatar, Russia, Serbia and Taiwan, China) and the information on the share of technical and vocational enrollment at the secondary level is missing in five countries (Latvia; Montenegro; Serbia; Taiwan, China; and the United States). Two other countries are excluded from the analysis: Liechtenstein and Luxembourg. The number of observations for Liechtenstein (339 examinees) makes the estimates of learning inequalities unreliable and Luxemburg is too much of an outlier in

28. Early tracking may also be costly in terms of the misallocation of students to tracks, and in terms of forgone versatility in the production of skills (Brunello and Checchi, 2007).

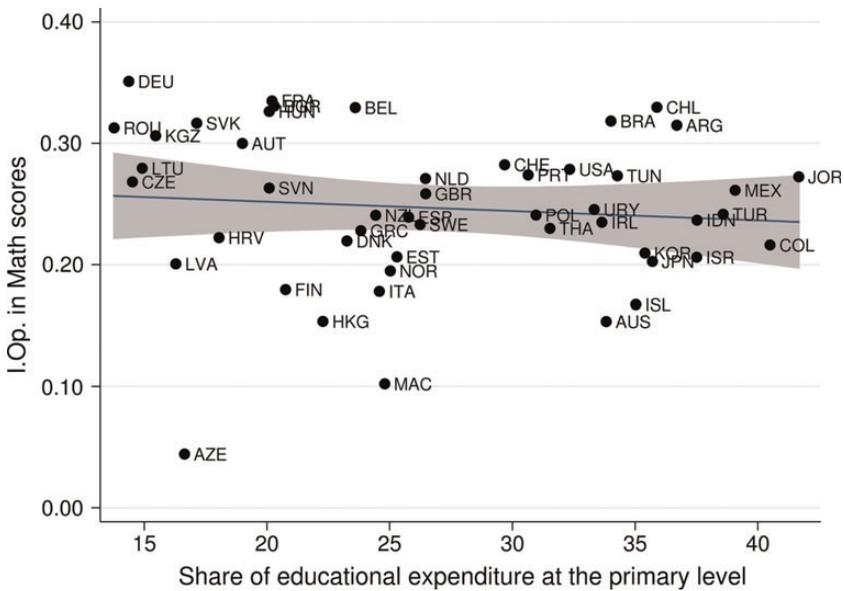
29. The data for 2006 correspond to the school year 2005-06 for countries where the school year laps over two calendar years.

30. As rightly noted by an anonymous referee, a preferable measure of tracking would focus exclusively on the proportion of students in vocational or technical tracks in lower (as opposed to both lower and upper) secondary education. Unfortunately, information on tracking at that more specific level is considerably scarcer: Of the 57 countries in PISA 2006 with information on enrollment at the lower secondary level in the UIS policy indicators data, 36 do not report enrollment in vocational or technical streams at that level.

terms of GDP per capita in 2006 (at about 69.000 US dollars, with the US in second place at 44.000 US dollars).

There is considerable variation in the share of expenditures allocated to the primary level of education in the remaining country sample. While the mean share is 27.0%, the lowest share is observed in Romania at 13.8% and the highest in Jordan at 41.7% (the first quartile is at 20.2% and the third quartile at 34.0%). Figure 1 provides an illustration of the relationship between the primary share of expenditures and IOp. The regression line and a 95% confidence interval for the mean are shown. Table 5 gives the tests of significance of

FIGURE 1. Inequality of Educational Opportunity and Public Expenditure at the Primary Level



Source: Authors' analysis based on data from PISA 2006.

TABLE 5. Coefficients on the Primary Share of Public Education Expenditure in Regressions of IOp on that Variable; with and without Controls

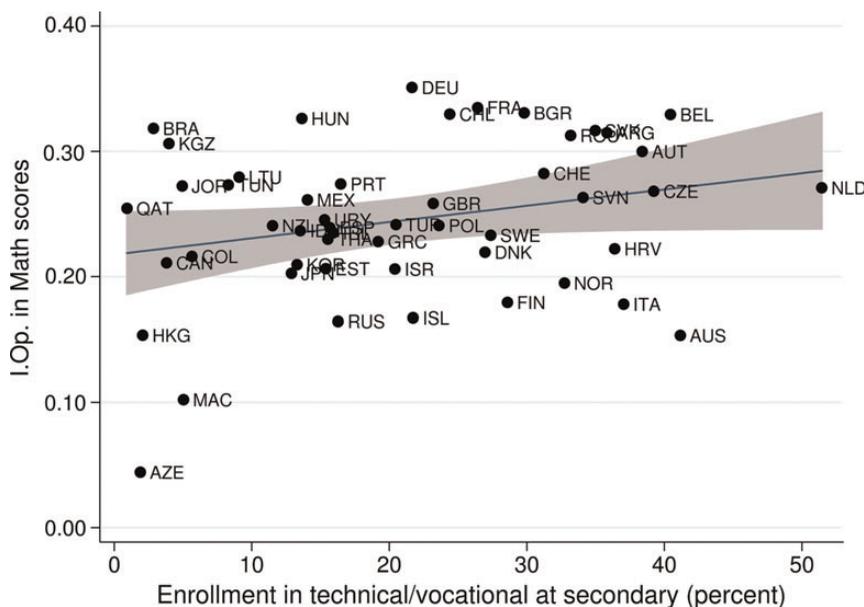
	Reading	Math	Science
No controls			
All countries	-0.00217*** (0.00092)	-0.00077 (0.00112)	-0.00152 (0.00105)
Excluding outliers	-0.00300*** (0.00078)	-0.00113 (0.00101)	-0.00172* (0.00101)
Controlling for GDP and public expenditure in education per pupil			
All countries	-0.00197** (0.00087)	-0.00013 (0.00120)	-0.00103 (0.00113)
Excluding outliers	-0.00184*** (0.00072)	-0.00181* (0.00102)	-0.00185* (0.00108)

Notes: Regression coefficients of the share of public expenditure in education allocated to the primary level. Dependent variable: IOp in the subject at column header. Standard errors in parentheses. Where indicated, outliers are identified using the method proposed by Besley, Kuh and Welsch (1980). Data source: UNESCO Institute for Statistics database; ***/**/*: significant at 1/5/10%.

this relationship both without any controls (first panel) and controlling for per capita GDP and public education expenditure per pupil (second panel). Once outliers are excluded, significant negative correlations exist both for reading and science, with or without controls. For math, the negative correlation is only significant with controls. The coefficients lie between -0.001 and -0.003 , indicating that an increase of 10 points in the share of resources allocated to primary schooling is associated with decreases of 1 to 3 points in inequality of educational opportunity.

There is also considerable heterogeneity in tracking in our country sample. The mean share is 20.8 percent and values range from 0.9% in Qatar to 51.4% percent in the Netherlands (the first quartile is at 12.9 and the third at 31.2). Figure 2 provides a scatter plot of the relationship between tracking and IOp in this sample, while Table 6 lists coefficients and standard errors, both without any controls (upper panel) and controlling for per capita GDP and public education expenditure per pupil (bottom panel). There is a clear pattern of significant positive relationships across all three subjects and both regression specifications, with the statistical significance being stronger in the specification with controls. Higher inequality of opportunity tends to be associated with higher shares of technical and vocational enrollment. The regression coefficients lie between 0.001 and 0.002, indicating that an increase of 10 points of the share of technical or vocational enrollments is associated with an increase of 1 to 2 points in inequality of opportunity.

FIGURE 2. Inequality of Educational Opportunity and Tracking



Note: Tracking is measured as the share of enrollment in technical or vocational curricula at the secondary level.

Source: Authors' analysis based on data from PISA 2006.

TABLE 6. Coefficients on Tracking in Regressions of IOp on that Variable; with and without Controls

	Reading		Math		Science	
No controls						
All countries	0.00106*	(0.00059)	0.00130*	(0.00070)	0.00179***	(0.00063)
Excluding outliers	0.00158**	(0.00060)	0.00109*	(0.00062)	0.00160***	(0.00059)
Controlling for GDP and public expenditure in education per pupil						
All countries	0.00148***	(0.00057)	0.00173***	(0.00074)	0.00214***	(0.00068)
Excluding outliers	0.00090*	(0.00047)	0.00175***	(0.00065)	0.00205***	(0.00067)

Notes: Regression coefficients of tracking (measured as the share of technical and vocational enrollment at the secondary level). Dependent variable: IOp in the subject at column header. Standard errors in parentheses. Where indicated, outliers are identified using the method proposed by Besley, Kuh and Welsch (1980). *Source:* UNESCO Institute for Statistics database; ***/**/*: significant at 1/5/10%.

These correlations suggest that our measure of inequality of opportunity is negatively associated with the share of public spending on primary education, and positively associated with tracking into general or technical/vocational schooling at the secondary level. These associations allow for absolutely no inference of causality, of course, but the results seem in line with and extend those of studies devoted to these relationships. For instance, while Hanushek and Woessman (2006) find tracking to be associated with higher levels of overall inequality in test scores, our results suggest it also tends to come with higher levels of inequality of learning opportunities. The analysis here is descriptive and is only meant to illustrate the potential use of indicators of inequality of opportunity for future studies of the distributive impacts of education policies. Future extensions—notably involving the use of panel data—might allow for causal analysis of these relationships.

V. CONCLUSIONS

Internationally comparable information on learning outcomes, such as the standardized test scores collected by PISA surveys, represents a revolution in the quality of data available for research on education. It allows for potentially much greater insight into the determinants of educational achievement, and might therefore contribute to the design of policies that raise average learning levels, or that reduce educational disparities.

The measurement of educational disparities using this kind of data is not, however, a trivial extension of inequality measurement in years of schooling, or in other variables like income. This paper has highlighted two issues that require special attention in the measurement of inequality in educational achievement, and which appear to have been overlooked so far. The first is the standardization of test scores, to which all meaningful measures of inequality are cardinally sensitive. More importantly, many common measures of inequality, including the

Gini coefficient and the Theil indices, are not even ordinally invariant to standardization, invalidating country rankings that are based on them.

We show that the simple variance (or the standard deviation) of test scores is ordinally invariant to standardization, and present estimates for all 57 countries that took part in the 2006 round of PISA surveys, in all three subjects for which tests are carried out: reading, mathematics and science. There is considerable international variation in educational inequality thus measured. The standard deviation in Math scores ranges from around 80 in Indonesia, Estonia and Finland, to nearly 110 in Belgium and Israel.

The second measurement issue that may compromise international inequality comparisons based on PISA test scores is the possibility of sample selection. The surveys are designed to be representative of the population of 15 year-olds enrolled in school, and attending grades 7 or above. While this stipulation covers most of the population of that age group in OECD countries, it purposively excludes substantial numbers in poorer countries. Selection into the sample is clearly correlated with determinants of test scores, leading to a classic problem of sample selection bias. Using information on characteristics of fifteen year-olds included in other, ancillary household surveys, we use sample re-weighting methods to assess the implications of the selection bias for our measures of educational inequality in achievement and opportunity. Results for Brazil, Indonesia, Mexico and Turkey suggest that the inequality measures are relatively robust to selection on the basis of three observed variables (gender, mother's education and father's occupation). Under a more stringent scenario of strong selection on unobservables with no common support, however, the current measures of educational inequality in these countries would appear to be substantially underestimated.

Finally, we also propose and compute a measure of inequality in educational opportunity. The measure is simply the share of the total variance in achievement that can be accounted for by pre-determined circumstance variables in a linear regression. The index is simple and intuitive, and provides a lower-bound estimate of the joint causal effect of all pre-determined circumstances on educational inequality. It is cardinally invariant to the standardization of test scores, and closely related to the origin-independence concept of inter-generational educational mobility.

Thus measured, inequality of opportunity in our sample of countries ranges from approximately 0.10 – 0.16 in Macao (China), Australia, and Hong Kong SAR, China, to 0.33 – 0.35 in Bulgaria, France and Germany. Although the measure is uncorrelated with average educational achievement and with GDP per capita, it appears to be higher in Latin America and parts of continental Europe (including France, Germany and Belgium). It is lower in Asia, the Nordic countries, and Australia. It is negatively correlated with the share of public educational spending allocated to primary schooling, and positively correlated with the extent of educational tracking, defined as the share of technical and/or vocational enrollment in secondary schools.

This paper sought to place the measurement of educational inequality on a sounder footing, given the specific characteristics of data on educational achievement. Some of our findings, however, raise new questions that may motivate future work. First, although we confine our work to PISA surveys, a number of the issues we addressed are also relevant for the TIMSS, PIRLS and IALS surveys, all of which provide standardized achievement measures, and many of which also provide family background variables.³¹ Since those surveys have different country coverage, applying these measures to those surveys may shed light on educational inequality in other parts of the world.³²

A second issue where additional research is clearly necessary is that of sample selection. The results presented in our Table 3 suggest that while inequality measures appear to be robust to selection on a small set of observables, there is scope for very large selection biases on other variables (which we are treating as unobservable). This serves as a cautionary remark, but also as a call for additional work on selection, either by closer examination of the sub-sample of fifteen year-olds who are not enrolled in school, or by combining and comparing different achievement surveys with distinct sampling frames, such as IALS (which samples households) and PISA (which samples schools). This might be done for countries which participate in both surveys, such as Chile.³³

Finally, the measures of inequality of achievement and opportunity proposed here could be used as dependent variables in the evaluation of different educational policy interventions, such as teacher training or school governance reforms. This would require achievement surveys that contain the right variables, and are representative at much more finely disaggregated spatial units, likely in combination with interventions that were randomized at the level of those spatial units. But this is not as difficult to achieve as it may appear: It is close in spirit, in fact, to the approach applied by [van de Gaer et al. \(2012\)](#) to the impact of Mexico's *Oportunidades* program on the distribution of health opportunities. To the extent that the effect of policy interventions on inequality of educational opportunity (or achievement) is a question of policy interest, the measurement tools developed in this paper may be useful.

REFERENCES

- Ariga, K., G. Brunello, R. Iwahashi, and L. Rocco. 2006. "On the Efficiency Costs of De-Tracking Secondary Schools." IZA Discussion Paper No. 2534. Institute for the Study of Labor, Bonn, Germany.
- Baker, F.B. 2001. *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.

31. Although it is important to recognize that the sample selection issues in the TIMSS and PIRLS surveys, which define their samples by grade, rather than age, are somewhat different. See [Jakubowski and Pokropek \(2011\)](#) for an attempt to address those issues.

32. Indeed, see [Salehi-Isfahani and Belhaj Hassine \(2012\)](#) for an application of our approach to TIMSS data for countries in the Middle East and North Africa.

33. We are grateful to an anonymous referee for this suggestion.

- Bedard, K., and C. Ferrall. 2003. "Wage and Test Score Dispersion: Some International Evidence." *Economics of Education Review* 22: 31–43.
- Bertocchi, G., and M. Spagat. 2004. "The Evolution of Modern Educational Systems: Technical vs. General Education, Distributional Conflict, and Growth." *Journal of Development Economics* 73 (2):559–582.
- Besley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Colinearity*. New York: Wiley.
- Birdsall, N. 1996. "Public Spending on Higher Education in Developing Countries: Too Much or Too Little?" *Economics of Education Review* 15 (4): 407–19.
- Blau, F., and L. Kahn. 2005. "Do Cognitive Test Scores Explain Higher US Wage Inequality?" *Review of Economics and Statistics* 87: 184–93.
- Bourguignon, F., F.H.G. Ferreira, and M. Menéndez. 2007. "Inequality of Opportunity in Brazil." *Review of Income Wealth* 53 (4): 585–618.
- Brown, G., J. Micklewright, S.V. Schnepf, and R. Waldmann. 2007. "International Surveys of Educational Achievement: How Robust are the Findings?" *Journal of the Royal Statistical Society* 170 (3): 623–46.
- Brunello, G., M. Giannini, and K. Ariga. 2006. "The Optimal Timing of School Tracking." In L. Wößmann, and P.E. Peterson, eds., *Schools and the Equal Opportunity Problem*. Cambridge, MA: MIT Press.
- Brunello, G., and D. Checchi. 2007. "Does School Tracking Affect Equality of Opportunity? New International Evidence." *Economic Policy* 22: 781–861.
- Castelló, A., and R. Doménech. 2002. "Human Capital Inequality and Economic Growth: Some New Evidence." *Economic Journal* 112: C187–200.
- Castro-Leal, F., J. Dayton, L. Demery, and K. Mehra. 1999. "Public Social Spending in Africa: Do the Poor Benefit?" *World Bank Research Observer* 14 (1): 49–72.
- Checchi, D., and V. Peragine. 2005. "Regional Disparities and Inequality of Opportunity: The Case of Italy." IZA Discussion Paper No. 1874. Institute for the Study of Labor, Bonn, Germany.
- . 2010. "Inequality of Opportunity in Italy." *Journal of Economic Inequality* 8 (4): 429–50.
- Cohen, G.A. 1989. "On the Currency of Egalitarian Justice." *Ethics* 99: 906–44.
- DiNardo, J., N. Fortin, and T. Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semi-Parametric Approach." *Econometrica* 64 (5): 1001–44.
- Dworkin, R. 1981. "What is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10 (4): 283–345.
- Fernández, R., and J. Galí. 1999. "To Each According to...? Markets, Tournaments, and the Matching Problem with Borrowing Constraints." *Review of Economic Studies* 66: 799–824.
- Ferreira, F.H.G., and J. Gignoux. 2011a. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *Review of Income Wealth* 57 (4):622–57.
- . 2011b. "The Measurement of Educational Inequality: Achievement and Opportunity." World Bank Policy Research Working Paper No. 5873. Washington, DC.
- Ferreira, F.H.G., J. Gignoux, and M. Aran. 2011. "Measuring Inequality of Opportunity with Imperfect Data: The Case of Turkey." *Journal of Economic Inequality* 9 (4): 651–80.
- Fields, G.S., and E.A. Ok. 1996. "The Meaning and Measurement of Income Mobility." *Journal of Economic Theory* 71 (2): 349–77.
- Fleurbaey, M. 2008. *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press.
- Fleurbaey, M., and V. Peragine. 2012. "Ex ante versus ex post equality of opportunity." *Economica* (published online 12 July 2012).
- Foster, J., and A. Shneyerov. 2000. "Path Independent Inequality Measures." *Journal of Economic Theory* 91: 199–222.

- Gamboa, L.F., and F.D. Waltenberg. 2012. "Inequality of Opportunity for Educational Achievement in Latin America: Evidence from PISA 2006–2009." *Economics of Education Review* 31 (5): 694–708.
- Hanushek, E., and L. Woessmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences-In-Differences Evidence across Countries." *Economic Journal* 116: C63–C76.
- Jakubowski, M., and A. Pokropek. 2011. "Measuring Progress in Reading Achievement between Primary and Secondary School across Countries." Working Paper n°20. University of Warsaw, Faculty of Economic Sciences, Warsaw, Poland.
- Jakubowski, M., H.A. Patrinos, E.E. Porta, and J. Wisniewski. 2010. "The Impact of the 1999 Education Reform in Poland." World Bank Policy Research Working Paper 5263. Washington, DC.
- Manning, A., and J.-S. Pischke. 2006. "Comprehensive versus Selective Schooling in England in Wales: What Do We Know?" IZA Discussion Paper 2072. Institute for the Study of Labor, Bonn, Germany.
- Macdonald, K., F. Barrera, J. Guaqueta, H. Patrinos, and E. Porta. 2010. "The Determinants of Wealth and Gender Inequity in Cognitive Skills in Latin America." World Bank Policy Research Working Paper 5189. Washington, DC.
- Marks, G.N. 2005. "Cross-National Differences in Accounting for Social Class Inequalities in Education." *International Sociology* 20 (4): 483–505.
- Micklewright, J., and S. Schnepf. 2007. "Inequality of Learning in Industrialized Countries." In S. Jenkins and J. Micklewright, eds., *Inequality and Poverty Re-examined*. Oxford: Oxford University Press.
- Mislevy, R.J. 1991. "Randomization Based Inference about Examinees in the Estimation of Item Parameters." *Psychometrika* 56: 177–96.
- Mislevy, R.J., A.E. Beaton, B. Kaplan, and K. Sheehan. 1992. "Estimating Population Characteristics from Sparse Matrix Samples of Item Responses." *Journal of Educational Measurement* 29 (2): 133–61.
- Morrisson, C., and F. Murtin. 2007. "Education Inequalities and the Kuznets Curves: A Global Perspective Since 1870." Paris School of Economics Working Paper 2007-12. Paris, France.
- OECD 2006. *PISA 2006 technical report*. Paris: OECD.
- . 2009. *PISA Data Analysis Manual*. Paris: OECD.
- Psacharopoulos, G. 1994. "Returns to Investment in Education: A Global Update." *World Development* 22:1325–43.
- Pekkarinen, T., R. Uusitalo, and S. Kerr. 2009. "School Tracking and Development of Cognitive Skills." IZA Discussion Paper No. 4058. Institute for the Study of Labor, Bonn, Germany.
- Roemer, J.E. 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
- Salehi-Isfahani, D., and N.B. Hassine. 2012. "Equality of Opportunity in Education in the Middle East and North Africa." Working Papers e07-33. Virginia Polytechnic Institute and State University, Department of Economics, Blacksburg, VA.
- Schultz, G., H.W. Ursprung, and L. Wossmann. 2008. "Education Policy and Equality of Opportunity." *Kyklos* 61 (2): 279–308.
- Sen, A. 1985. *Commodities and Capabilities*. Amsterdam, The Netherlands: North-Holland.
- Shorrocks, A. 1978. "The Measurement of Mobility." *Econometrica* 46: 1013–24.
- . 1999. "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value." unpublished manuscript, University of Essex, Colchester, UK.
- Thomas, V., Y. Wang, and X. Fan. 2001. "Measuring Education Inequality: Gini Coefficients of Education." World Bank Policy Research Working Paper 2525. Washington, DC.
- van de Gaer, D. 1993. "Equality of Opportunity and Investment in Human Capital." PhD dissertation, Catholic University of Leuven, Leuven, Belgium.
- van de Gaer, D., J. Vandebosche, and J.L. Figueroa. 2012. "Children's health opportunities and project evaluation: Mexico's *Oportunidades* program." *World Bank Economic Review* 28 (2): 282–310.

- van de Walle, D., and K. Nead. 1995. *Public Spending and the Poor: Theory and Evidence*. Washington, DC: Johns Hopkins and World Bank.
- Zheng, B. 1994. "Can a Poverty Index be Both Relative and Absolute?" *Econometrica* 62 (6): 1453–58.

Economic Growth and Equality of Opportunity

Vito Peragine¹, Flavia Palmisano, and Paolo Brunori

In this paper, we argue that a better understanding of the relationship between inequality and economic growth can be obtained by shifting the analysis from the space of final achievements to the space of opportunities. To this end, we introduce a formal framework based on the concept of the Opportunity Growth Incidence Curve. This framework can be used to evaluate the income dynamics of specific groups of the population and to infer the role of growth in the evolution of inequality of opportunity over time. We show the relevance of the introduced framework by providing two empirical analyses, one for Italy and the other for Brazil. These analyses show the distributional impact of the recent growth experienced by Brazil and the recent crisis suffered by Italy from both the income inequality and opportunity inequality perspectives. JEL codes: D63, E24, O15, O40

In recent years, a central topic in the economic development literature has been the measurement of the distributive impact of growth (see [Ferreira 2010](#)). This literature has provided analytical tools to identify and quantify the effect of growth on distributional phenomena such as income poverty and income inequality. Indices for measuring the pro-pooriness of growth have been proposed,² and the Growth Incidence Curve (GIC), measuring the quantile-specific rate of economic growth in a given period of time ([Ravallion and Chen 2003](#); [Son 2004](#)), has become a standard tool in evaluating growth from a distributional viewpoint. The interplay among growth, inequality, and poverty reduction has

1. Vito Peragine (corresponding author) is a professor at Università di Bari, Italy; his email address is v.peragine@dse.uniba.it. Flavia Palmisano is postdoctoral fellow at the Università di Bari; her email address is flaviana.palmisano@gmail.com. Paolo Brunori is assistant professor at Università di Bari; his mail address is paolo.brunori@uniba.it. The authors thank Francisco Ferreira, Dirk Van de Gaer, the editors and three anonymous referees for helpful comments on earlier drafts. The authors also wish to thank Jean-Yves Duclos, Michael Lokshin, and Laura Serlenga. Insightful comments were received at conferences or seminars at the World Bank ABCDE Conference, the University of Rome Tor Vergata, VI Academia Belgica-Francqui Foundation Rome Conference and GRASS workshop, and the College d'Etudes Mondiale, Paris. The authors also thank Francisco Ferreira and Maria Ana Lugo for kindly providing them with access to data.

2. See [Essama-Nssah and Lambert \(2009\)](#) for a comprehensive survey.

THE WORLD BANK ECONOMIC REVIEW, VOL. 28, NO. 2, pp. 247–281
Advance Access Publication October 14, 2013

doi:10.1093/wber/lht030

© The Author 2013. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

also been investigated (Bourguignon 2004). All of these tools are now used extensively in the field of development economics to evaluate and compare different growth processes in terms of social desirability and social welfare (see Atkinson and Brandolini 2010; Datt and Ravallion 2011).

A common feature of this literature is the focus on individual achievements, such as (equivalent) income or consumption, as the proper “space” of distributional assessments.

In contrast, recent literature in the field of normative economics has argued that equity judgments should be based on opportunities rather than on observed outcomes (see Dworkin 1981a, b; Cohen 1989; Arneson 1989; Roemer 1998; Fleurbaey 2008). The equal-opportunity framework stresses the link between the opportunities available to an agent and the initial conditions that are inherited or beyond the control of this agent. Proponents of equality of opportunity (EOp) accept the inequality of outcomes that arises from individual choices and effort, but they do not accept the inequality of outcomes caused by circumstances beyond individual control. This literature has motivated a rapidly growing number of empirical applications interested in measuring the degree of inequality of opportunity (IOp) in a distribution and evaluating public policies in terms of equality of opportunity (see, among others, Aaberge et al. 2011; Bourguignon et al. 2007; Checchi and Peragine 2010; LeFranc et al. 2009; Roemer et al. 2003). Book-length collections of empirical analyses of EOp in developing countries can be found in World Bank (2006) and de Barros et al. (2009).

The growing interest in EOp, in addition to the intrinsic normative justifications, is motivated by instrumental reasons: it has been convincingly argued (see World Bank 2006, among others) that the degree of opportunity inequality in an economy may be related to the potential for future growth. The idea is that when exogenous circumstances such as gender, race, or parental background play a strong role in determining individual income and occupation prospects, there is a suboptimal allocation of resources and lower potential for growth. The existence of *inequality traps*, which systematically exclude some groups of the population from participation in economic activity, is harmful to growth.

We share this view, and we believe that a better understanding of the relationship between inequality and growth can be obtained by shifting the analysis from the space of final achievements to the space of opportunities. If two growth processes have, say, the same impact in terms of poverty and inequality reduction, but in the first case, all members of a certain ethnic minority - or all people whose parents are illiterate - experience the lowest growth rate whereas poverty reduction in another case is uncorrelated with differences in race or family background, our current arsenal of measures does not readily allow us to distinguish them. Moreover, although a set of tools has been proposed to explain changes in outcome inequality as the result of differences in growth for individuals with different initial outcomes, to the best of our knowledge, the relationship between the change in IOp and growth has never been investigated.

Our aim is to address this measurement problem³ by proposing a framework and a set of simple tools that can be used to investigate the distributional effects of growth from an opportunity egalitarian viewpoint. In particular, with reference to a given growth episode, we address the following questions: is growth reducing or increasing the degree of IOp? Are some socio-economic groups systematically excluded from growth?

To answer these questions, we depart from the concept of the GIC provided by Ravallion and Chen (2003) and further developed by Son (2004) and Essama-Nssah (2005), and we extend it to the space of opportunities. Hence, we introduce the concept of the Opportunity Growth Incidence Curve (OGIC), which is intended to capture the effect of growth from the EOp perspective. We distinguish between an *individual OGIC* and a *type OGIC*: the former plots the rate of growth of the (value of the) opportunity set given to individuals in the same position in the distributions of opportunities. The latter plots the rate of income growth for each sub-group of the population, where the sub-groups are defined in terms of initial exogenous circumstances. As shown in the paper, these tools capture distinct phenomena: the individual OGIC enables us to assess the pure distributional effect of growth in terms of increasing or reducing aggregate IOp; the type OGIC, in contrast, allows us to track the evolution of specific groups of the population in the growth process to detect the existence of possible inequality traps. For each of the two, we also provide summary measures of growth.

These tools can be used as complements to the standard analysis of the pro-poorness of growth and may provide interesting insights for the design of public policies. In particular, they may help target specific groups of the population and/or identify priorities in redistributive and social policies. Moreover, these tools can be used for the evaluation of public policies in terms of equality of opportunity. In fact, the two-period framework could easily be adapted for the comparison of pre- and post-public intervention distributions—for instance, if one is interested in evaluating the distributive impact of a certain fiscal reform in the space of opportunities.

In this paper, we adopt this theoretical framework to analyze the distributional impact of growth in two different countries, Italy and Brazil, in recent years. These two countries experienced very different patterns of growth in the last decade. On the one hand, Italy experienced a period of very limited growth. According to the Bank of Italy, in the 2002–04 and 2004–06 periods, the average household income increased by 2 percent and 2.6 percent, respectively, whereas the equivalent disposable income of Italian households was characterized by a long spell of negative growth during the recent economic crisis: it decreased by 2.6 percent in

3. Hence, we investigate the relationship between growth and inequality of opportunity using a “micro approach”; an alternative “macro approach” would also be possible by investigating the relationship between growth and IOp from a cross-country or longitudinal perspective (see Marrero and Rodriguez 2010).

the 2006–10 period and by 0.6 percent between 2008 and 2010 (*Banca d'Italia* 2008, 2012). Inequality in the same period increased, but only slightly. On the other hand, Brazil faced a period of sustained growth (with an average 5 percent GDP yearly growth in the last decade), and this growth, as shown in the literature, was markedly progressive. In fact, the Gini index for the entire distribution decreased during the period considered from 60.01 in 2001 to 54.7 in 2009 (see contributions by *Ferreira et al.* 2008, *World Bank* 2012).

Therefore, it is interesting to examine how the perspective of opportunity inequality can add elements of knowledge to the analysis of two markedly different distributional dynamics.

We use the Bank of Italy's "Survey on Household Income and Wealth" (SHIW) to assess the distributional impact of growth in Italy. In particular, we consider four of the most recent available waves to compare the 2002–06 growth episode with the 2006–10 episode. We use the "Pesquisa Nacional por Amostra de Domicílios" (PNAD), provided by the Instituto Brasileiro de Geografia e Estatística, to analyze growth in Brazil, and we focus on the 2002–05 growth episode against the 2005–08 episode.

As far as Italy is concerned, when we focus on each single growth episode, some relevant insights arise. For instance, when the 2002–06 growth period is considered, the standard GIC shows a clear progressive pattern, but this pattern is reversed when the individual OGIC is adopted. When the 2006–10 period is considered, the regressive pattern shown by both the individual OGIC and the type OGIC demonstrates that the burden of the economic crisis has been borne by the weak groups in the population. Important information can be gained when we compare the two periods. The first period dominates the second according to the GIC and the individual OGIC, but this dominance does not hold when the type OGIC is adopted. We suggest that these results may be interpreted as the consequence of differences in per capita income growth between regions and some structural changes introduced in the Italian labor market in the recent past.

With respect to Brazil, it is interesting to note that although the growth experienced by the individual outcome in 2002–05 appears considerable for the whole distribution (with the exception of the top 15 percent), the growth experienced in terms of opportunities is less prominent. Indeed, most of the types suffer a reduction in the value of the opportunity during the growth process.⁴ In contrast, the 2005–08 growth episode appears to be beneficial for the whole population regardless of the focus of the analysis (whether outcome or opportunity). Our analysis shows that the 2005–08 growth process is not only generally progressive but that it also leads to a reduction in the IOp (progressive individual OGIC). Furthermore, the initially disadvantaged groups of the population seem to benefit more from growth than those that were initially advantaged (decreasing type OGIC). When the two processes are compared, the dominance of the

4. To obtain this conflict between type OGIC and GIC, it is necessary that rich individuals experiencing losses are spread across the majority of socioeconomic groups.

2002–05 growth episode over the 2005–08 episode is evident for every perspective adopted.

Hence, we contribute to the literature by showing how it is possible to extend the existing frameworks proposed for the distributional assessment of growth to make them consistent with the EOp approach. The empirical analyses conducted in the paper show that the evaluation of growth may differ if the opportunity inequality perspective is adopted instead of the standard income inequality perspective.

The rest of this paper is organized as follows. Section I introduces the models used in the literature on the distributional effect of growth and in the EOp literature. It then proposes the opportunity growth incidence curves and summary indexes to assess the distributional impact of growth in terms of opportunity. Section II provides the empirical analyses based on Italian and Brazilian data. Section III concludes.

THE INCIDENCE OF GROWTH IN THE SPACE OF OPPORTUNITIES

A well-developed body of literature has proposed a number of tools that can be used to evaluate the distributive impact of growth⁵ in the space of final achievements. After a brief survey of these tools, this section will propose a set of formal tools that can be used to evaluate the impact of growth in the space of opportunities.

Growth and Income Inequality

Let $F(y_t)$ be the cumulative distribution function of income at time t , with mean income $\mu(y_t)$, and let $y_t(p)$ be the quantile function of $F(y_t)$, representing the income corresponding to quantile p in $F(y_t)$. To evaluate the growth taking place from t to $t + 1$, Ravallion and Chen (2003) define the Growth Incidence Curve (GIC) as follows⁶:

$$g(p) = \frac{y_{t+1}(p)}{y_t(p)} - 1 = \frac{L'_{t+1}(p)}{L'_t(p)}(\gamma + 1) - 1, \text{ for all } p \in [0, 1] \quad (1)$$

where $L'(p)$ is the first derivative of the Lorenz curve at percentile p and $\gamma = \mu(y_{t+1})/\mu(y_t) - 1$ is the overall mean income growth rate. The GIC plots the percentile-specific rate of income growth in a given period of time. Clearly, $g(p) \geq 0$ ($g(p) < 0$) indicates positive (negative) growth at p . A downward-sloping GIC indicates that growth contributes to equalize the distribution of

5. In what follows, we focus, in particular, on those tools that will be extended to the EOp model in the next section. For a detailed survey of other existing measures of growth, see Essama-Nsaah and Lambert (2009) and Ferreira (2010).

6. For a longitudinal perspective on the evaluation of growth, see Bourguignon (2011) and Jenkins and Van Kerm (2011).

income (i.e., $g(p)$ decreases as p increases), whereas an upward-sloping GIC indicates non-equalizing growth (i.e., $g(p)$ increases as p increases). When the GIC is a horizontal line, inequality does not change over time, and the rate of growth experienced by each quantile is equal to the rate of growth in the overall mean income.

Growth incidence curves are used to detect how a given growth spell affects the different parts of the distribution. In addition, they are used as criteria to rank different growth episodes. Ravallion and Chen (2003) apply first-order dominance criteria based on the GIC: first-order dominance implies that the GIC of a growth spell is everywhere above the GIC of another growth spell. Son (2004) elaborates on this concept by proposing weaker second-order dominance conditions, requiring that the mean growth rate up to the p poorest percentile in a growth episode - or the “cumulative GIC” - be everywhere larger than in another. In this case, the cumulative GIC is given by $G(p) = \int_0^p g(q)y_t(q)dq / \int_0^p y_t(q)dq$ for all $p \in [0,1]$.

Building on the concept of the GIC, the literature has provided a variety of aggregate measures of growth. We recall, among these, the rate of pro-poor growth proposed⁷ by Essama-Nssah (2005): $RPPG_{EN} = \int_0^1 v(p)g(p)dp$, where $v(p) > 0$, and $v'(p) \leq 0$ is a normalized social weight, decreasing with the rank in the income distribution. Hence, $RPPG_{EN}$ represents a rank-dependent aggregation of each point of the GIC and measures the overall extent of growth, giving more importance to the growth experienced by the income of the poorest individuals.⁸ We enrich this framework by looking at the literature on EOp measurement.

From Income to Opportunity Inequality

In the EOp model (see Roemer 1998, Van de Gaer 1993, Peragine 2002), the individual income at a given time, $t \in \{1, \dots, T\}$, y_t , is assumed to be a function of two sets of characteristics: the circumstances, \mathbf{c} , belonging to a finite set ω and the level of effort, $e_t \in \Theta \subseteq \mathbb{R}_+$. The individual cannot be held responsible for \mathbf{c} , which is fixed over time; he is, instead, responsible for the effort e_t that he autonomously decides to exert in every period of time. Income is generated by a production function $g: \Omega \times \Theta \rightarrow \mathbb{R}_+$:

$$y_t = g(\mathbf{c}, e_t). \quad (2)$$

This is a reduced form model in which circumstances and effort are assumed to be orthogonal, and the function g is assumed to be monotonic in both arguments. Although the monotonicity of g is a fairly reasonable assumption, the orthogonality assumption rests on the theoretical argument that it would be hardly

7. In the original paper, $RPPG_{EN}$ is applied to discrete distributions. Here, we use a continuous version of the same index to be consistent with our notation.

8. Ravallion and Chen (2003) also propose the $RPPG_{RC} = \int_0^{H_t} g(p)dp/H_t$ where H_t is the initial poverty headcount ratio. $RPPG_{RC}$ measures the proportionate income change of the poorest individuals.

sustainable to hold people accountable for factor e_t if it were dependent on exogenous circumstances.

In line with this model, a partition of the total population is now introduced. Each group in this partition is called a *type* and includes all individuals sharing the same circumstances. For example, if the only two circumstances were gender (male or female) and race (black or white), then there would be four types in the population: white men, black men, white women, and black women. Hence, considering n types, for all $i = 1, \dots, n$, the outcome distribution of type i at time t is represented by a cdf $F_i(y_t)$, with population size m_{it} , population share q_{it} , and mean $\mu_i(y_t)$.

Given this analytical framework, the focus is on the income prospects of individuals of the same type, represented by the type-specific income distribution $F_i(y_t)$. This distribution is interpreted as the set of opportunities open to each individual in type i . In other words, the observable actual incomes of all individuals in a given type is used to proxy the unobservable *ex ante* opportunities of all individuals in that type.

Let us underline here a dual interpretation of the types in the EOp model: on the one hand, the type is a component of a model that, starting from a multivariate distribution of income and circumstances, allows us to obtain a distribution of (the value of) opportunity sets enjoyed by each individual in the population. On the other hand, given the nature of the circumstances typically observed and used in empirical application, the partition in types may be of interest *per se*: they can often identify well-defined socio-economic groups that may deserve special attention by the policy makers. As we will see, this dual interpretation of the types will be exploited in the analysis of the impact of growth on EOp.

A specific version of the EOp model, which is called “utilitarian”, further assumes that the value of the opportunity set $F_i(y_t)$ can be summarized by the mean $\mu_i(y_t)$. This is clearly a strong assumption because it implies neutrality with respect to the inequality within types. Assuming within-type neutrality, the next step consists of constructing an artificial distribution in which each individual income is substituted with the value of the opportunity set of that individual, that is, the mean income of the type to which the individual belongs. More formally, by ordering the types on the basis of their mean such that $\mu_1(y_t) \leq \dots \leq \mu_j(y_t) \leq \dots \leq \mu_n(y_t)$, the *smoothed distribution* corresponding to $F(y_t)$ is defined as $Y_t^s = (\mu_1^t, \dots, \mu_j^t, \dots, \mu_N^t)$. N is the total size of the population, and μ_j^t is the smoothed income, interpreted as the value of the opportunity set, of the individual ranked j/N in Y_t^s . Hence, in this model, measuring opportunity inequality simply amounts to measuring inequality in the smoothed distribution Y_t^s .

Some authors have questioned this “utilitarian” approach (see [Fleurbaey 2008](#) for a discussion of the issue). For instance, some authors argue that in addition to circumstances and effort, an additional factor, luck, plays a role in determining the individual outcome (see, inter alia, [Van de Gaer 1993](#); [LeFranc et al. 2008, 2009](#)). Therefore, they argue, only part of within-type heterogeneity can be directly attributable to differences in effort. In particular, the unequal

outcomes resulting from “brute” luck should be compensated for.⁹ Furthermore, these authors argue, individuals may be risk averse; hence, the within-type inequality may have a cost for them. Following this line of reasoning, alternative models of EOp that consider within-type heterogeneity have been proposed in the literature.¹⁰

The model adopted in this paper, based on the assumption of within-type inequality neutrality and the use of the mean income conditional on each type as the value of the opportunity set, is well grounded on normative reasons and, in particular, is consistent with a strong version of the reward principle; see Fleurbaey (2008) and Fleurbaey and Peragine (2013) for a discussion. However, it is also motivated by practical reasons; accounting for within-type heterogeneity is very demanding in terms of data. It is often the case that the small size of the samples used makes it difficult to obtain easily comparable within-type distributions. This approach makes our empirical analysis fully consistent with most of the analyses performed in the existing literature.¹¹ Nevertheless, although our theoretical model is built on the assumption of within-type neutrality, we explore the issue of within-type heterogeneity in the empirical section by looking at growth within each type. It is shown that the dynamic of inequality within types can be a source of divergence between the standard approach based on income inequality and the opportunity egalitarian approach.

A final methodological consideration is in order here and concerns the issue of omitted circumstance variables. We use a pure deterministic model where, given a set of selected circumstances, any residual variation in individual income is attributed to personal effort. This amounts to saying that once the vector of circumstances has been defined, on the basis of normative grounds and observability constraints, all other factors are implicitly classified as within the sphere of individual responsibility. However, the vector \mathbf{c} observed in any particular dataset is likely to be a sub-vector of the theoretical vector of all possible circumstances that determine a person’s outcome. Whenever the dimension of the observed vector \mathbf{c} is less than the dimension of the “true” vector, then we obtain lower-bound estimators of true inequality of opportunity; that is, the inequality

9. The literature distinguishes between *brute luck*, which is unrelated to individual choices and hence deserves compensation, and *option luck*, which is a risk that individuals deliberately assume and does not call for compensation. See Ramos and Van de Gaer (2012), Fleurbaey (2008), and LeFranc et al. (2009) for a detailed discussion of the different meanings of luck.

10. For example, LeFranc et al. (2008) and Peragine and Serlenga (2008) use stochastic dominance conditions to compare the different type distributions. Moreover, LeFranc et al. (2008) measure the opportunity set as (twice) the surface under the generalized Lorenz curve of the income distribution of the individual’s type, that is $\mu_i(1 - G_i)$, where the type mean income μ_i and $(1 - G_i)$ represent, respectively, the return component and the risk component, with G_i denoting the Gini inequality index within type i . See also O’Neill et al. (2000) and Nilsson (2005) for empirical analyses that attempt to provide alternative evaluations of opportunity sets using parametric estimates.

11. As discussed in Brunori et al. (2013), the (ex ante) utilitarian approach has been by now adopted by several authors to assess IOp in about 41 different countries, making an international comparison of inequality of opportunity estimates across the world possible.

that would be captured by observing the full vector of circumstances. The implication is that the empirical estimates obtained using this model should be interpreted as lower-bound estimates of IOp.¹² Similarly, it is worth underlining that whenever circumstances are partially unobservable, the change in IOp due to growth should be interpreted as the change in the lower bound IOp conditioned to the observable circumstances. An evaluation of change in IOp based on a different set of variables could lead to different conclusions.

The Opportunity Growth Incidence Curve

In this section, we introduce the *two versions* of the Opportunity Growth Incidence Curve (OGIC), which can be considered complementary tools to the GIC, to improve the understanding of the distributional features of growth when an opportunity egalitarian perspective is adopted. *The two versions, the individual OGIC and the type OGIC*, capture two different intuitions about the relationship between growth and EOp. The first focuses on the impact of growth on the distribution of opportunities. The second focuses on the relationship between overall economic growth and type-specific growth.

Given an initial distribution of income Y_t and the corresponding smoothed distribution Y_t^s introduced in the previous section, the individual OGIC can simply be obtained by applying the GIC proposed by Ravallion and Chen (2003) to the smoothed distribution. Hence, the *individual OGIC* can be defined as follows:

$$g_{Y_t^s}^o\left(\frac{j}{N}\right) = \frac{\mu_j^{t+1}}{\mu_j^t} - 1, \forall j \in \{1, \dots, N\}. \tag{3}$$

$g_{Y_t^s}^o(j/N)$ measures the proportionate change in the value of opportunities of the individuals ranked j/N in the smoothed distributions. Obviously, $g_{Y_t^s}^o(j/N) \geq 0$ ($g_{Y_t^s}^o(j/N) < 0$) means that there has been positive (negative) growth in the value of the opportunity set given to the individuals ranked j/N respectively in Y_t^s and in¹³ Y_{t+1}^s .

The individual OGIC provides information on the impact of growth on IOp. Consider the Lorenz curve of Y_t^s :

$$L_{Y_t^s}\left(\frac{j}{N}\right) = \frac{\sum_{k=1}^j \mu_k^t}{\sum_{k=1}^N \mu_k^t}, \forall k \in \{1, \dots, N\}, \forall t \in \{1, \dots, T\}. \tag{4}$$

12. For a discussion of this issue with reference to a non deterministic, parametric model of EOp, see Ferreira and Gignoux (2011) and Luongo (2011).

13. Note that, given the assumption of anonymity implicit in this framework, the individuals ranked j/N in t can be different from those ranked j/N in $t + 1$.

The individual OGIC defined in eq. (3) can be decomposed in such a way that it becomes a function of the Lorenz curve defined in eq. (4), as follows:

$$g_{Y_t^s}^o\left(\frac{j}{N}\right) = \frac{\Delta L_{Y_{t+1}^s}\left(\frac{j}{N}\right)}{\Delta L_{Y_t^s}\left(\frac{j}{N}\right)}(\gamma + 1) - 1, \forall j \in \{1, \dots, N\} \quad (5)$$

where $\Delta L_{Y_t^s}(j/N) = \mu_j^t/\mu(y_t)$ is the first derivative of $L_{Y_t^s}(j/N)$ with respect to j/N , and $\gamma = \mu(Y_{t+1})/\mu(y_t) - 1$ is the overall mean income growth rate.

Thus, when growth is proportional, it does not have any impact on the level of IOp: $\Delta L_{Y_{t+1}^s}(j/N)/\Delta L_{Y_t^s}(j/N) = 1$, and $g_{Y_t^s}^o(j/N)$ will just be an horizontal line, with $g_{Y_t^s}^o(j/N) = \gamma$ for all j . On the contrary, when growth is progressive (regressive) in terms of opportunity, growth acts by reducing (worsening) IOp: $\Delta L_{Y_{t+1}^s}(j/N)/\Delta L_{Y_t^s}(j/N) \neq 1$, and $g_{Y_t^s}^o(j/N)$ will be a decreasing (increasing) curve.

The main aspect that distinguishes the individual OGIC from the standard GIC is represented by the distributions used to construct that curve. This variation allows us to establish a link between growth and IOp. Note that the smoothed distribution at the base of the individual OGIC is the same used by [Checchi and Peragine \(2010\)](#) and [Ferreira and Gignoux \(2011\)](#) to measure ex ante IOp. Therefore, our evaluation of growth based on the individual OGIC is, by construction, consistent with the IOp index they proposed; other things being equal, an individual OGIC curve that is downward sloping in all of its domain implies a reduction in IOp.

However, the individual OGIC is unable to track the evolution of each type during the growth process. In the smoothed distribution, types are ranked according to the value of their opportunity set at each point in time. Thus, the shape of the curve depends not only on the change in the type-specific mean income but also on the type-specific population share and the reranking of types taking place during the growth process. Now, although these features are desirable when one is interested in studying the evolution of IOp over time, the same characteristics make it impossible to detect the individual OGIC if there are groups of the population that are systematically excluded from growth. However, this can provide valuable information for analysts and policy makers. For example, consider a very small type that suffers a deterioration of its condition over time. This information could be irrelevant for the evolution of the overall opportunity inequality, but it would be extremely important for the design of tailored policy interventions toward that group.

To address this specific issue and to investigate the relationship between overall economic growth and type-specific growth, we introduce a second version of the OGIC, which we label the *type OGIC*.

Letting $Y_{\mu t} = (\mu_1(y_t), \dots, \mu_n(y_t))$ be the distribution of type mean income at time t , where types are ordered increasingly according to their mean,

i.e., $\mu_1(y_t) \leq \dots \leq \mu_n(y_t)$, and $\tilde{Y}_{\mu t+1} = (\tilde{\mu}_1(y_{t+1}), \dots, \tilde{\mu}_n(y_{t+1}))$ is the distribution of type mean income at time $t + 1$, where types are ordered according to their position at time¹⁴ t , we define the *type OGIC* as follows:

$$\tilde{g}^o\left(\frac{i}{n}\right) = \frac{\tilde{\mu}_i(y_{t+1}) - \mu_i(y_t)}{\mu_i(y_t)}, \forall i \in \{1, \dots, n\}. \tag{6}$$

The type OGIC plots, against each type, the variation of the opportunity set of that type. This can be interpreted as the rate of economic development of each social group in the population, where these groups are defined on the basis of initial circumstances. $\tilde{g}^o(i/n)$ is horizontal if each type benefits (loses) in the same measure from growth. It is negatively (positively) sloped if the initially disadvantaged types get higher (lower) benefit from growth than those initially advantaged.¹⁵

The type OGIC differs from the standard GIC in two aspects. The first is represented by the distribution used to plot the curve: the GIC is based on the income distribution, whereas the OGIC is based on the distribution of opportunity sets. The second is represented by the weakening of the anonymity assumption for types. Thus, the type OGIC, tracking the same type over time, provides information on the temporal evolution of the opportunity set.

The OGIC, in both the individual and the type versions, can be used to rank different growth episodes. Analogously with the literature on the standard GIC, we can apply first-order dominance criteria based on the OGIC.¹⁶ First-order dominance implies that the OGIC of a growth spell is everywhere above that of another.

However, the two approaches (individual and type OGIC) are generally not equivalent, and they can generate a different ranking of growth processes. In fact, beyond their interpretation and the fact that they can be used to investigate different aspects of the relationship between economic growth and EOP, the differences between the individual and the type OGIC are mainly due to demographic and reranking issues. The following remark makes this point clear.

Remark 1. *Let Y_t^A and Y_t^B be two initial distributions of income, and let G^A and G^B be two different growth processes taking place, respectively, on Y_t^A and Y_t^B and generating, respectively, two final distributions of income, Y_{t+1}^A and Y_{t+1}^B . Moreover, let n^A and n^B be the number of types, respectively, in Y_t^A and Y_t^B and m_{Ai} and m_{Bi} be the number of individuals in each type $i = 1, \dots, n$, respectively, in Y_t^A and Y_t^B . If (i) $n_{At} = n_{Bt}, \forall t = 1, \dots, T$, (ii) $m_{Ait} = m_{Bit} \forall i \in \{1, \dots, n\}, \forall t = 1, \dots, T$, (iii) no reranking of types, then $\tilde{g}^{Ao}(i/n) \succeq \tilde{g}^{Bo}(i/n) \forall i \in \{1, \dots, n\}$ if and only if $g_{Y_t^A}^{Ao}(j/N) \succeq g_{Y_t^B}^{Bo}(j/N) \forall j \in \{1, \dots, N\}$.*

14. Note that we track the same type but do not track the same individuals.

15. Note that the type OGIC is a generalization of the idea underlying the first component of Roemer’s (2011) index of development, that is, “how well the most disadvantaged type is doing”.

16. For a normative justification of these dominance conditions based on a rank-dependent social welfare function, see the working paper version of the paper: Peragine et al. (2011).

Proof. See appendix.

This remark establishes that when the two distributions have, at each point in time (*i*), the same number of types and (*ii*) the same type-specific population size, and when (*iii*) types keep their relative position in the type mean income distribution over time, ranking income distributions according to the individual OGIC is equivalent to ranking income distributions according to the type OGIC. Because conditions (*i*) and (*ii*) basically impose restrictions on the types' demography and condition (*iii*) imposes restrictions on the rank of the types, it is clear that possible differences in the ordering provided by the two OGICs are determined by variations in the type's population shares, between the two distributions and the two periods compared, and by the reranking of types over time.

Although the conditions in Remark 1 may seem demanding, an interesting case in which they are met is the comparison of growth processes taking place on the same initial distribution. This is the standard case in the literature on microsimulation analyses¹⁷ and, in general, in the case of an evaluation of policy interventions.

The Cumulative OGIC

So far, we have focused on first-order OGIC dominance, which is a strong condition that is rarely verified with real data. A weaker condition is obtained by second-order dominance. This order of dominance builds on the definition of the cumulative¹⁸ OGIC.

To obtain the cumulative OGIC, one should look at the proportionate difference between the generalized Lorenz curves applied to the smoothed distribution at time t and $t + 1$, which, after rearranging, gives the following expression for the individual version:

$$G_{Y^s}^o\left(\frac{j}{N}\right) = \frac{\sum_{k=1}^j g_{Y^s}^o\left(\frac{k}{N}\right) \mu_k^t}{\sum_{k=1}^j \mu_k^t} = \left(\frac{L_{Y^s}_{t+1}\left(\frac{j}{N}\right)}{L_{Y^s}_t\left(\frac{j}{N}\right)} (\gamma + 1) \right) - 1, \forall j \in \{1, \dots, N\}. \quad (7)$$

The cumulative individual OGIC plots the mean income growth rate up to the j th poorest individual in Y^s . It can be downward or upward sloping depending on the pattern of growth among smoothed incomes. Clearly, at $j/N = 1$, $G_{Y^s}^o(j/N)$ equals the overall mean income growth rate, γ .

The above decomposition allows to express the cumulative OGIC as depending on two components: the overall mean income change and the variation in the

17. See, *inter alia*, Sutherland et al. (1999).

18. Similar to the OGIC, the derivation of its cumulative version closely follows the methodology proposed by Son (2004), adequately adapted to be consistent with the EOp theory.

level of the IOP. In case of proportional growth, the Lorenz curves do not change, and the cumulative OGIC is equal to overall mean income growth rate.

On the other hand, the cumulative type OGIC is defined as follows¹⁹:

$$\tilde{G}_{Y_\mu}^o\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^i \tilde{g}^o\left(\frac{j}{n}\right) \mu_j(y_t)}{\sum_{j=1}^i \mu_j(y_t)}, \forall i \in \{1, \dots, n\}. \tag{8}$$

The cumulative type OGIC plots the mean income growth rate up to the type ranked i in the initial type mean distribution against each type in the population. It can be downward or upward sloping, depending on the pattern of growth among types. At $i = n$, $\tilde{G}_{Y_\mu}^o(i/n)$ equals the overall mean growth rate of Y_μ .

OGIC Indexes

To avoid inconclusive results because of the partiality of the dominance conditions based on the curves presented so far, we propose aggregate measures of growth that incorporate some basic EOp principles.

From the individual perspective, adopting a rank-dependent approach to the evaluation of growth, an aggregate measure of growth consistent with the EOp theory can be expressed as follows:²⁰

$$G_{Ys} = \frac{1}{N} \frac{\sum_{j=1}^N v\left(\frac{j}{N}\right) g_{Ys}^o\left(\frac{j}{N}\right)}{\sum_{j=1}^N v\left(\frac{j}{N}\right)}. \tag{9}$$

Given the assumption of anonymity of the individual OGIC, the weight $v(j/N)$ depends on the relative position of individuals in the smoothed distribution, respectively, in t and $t + 1$. Thus, the same weight is given to the value of the opportunity set of individuals ranked the same in the smoothed distribution of the two periods²¹. $v(j/N)$ represents the social evaluation of the growth in the opportunity enjoyed by individuals in the same position in t and $t + 1$.

Thus, eq. (9) represents a rank-dependgent aggregation of the information provided by each single point of the individual OGIC. In particular, imposing *monotonicity*, $v(j/N) \geq 0, \forall j \in \{1, \dots, N\}$, and *opportunity inequality aversion*,

19. Similar to the cumulative individual OGIC, the cumulative type OGIC is obtained by rearranging the difference between the Generalized Lorenz curves applied to the type mean distributions Y_{μ_t} and $\tilde{Y}_{\mu_{t+1}}$.

20. The approach is close in spirit to [Essama-Nssah \(2005\)](#), reviewed in a previous section. For a normative justification of the rank-dependent approach to IOp analyses, see [Peragine \(2002\)](#), [Aaberge et al. \(2011\)](#), and [Palmisano \(2011\)](#)

21. See endnote 12.

$v(j/N) \geq v(j+1/N)$, $\forall j \in \{1, \dots, N-1\}$, we obtain a measure of opportunity-sensitive growth. This measure is increasing in each individual opportunity growth and is more sensitive to the growth in the opportunity experienced by those individuals with the lowest opportunities. Using the specification $v(j/N) = 2(1-j/N)$, we obtain a Gini-type measure of opportunity-sensitive growth.

If, instead, one is interested in assessing the pure progressivity of growth without concern for the aggregate growth, then the following index can be adopted:

$$OG_{Y^s} = G_{Y^s} - \bar{G}_{Y^s} \quad (10)$$

where $\bar{G}_{Y^s} = \frac{1}{N} \sum_{j=1}^N g_{Y^s}^o \left(\frac{j}{N} \right)$. $OG_{Y^s} = 0$ if growth is proportional; it is positive (negative) if growth is progressive (regressive).

An alternative expression can be obtained by using a weighted average of the growth experienced by each type, with weights incorporating a concern for the initial condition of the types:

$$G_{Y^\mu} = \frac{1 \sum_{i=1}^n w \left(\frac{i}{n} \right) \tilde{g}^o \left(\frac{i}{n} \right)}{n \sum_{i=1}^n w \left(\frac{i}{n} \right)}. \quad (11)$$

The function $w(i/n)$ is the social weight associated to type i and depends on the rank of the type in the initial distribution of income. As before, this index satisfies *monotonicity*: $w(i/n) \geq 0$, $i \in \{1, \dots, n\}$ (that is, aggregate growth is not decreasing in each type growth) and *opportunity inequality aversion*: $w(i/n) \geq w(i+1/n)$, $i \in \{1, \dots, n-1\}$ (that is, more weight is given to the income growth experienced by the most disadvantaged types).

Following Aaberge et al. (2011) and choosing $w \left(\frac{i}{n} \right) = 1 - \sum_{j=1}^i q_{jt}$, a Gini-type index of opportunity-sensitive growth results.

THE EMPIRICAL ANALYSES

This section investigates the distributional changes that occurred in Italy and Brazil in the last decade. These analyses pursue two additional aims: (i) assessing the main consequences of the actual economic crisis on the Italian distribution of income according to the EOp perspective and (ii) assessing the distributional implications of the most recent economic development experienced by Brazil in terms of EOp.

For both applications, we first provide an assessment of growth according to the equality of outcome perspective. We then move to the analysis of growth according to the EOp perspective.²²

Opportunity and Growth in Italy: The Data

Italy is the first country considered in this section. This analysis is developed using the Bank of Italy's "Survey on Household Income and Wealth" (SHIW), a representative sample of the Italian resident population interviewed every two years. Three waves of the survey are considered: 2002, 2006, and 2010 (the latest available).

The unit of observation is the household, defined as all persons sharing the same dwelling. The individual outcome is, then, measured as the household equivalent income in 2010 euro.²³ Income includes all household earnings, transfers, pensions, and capital incomes, net taxes, and social security contributions. The richest and poorest 1 percent of the households in each wave are dropped to avoid the effect of outliers. To identify the types, the distribution is partitioned into 18 types using information about three characteristics of the head of the family: the highest educational attainment of her parents (three levels: up to elementary school, lower secondary, and higher), the highest occupational status of her parents (two levels: not in the labor force/blue collar and white collar) and the geographical area of birth (three areas: North, Centre, and South). Note, however, that those households for which the identification of the type is not possible because of missing information about one or more circumstances are excluded. The sample sizes of each wave considered are 6,428 in 2002, 6,354 in 2006, and 6,579 in 2010.

The list of types with their respective opportunity profiles²⁴ is reported in Table 1 for each wave. Types are ranked according to their average income. Rankings are clearly driven by the regional origin of the household head. In particular, although some reranking takes place for types of other regions, five of the six types from the South of Italy are the lowest-ranked at all times.

22. We calculate confidence intervals for the difference between individual OGIC, type OGIC, and indexes in the two growth processes. The resampling procedure that we use is in line with the approach proposed by Lokshin (2008) for the GIC. We assume that the income distributions observed at the two points in time, y^t , y^{t+1} , are independent and identically distributed observations of the unknown probability distributions $F(y^t), F(y^{t+1})$. γ is the statistic of interest, and its standard error is $\sigma(F(y^t), F(y^{t+1})) = \sqrt{\text{Var}\gamma(y^t, y^{t+1})}$. Our bootstrap estimate of the standard error is $\hat{\sigma} = \sigma(\hat{F}(y^t), \hat{F}(y^{t+1}))$, where $\hat{F}(y^t), \hat{F}(y^{t+1})$ are the empirical distributions observed. The 95 percent confidence interval is obtained by resampling $B = 1,000$ ordinary non parametric bootstrap replications of the two distributions y_t^*, y_{t+1}^* . The standard error of parameter $\hat{\gamma}$ is obtained using $\hat{\sigma}_B = \sqrt{\sum_{b=1}^B \{\hat{\gamma}^*(b) - \hat{\gamma}(\cdot)\}^2 / (B - 1)}$, where $\hat{\gamma}(\cdot) = \sum_{b=1}^B \gamma^*(b) / B$. We know that $\hat{\sigma}_B \rightarrow \hat{\sigma}$ when $B \rightarrow \infty$, and, under the assumption that γ is approximately normally distributed, we calculate confidence intervals: $\hat{\gamma} = \hat{\gamma} \pm z_{1-\alpha/2} \hat{\sigma}_B$. Our estimate quality relies on strong assumptions. However, as will be clear in the discussion of the results, dominances appear rather reliable for the illustrative purpose of the exercise.

23. We use the OECD equivalence scale given by the square root of the household size.

24. All standard errors are obtained using the sample weights according to the suggestion in Banca d'Italia (2012).

TABLE 1. Italy 2002-2006-2010: Descriptive Statistics and Partition in Types

Area	Education	Occupation	rank ⁰²	sample ⁰²	q _t ⁰²	μ _t ⁰²	rank ⁰⁶	sample ⁰⁶	q _t ⁰⁶	μ _t ⁰⁶	rank ¹⁰	sample ¹⁰	q _t ¹⁰	μ _t ¹⁰
South	No-edu/Elementary	Blue c./not in l.f.	1	1241	0.2174	14065.82	2	1273	0.2291	15279.71	3	1512	0.2385	14974.97
South	Lower secondary	Blue c./not in l.f.	2	110	0.0214	14386.26	4	124	0.0214	17783.99	1	198	0.0408	13593.33
South	Higher	Blue c./not in l.f.	3	137	0.0233	15673.90	1	104	0.0150	14800.64	2	126	0.0214	14749.59
South	No-edu/Elementary	White c.	4	682	0.1130	16949.30	3	604	0.1098	17149.07	4	594	0.0990	17021.24
South	Lower secondary	White c.	5	213	0.0324	17917.02	6	230	0.0421	20127.67	5	228	0.0372	17903.09
Centre	No-edu/Elementary	Blue c./not in l.f.	6	657	0.0822	19477.92	7	604	0.0755	21970.48	9	622	0.0729	23528.86
Centre	Lower secondary/ Higher	Blue c./not in l.f.	7	51	0.0068	20106.76	12	49	0.0082	26077.04	13	60	0.0111	26010.30
North	Lower secondary	Blue c./not in l.f.	8	135	0.0237	20910.44	10	182	0.0301	24799.79	10	162	0.0294	23548.54
North	No-edu/Elementary	Blue c./not in l.f.	9	1137	0.1623	22095.60	8	1121	0.1591	23292.56	8	1022	0.1465	23063.41
Centre	No-edu/Elementary	White c.	10	316	0.0384	22579.76	9	287	0.0401	23873.59	14	260	0.0268	26348.91
South	Higher	White c.	11	270	0.0406	22828.57	13	239	0.0356	26290.72	11	295	0.0375	24052.45
North	No-edu/Elementary	White c.	12	594	0.0996	23922.43	11	543	0.0839	25240.80	12	474	0.0709	25209.78
Centre	Lower secondary	White c.	13	107	0.0187	24702.06	16	93	0.0128	30371.49	16	119	0.0202	28257.28
North	Higher	Blue c./not in l.f.	14	71	0.0094	25625.36	14	94	0.0140	27060.96	7	100	0.0160	22652.13
Centre	Higher	Blue c./not in l.f.	15	32	0.0039	25664.17	5	45	0.0059	20096.12	6	30	0.0034	21798.12
North	Lower secondary	White c.	16	253	0.0421	26890.26	15	250	0.0387	27748.28	15	247	0.0471	27114.15
North	Higher	White c.	17	296	0.0452	29955.46	17	363	0.0519	32143.62	18	343	0.0543	32106.09
Centre	Higher	White c.	18	126	0.0197	30786.71	18	149	0.0268	33395.35	17	187	0.0268	30670.72

Note: Types are ranked in ascending order according to the average income at the beginning of each growth period.

Source: Authors' calculations on SHIW (Banca d'Italia).

To analyze growth, we consider two four-year periods: 2002–06 and 2006–10. The exercise is appealing because it compares two periods during which Italy faced two different economic slowdowns. The former was characterized by the almost total absence of growth in 2002 and 2003. The latter, triggered by the 2008 financial crisis, was characterized by a deep fall in the GDP growth rate in 2008 and, after a slight respite between 2009 and 2010, is ongoing.

Opportunity and Growth in Italy: The Results

The GICs for the two periods are reported in Figure 1. These curves are obtained by partitioning the distribution into percentiles and by plotting against each percentile its specific growth rate, expressed in yearly percentage points.

Two features stand out. First, the GICs for the two periods lie in two different domains: positive for the first period and negative for the second period, with the exception of the last percentile. This feature is further captured by the mean income growth rate relative to each period, which is 1.96 percent for 2002–06 and –0.66 percent for 2006–10. Second, the two growth processes show very different and symmetric patterns. The income dynamic is progressive between 2002 and 2006, but it becomes quite regressive between 2006 and 2010. Their symmetrical shape suggests that the two processes might have an equally opposed redistributive impact. The sign of the variation over time of their respective aggregate indexes of inequality confirms this supposition: income inequality decreases during the first period and increases during the second period²⁵ (see Table 2 in the appendix).

We proceed in our analysis with the assessment of the distributional effects of growth in the space of “opportunities”. The individual OGIC for the periods considered are reported in Figure 2.

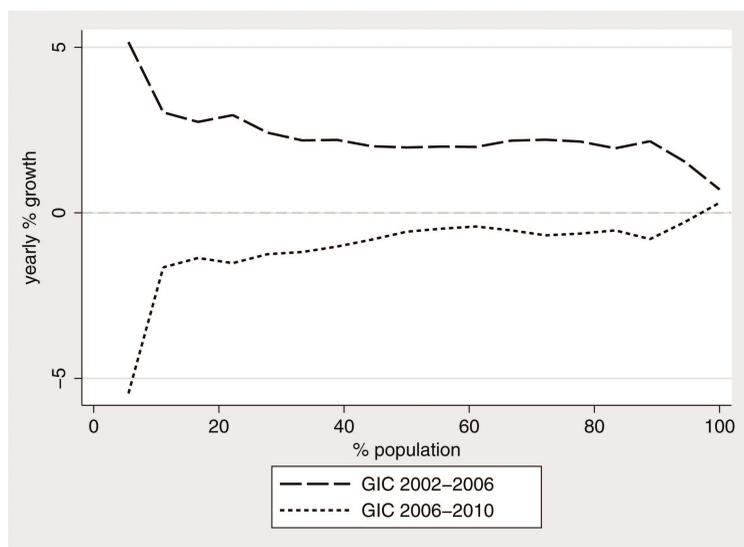
The individual OGIC of 2002–06 shows that growth acts by increasing the value of the opportunities for all quantiles of the smoothed distributions.²⁶ However, the growth rate is not stable across quantiles. In particular, the slightly increasing pattern of the individual OGIC over the whole distribution demonstrates an opportunity-regressive impact of growth.

The peculiarities of this growth process are confirmed by the value of the synthetic measures of growth (see Table 2 in the appendix). The first index, measuring the extent of the opportunity-sensitive growth, is positive, as expected because the individual OGIC lies above 0. The second index, exclusively capturing the equal opportunity-enhancing effect of growth is negative, demonstrating that growth might have failed in its role as an instrument to reduce IOp. These results emphasize the relevance of extending standard analyses of growth to the space of “opportunity”. For instance, the different shapes characterizing the GIC

25. The results for the second period are consistent with other empirical evidence on the effect of the last financial and economic crisis. See, for example, [Jenkins et al. \(2013\)](#).

26. To make the individual OGIC and the type OGIC graphically comparable, we partitioned the smoothed distributions into 18 quantiles.

FIGURE 1. Italy 2002–2006–2010: Growth Incidence Curve



Source: Authors' calculation from SHIW (Bank of Italy).

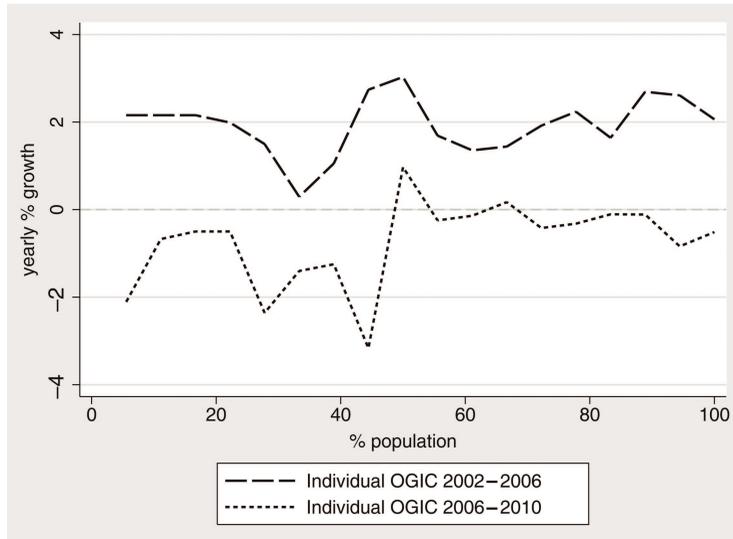
TABLE 2. Italy: 2002–2006–2010 Dominance Conditions

quantiles/ types rank	GIC	type OGIC	cum. type OGIC	individual OGIC	cum. individual OGIC
1	10.5691***	2.6484***	2.6180***	3.9839***	3.9744***
2	4.6810***	11.5799***	7.3428***	2.6985***	3.3317***
3	4.1114***	-1.5181	4.3373***	2.6562***	3.1058***
4	4.4694***	0.5413	3.3125***	2.5996***	2.9757***
5	3.6610***	5.9404***	3.9061***	3.4201***	3.0512***
6	3.3625***	1.3937	3.3944***	1.6977***	2.7881***
7	3.2277***	7.8721**	4.0511***	2.8942***	2.8017***
8	2.8174***	6.0244***	4.3561***	5.4506***	3.1885***
9	2.5479***	1.6141**	3.9883***	1.8843***	2.9947***
10	2.4750***	-1.1908	3.3700***	2.1158***	2.8801***
11	2.3956***	5.7042***	3.6263***	1.5239***	2.7224***
12	2.7012***	1.4691	3.3977***	1.3037***	2.5751***
13	2.8946***	7.2706**	3.8027***	2.6333***	2.5808***
14	2.7802***	5.3008**	3.9270***	2.6164***	2.5835***
15	2.4743***	-7.5717*	3.0613***	1.8334***	2.5185***
16	2.9552***	1.4023	2.9156***	2.6758***	2.5292***
17	1.8412***	1.9006	2.8161***	3.4850***	2.6090***
18	0.3548	4.2672**	2.9169***	2.5781***	2.6063***

* = 90 percent, ** = 95 percent, *** = 99 percent are significance levels for the difference between curves obtained from 1,000 bootstrap replications of the statistics.

Source: Authors' calculations on SHIW (Banca d'Italia).

FIGURE 2. Italy 2002–2006–2010: Individual Opportunity Growth Incidence Curve



Source: Authors' calculation from SHIW (Bank of Italy).

and the individual OGIC explain the diverging trends of inequality of outcome compared to the trend of IOp: inequality of outcome decreases, whereas IOp increases.

For the second period, the 2006–10 individual OGIC lies below zero for most of the distribution, suggesting that growth generates a reduction in the values of the opportunities enjoyed by individuals. In particular, it appears that the highest cost of the recession is borne by the individuals in the poorest quantiles of the smoothed distributions. Furthermore, similar to the previous period, the individual OGIC for 2006–10 shows an increasing trend, implying that growth might have acted by worsening opportunity inequality. The severe consequences of the recession are also captured by the two synthetic measures of growth, which both take a negative value.

Turning now to the comparison of the two episodes, the results are clear. The individual OGIC of 2002–06 lies always above the individual OGIC of 2006–10, and the dominance is statistically significant at all points of the curves.²⁷

Hence, the growth process in 2002–06 dominates the growth process in 2006–10 when both the extent of growth and progressivity components are considered. However, if we want to focus exclusively on their opportunity-redistributive impact (that is, on the extent to which these processes act by increasing or reducing IOp), the dominance is not clear because they both show a

27. This dominance is confirmed by the comparison of their cumulative individual OGICs (figures and data available upon request).

regressive pattern. It can be helpful, in this case, to compare the values of their respective opportunity-equalizing indexes, which show that 2002–06 is, with statistical significance, less regressive than 2006–10.

We can conclude that both of the income dynamics under scrutiny act by increasing IOp. However, whereas this trend is consistent with the change in outcome inequality in the second period, in the first period, the variation of outcome inequality and the variation of opportunity inequality are in the opposite direction. This result reveals that a conflict may arise in the evaluation of growth when these two different perspectives are adopted for the assessment of the same growth process.

It is interesting to examine why such a conflict arises. If inequality between types increases while overall outcome inequality declines, the within-type share of total inequality must necessarily decline.²⁸ From this perspective, it may be helpful to look at Figure 3, which reports the GICs within types for the nine poorest and the nine richest types in each process. As expected, growth is progressive in both the poorest and richest types, with an higher average growth in the richest type.²⁹ This within-type dynamic explains the divergence between the income- and opportunity-based distributional assessments.

Turning the focus to the type-specific growth, the picture changes dramatically. The type OGIC for 2002–06, reported in Figure 4, does not always lie above zero for the whole distribution; in particular, the types ranked 3 and 15 experience a loss. Most importantly, the shape of the type OGIC differs significantly from the shape of the individual OGIC. According to this perspective, growth can no longer be classified as regressive. For the Italian case, this is equivalent to saying that households whose heads were born in the South grow, on average, less than households with different geographical origins.³⁰

The type population share and the anonymity implicit in the individual OGIC explain why a regressive individual OGIC is coupled with a non-regressive type OGIC. The smoothed distribution, constructed to evaluate distributional phenomena from an EOp perspective, ranks the types according to their average income at each point in time. Hence, growth is evaluated by comparing the average of different types whenever there is a reranking of types over time. In contrast, the type OGIC tracks types over time. Hence, types are ranked according to their average income at the initial period of time. Whenever there is a reranking of types over time, some GIC-OGIC divergence may emerge.

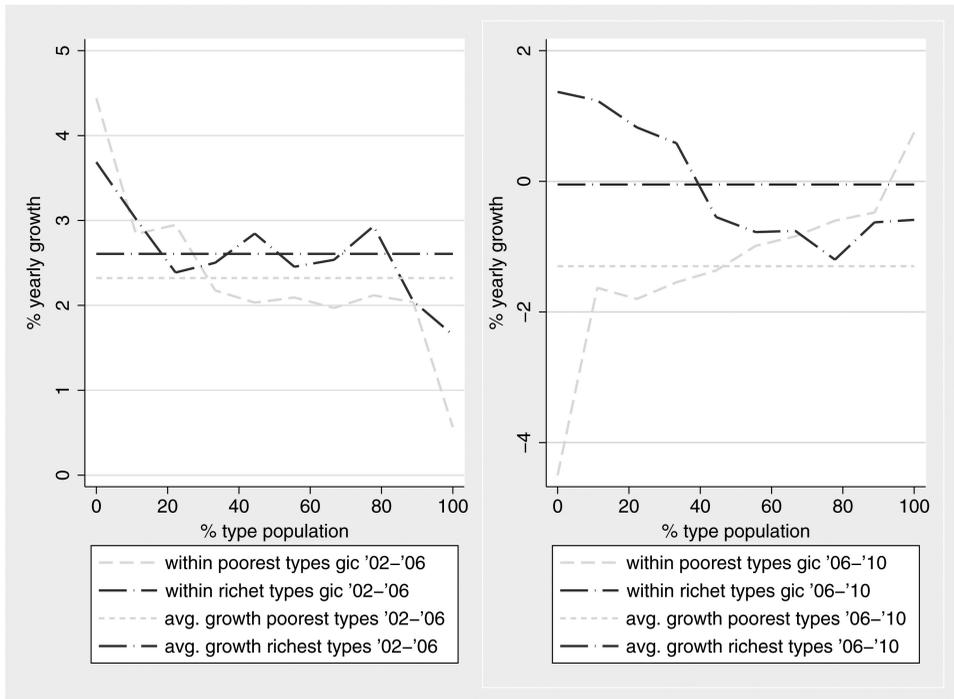
For the second growth process, the 2006–10 type OGIC shows some similarity to the individual OGIC of the same period. In particular, most of the types experience a reduction in the value of their opportunity set, and this reduction is higher for the disadvantaged types. In sum, both the individual and the type

28. Note that in these empirical applications, the inequality measure used is additively decomposable for within and between groups.

29. We aggregate types to have sufficient observations in each quantile of the within-type GIC.

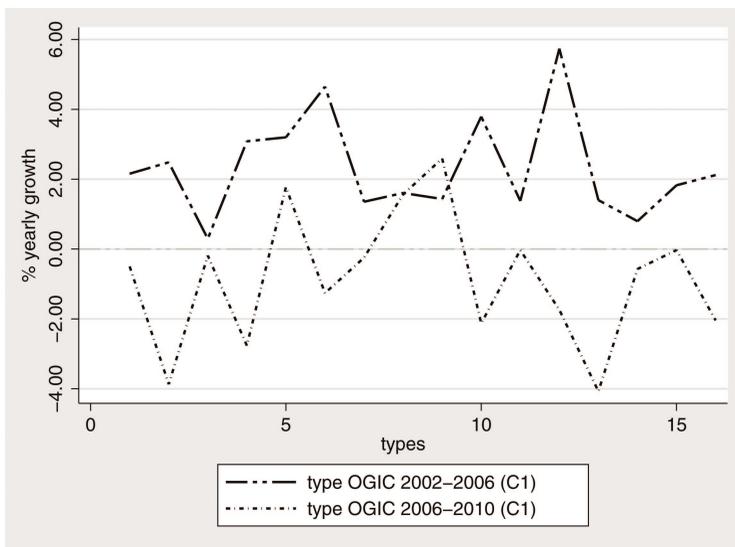
30. As reported in Table 1, the circumstance “head born in the South” appears in the five poorest types in 2002 and 2010 and in the four poorest types in 2006.

FIGURE 3. Italy 2002–2006–2010: Within-Types Growth Incidence Curve



Source: Authors' calculation from SHIW (Bank of Italy).

FIGURE 4. Italy 2002–2006–2010: Type Opportunity Growth Incidence Curve



Source: Authors' calculation from SHIW (Bank of Italy).

OGIC confirm the negative impact of the crisis in terms of the extent of opportunity and the distribution of opportunity.

Interestingly, the only three types that demonstrate positive growth in this period share the circumstance of coming from central Italy. This finding is consistent with the reduction of between-region inequality in Italy due to their different rates of income decline during the recent economic recession. Whereas the North-South gap remained stable, the recession narrowed the gap between the North and the Centre. Among the reasons that may explain this trend is the negative performance of incomes in the North during the recent slowdown, which is generally attributed to the decline of the car industry and other manufacturing sectors, largely developed in Piedmont and Friuli-Venezia-Giulia (Istat, 2012). A severe crisis in the agricultural sector and a growing service industry (especially in the health care sector) may explain, at least in part, the diverging trend of the Southern and Central regions.

The comparison of the two growth episodes is less clear because they have a specular shape: types that benefit most from growth during the first process are those that lose more during the second. The two type OGICs intersect more than once; hence, it is not possible to establish a ranking between the two growth processes.³¹ It is possible to obtain an unambiguous ordering by weakening the dominance conditions and comparing the cumulative type OGICs. We find that the first process dominates the second and that this dominance is always statistically significant. This result is also supported by the comparison of the synthetic measures of growth between the two periods. The index evaluating the extent of growth, with concern for the growth experienced by the initially disadvantaged types, is positive for the first period and negative for the second, and their difference is statistically significant (see Table 3).

It is not an easy task to understand the driving forces of these transformations. Given that, by definition, the rank of types and income are correlated, it is extremely difficult to disentangle the changes that may have affected, in opposite directions, the distribution of outcome and the distribution of opportunities.³² However, the trend of the North-South divide and labor market reforms may be considered among the determinants of redistribution since 2002. First, the different reforms realized in the recent past to reduce the gap in the opportunities accessible to different individuals have not been able to fulfill the desired goal. In particular, as shown by Pavolini (2011), among others, different public services, particularly different measures and interventions of the welfare state, are still suffering from territorial divergences with consequences in terms of an increase in IOp over time, as witnessed by the lower growth rates experienced by the Southern types.

31. Although the first process is better than the second and the dominance is statistically significant for most of the types, for type 15, the second process is preferred to the first one with statistical significance.

32. This may be a challenging question for future research.

TABLE 3. Italy: 2002–2006–2008 Complete Rankings and Inequality

	2002	2006	2010
$\mu(y)$ eq.	20116.82 (4735.42)	21692.12 (5275.08)	21117.34 (5445.91)
mld (all)	0.1422 (0.0026)	0.1301 (0.0021)	0.1437 (0.0027)
mld (between)	0.0256 (0.0006)	0.0274 (0.0001)	0.0313 (0.0007)
G_{Ys}			'06-'10
OG_{Ys}			1.821 (0.0145)
$G_{Y\mu}$			-0.9532 (0.0155)
			-2.869 (0.0244)
			-0.2340 (0.3707)

Note: *mld* = mean logarithmic deviation or generalized entropy index with parameter 0, G_{Ys} = EOp consistent aggregate measures of growth (eq. 9), OG_{Ys} = EOp consistent aggregate measures of growth progressivity (eq. 10), $G_{Y\mu}$ = Aggregate measure of between-type inequality of growth (eq. 11); 95 percent bootstrapped standard errors are reported in parenthesis.

Source: Authors' calculations on SHIW (Banca d'Italia).

Second, the labor market reforms introduced in 1998 and extended in 2000 and 2003, which mainly aimed to reduce the labor protection legislation (particularly for temporary workers), have increased wage flexibility and job turnover, increasing the “instability” in the opportunity faced by individuals (Jappelli and Pistaferri, 2009). This instability may explain why growth appears more opportunity regressive in the second period, a period of crisis. Boeri and Garibaldi (2007) suggest that although job flexibility generates instability, it may provide more job opportunities during periods of positive growth. This is not the case during recessions because these workers, in all categories of atypical job contracts, are more likely to be fired and are often excluded from social security benefits. We suggest that such an effect has been stronger in the southern regions, thereby explaining the territorial gradient in the diverging trends of different types.

Opportunity and Growth in Brazil: The Data

Our theoretical framework may be of particular interest in the analysis of developing and emerging economies that experience lively growth processes with a dramatic impact on poverty and redistribution. For this reason, the second country considered in this paper is Brazil. To perform this analysis, the 2002, 2005, and 2008 waves of the Brazilian Pesquisa Nacional por Amostra de Domicílios (PNAD), a representative survey of the Brazilian population, are used.

The unit of observation is the household, and the individual outcome is measured as the monthly household equivalent income, expressed in 2008 Brazilian real.³³ Household income is computed as the sum of all household members’ individual incomes, including earnings from all jobs, and all other reported income, including income from assets, pensions, and transfers.

The population is partitioned into 15 types using the information on two circumstances: region of birth and race. Region of birth is coded in five categories (North, Northeast, Southeast, South, Center-west), and race is coded in three categories (white/east Asian, black/mixed race, and indigenous). Individuals who were born abroad and those classified as “other” for the variable race are excluded because the number of observations is too low to make appropriate inference. Hence, the sample sizes of each wave considered in this analysis are as follows: 366,388 households in 2002, 390,046 in 2005, and 372,581 in 2008.³⁴

The full opportunity profiles for the three waves are reported in Table 4 in the appendix.³⁵ In this table, it is clear that race is the main determinant of the disparity in opportunities. Consistent with a number of contributions on socioeconomic inequality in Brazil, racial relationships appear to be the major source of outcome and opportunity inequality in Brazil (Telles 2004; Bourguignon et al. 2007; among others).

33. Equivalent income is obtained by dividing total income by the square root of the household size.

34. Again, the richest and poorest 1 percent of the household distribution in each wave are dropped.

35. All estimates are based on the sample weights according to Silva et al. (2002).

TABLE 4. Brazil: 2002 – 2005 – 2008 Descriptive Statistics and Partition in Types

Region	Race	rank ⁰²	sample ⁰²	q _i ⁰²	μ _i ⁰²	rank ⁰⁵	sample ⁰⁵	q _i ⁰⁵	μ _i ⁰⁵	rank ⁰⁸	sample ⁰⁸	q _i ⁰⁸	μ _i ⁰⁸
Northeast	black-mixed	1	91118	0.2227	516.73	2	97846	0.2229	550.09	1	93547	0.2272	695.64
Northeast	indigenous	2	299	0.0007	576.47	6	309	0.0006	702.42	2	398	0.0010	715.49
North	black-mixed	3	25874	0.0381	631.47	3	35053	0.0542	604.64	3	33200	0.0556	769.59
South	black-mixed	4	10121	0.0270	683.06	7	11549	0.0292	748.19	6	12006	0.0319	937.98
Southeast	black-mixed	5	42007	0.1448	768.61	9	48800	0.1606	806.90	8	47725	0.1633	969.41
Center-west	black-mixed	6	16052	0.0300	777.33	8	17223	0.0306	799.66	10	17472	0.0321	1006.28
Center-west	indigenous	7	154	0.0003	806.05	1	136	0.0002	444.41	4	175	0.0003	859.83
Northeast	white-east asian	8	42720	0.1094	821.07	10	42911	0.1017	823.36	9	40880	0.1018	975.68
South	indigenous	9	119	0.0002	866.19	5	128	0.0003	628.87	7	183	0.0005	940.65
North	indigenous	10	98	0.0002	879.60	4	206	0.0002	622.59	5	236	0.0003	861.65
North	white-east asian	11	9916	0.0146	970.79	11	11088	0.0167	903.47	11	9942	0.0164	1102.10
Southeast	indigenous	12	117	0.0004	1082.98	12	105	0.0004	1011.33	12	153	0.0005	1192.87
South	white-east asian	13	49021	0.1311	1169.46	14	49133	0.1244	1229.42	14	44957	0.1198	1456.16
Center-west	white-east asian	14	12717	0.0244	1179.96	13	13147	0.0238	1176.54	13	12642	0.0236	1433.06
Southeast	white-east asian	15	66055	0.2561	1385.93	15	62412	0.2341	1387.16	15	59065	0.2255	1613.84

Note: Types are ranked in ascending order according to the average income at the beginning of each growth period.

Source: Authors' calculations on PNAD (Instituto Brasileiro de Geografia e Estatística).

To analyze the distributional impact of growth in Brazil according to the EOp perspective, two three-year period growth processes are considered: 2002–05 and 2005–08. The choice of these particular periods is driven by the observation that during these years, Brazil experienced quite diverging economic trends. The former was a period of economic slowdown; the PNAD data record an increase in the overall mean income of only 0.26 percent. In contrast, the latter period was a period of pronounced growth, with an overall mean income growth of approximately 6.36 percent.

Opportunity and Growth in Brazil: The Results

As in the first illustration, we begin this analysis with the assessment of growth according to the equality of outcome perspective. The GICs for the two periods considered are reported in Figure 5.

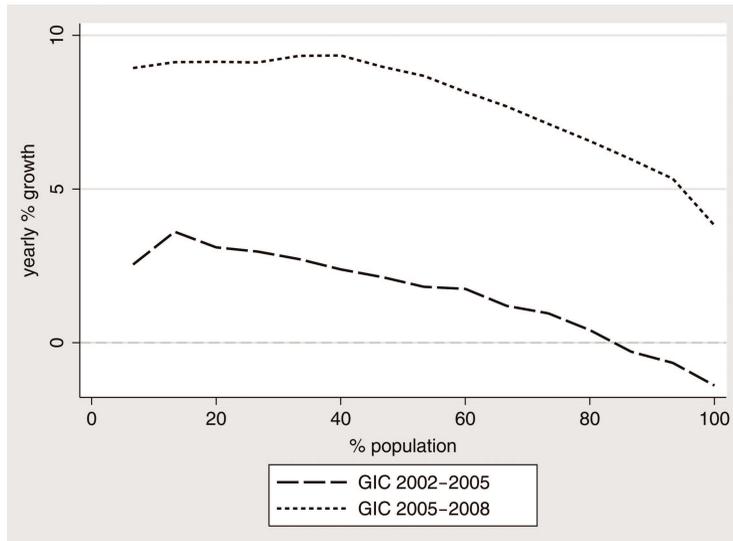
Although both curves lie almost always above zero, growth is outstanding in the second period. In fact, it is possible to unambiguously order the two growth processes because the difference between the GIC coordinates in the two periods is always statistically significant (see Table 5 in the appendix). The redistributive impact of the two processes is very similar. The respective curves are both neatly decreasing, demonstrating that growth acts by alleviating outcome inequality.

We now proceed in the evaluation of the Brazilian growth by endorsing an opportunity-egalitarian perspective. The individual OGICs for the two growth episodes are reported in Figure 6.

One feature stands out. For the 2002–05 growth episode, although the GIC lies almost always above zero, the individual OGIC is positive only for half of the smoothed distribution. This conflict indicates that although the majority of households experience positive growth, the extent of the losses borne by the richest 15 percent is substantial in determining the change in the value of the opportunity sets. This effect is plausible whenever the richest households are not concentrated only in the richest type; that is, income quantiles and types are not perfectly correlated, as for the case of Brazil during 2002–05.

This does not happen during 2005–08, when the individual OGIC lies above zero, implying that growth plays a positive role in determining an improvement of the opportunities faced by the entire population. As a result, the second process also dominates the first when an opportunity-egalitarian perspective is adopted, and the dominance is statistically significant (see Table 5 in the appendix). The sign of the dominance is also confirmed by the plot of the cumulative individual OGIC. The progressivity of the two growth episodes is clarified by the decreasing shape of the two curves. These results are further supported by the estimation of the synthetic measures of growth. The index capturing the opportunity-sensitive extent of growth is positive for both the 2002–05 and 2005–08 processes, but it is higher for 2005–08. In the same way, the value of the index capturing the progressivity of growth, in terms of equality of

FIGURE 5. Brazil: 2002–2005–2008 Growth Incidence Curve



Source: Authors’ calculation from PNAD (Instituto Brasileiro de Geografia e Estatística).

TABLE 5. Brazil: 2002–2005–2008 Dominance Conditions

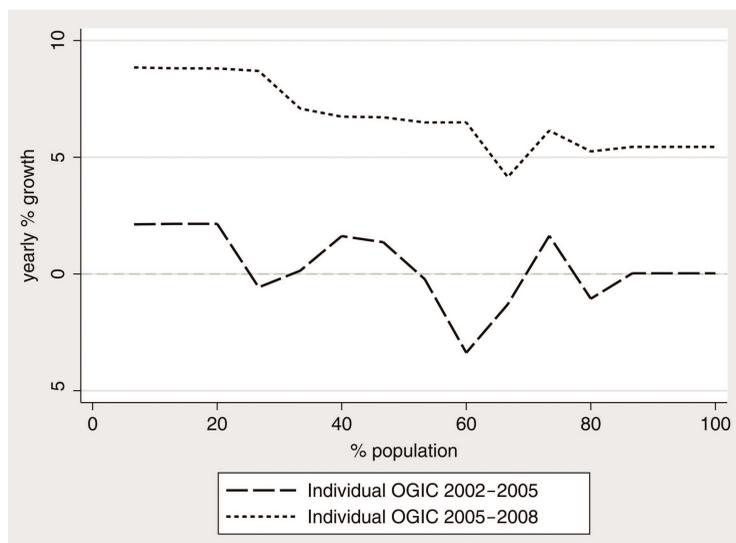
quantiles/ types rank	GIC	type OGIC	cum. type OGIC	individual OGIC	cum. individual OGIC
1	5.9040***	29.4150***	29.4150***	9.7517***	9.7517***
2	6.3070***	0.9296	13.9522***	5.3240***	7.3602***
3	6.5042***	10.7992***	12.6996***	5.4259***	6.6773***
4	6.6490***	8.7660**	11.5471***	6.5167***	6.6375***
5	6.7888***	18.3645***	12.9442***	8.3194***	7.0596***
6	6.9937***	-1.1505	10.0393***	6.6153***	6.9932***
7	6.6517***	23.7151***	12.5330***	6.6142***	6.9419***
8	6.6463***	9.3222***	12.0397***	6.6130***	6.9018***
9	6.6600***	16.2690***	12.5658***	6.6119***	6.8700***
10	6.3123***	15.8542***	12.9423***	6.8518***	6.8707***
11	6.2099***	8.8956***	12.4650***	6.6791***	6.8519***
12	6.0596***	5.1427	11.5463***	5.3881***	6.7002***
13	6.0661***	5.6651***	10.8799***	5.3695***	6.5675***
14	6.2796***	6.6171***	10.4224***	5.2087***	6.4396***
15	5.7342***	5.4429***	9.8781***	5.2070***	6.3336***

* = 90 percent, ** = 95 percent, *** = 99 percent are significance levels for the difference between curves obtained by 1,000 bootstrap replications of the statistics.

Source: Authors’ calculations on PNAD (Instituto Brasileiro de Geografia e Estatística).

opportunity, is positive for both processes. This means that during the two periods, growth acts by alleviating the disparities in opportunities, but this effect is stronger for the 2005–08 process (see Table 6 in the appendix).

FIGURE 6. Brazil: 2002–2005–2008 Individual Opportunity Growth Incidence Curve



Source: Authors' calculation from PNAD (Instituto Brasileiro de Geografia e Estatística).

Similar features characterize the assessment of growth when the focus is on the type-specific growth. Figure 7 reports the type OGICs for the 2002–05 and 2005–08 periods.

Regarding the first period, it is possible to observe that, consistent with the individual OGIC, most of the types experience a reduction in the value of their opportunity set. These types particularly include households with an indigenous head.³⁶ However, the curve does not appear to show a clear pattern; it is progressive for the lowest part of the distribution up to type 7 and then takes a clear regressive shape. The unstable trend is confirmed by the negative value of the opportunity-sensitive growth measure. It can thus be inferred that the negative growth experienced by certain types more than compensates for the positive growth experienced by the poorest types.

For the second period, the positive distributional implications of the growth process are again confirmed by the type-specific OGIC. All types experience an increase in the values of their opportunity set with a quite progressive trend. These results are supported by the positive value of the index measuring the extent of opportunity-sensitive growth (see Table 6 in the appendix). Thus, we can conclude that this growth process is beneficial in terms of opportunity when both size and distributional aspects are considered.

36. However, recall that this curve does not take into account the relative size of types. In this specific case, in fact, the types that experience an increase in the value of their opportunity set represent over 90 percent of the population.

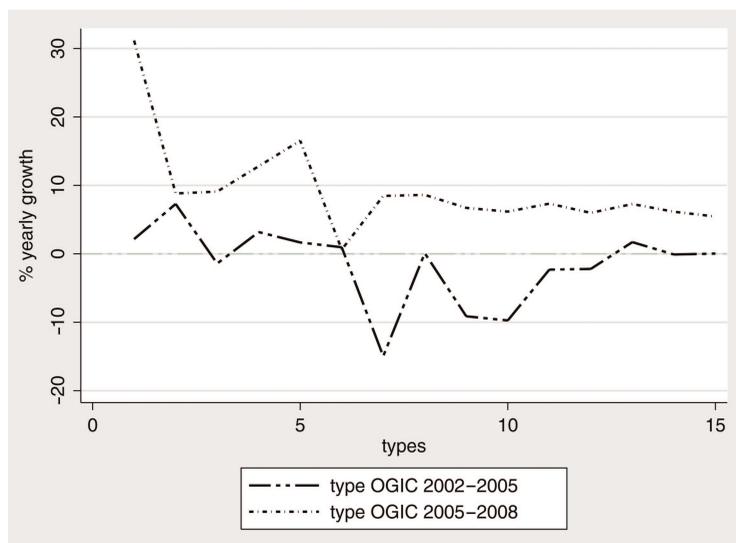
TABLE 6. Brazil: 2002–2005–2008 Complete Rankings and Inequality

	2002	2005	2008
N	366,388	390,046	372,581
$\mu(y)$ eq.	934.66 (333.55)	937.057 (324.13)	1113.48 (355.26)
mld (all)	0.4738 (0.0014)	0.4327 (0.00131)	0.3922 (0.0010)
mld (between)	0.0672 (0.0004)	0.0618 (0.0004)	0.0512 (0.0003)
avg. growth.			'02-'05
G_{Ys}			0.26%
OG_{Ys}			0.7547 (0.0910)
G_Y			0.4384 (0.0559)
			-1.0213 (0.6120)
			'05-'08
			6.36%
			7.4937 (0.1169)
			0.7891 (0.0566)
			9.5304 (1.0758)

Note: *mld* = mean logarithmic deviation or generalized entropy index with parameter 0, G_{Ys} = EOp consistent aggregate measures of growth (eq. 9), OG_{Ys} = EOp consistent aggregate measures of growth progressivity (eq. 10), $G_{Y\mu}$ = Aggregate measure of between type inequality of growth (eq. 11), 95 percent bootstrapped standard errors are reported in parenthesis.

Source: Authors' calculations on PNAD (Instituto Brasileiro de Geografia e Estatística).

FIGURE 7. Brazil: 2002–2005–2008 Type Opportunity Growth Incidence Curve



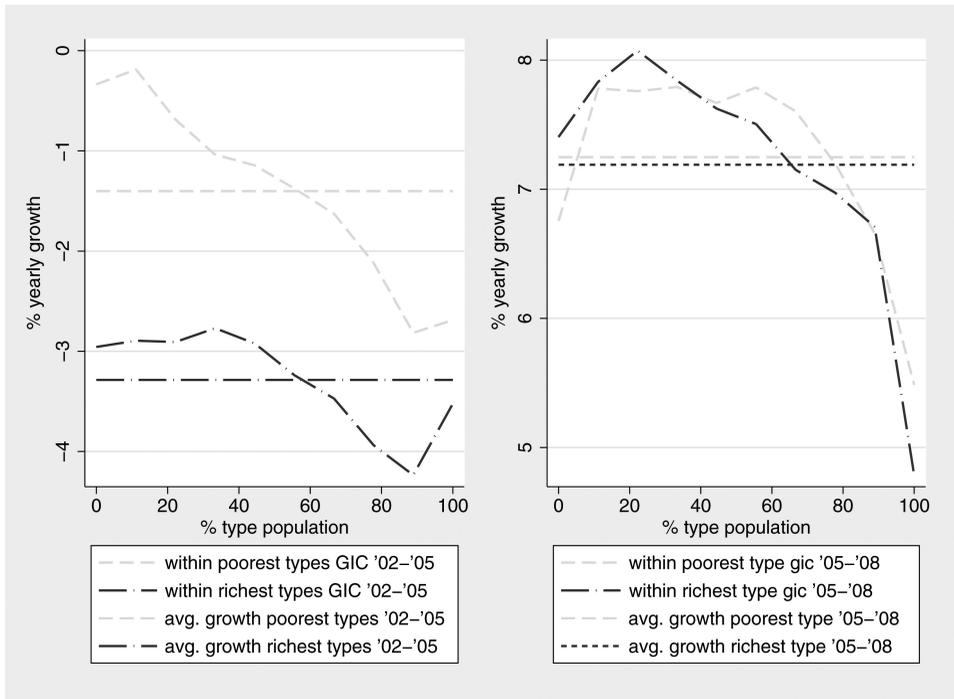
Source: Authors' calculation from PNAD (Instituto Brasileiro de Geografia e Estatística).

As is reasonable to expect, the comparison of the two processes highlights an unambiguous dominance of the second period growth over the first. The difference in the OGIC coordinates is statistically significant for almost all types, as shown in Table 5 in the appendix. For robustness purposes, we also test the difference of the respective cumulative type OGICs coordinates, which is clearly statistically significant for all types, and the difference, which is again significant, of their aggregate index of growth (see Table 6 in the appendix).

Finally, Figure 8, reporting the within-type GICs, explains how the progressive growth of Brazil between 2002 and 2005 is the joint effect of a reduction of between- and within type inequality. The four within-type GICs are downward sloping, and the average growth rate in the poorest seven types is higher in both cases than the same rate in the eight richest types.

This considerable change in the overall inequality for the time span considered is well known in the literature. Ferreira et al. (2008) suggest a number of determinants of this change: the decline in inequality between educational subgroups, a reduction in the urban-rural gap, a reduction of inequalities between racial groups, a dramatic increase in the minimum wage, and improvements in social protection programs. Clearly, these variables have a direct impact on inequality of outcome and on the distribution of opportunities. Moreover, our analysis shows that these growth processes have been beneficial in terms of improving opportunities and that Brazil has experienced an impressive increase in the degree of EOp, particularly during the 2002–05 period. Our conclusions complement

FIGURE 8. Brazil: 2002–2005–2008 Within-Types Growth Incidence Curve



Source: Authors' calculation from PNAD (Instituto Brasileiro de Geografia e Estatística).

the findings of [Molinas et al. \(2011\)](#), who look at the development of IOp in Brazil with a specific focus on the opportunities of children.

CONCLUSIONS

In this paper, we have argued that a better understanding of the relationship between inequality and growth can be obtained by shifting the analysis from final achievements to opportunities.

To this end, we have introduced the individual OGIC and the type OGIC. The former can be used to infer the role of growth in the evolution of IOp over time. The latter can be used to evaluate the income dynamics of specific groups of the population. For both versions of the OGIC, we have also proposed aggregate indices that can be used to measure the distributional impact of growth from the EOp perspective when it is not possible to rank growth episodes through the use of curves. We have shown that possible divergences in the rankings obtained through the use of the individual OGIC and the type OGIC are mostly due to demographic issues.

We have provided two empirical applications, for Italy and for Brazil. These analyses show that the measurement framework we have introduced can be used to complement existing tools for the evaluation of the distributional implications of growth. Moreover, our tools appear to be potentially relevant for the

understanding of the joint dynamic of income inequality and inequality of opportunity. Another field of application of our framework is the analysis of tax-benefit systems of reforms. Typically, the distributional aspects of these reforms are analyzed through microsimulation techniques and are evaluated in terms of income inequality reduction. Comparing reforms with the help of the tools developed in this paper, which allow the evaluation of the IOP reduction, seems a promising path for future research.

APPENDIX

Proof of Remark 1 *We start by showing the sufficiency that the individual OGIC implies the type OGIC dominance. Let the two type OGICs be defined as follows: $\tilde{g}^{Ao}(i/n_A) = \tilde{\mu}_i^A(y_{t+1})/\mu_i^A(y_t) - 1 \quad \forall i \in \{1, \dots, n_A\}$ and $\tilde{g}^{Bo}(i/n_B) = \tilde{\mu}_i^B(y_{t+1})/\mu_i^B(y_t) - 1 \quad \forall i \in \{1, \dots, n_B\}$. If (i) holds and there is type OGIC dominance between the two growth processes G^A and G^B , we will have the following:*

$$\tilde{g}^{Ao}\left(\frac{i}{n}\right) \geq \tilde{g}^{Bo}\left(\frac{i}{n}\right) \Leftrightarrow \frac{\tilde{\mu}_i^A(y_{t+1})}{\mu_i^A(y_t)} \geq \frac{\tilde{\mu}_i^B(y_{t+1})}{\mu_i^B(y_t)}, \forall i \in \{1, \dots, n\}. \tag{12}$$

If (iii) holds, the type OGIC dominance of the growth processes G^A over G^B will be

$$\tilde{g}^{Ao}\left(\frac{i}{n}\right) \geq \tilde{g}^{Bo}\left(\frac{i}{n}\right) \Leftrightarrow \frac{\mu_i^A(y_{t+1})}{\mu_i^A(y_t)} \geq \frac{\mu_i^B(y_{t+1})}{\mu_i^B(y_t)}, \forall i \in \{1, \dots, n\} \tag{13}$$

where $\mu_i(y_{t+1})$ is the mean income of the type ranked i in the final distribution of the types' mean income, which, under (iii), corresponds to $\tilde{\mu}_i(y_{t+1})$. Now, let the two individual OGICs be defined as follows: $g_{Y^s}^{Ao}(j/N_A) = \mu_j^{A_{t+1}}/\mu_j^{A_t} - 1 \quad \forall j = 1, \dots, N_A$ and $g_{Y^s}^{Bo}(j/N_B) = \mu_j^{B_{t+1}}/\mu_j^{B_t} - 1 \quad \forall j = 1, \dots, N_B$. (i) and (ii) implies $N_A = N_B$. Hence, if there is individual OGIC dominance of the growth process G^A over G^B , we will have the following:

$$g_{Y^s}^{Ao}\left(\frac{j}{N}\right) \geq g_{Y^s}^{Bo}\left(\frac{j}{N}\right) \Leftrightarrow \frac{\mu_j^{A_{t+1}}}{\mu_j^{A_t}} \geq \frac{\mu_j^{B_{t+1}}}{\mu_j^{B_t}} \forall j \in \{1, \dots, N\}. \tag{14}$$

Now, for the individuals j belonging to type i , given (ii) and because we use smoothed income, we can write eq. (13) in terms of (14):

$$\tilde{g}^{Ao}\left(\frac{i}{n}\right) \geq \tilde{g}^{Bo}\left(\frac{i}{n}\right) \Leftrightarrow \sum_{j=1}^{m_{i+1}} \frac{\mu_j^{A_{t+1}}}{\mu_j^{A_t}} \geq \sum_{j=1}^{m_{i+1}} \frac{\mu_j^{B_{t+1}}}{\mu_j^{B_t}} \forall i \in \{1, \dots, n\}. \tag{15}$$

If eq. (14) holds, than it must be the case that the dominance in their type aggregation holds, providing the dominance in eq. (15). Hence we have proved the sufficiency of the remark.

We now prove the necessary condition by contradiction.

Suppose that eq. (15) holds. Now, pick a type $i \in \{1, \dots, n\}$. Assume that for that type $\exists k \{1, \dots, m_i\}$ such that $\mu_k^{At+1}/\mu_k^{At} < \mu_k^{Bt+1}/\mu_k^{Bt}$, then because all individuals in the same type are given the same mean income, $\sum_{j=1}^{m_i} \frac{\mu_j^{At+1}}{\mu_j^{At}} - \sum_{j=1}^{m_i} \frac{\mu_j^{Bt+1}}{\mu_j^{Bt}} < 0$ for a given type i , contradicting eq. (15). QED

REFERENCES

- Aaberge, R., M. Mogstad, and V. Peragine. 2011. "Measuring Long-term Inequality of Opportunity." *Journal of Public Economics* 95 (3–4): 193–204.
- Arneson, R.J. 1989. "Equality of Opportunity for Welfare." *Philos. Stud* 56: 77–93.
- Atkinson, A.B., and A. Brandolini. 2010. "On Analyzing the World Distribution of Income." *World Bank Economic Review* 24 (1): 1–37.
- Banca d'Italia. 2008. "Supplemento al Bollettino Statistico Note Metodologiche e informazioni Statistiche." *I Bilanci delle Famiglie Italiane nell'Anno 2006 Anno XVIII* 7.
- . 2012. "Supplemento al Bollettino Statistico Note Metodologiche e informazioni Statistiche." *I Bilanci delle Famiglie Italiane nell'Anno 2010 Anno XXII* 25.
- de Barros, R.P., F.H.G. Ferreira, J.R. Vega, and J.S. Chanduvi. 2009. *Measuring Inequality of Opportunity in Latin America and the Caribbean*. Washington, DC: World Bank.
- Boeri, T., and P. Garibaldi. 2007. "Two Tier Reforms of Employment Protection: a Honeymoon Effect?" *Economic Journal* 117 (521): 357–85.
- Bourguignon, F. 2004. "The Poverty-Growth-Inequality Triangle." Working Paper 125. Indian Council for Research on International Economic Relations, New Delhi, India.
- . 2011. "Non-anonymous Growth Incidence Curves, Income Mobility and Social Welfare Dominance." *Journal of Economic Inequality* 9 (4): 605–27.
- Bourguignon, F., F.H.G. Ferreira, and M. Menéndez. 2007. "Inequality Of Opportunity In Brazil." *Review of Income and Wealth* 53 (4): 585–18.
- Brunori, P., F.F.G. Ferreira, and V. Peragine. 2013. "Inequality of Opportunity, Income Inequality and Economic Mobility: Some International Comparisons" Policy Research Working Paper 6304. World Bank, Washington, DC.
- Cecchi, D., and V. Peragine. 2010. "Inequality of Opportunity in Italy." *Journal of Economic Inequality* 8 (4): 429–50.
- Cohen, G.A. 1989. "On the Currency of Egalitarian Justice." *Ethics* 99: 906–44.
- Datt, G., and M. Ravallion. 2011. "Has India's Economic Growth Become More Pro-Poor in the Wake of Economic Reforms?" *World Bank Economic Review* 25 (2): 157–89.
- Dworkin, R. 1981a. "What is Equality? Part 1: Equality of Welfare." *Philos. Public Affairs* 10: 185–246.
- . 1981b. "What is Equality? Part 2: Equality of Resources." *Philos. Public Affairs* 10: 283–345.
- Essama-Nssah, B. 2005. "A Unified Framework for Pro-poor Growth Analysis." *Economics Letters* 89: 216–21.
- Essama-Nssah, B., and P. Lambert. 2009. "Measuring Pro-pooriness: a Unifying Approach with New Results." *Review of Income and Wealth* 55 (3): 752–78.
- Ferreira, F.H.G. 2010. "Distributions in Motion. Economic Growth, Inequality, and Poverty Dynamics." Policy Research Working paper 5424. World Bank, Washington, DC.

- Ferreira, F.H.G., and J. Gignoux. 2011. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *The Review of Income and Wealth* 57 (4): 622–57.
- Ferreira, F.H.G., P.G. Leite, and J.A. Litchfield. 2008. "The Rise And Fall Of Brazilian Inequality: 1981–2004." *Macroeconomic Dynamics* 12 (S2): 199–230.
- Fleurbaey, M. 2008. *Fairness, Responsibility and Welfare*. 1st Edition. New York: Oxford University Press.
- Fleurbaey, M., and V. Peragine. 2013. "Ex ante versus ex post Equality of Opportunity." *Economica* 8 (317): 118–30.
- Istat. 2012. "Statistiche Report, Conti Commerciali Regionali 2007-2009." Istat.
- Jappelli, T., and L. Pistaferri. 2009. "Does Consumption Inequality Track Income Inequality in Italy?" *Review of Economic Dynamics* 13 (1): 133–53.
- Jenkins, S., A. Brandolini, J. Micklewright, and B. Nolan. 2013. *The Great Recession and the Distribution of Household Income*. Oxford: Oxford University Press.
- Jenkins, S.P., and P. Van Kerm. 2011. "Trends in Individual Income Growth: Measurement Methods and British Evidence." Discussion Paper 5510. IZA, Bonn, Germany.
- LeFranc, A., N. Pistoiesi, and A. Trannoy. 2008. "Inequality Of Opportunities Vs. Inequality Of Outcomes: Are Western Societies All Alike?" *Review of Income and Wealth* 54 (4): 513–46.
- . 2009. "Equality Of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France." *Journal of Public Economics* 93 (11–12): 1189–1207.
- Lokshin, M. 2008. GICURVE version 2.15 [Computer program]. Available at <http://adeptanalytics.org/download/ado/gicurve/gicurve.ado>
- Luongo, P. 2011. "The Implication of Partial Observability of Circumstances on the Measurement of IOp." *Research on Economic Inequality* 19 (2): 23–49.
- Marrero, G.A., and J.G. Rodríguez. 2010. "Inequality of Opportunity and Growth." Working Papers 154. ECINEQ, Society for the Study of Economic Inequality. University of Catania, Italy.
- Molinas Vega, J., R.P. de Barros, J. Chanduvi, and M. Giugale. 2011. *Do Our Children Have a Chance?* Washington, DC: World Bank.
- Nilsson, W. 2005. "Opportunities, Preferences and Incomes." *Umea Economic Studies*, 649.
- O'Neill, D., O. Sweetman, and D. Van de Gaer. 2000. "Equality of Opportunity and Kernel Density Estimation." In T.B. Fomby and R. Carter Hill, eds., *Advances in Econometrics* 14: 259–274. JAI Press, Stamford.
- Palmisano, F. 2011. "Mobility and Long Term Equality of Opportunity." *Research on Economic Inequality* 19: 51–75.
- Pavolini, E. 2011. "Welfare e dualizzazione dei diritti sociali." in U. Ascoli, ed., *Il Welfare in Italia*. Bologna: il Mulino.
- Peragine, V. 2002. "Opportunity Egalitarianism and Income Inequality: the Rank-Dependent Approach." *Mathematical Social Sciences* 44: 45–64.
- Peragine, V., F. Palmisano, and P. Brunori. 2011. "Economic Growth and Equality of Opportunity." Working Paper 232. ECINEQ, University of Catania, Italy.
- Peragine, V., and L. Serlenga. 2008. "Higher Education and Equality of Opportunity in Italy." In J. Bishop, and B. Zheng, eds., *Research in Economic Inequality* XIV.
- Ramos, X., and D. Van de Gaer. 2012. "Empirical Approaches to Inequality of Opportunity: Principles, Measures and Evidences." Working Paper 259. ECINEQ, University of Catania, Italy.
- Ravallion, M., and S. Chen. 2003. "Measuring Pro-poor Growth." *Economics Letters* 78 (1): 93–99.
- Roemer, J.E. 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
- . 2011. "Economic Development and Opportunity Equalization." Paper presented at the ABCDE World Bank Conference, Paris 2011.

- Roemer, J.E., R. Aaberge, U. Colombino, J. Fritzell, S. Jenkins, and A. LeFrancet al. 2003. "To What Extent do Fiscal Systems Equalize Opportunities for Income Acquisition among Citizens?" *Journal of Public Economics* 87: 539–65.
- Silva, P.L., D.G.C. Pessoa, and M.F. Lila. 2002. "Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral." *Ciência saúde coletiva* 7 (4): 659–70.
- Son, H. 2004. "A Note on Pro-poor Growth." *Economics Letters* 82 (3): 307–14.
- Sutherland, H., H. Immervoll, and C. O'Donoghue. 1999. "An Introduction to EUROMOD." Working Papers EM0/99. EUROMOD, Institute for Social and Economic Research, Essex, UK.
- Telles, E.E. 2004. *Race in Another America: The Significance of Skin Color in Brazil*. Princeton, NJ: Princeton University Press.
- Van de Gaer, D. 1993. *Equality of Opportunity and Investment in Human Capital*. Ph.D. Dissertation, Katholieke Universiteit Leuven, Belgium.
- World Bank. 2006. *World Development Report 2006: Equity and Development*. Washington, DC: World Bank.
- . 2012. *World Development Report 2012: Jobs*. Washington, DC: World Bank.

Children's Health Opportunities and Project Evaluation: Mexico's Oportunidades Program

Dirk Van de gaer, Joost Vandenbossche, and José Luis Figueroa

We propose a methodology to evaluate social projects from the perspective of children's opportunities on the basis of the effects of these projects on the distribution of outcomes. We condition our evaluation on characteristics for which individuals are not responsible; in this case, we use parental education level and indigenous background. The methodology is applied to evaluate the effects on children's health opportunities of Mexico's Oportunidades program, one of the largest conditional cash transfer programs for poor households in the world. The evidence from this program shows that gains in health opportunities for children from indigenous backgrounds are substantial and are situated in crucial parts of the distribution, whereas gains for children from nonindigenous backgrounds are more limited. JEL codes: I18, I38, D63

Dirk Van de gaer (corresponding author) is Professor in Economics, Vakgroep Sociale Economie and SHERPPA, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and Associate Fellow at Université Catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium. The research was completed while he was visiting IAE - CSIC, Campus UAB, 08193 - Bellaterra, Barcelona, Spain. Tel: +32-(0)9-2643490. Fax: +32-(0)9-2648996. E-mail: Dirk.Vandegaer@ugent.be. Joost Vandenbossche is a PhD student in Economics, SHERPPA, Vakgroep Sociale Economie, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and Aspirant FWO - Flanders. E-mail: Joost.Vandenbossche@UGent.be. José Luis Figueroa is a PhD student in Economics, SHERPPA, Vakgroep Sociale Economie, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and CES, Katholieke Universiteit Leuven. E-mail: joseluis.figueroaoropeza@ugent.be. This work was supported by the Belgian Program on Inter University Poles of Attraction, initiated by the Belgian State, Prime Minister's Office, Science Policy Programming [Contract No. P6/07] and by the FWO Flanders, project number 3G079112. We thank the editor, two referees, Bart Cockx, Aitor Calo Blanco, Gaston Yalonetzky, Alain Trannoy, Stefan Dercon, Francisco Ferreira, Vito Peragine, and Nicolas Van de Sijpe for many useful comments and suggestions and Jean-Yves Duclos for showing us how to incorporate the survey design into the bootstrap procedure. We gratefully acknowledge comments received on preliminary versions presented at the GREQAM-IDEF workshop "The Multiple Dimensions of Equality and Fairness" (Marseille, France, November 17, 2010), the OPHI workshop "Inequalities of Opportunities" (Oxford, UK, November 22–23, 2010), the UAB workshop "Equality of Opportunity and Intergenerational Mobility" (Barcelona, Spain, December 17, 2010), the winter school on "Inequality and Social Welfare Theory" (Canazei, Italy, January 10–13, 2011), the faculty seminar in Caen (France, March 28, 2011), the workshop "Equity in Health" (Louvain la Neuve, Belgium, May 11–13, 2011), the ABCDE conference (Paris, France, May 30–June 01, 2011), the conference "Mind the Gap: from Evidence to Policy" (Cuernavaca, Mexico, June 15–17, 2011), the conference "Micro Evidence on Innovation in Developing Countries" (San Jose, Costa Rica, June 27–28, 2011), the ECINEQ conference (Catania, Italy, July 18–20, 2011), and the EEA conference (Oslo, Norway, August 25–29, 2011). A supplemental appendix to this paper is available at <http://wber.oxfordjournals.org/>.

THE WORLD BANK ECONOMIC REVIEW, VOL. 28, NO. 2, pp. 282–310
Advance Access Publication January 7, 2013

doi:10.1093/wber/lhs032

© The Author 2013. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

This paper evaluates the change in health opportunities for children aged two to six years who participate in the Mexican Oportunidades program. Oportunidades is a large-scale, conditional cash transfer program initiated in 1998 through which poor rural households receive cash in exchange for their compliance with preventive health care requirements, nutrition supplementation, education, and monitoring. In 2010, approximately 5.8 million families participated in the program, and cash transfers to the participants totaled \$4.8 billion. The average treatment effects of the program on the health of young children have been shown to be positive (see the literature surveyed in [Parker et al. 2008](#)). We propose a methodology that focuses on the conditional cumulative distribution functions of health outcomes to identify whether and where in the distribution the program is effective for children whose parents have certain characteristics. Our methodology evaluates the program from the perspective of children's opportunities rather than average treatment effects.

[Fiszbein et al. \(2009\)](#) report that in 1997, only three developing countries (Mexico, Brazil, and Bangladesh) had conditional cash transfer programs in place; by 2008, this number had increased to 29, with many more countries planning to implement such programs. It is important to develop techniques to evaluate the effects of these programs on children's opportunities because these programs are increasingly popular in developing countries, they are sometimes conducted on a large scale, and their focus is on breaking the intergenerational poverty cycle. Despite the recent emergence of substantial empirical literature measuring inequality of opportunity (e.g., [Paes et al. 2009](#) and the references below), no such techniques currently exist.

In the recent literature on equality of opportunity (e.g., [Bossert 1995](#); [Fleurbaey 1995, 2008](#); [Roemer 1993](#)), a distinction is generally drawn between two types of factors that influence the outcome under consideration. On the one hand, there are circumstances and characteristics for which an individual is not responsible, such as race, sex, and parental background; these are the characteristics upon which we condition the cumulative distribution function. On the other hand, there are other characteristics for which individuals are considered responsible, such as having a good work ethic. The idea is that public policies, including conditional cash transfer programs, should compensate for the former while respecting the influence of the latter.¹

We apply the framework to health outcomes of children aged two to six years. We consider the following circumstances for which parents are not responsible: race, in particular, whether either parent is indigenous; educational level, determined by whether either parent had primary education; and participation in the program. Each possible combination of circumstances corresponds to a "type," in Roemer's terminology ([Roemer 1993](#)). Therefore, we

1. Recently, [Lefranc et al. \(2009\)](#) extended this framework with a third factor, random factors that are legitimate sources of inequality "as long as they affect individual outcomes and circumstances in a neutral way" (p. 1192).

have eight types. To evaluate the program, we take the health outcomes of children who belong to families enrolled in the program for each of the four types, which are defined on the basis of the parents' race and education level, and we compare those outcomes with the health outcomes of children whose parents belong to the corresponding type that was not enrolled in the program. Within each type, outcomes can (and will) differ because of factors that are unobserved and ascribed to parental responsibility, such as parental health investments in children. In section II, we argue that an opportunity perspective implies that the comparison of treatment and control types must be based on first- or second-order stochastic dominance.

The idea of using first- or second-order stochastic dominance to investigate equality of opportunity for a particular outcome is not novel. However, until now, this method has been applied only to study whether opportunities are equal within a particular population (see O'Neill et al. 2000 and Lefranc et al. 2009 for studies in which the outcome is income; see Rosa Dias 2009 and Trannoy et al. 2010 for adults' self-assessed health studies; for comparisons between different countries, see Lefranc et al. 2008 for income-based outcomes; for comparisons between regions, see Peragine and Serlenga 2008 for education-based outcomes). Our paper makes three primary contributions to this literature. First, and most important, we conduct our evaluation by establishing the effect of Oportunidades on children's health opportunities. Second, we consider opportunity in the health of young children because their health is crucial for their adult outcomes (see, e.g., Black et al. 2007 and Alderman et al. 2006) and because it is important in its own right. Third, in contrast to previous literature that tested for stochastic dominance in the context of equality of opportunity, our test procedure is based on Davidson and Duclos (2009) and Davidson (2009). Thus, we test the null of nondominance against the alternative of dominance so that rejection of the null logically entails dominance.

Most of the literature on program evaluation focuses on estimating average treatment effects. However, we are interested in establishing or rejecting stochastic dominance between the distributions of health outcomes of children when their parents are either in or out of the program. This exercise is not trivial because we cannot observe the same child both in and out of the program; in other words, we cannot simply resort to a comparison of the cumulative distributions of treatment and control types without making additional assumptions (Heckman 1992). One such assumption is perfect positive quantile dependence (see Heckman et al. 1997), which stipulates that those who are at the q th quantile in the distribution with treatment would have been at the q th quantile in the distribution without treatment. Roemer's identification axiom (Roemer 1993) is usually invoked in empirical applications of equality of opportunity when responsibility characteristics are unobserved. This axiom posits that the parents of children who are at the same percentile of their type distribution have exercised comparable responsibility. We argue below that this axiom provides a normatively inspired alternative to perfect positive quantile dependence by reducing

the problem to a comparison of the cumulative distribution functions of the corresponding treatment and control types. The literature on average treatment effects stresses that treatment and control samples must be comparable in terms of preprogram characteristics. We show that this is also imperative when testing for stochastic dominance. Following the literature on average treatment effects, we propose a propensity score matching technique on the basis of preprogram characteristics to better compare treatment and control types. Finally, it is noteworthy that two authors recently suggested incorporating stochastic dominance into project evaluation: Verme (2010) proposed a stochastic dominance approach to determine the effect of a perfectly randomized experiment based on the measures establishing poverty line dominance (i.e., dominance for a range of poverty lines) developed by Foster et al. (1984). Our approach, based on equality of opportunity, stresses that we should focus on the distributions that are conditional on circumstances instead of comparing the distributions of all treatment and control samples. Therefore, we compare the distributions of corresponding treatment and control types. Moreover, our propensity score matching technique makes this approach effective for imperfectly randomized experiments. Naschold and Barrett (2010) allow for nonrandomized treatment by focusing on stochastic dominance between treatment and control samples of the distribution of the difference in outcome, both before and after treatment. They do not focus on types, and the results are difficult to interpret because dominance in terms of differences does not imply that treatment leads to a dominating distribution, which fundamentally depends on who gains and who loses.

Our main findings are that the treatment has substantial positive effects on the health opportunities of children from indigenous families. The effects on children growing up in nonindigenous families are weaker, although we still find significant positive treatment effects for that group.

The paper is structured as follows. Section I provides definitions and explains the methodology. The data are described in section II. Section III presents the empirical results, including a discussion of the relationship with previous studies. Section IV concludes.

I. DEFINITIONS AND METHODOLOGY

Let a child's health outcome be represented by the variable $h \in H = [\underline{h}, \bar{h}] \subseteq \mathbb{R}$, and let higher values for h mean better health. A child's health is the result of two types of variables. The first variable, $c \in C$, represents circumstances and characteristics for which the child's parents are not responsible, such as race, educational background, and whether the family participates in the program.² The second

2. Race and educational background are circumstances because they should not influence the health opportunities parents can obtain for their children. Whether the family participates in the program is largely determined by the locality in which they lived at the time the program began; therefore, this is outside of parental control.

variable, $r \in R$, represents characteristics for which parents are responsible, such as health investments in children. Each combination of circumstances corresponds to a type. Social programs should improve children's opportunities, and from the perspective of the equality of opportunity literature, they should compensate for health differences that are caused by circumstances. Moreover, they should respect the influence of parental responsibility, at least to some extent (see, e.g., [Swift 2005](#) for a defense of this position).

In many empirical applications, responsibility is unobserved, as it is here. In such cases, the equality of opportunity framework is usually operationalized using the identification axiom proposed by [Roemer \(1993\)](#), which states that the parents of two children who are at the same percentile of their type distribution of health have exercised identical responsibility.³ Thus, if the cumulative distribution function of health for a type whose family participated in the program lies below the cumulative distribution function of health for the corresponding type who did not participate in the program, the type in the program needs less parental effort to obtain a particular level of child health than the type not in the program. If this holds for all levels of health, program participation unambiguously improves the opportunities for this type. Consequently, if the distribution of a type with treatment first-order stochastically dominates the distribution of the corresponding type that did not receive treatment, the program improves this type's opportunities. Similar reasoning applies to second-order stochastic dominance, with the caveat that second-order stochastic dominance can also be obtained by within-type, inequality-reducing transfers of health that do not fully respect the influence of parental responsibility.⁴ Roemer's identification axiom does not necessarily imply that we would find children with and without treatment at exactly the same q th quantile (which is the perfect positive quantile dependence found in [Heckman et al. 1997](#)); instead, it merely states that the comparison of the quantiles of the treated and corresponding untreated type is normatively relevant because it compares the health outcomes of children of parents who behaved equally responsibly.

Let $F^c(b|c)$ denote the conditional distribution of children's health for parents with circumstances c in the control sample, and let $F^T(b|c)$ denote the same distribution in the treatment sample. We say that the project improves the opportunities for the health of children with parental circumstances c if the conditional distribution $F^T(b|c)$ first-order stochastically dominates the conditional distribution $F^c(b|c)$, and we test whether first-order stochastic dominance occurs. Thus, the issue of statistical inference arises. We follow [Davidson and Duclos \(2009\)](#), starting from nondominance as the null hypothesis. To

3. See [Roemer \(1993\)](#) and [Roemer \(1998\)](#) for a defense of this principle, and see [Fleurbaey \(1998\)](#) for a discussion of the assumptions involved.

4. Fully respecting the influence of responsibility means that the health differences caused by responsibility are fully preserved by the program. Alternative notions of responsibility are weaker and require, for instance, that the program does not change the rank order of children's health. This weaker requirement is compatible with second-order stochastic dominance.

illustrate the procedure for testing first-order dominance and to describe the test more formally, let $U \subseteq H$ be the union of the supports of $F^c(h|c)$ and $F^T(h|c)$. We test the null hypothesis of nondominance of $F^c(h|c)$ by $F^T(h|c)$,

$$\max_{z \in U} (F^T(z|c) - F^c(z|c)) \geq 0$$

against the alternative hypothesis that $F^T(h|c)$ first-order stochastically dominates $F^c(h|c)$,

$$\max_{z \in U} (F^T(z|c) - F^c(z|c)) < 0.$$

This approach has the advantage of allowing us to draw the conclusion of dominance if we succeed in rejecting the null hypothesis; in other words, when the null is rejected, the only other possibility is dominance. By contrast, if dominance is the null hypothesis, as is the case in most empirical work to date, failure to reject dominance does not allow us to accept dominance. As Davidson and Duclos (2009) point out, taking nondominance as the null with continuous distributions comes at the cost that it is not possible to reject nondominance in favor of dominance over the entire support of the distribution.⁵ Rejecting nondominance is normally possible only over *restricted* ranges of the observed variable. Thus, another merit of this approach is that it allows us to identify the maximal range over the supports of the distribution for which we are able to reject the null of nondominance and, therefore, to accept dominance in favor of the project. In this way, we can check whether we have dominance over ranges of the observed variable that are of special importance, such as the range below minus two standard deviations from the reference height for standardized height, which indicates stunting.

Of course, we must use the identical procedure to test the null of nondominance of $F^T(h|c)$ by $F^C(h|c)$ against the alternative hypothesis that $F^C(h|c)$ dominates $F^T(h|c)$. If rejection occurs, we identify the maximal range over the support of the distribution for which we are able to reject the null of nondominance and to accept dominance against the project.⁶ These elements are incorporated in the following weak version of improvements in opportunities, which encompasses most of the work in this paper.

First-Order Improvements: The project leads to a first-order improvement of the opportunities of children with parental circumstances c if (a) there exists

5. Let \underline{h} be the lower bound of U . Evidently, $F^T(\underline{h}|c) - F^C(\underline{h}|c) = 0$; therefore, the maximum over U is never less than zero. Moreover, close to the boundaries of the support, there may be too little information to reject nondominance.

6. Supplemental appendix S1 contains more details about stochastic dominance tests. The appendix is available at <http://wber.oxfordjournals.org/>.

$U^0 \subseteq U$ such that we can reject the null of nondominance of $F^c(h|c)$ by $F^T(h|c)$ against the alternative that $F^T(h|c)$ dominates $F^c(h|c)$ over U^0 , and (b) there exists no $U^1 \subseteq U$ such that we can reject the null of nondominance of $F^T(h|c)$ by $F^c(h|c)$ against the alternative that $F^c(h|c)$ dominates $F^T(h|c)$ over U^1 .

Assuming that the influence of parental responsibility on children's health need not be fully respected and that health is cardinally measurable, equalizing health outcomes within type becomes desirable, and it becomes meaningful to ask whether the conditional distribution $F^T(h|c)$ second-order stochastically dominates the conditional distribution $F^c(h|c)$, if the project does not lead to a first-order improvement. Similar statistical issues arise here as for first-order stochastic dominance (see Davidson 2009), leading to the following definition.

Second-Order Improvements: The project leads to a second-order improvement of the opportunities of children with parental circumstances c if (a) the project does not lead to a first-order improvement, (b) there exists $U^0 \subseteq U$ such that we can reject the null of absence of second-order dominance $F^c(h|c)$ by $F^T(h|c)$ against the alternative that $F^T(h|c)$ second-order stochastically dominates $F^c(h|c)$ over U^0 , and (c) there exists no $U^1 \subseteq U$ such that we can reject the null of absence of second-order stochastic dominance of $F^T(h|c)$ by $F^c(h|c)$ against the alternative that $F^c(h|c)$ second-order stochastically dominates $F^T(h|c)$ over U^1 .

Finally, when comparing conditional distribution functions to evaluate a program, it is important to note that inaccurate conclusions may be drawn when preprogram characteristics are not accounted for and when they differ for the treatment types in comparison with the control types (including compensation characteristics). Suppose that we have two sets of characteristics, preprogram characteristics $x \in X$, which are not accounted for, and observable circumstances c . For the type with observed circumstances c_1 , we then have

$$\begin{aligned} F(h|c_1) &= \frac{\int_{\underline{h}}^h f(\tilde{h}, c_1) d\tilde{h}}{f(c_1)} = \frac{\int_X \int_{\underline{h}}^h f(\tilde{h}, c_1, x) d\tilde{h} dx}{f(c_1)} \\ &= \int_X \int_{\underline{h}}^h f(\tilde{h}|c_1, x) \frac{f(c_1, x)}{f(c_1)} d\tilde{h} dx = \int_X F(h|c_1, x) f(x|c_1) dx. \end{aligned}$$

This equation clearly shows that the composition of the c_1 type in terms of x matters. Indeed, suppose that the treatment has no effect ($F^c(h|c_1, x) = F^T(h|c_1, x)$), but the composition of those with circumstances c_1 differs between the control and treatment types. Suppose that $f^c(x|c_1)$ is higher than $f^T(x|c_1)$ for favorable preprogram characteristics x , or characteristics for which $F^c(h|c_1, x)$ is lower, and that $f^c(x|c_1)$ is lower than $f^T(x|c_1)$ for unfavorable preprogram characteristics. As a result, $F^c(h|c_1)$ is smaller than $F^T(h|c_1)$, and we might erroneously

infer that the treatment had an adverse effect on the opportunities of those with circumstances c_1 .

II. DATA DESCRIPTION

In this section, we describe the Oportunidades program and the construction of treatment and control samples. We describe the selection of circumstances and outcomes and examine the data used to evaluate the program.

The Oportunidades Program

The Oportunidades program is a conditional cash transfer program in which bimonthly cash transfers are provided to households in extreme poverty. The cash transfers are conditioned on the attendance of children in school, health care visits for all members of the household, and attendance at information sessions on primary health care and nutrition. Money for schooling constitutes the largest part of the conditional cash transfer. The total amount that a household receives depends on the number, age, and sex of its children. On average, households receive approximately 20 percent of their household consumption from such cash transfers.

Interventions for young children and their mothers are particularly emphasized. Prenatal and postpartum care visits, growth monitoring, immunization, and management of diarrhea and antiparasitic treatments are provided to mothers and young children. Children between the ages of four months and 23 months must have nine periodic medical check ups. From the age of 23 months until the child turns 19 years old, household members must have at least two check ups per year. Children between the ages of six and 23 months, lactating women and low-weight children between the ages of two and four years receive milk-based and micronutrient fortified foods containing the daily recommended intake of zinc, iron, and essential vitamins.⁷

Sample Design

The selection of immediate and delayed treatment samples was undertaken in several steps (see, e.g., [INSP 2005](#)). Highly deprived localities were identified by using a deprivation index computed on the basis of relevant sociodemographic data available from national censuses. Localities with at least 500 and not more than 2,500 inhabitants, that were categorized as having high or very high deprivation and that had access to an elementary school, a middle school and a health clinic were eligible for treatment. Localities were identified, and a random sample was constructed that was stratified by locality size. Within each state, localities were randomly assigned into treatment and control groups.

7. These supplements may also be given to children in households that are not receiving treatment (including children in the control sample) if signs of malnutrition are detected. This may lead to a downward bias of the estimated effect of Oportunidades (see also [Behrman et al. 2009b](#), footnote 8).

A sample of 506 localities was finally selected for the study. A random procedure assigned 320 of these localities to receive immediate treatment; the remaining 186 began receiving treatment approximately 18 months later. In the selected localities, the poverty conditions of all households were evaluated, and households categorized as experiencing extreme poverty were included in the program. This categorization was based on household income, characteristics of the head of household, and variables related to dwelling conditions. Comments by a community assembly on the inclusion and exclusion of households were considered if they met certain criteria to identify beneficiary families. The randomized design enabled us to use the immediate treatment sample as the treatment group and the delayed treatment sample as the control group.⁸ However, when we consider the effect of the program on the health outcomes of children between the ages of two and six years in 2003, most of these children grew up in families that were in the program for their entire lives. For children born before the delayed treatment began, this comparison can only show the effect of the difference in exposure when the children were young.⁹ Therefore, and because we want to limit our study to an analysis of households that actually received cash transfers (this information is not available for the initial treatment sample), our treatment sample is a subset of the delayed treatment sample.¹⁰ Once the delayed treatment sample began receiving treatment, we had to construct a new control sample, with the intention of making it as similar as possible to the treatment samples (see, e.g., Todd 2004 and Behrman et al. 2006). First, localities that did not meet the criteria for access to an elementary school, a middle school, and a health clinic were excluded. Next, a propensity score method was used that was based on data at the local level as a function of observed characteristics from the 2000 Census that permitted comparison with the localities of the original sample. This procedure led to a selection of 151 localities in which households that met the criteria for program eligibility were included in the control sample. We compare this control sample to the subset of the delayed treatment sample, as described above.

As we explained at the end of section I, the households in the treatment and control samples must be comparable in terms of preprogram characteristics.

8. Most studies focus on a comparison of the immediate and delayed treatment samples and therefore evaluate the effect of differences in duration of program participation; see, e.g., Schultz (2004), Behrman et al. (2005), or Behrman et al. (2009a).

9. In the working paper version, we repeat the analysis for children born after April 1998 (when the original treatment started) and before October 1999 (when delayed treatment started), taking the original treatment sample as the treatment sample and the delayed treatment sample as the control. The program effects are less clearly shown, but some positive treatment effects remain; see also note 21.

10. Sensitivity analysis (reported in the working paper version, available at http://www.feb.ugent.be/nl/Ondz/WP/Papers/wp_11_749.pdf) shows that the results are similar when we compare the entire delayed treatment sample (including those for which no positive transfers were reported) and the control sample.

There are important problems with the way the control sample was selected.¹¹ Matching at the local level was performed on the basis of a comparison with observable characteristics in 2000. By this time, the treatment sample had already received treatment. However, matching should have been performed on the basis of characteristics before treatment began. In addition, matching at the local level does not imply matching at the household level (see also [Behrman and Todd 1999](#)). Moreover, we do not have data on all children of the households that were in the delayed treatment sample for three reasons (see table A.1 in appendix 1). First, some households dropped out of the sample because of sample attrition. Second, health data were only collected for a subsample of children. Third, because of problems with household identifiers, it was impossible to match all of the children for whom health data were available with only one household each. We only included unique matches in our samples (accounting for more than 80 percent of the children, fortunately). The second and third problems were also present in the control sample. As a result, the treatment and control samples may have differences in terms of preprogram characteristics.

For our empirical strategy in section III, we first use a logistic regression approach to test whether there are statistically significant differences in composition between the treatment and control samples in 1997 for the households with children that were observed in 2003.¹² We use a propensity score matching technique to match the four treatment types with the corresponding control types to correct for possible under- and overrepresentation of households with certain preprogram characteristics. This technique entails weighted sampling (see appendix 3). We compare the resulting weighted distributions at crucial points (such as standardized height below minus two standard deviations from the reference height, indicating stunting) to establish whether the treatment led to first- or second-order improvements of opportunities for each type by performing stochastic dominance tests on the weighted distribution functions.

Circumstances and Outcomes

Ideally, normative theory requires us to obtain a full description of parental circumstances. In reality, an exhaustive description is not available from surveys, and the inclusion of an extensive set of circumstances is statistically unworkable for nonparametric procedures such as ours because of the limited number of observations. For these reasons, we limit ourselves to program participation and two additional circumstances.

11. This may explain why the control sample has rarely been used in academic papers. Recently, however, matched sampling was used to compare schooling ([Behrman et al. 2009b](#) and [Behrman et al. 2010](#)) and work outcomes ([Behrman et al. 2010](#)) in immediate treatment, delayed treatment, and control samples.

12. In 2003, in addition to the regular household data, a questionnaire with recall data was collected. The purpose of these retrospective questions was to compare the preprogram characteristics for the treatment samples with the new control sample.

TABLE 1. Composition of the Samples

	Control sample		Treatment sample	
	#	%	#	%
All	1859	100	1125	100
IL	241	13.0	274	24.4
IP	173	9.3	209	18.6
NL	621	33.4	321	28.5
NP	824	44.3	321	28.5

Note: The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education.

Source: Authors' analysis based on data sources discussed in the text.

The first circumstance refers to parental educational background. In the literature on equality of opportunity, this variable is used most frequently, is always statistically significant, and has been shown to be the most important circumstance in Latin American countries (see, e.g., Bourguignon et al. 2007 and Ferreira and Gignoux 2011). We measure educational background with a dichotomous variable indicating whether at least one parent completed primary education.¹³ The second circumstance variable refers to parents' indigenous background. There is substantial literature indicating that indigenous people remain disadvantaged in Mexico (Olaiz et al. 2006; Psacharopoulos and Patrinos 1994; Rivera et al. 2003; SEDESOL 2008). We consider parents to have an indigenous background if at least one of them can speak or understand an indigenous language.

Combining these two binary characteristics with a binary characteristic indicating program participation yields eight types in Roemer's terminology. We partition the samples on the basis of parental indigenous origin (indigenous or nonindigenous) and parental level of education (primary or less than primary) to form the following types: indigenous, less than primary education (IL); indigenous, primary education (IP); nonindigenous, less than primary education (NL); nonindigenous, primary education (NP). Table 1 shows that there are remarkable differences in the composition of the control sample and the treatment sample among these groups. Clearly, the control sample contains fewer indigenous children and more nonindigenous children with at least one parent who completed primary education than the treatment sample. Because we are comparing cumulative distribution functions of types in the control sample with the corresponding types in the treatment sample, this creates no problem for our analysis. However, as shown in section I, problems arise when there are

13. In the working paper version, we report the results when parental background is measured on the basis of mother's education only. The results are similar to the ones we present here.

important differences in terms of preprogram characteristics between the treatment and control types that are compared.

We focus on several health outcomes. Two important measures of malnutrition for children are anemia, which is defined as hemoglobin levels lower than 11 grams per deciliter, and stunting, which covers a wider range of nutritional deficiencies and is defined as height for age below minus two standard deviations from the WHO International Growth Reference. The latter implies that in a reference population, approximately 2.3 percent of the population is stunted. As reviewed by [Grantham-McGregor and Ani \(2001\)](#), anemia (iron deficiency) in infancy has been associated with poorer cognition, school achievement, and behavioral problems into middle childhood. [Branca and Ferrari \(2002\)](#) point out that stunting is associated with developmental delay, delayed achievement of developmental milestones (such as walking), later deficiencies in cognitive ability, reduced school performance, increased child morbidity and mortality, higher risk of developing chronic diseases, impaired fat oxidation (stimulating the development of obesity), small stature later in life, and reduced productivity and chronic poverty in adulthood. In addition to actual stunting, height has a positive effect on completed years of schooling, earnings (see, e.g., [Alderman et al. 2006](#)), and cognitive and noncognitive abilities (see, e.g., [Case and Paxson 2008](#) and [Schick and Steckel 2010](#)) throughout the distribution. Therefore, we treat our two measures of malnutrition as dichotomous and continuous variables, focusing on the fraction of anemic (stunted) children and on the entire distribution of hemoglobin levels (standardized height). Another health outcome is based on the standardized Body Mass Index (BMI); children are at risk of being overweight if their standardized BMI is larger than 1.15.¹⁴ In a reference population, this cutoff value indicates that 15 percent of children are at risk of being overweight. Overweight children have delayed skill acquisition at young ages ([Cawley and Spiess 2008](#)), are more likely to have psychological or psychiatric problems, have increased cardiovascular risk factors, have increased incidence of asthma and diabetes ([Reilly et al. 2003](#)), are more likely to be obese as adults ([Serdula et al. 1993](#)), and may earn lower wages ([Cawley 2004](#)). A final health outcome is based on the number of days parents reported that the child was sick during the previous four-week period. We consider the percentage of children reporting zero days and more than three days. [Table 2](#) provides information on the outcome variables of the control and treatment samples.

Considering all households, it is striking that the different entries are similar for all health outcomes in the control and treatment samples, with the exception of the number of days sick; fewer sick days were reported for children in the treatment sample than in the control sample. Approximately one child in four is anemic, and one in three is stunted. Compared with the reference

14. The incidence of underweightness is lower than in a reference population.

TABLE 2. Health Outcomes of Two- to Six-Year-Old Children in 2003

	A. Control sample						
	Hemoglobin		zheight		zBMI	Days sick	
	Anemic	Median	Stunted	Median	ROW	0	>3
All	0.24	12.00	0.32	-1.46	0.24	0.58	0.17
IL	0.30	11.90	0.64	-2.40	0.30	0.64	0.13
IP	0.36	11.60	0.50	-1.99	0.23	0.57	0.19
NL	0.25	12.00	0.32	-1.47	0.25	0.58	0.18
NP	0.18	12.20	0.20	-1.13	0.22	0.56	0.18
	B. Treatment sample						
	Hemoglobin		zheight		zBMI	Days sick	
	Anemic	Median	Stunted	Median	ROW	0	>3
All	0.23	12.10	0.34	-1.58	0.20	0.67	0.12
IL	0.29	11.70	0.43	-1.82	0.16	0.72	0.11
IP	0.27	12.00	0.35	-1.63	0.14	0.64	0.14
NL	0.24	12.20	0.33	-1.58	0.22	0.63	0.16
NP	0.13	12.50	0.26	-1.32	0.24	0.68	0.10

Note: The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education. ROW, risk of being overweight.

Source: Authors' analysis based on data sources discussed in the text.

population, our sample contains far too many stunted children and too many children at risk of being overweight.

Interesting but predictable patterns emerge when considering the distribution of health outcomes over the types.¹⁵ Comparing the IL type with the NL type and the IP type with the NP type, indigenous children have worse health outcomes than nonindigenous children, except for the risk of being overweight in the treatment sample. The differences are substantial, particularly for hemoglobin concentration and standardized height in the control sample. Comparing the IL type with the IP type and the NL type with the NP type, the differences between children who had at least one parent who completed primary education and children whose parents had less than primary education are less obvious. The largest differences occur for standardized height; here having a parent who completed primary education is a clear advantage. Overall, these results are in line with the previous literature (see, e.g., Backstrand et al. 1997; Fernald and Neufeld 2006; González de Cossío et al. 2009; Rivera and Sepúlveda 2003; Rivera et al. 2003).

15. The types may differ in terms of characteristics that do not enter the definition of type and in terms of preprogram characteristics.

III. EMPIRICAL RESULTS

We now use the data described in the previous section to evaluate the Oportunidades program. We show that the treatment and control samples are not comparable in terms of preprogram characteristics, and we apply a propensity score matching technique to make them comparable. We apply the methodology presented in section I on the resulting samples to evaluate the program. We then compare the results to previous studies.

Comparison of Weighted Treatment and Control Types

As stated at the end of section I, a crucial assumption in the identification of treatment effects on the basis of a simple comparison of the outcomes of treatment and control samples is that $F^c(x|c_1) = F^T(x|c_1)$, implying that the two samples must be similar in terms of preprogram characteristics. If that is the case, after conditioning on c_1 , observing x does not provide any information about whether an observation belongs to the treatment or control sample. We test this hypothesis as described below.

We construct a sample containing members of both the control and treatment samples. Next, we perform a logistic regression in which the dependent variable takes the value one if the observation belongs to the control sample and the value zero if it belongs to the treatment sample.

Explanatory variables are characteristics of the family, characteristics of the family's dwelling, family assets, and state of residence (see appendix 2 for more details). These characteristics were measured in 1997, before the program started.¹⁶ The results are reported in table A.2 in appendix 2. We find that many of the characteristics significantly affect the probability that the observation comes from the control sample, indicating that the hypothesis that treatment and control samples are comparable in terms of the composition of their preprogram characteristics must be rejected.

In the identification of average treatment effects, a standard way to address differences in the composition of the treatment and control samples is to use propensity score matching techniques. The goal is to make the treatment and control samples more comparable by weighting different observations based on the estimated probability that the observation belongs to the control sample, as determined by the logistic regression discussed in the previous paragraph. Appendix 3 explains this procedure and how the weighting is used to obtain estimates of the relevant distribution functions. The weighting procedure has a substantial effect on the Roemer motivation for considering cumulative distribution functions (Roemer's identification axiom), as we discuss in

16. For the control sample, this is based on recall data (see also note 12).

TABLE 3. Difference between Control and Treatment Groups in the Fraction of Anemic, Stunted, at Risk of Overweight Children and Days Sick. Weighted Samples

	Anemic	Stunted	Risk overweight	0 days sick	>3 days sick
All	-0.03	0.01	-0.04	0.09**	-0.06**
IL	-0.05	-0.18*	-0.11**	0.10*	-0.05*
IP	-0.17**	-0.17**	-0.08	0.09	-0.06
NL	0.00	-0.01	-0.04	0.06	-0.02
NP	-0.08**	0.05	0.03	0.07	-0.09**

Note: The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education.

Source: Authors' analysis based on data sources discussed in the text.

appendix S2.¹⁷ Appendix S3 provides the equivalent of table 2 for the weighted (matched) samples. Supplemental appendices S2 and S3 are available at <http://wber.oxfordjournals.org/>.

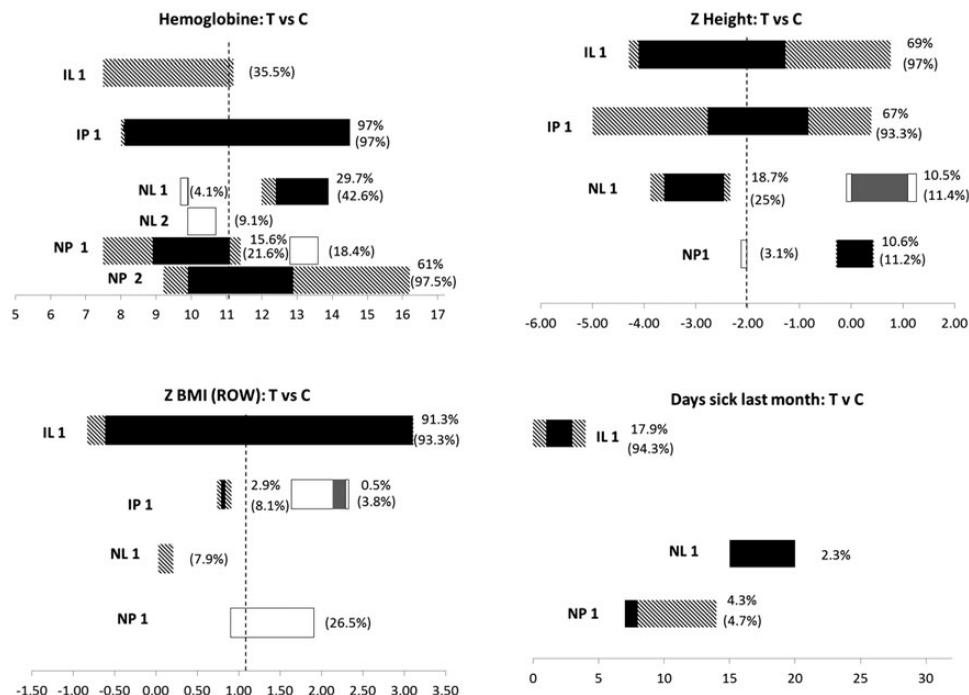
In table 3, we use the weighted samples to consider the effect of the treatments on the fraction of children who are anemic, stunted, or at risk of being overweight. We use the same samples to examine the fraction of children for whom zero sick days or more than three sick days during the previous four weeks were reported. Effects that are statistically significantly different from zero at the 5 percent level of significance are indicated by “**,” and effects that are statistically significantly different from zero at the 10 percent level of significance are indicated by one “*.” Each entry provides the effect of the treatment. From an opportunity perspective, a desirable effect on these fractions indicates that less responsibility allows parents to prevent their children from being anemic, stunted, at risk of being overweight, or sick for more than three days in the previous four-week period.

We see that the treatment effects reported in table 3 are substantial, and all significant effects of the program are in a desirable direction. For each health indicator, we find at least one significant desirable treatment effect for one of the types. The table suggests that the program works well, particularly for children of indigenous origin without a parent who completed primary education. This type is likely to be the most disadvantaged, as table 2 suggests.

Children of indigenous origin with a parent who completed primary education have an improvement in all indicators, although the effects are only

17. Because health is also influenced by preprogram characteristics, we can no longer infer from the percentile in the distribution of health for each type the corresponding responsibility; the same percentile will be obtained by people with different combinations of responsibility and preprogram characteristics. In the supplemental appendix S2, we show that, under certain assumptions, the weighting procedure guarantees that individuals at the same percentile in the weighted treatment and the control sample have identical expected responsibility.

FIGURE 1. Stochastic Dominance Intervals for Health Outcomes among IL, IP, NL, and NP Groups



significant for the fraction of anemic and stunted children. For nonindigenous children, the results are less obvious. The fraction of nonindigenous children who are anemic decreases because of the program, but the results presented in table 3 identify no other significant treatment effects for nonindigenous children.

Figure 1 presents the results of the stochastic dominance tests, using the procedure explained in section I.¹⁸ The horizontal axis denotes the numerical value of the variable of interest (hemoglobin concentration, standardized height, standardized BMI, and reported days sick).

The black (grey) boxes depict the maximal range over the support of the distributions for which the null of nondominance is rejected at the 5 percent level of significance in favor of a desirable (undesirable) effect of the treatment. Hatched (white) boxes indicate the same at a significance level of 10 percent. When hatched (white) boxes are adjacent to a black (grey) box, they show how far the rejection range of the null can be extended for the 10 percent level of

18. Because of the many zero observations, this test procedure cannot be used for the number of days sick. Here, the stochastic dominance test is based on a standard test for the difference between the cumulative distribution functions at the natural numbers between 0 and 30. The intervals shown for this health outcome connect the points in the support where the difference between the cumulative distribution functions is statistically significant.

significance. Each row contains an acronym “XY_i,” of which the first two characters, “XY,” indicate the name of the types that are compared (XY = IL, IP, NL, or NP), and the character “i” indicates whether the test refers to first- (i = 1) or second- (i = 2) order stochastic dominance. The numbers in parentheses behind the boxes show the percentage of observations of the treated type within the black or grey (hatched or white) box.

For example, in the top left panel of figure 1, the hatched box labeled “IL1” shows that, using a 10 percent level of significance, the null hypothesis that the cumulative distribution of the treatment type does not first-order stochastically dominate the distribution of the control type must be rejected against the alternative, that the distribution of the treatment type first-order stochastically dominates the distribution of the control type over the range [7.5, 11.2], which contains 35.5 percent of the treated type. The hypothesis of nondominance can only be rejected at the 10 percent level of significance. Thus, we tested the null hypothesis of the absence of second-order stochastic dominance in favor of the treatment against the alternative, that the distribution of the treatment type second-order stochastically dominates the distribution of the control type at the 5 percent level of significance. We failed to reject the null, such that no box “IL2” is drawn. For IP types, the black box labeled “IP1” indicates that the null hypothesis of nondominance can be rejected at the 5 percent level of significance over the range [8.1, 14.5], which contains 97 percent of the treated IP type. When we increase the level of significance to 10 percent, the hatched box shows that the rejection interval enlarges only marginally, to [8.0, 14.5]. For NL types, when testing for first-order stochastic dominance, we find a white box over the small range of [9.7, 9.9] with very few observations of the treatment type and a solid black box further up in the distribution. When testing NL types for second-order stochastic dominance, we find a small white box. On balance, the evidence for this type against treatment is not strong. Finally, for NP types, we have first a solid black and then a white box. The latter is only significant at the 10 percent level of significance and occurs at a less important part of the distribution (above 11, when children are no longer anemic). When testing for second-order stochastic dominance, we see a solid black box labeled “NP2,” indicating that the project leads to second-order improvement,¹⁹ and this type is also positively affected by the program.

19. Observe that the “NP1” interval is not a subset of the “NP2” interval. This is because the test procedure for first-order (second-order) stochastic dominance identifies the point in the support where the difference between the cumulative (cumulated) distribution functions is most significant and then constructs the interval around this point. There is no reason why the point (and, hence, the intervals) identified should be the same or why the intervals should be related by set inclusion. Moreover, first-order stochastic dominance over a particular interval does not imply second-order stochastic dominance over that same interval because, for second-order dominance, the values of the cumulative distribution functions to the left of the first interval are also relevant. Hence, it may occur that we find an interval over which we reject non-first-order stochastic dominance, but we cannot find an interval over which we reject non-second-order stochastic dominance.

The other panels in figure 1 can be similarly interpreted. In the top right panel, we see that the treatment leads to first-order improvements in the standardized height for IL and IP types over large and crucial parts of the support (standardized height below minus two standard deviations from the reference height). For NL types, we find a first-order stochastic dominance effect in favor of the treatment in an important part of the distribution (standardized height below minus two standard deviations from the reference height) and an adverse effect higher up in the distribution. There is evidence of a marginal perverse first-order treatment effect at a significance level of 10 percent on standardized height for NP types over a small range of $[-2.11, -2.00]$, which contains only 3 percent of the observations of the treated type, and a positive effect higher up in the distribution. No second-order stochastic dominance effects can be established for the nonindigenous types. In the bottom left panel, we concentrate on what occurs at the right of the dotted vertical line, which represents children at risk of being overweight. We see positive, first-order stochastic dominance effects at the 5 percent level of significance for IL types and some evidence of marginally significant perverse treatment effects for IP and NP types. The bottom right panel shows first-order improvements for IL, NL, and NP types. The intervals reported here, except for IL, contain few observations, because of the high frequency of zero reported sick days (see table 2).

The results reported in table 3 and figure 1 are consistent. The stochastic dominance results provide more detail and identify effects in important parts of the distribution that would otherwise go unnoticed, such as the positive first-order stochastic dominance effect on standardized height for NL children. If first-order improvements cannot be found and the influence of parental responsibility is not to be fully respected, then second-order stochastic dominance provides a way to determine whether the program has positive effects. Second-order improvements occur only once in our application, for the hemoglobin concentration of NP types. In summary, we find strong evidence of positive treatment effects for children of indigenous origin, particularly for those without a parent who completed primary education. The evidence for children from nonindigenous origin is not as strong, but enrollment in the program also seems to have positive effects on health opportunities for these children, on balance.

Comparison to Previous Studies

Diaz and Handa (2006) use propensity score matching techniques to construct alternative control samples from the Mexican national household survey. They compute average treatment effects by comparing the immediate treatment sample after eight months of receiving program benefits with the delayed treatment sample (who had not yet received benefits), on the one hand, and their newly constructed control samples, on the other. They conclude, “The PSM [propensity score matching] technique requires an extremely rich set of covariates, detailed knowledge of the beneficiary selection process, and the outcomes

of interest need to be measured as comparably as possible in order to produce viable estimates of impact” (p. 341). In our case, the outcomes are measured in identical ways in the delayed treatment and control samples, and the control sample is constructed following the beneficiary selection process as closely as possible. Our selection of covariates for the propensity score matching closely follows [Behrman et al. \(2009b\)](#), who use almost identical covariates in comparing the effects on schooling outcomes of the short-run differential exposure (between the immediate and delayed treatment samples) with the long-run differential exposure (between the immediate treatment and control samples). They find that longer exposure produces larger effects, and the differences between the order of magnitude of the short- and long-run effects are reasonable. This finding suggests that the propensity score matching technique we use can produce reliable estimates of average treatment effects.

The interpretation of the difference between the distributions of the weighted treatment and control samples as a treatment effect depends on the extent to which the weighting procedure manages to correct for possibly unobserved heterogeneity caused by the imperfect randomness of the assignment to treatment and control groups. Of course, it is not possible to test this directly, but we can compare our results to the findings in the literature that consider differences in children’s health outcomes between immediate and delayed treatment samples. [Rivera et al. \(2004\)](#) compare the health outcomes of children younger than 12 months old in 1997. They find that in 1999 after 12 months of treatment, children in the immediate treatment sample had higher mean hemoglobin values than the children from the delayed treatment sample, who were untreated up to that point. After the immediate treatment sample had received 24 months of treatment and the delayed treatment sample had received approximately six months of treatment, children from the immediate treatment sample had grown more than children in the delayed treatment sample, and the differences in height were significantly larger for households with low socioeconomic status (a score based on dwelling characteristics, possession of durable goods, and access to water and sanitation). [Gertler \(2004\)](#) finds similar results for children aged 0 to 35 months in 1997, stating that “treatment children were 25.3 percent less likely to be anemic and grew about 1 centimeter more during the first year of the program” (p. 340). Both of these differences are statistically significant at the 1 percent level. Unfortunately, Gertler does not report whether the effect differs for different subgroups, such as our types. Hemoglobin levels, unlike height, were not observed before the program started. Therefore, the results for hemoglobin levels do not control for child fixed effects as opposed to growth effects, as noted by [Behrman and Hoddinott \(2005\)](#). They investigate the effect on the height of children who were between 4 and 48 months of age when treatment began in August 1998. They find that when child fixed effects are not included, treatment has a significant negative effect on child height for children between 4 and 36 months of age. However, if child fixed effects are controlled (by considering the difference between 1999

and 1998), the treatment effect becomes significantly positive at approximately one centimeter, as in [Gertler \(2004\)](#).²⁰ Notably, program effects are larger for children in households in which the head of the household speaks an indigenous language and the mother is more educated.²¹

Finally, [Fernald et al. \(2008\)](#) use a different approach. They combine the data of both the immediate and delayed treatment samples to estimate the effect of the size of the conditional cash transfer received on children between 24 and 68 months of age in 2003, when the children's height was measured. Increasing the size of the transfer leads to higher height-for-age scores, a lower prevalence of stunting and a lower prevalence of obesity. Parental level of education and whether the head of the household spoke an indigenous language were not significant controls in their model.

Overall, these findings are in line with ours. The program has significant positive effects on children's height and hemoglobin concentration levels. Larger effects tend to be found for households in which an indigenous language is spoken. This finding is compatible with [Fernald et al. \(2008\)](#) because, in general, indigenous families receive larger cash transfers than nonindigenous families based on the finding that they tend to have more children. Our results indicate where in the distribution the program is most effective for the different types, and we can see that the program is most powerful for the most disadvantaged types, children of indigenous origin.

IV. CONCLUSION

There is a growing body of literature on the measurement of inequality of opportunity (for an overview, see, e.g., [Ramos and Van de gaer 2012](#)). Thus far, the ideas in the literature have not been applied to evaluate social programs. We propose a methodology to do so.

We bring together insights from the literature on equality of opportunity, the literature on program evaluation, and the literature on testing for stochastic dominance. [Roemer's \(1993\)](#) normative approach to equality of opportunity indicates that we should focus on types and that, if responsibility characteristics are unobserved, individuals at the same percentile of the distribution of the

20. [Behrman and Hoddinott \(2005\)](#) obtain the same pattern when considering standardized height-for-age scores.

21. We compare the health outcomes of immediate and delayed treatment in the working paper version of the paper for children born between the beginning of the initial treatment and the beginning of the delayed treatment. This substantially limits the size of the sample. Moreover, because all of these children received at least three years of treatment by the time their health outcomes were measured, few significant effects can be found, particularly for hemoglobin concentration and reported days sick. This indicates that these variables are more sensitive to nutritional status in the immediate past than in the more distant past. We find a significant positive effect on standardized height for indigenous children without parental primary education over a large range of the support of the distribution and for nonindigenous children with parental primary education over a limited support of the distribution. Again, the evidence is in favor of the program.

outcome within their type have exercised a comparable degree of responsibility. This approach provides a normative foundation for the comparison of cumulative distribution functions of corresponding treatment and control types. The literature on program evaluation stresses that care should be taken to ensure that the treatment and control samples are comparable in terms of preprogram characteristics. If they are not, propensity score matching techniques can be used to make the samples more comparable. Hence, we test whether the treatment and control samples are comparable in terms of preprogram characteristics and since the test fails, we propose a weighted sampling method based on standard propensity score matching techniques to make the treatment and control types comparable. Finally, Davidson and Duclos (2009) and Davidson (2009) propose a new technique to test for stochastic dominance, taking non-dominance as the null so that rejection of the null implies dominance. Their test procedure is particularly suited to our study because it allows us to see where dominance can be established along the distribution.

We applied our procedure to study the effect of the Mexican Oportunidades program on children's health opportunities. We can draw two conclusions about the proposed methodology. First, in our application (as in the applications by Lefranc et al. 2008, Lefranc et al. 2009, Peragine and Serlenga 2008, and Rosa Dias 2009), looking for second-order stochastic dominance does not significantly add to the conclusions drawn from first-order stochastic dominance. Thus, whether the influence of parental responsibility is to be fully respected does not substantially affect the conclusions. Second, the treatment and control samples differed substantially in terms of preprogram characteristics. Therefore, it is important to use weighted sampling based on techniques such as propensity score matching to make the samples (more) comparable. Concerning the actual effects of the program, our results indicate that the Oportunidades program has a substantially favorable effect on the health opportunities of the most disadvantaged children, that is, those with parents of indigenous origin and without a parent who completed primary education. Additionally, the effects on children of indigenous origin with a parent who completed primary education are sizable and important. The effects on nonindigenous children are less obvious, but the overall evidence in this paper indicates that the program also results in better health opportunities for these children.

APPENDIX 1. SAMPLING PROCEDURE

When we compare the sample sizes in the column "1997 data available" with the sizes in table 1 in the main text, we see that 12 (three) observations dropped out in the final control (treatment) sample because of missing observations on circumstances.

TABLE A.1. Sampling Process

	Original number of children (a)	Matched children		1997 data available	
		number (b)	% of (a)	number	% of (b)
Control	2,247	1,871	83	1,871	100
Treatment	2,615	2,200	84	1,128	51
Total	4,862	4,071	84	2,999	73

Source: Authors' analysis based on data sources discussed in the text.

APPENDIX 2. RESULTS OF THE LOGISTIC REGRESSION

Our specification for the logistic regression is close to the specification used for propensity score matching by [Behrman et al. \(2009b\)](#) and [Behrman and Parker \(2011\)](#). The dependent variable equals one if the observation comes from the control sample and zero otherwise. Explanatory variables are based on preprogram characteristics of the treatment sample and the 1997 recall characteristics of the control sample. We have five types of explanatory variables:

- (1) Household characteristics, which include the ages of the head of the household and spouse (in years); the sex of the head of the household; whether the head of the household and spouse speak an indigenous language; whether the parents completed primary education; whether the parents work; and the composition of the household (number of children and women and men of different ages)
- (2) Dwelling conditions of the household, which include the number of rooms in the house and a list of dummy variables indicating the presence of electric light, running water on the property, running water in the house (which implies the presence of running water on the property), a dirt floor, and whether the roof and walls are of poor quality
- (3) Asset information, which includes dummy variables indicating whether the family owns animals or land and whether the family possesses a blender, refrigerator, fan, gas stove, gas heater, radio, stereo, TV, video, washing machine, car, or truck
- (4) State of residence, which includes a list of dummy variables indicating the state in which the family lives, with the reference state (all state of residence dummies equal to zero) of Veracruz
- (5) Dummy variables for missing characteristics whose effects could be meaningfully estimated, following [Behrman et al. \(2009b\)](#) and [Behrman and Parker \(2011\)](#); the variable “Miss Asset” takes the value of one if any of the assets listed in the table between “Animals” and “Truck” is missing

Table [A.2](#) gives the estimated coefficients.

TABLE A.2. Logistic Regression Results.

Variable	Coef.	SE	z	Variable	Coef.	SE	z
Age Hh. head	-0.013	0.007	-1.96	Blender	-0.169	0.132	-1.27
Age spouse	-0.012	0.007	-0.61	Fridge	0.054	0.200	0.27
Sex Hh. head	-2.197	0.351	-6.25	Fan	0.142	0.120	0.71
Indig. Hh. head	-0.718	0.272	-2.64	Gas stove	0.377	0.145	2.60
Indig. spouse	0.249	0.278	0.90	Gas heater	0.709	0.360	1.97
Educ. Hh. head	-0.229	0.114	-2.01	Radio	-0.600	0.100	-5.96
Educ. spouse	-0.386	0.116	-3.32	Hifi	-0.361	0.251	-1.44
Work Hh. head	1.124	0.262	4.29	TV	-0.635	0.188	-5.53
Work spouse	0.623	0.161	3.86	Video	0.498	0.345	1.44
# Children 0-5	-0.090	0.048	-1.89	Washing machine	-0.35	0.330	-0.11
# Children 6-12	-0.211	0.042	-5.06	Car	1.229	0.465	2.64
# Children 13-15	-0.160	0.084	-1.91	Truck	0.243	0.282	0.86
# Children 16-20	-0.016	0.073	-0.22	Guerrero	-0.548	0.190	-2.88
# Women 20-39	-0.014	0.119	-0.12	Hidalgo	-0.937	0.209	-4.48
# Women 40-59	0.040	0.155	0.26	Michoacán	-0.582	0.176	-3.30
# Women 60 +	0.040	0.185	0.22	Puebla	-1.097	0.150	-7.33
# Men 20-39	-0.162	0.106	-1.54	Querétaro	0.119	0.219	0.54
# Men 40-59	0.366	0.161	2.28	San Luis	-0.462	0.153	-3.02
# Men 60 +	0.698	0.234	2.99	Miss Age Sp.	-4.297	0.713	-6.03
# Rooms	-0.006	0.010	-0.58	Miss Indg. Hh.	0.799	1.959	0.41
Electrical light	0.036	0.115	0.32	Miss Indg. Sp.	-2.102	1.894	-1.11
Running water land	0.879	0.115	7.67	Miss Work Hh.	3.461	1.871	1.85
Running water house	-0.435	0.208	-2.10	Miss Work Sp.	3.817	1.844	2.07
Dirt floor	0.096	0.118	0.81	Miss water land	0.871	1.640	0.53
Poor quality roof	-0.026	0.108	-0.24	Miss water house	0.699	0.827	0.84
Poor quality wall	-0.483	0.126	-3.82	Miss Assets	-4.121	2.398	-1.72
Animals	-0.168	0.113	-1.48	Constant	3.860	0.422	9.13
Land	-0.545	0.105	-5.17				
Number of obs.	2,741						
LR χ^2 (54)	730.0		Pseudo R^2	0.198			
Prob. > χ^2	0.000		Log Likelihood	-1478.75			

Note: Dependent variable equals one if the observation is in control and zero if the observation is in treatment group.

Source: Authors' analysis based on data sources discussed in the text.

APPENDIX 3. MATCHING ESTIMATOR AND CONSTRUCTION OF THE CORRESPONDING DISTRIBUTION FUNCTION

Step 1: Propensity score matching

The estimated logistic regressions allow us to compute, for each observation, the propensity score P_i , the probability that the observation is in the control sample given its preprogram characteristics x_i . Figure A.1 depicts the estimated propensity scores because we matched the treatment into the control sample for each of the four combinations of race and parental level of education, and we determined the common support for each of these four comparisons as the overlap of the support of the control and treatment samples. Table A.3 above gives the common support and the number of observations in the common support for each of the types.

We tested the balancing property score using Stata. The optimal number of blocks was 11, and we had 54 explanatory variables, resulting in 594 tests. In 14 cases, the balancing property was rejected. As an additional test, we reran the logistic equation from table A.2 using the weighted sample. Only four coefficients out of 54 were significant. These results are encouraging.

Step 2: Construction of the cumulative distribution function

Let I_1 denote the set of individuals in the treatment sample, I_0 denote the set of individuals in the control sample, and S_P denote the region of common support. The number n_0 gives the number of individuals in the set $I_0 \cap S_P$. The outcome of individual j in the control sample is Y_{0j} , and the outcome of individual i in the treatment sample is Y_{1i} . Let $D = 1$ for program participants and $D = 0$ for those who do not participate in the program.

The purpose is to match each individual in the control sample with a weighted average of individuals in the treatment sample. The usual estimator of the average treatment effect thus becomes

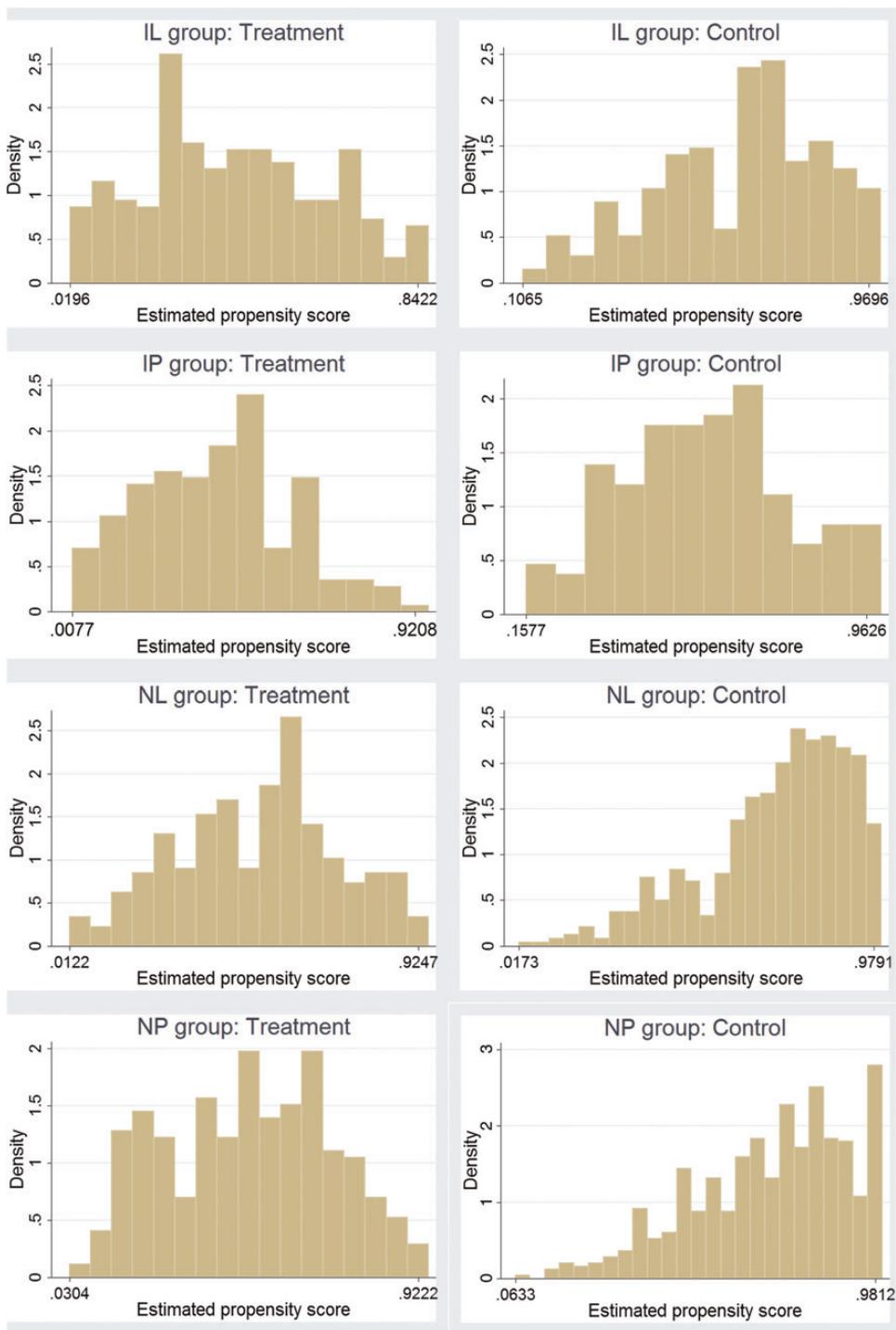
$$T = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} [E(Y_{1j}|D = 1, P_j) - Y_{0j}]$$

$$\text{with } E(Y_{1j}|D = 1, P_j) = \sum_{i \in I_1} W(i, j) Y_{1i}.$$

The construct $E(Y_{1j}|D = 1, P_j)$ is the outcome of the hypothetical individual matched to individual j . The average treatment effect can be written as

$$T = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} \sum_{i \in I_1} W(i, j) Y_{1i} - \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} Y_{0j}.$$

FIGURE A.1. Estimated Propensity Scores



The first term is the average of the matched observations, which attaches to each of the original observations Y_{1i} a weight

$$\omega_i = \frac{1}{n_0} \sum_{j \in I_0 \cap S_p} W(i, j).$$

It is therefore natural (and consistent with the standard model of the estimation of average treatment effects) to use for each observation Y_{1i} the weight ω_i to construct the cumulative distribution function.

Many possible ways exist to determine the weights $W(i, j)$. We use a Kernel estimator, such that

$$W(i, j) = \frac{G\left(\frac{P_i - P_j}{\alpha}\right)}{\sum_{k \in I_1} G\left(\frac{P_k - P_j}{\alpha}\right)},$$

where $G(\cdot)$ is the Epanechnikov kernel function and α is a bandwidth parameter. The bandwidth parameter was chosen in an optimal way using the formula in Silverman (1986, 45–47):

$$\alpha = 1.06 \min\left(\sigma, \frac{\rho}{1.34}\right),$$

where σ is the standard deviation and ρ is the interquartile range of the distribution of propensity scores. The resulting bandwidths for each of the types are given in the last column of table A.3.

TABLE A.3. Propensity Score Matching: Common Support and Number of Observations in the Common Support

	Common support	Control #	Treatment #	Bandwidth
IL	[0.106, 0.868]	228	260	0.074
IP	[0.158, 0.957]	155	193	0.074
NL	[0.017, 0.952]	586	318	0.071
NP	[0.063, 0.949]	668	318	0.071
Total		1,637	1,089	

Note: The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education.

Source: Authors’ analysis based on data sources discussed in the text.

REFERENCES

- Alderman, H., J. Hoddinott, and B. Kinsey. 2006. "Long-term Consequences of Early Childhood Malnutrition." *Oxford Economic Papers* 58 (3): 450–74.
- Backstrand, J.R., L.H. Allen, G.H. Pelto, and A. Chávez. 1997. "Examining the Gender Gap in Nutrition: An Example from Rural Mexico." *Social Science & Medicine* 44 (11): 1751–9.
- Behrman, J.R., and J. Hoddinott. 2005. "Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition." *Oxford Bulletin of Economics and Statistics* 67 (4): 547–69.
- Behrman, J.R., S.W. Parker, and P.E. Todd. 2009a. "Medium-Term Impact of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico." In S. Klasen, and F. Nowak Lehmann, eds., *Poverty, Inequality and Policy in Latin America*. Cambridge: MIT Press.
- . 2009b. "Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico." *Economic Development and Cultural Change* 57 (3): 439–77.
- . 2011. "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five Year Follow-up of ROGRESA/Oportunidades." *Journal of Human Resources* 46: 93–122.
- Behrman, J.R., P. Sengupta, and P.E. Todd. 2005. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change* 54 (1): 237–75.
- Behrman, J.R., and P.E. Todd. 1999. *Randomness in the Experimental Samples of PROGRESA – Education, Health, and Nutrition Program*. International Food Policy Research Institute.
- Behrman, J.R., P.E. Todd, B. Hernández, J. Urquieta, O. Attanasio, M. Angelucci, and M. Hernández. 2006. *Evaluación externa de impacto del programa Oportunidades 2006*. Instituto Nacional de Salud Pública.
- Black, S.E., P. Devereux, and K. Salvanes. 2007. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *The Quarterly Journal of Economics* 122 (1): 409–39.
- Bossert, W. 1995. "Redistribution Mechanisms Based on Individual Characteristics." *Mathematical Social Sciences* 29 (1): 1–17.
- Bourguignon, F., F.H.G. Ferreira, and M. Menéndez. 2007. "Inequality of Opportunity in Brazil." *Review of Income and Wealth* 53 (4): 585–618.
- Branca, F., and M. Ferrari. 2002. "Impact of Micronutrient Deficiencies on Growth: The Stunting Syndrome." *Annals of Nutrition and Metabolism* 46 (Suppl. 1): 8–17.
- Case, A., and C. Paxson. 2008. "Stature and Status: Height, Ability and Labor Market Outcomes." *Journal of Political Economy* 116 (3): 499–532.
- Cawley, J. 2004. "The Impact of Obesity on Wages." *Journal of Human Resources* 39 (2): 451–74.
- Cawley, J., and C.K. Spiess. 2008. "Obesity and Skill Attainment in Early Childhood." *Economics and Human Biology* 6: 388–97.
- Chen, W.H., and J.Y. Duclos. 2008. *Testing for Poverty Dominance: An Application to Canada*. IZA Discussion Paper N 2829.
- Davidson, R. 2009. "Testing for Restricted Stochastic Dominance: Some Further Results." *Review of Economic Analysis* 1 (1): 34–59.
- Davidson, R., and J.Y. Duclos. 2009. *Testing for Restricted Stochastic Dominance*. GREQAM Document de Travail 2009-38 (06-09).
- Diaz, J.J., and S. Handa. 2006. "An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico's PROGRESA Program." *Journal of Human Resources* 41 (2): 319–45.
- Fernald, L.C.H., P.J. Gertler, and L.M. Neufeld. 2008. "Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's Oportunidades." *The Lancet* 371 (9615): 828–37.

- Fernald, L.C.H., and L.M. Neufeld. 2006. "Overweight with Concurrent Stunting in Very Young Children from Rural Mexico: Prevalence and Associated Factors." *European Journal of Clinical Nutrition* 61 (5): 623–32.
- Ferreira, F.H.G., and J. Gignoux. 2011. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *Review of Income and Wealth* 57 (4): 622–54.
- Fiszbein, A., N. Schady, F.H.G. Ferreira, M. Grosh, N. Kelleher, P. Olinto, and E. Skoufias. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty, a World Bank Policy Research Report*. Washington, DC: The World Bank.
- Fleurbaey, M. 1995. "The Requisites of Equal Opportunity." In M. Salles and N. Schofield, eds., *Social Choice, Welfare and Ethics*. Cambridge University Press.
- . 1998. "Equality among responsible individuals." In J. Laslier, M. Fleurbaey, N. Gravel, and A. Trannoy, eds., *Freedom in Economics: New Perspectives in Normative Economics*. London: Routledge.
- . 2008. *Fairness, Responsibility and Welfare*. Oxford: Oxford University Press.
- Foster, J., J. Greer, and E. Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52 (3): 761–66.
- Gertler, P.J. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." *American Economic Review* 94 (2): 336–41.
- Gonzá de Cossiolez, T., J.A. Rivera, D. González Castell, M. Unar Munguía, and E.A. Monterrubio. 2009. "Child Malnutrition in Mexico in the Last Two Decades: Prevalence Using the New WHO 2006 Growth Standards." *Salud Pública de México* 51 (Supp 4): S494–S506.
- Grantham-McGregor, S., and C. Ani. 2001. "A Review of Studies on the Effect of Iron Deficiency on Cognitive Development in Children." *The Journal of Nutrition* 131 (2): 649S–68S.
- Heckman, J.J. 1992. "Randomization and social policy evaluation." In C. Manski, and I Garfinkel, eds., *Evaluating Welfare and Training Programs*. Cambridge: Harvard University Press.
- Heckman, J.J., J. Smith, and N. Clements. 1997. "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies* 64 (4): 487–535.
- INSP (Instituto Nacional de Salud Pública). 2005. *General Rural Methodology Note*. Cuernavaca, Mexico.
- Lefranc, A., N. Pistoletti, and A. Trannoy. 2008. "Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike?" *Review of Income and Wealth* 54 (4): 513–46.
- . 2009. "Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France." *Journal of Public Economics* 93 (11-12): 1189–1207.
- Naschold, F., and C.B. Barrett. 2010. *A Stochastic Dominance Approach to Program Evaluation with an Application to Child Nutritional Status in Kenya*. Working Paper.
- Olaiz, G., J.A. Rivera, T. Shamah, R. Rojas, S. Villalpando, M. Hernández, and Jaime Sepúlveda. 2006. *Encuesta Nacional de Salud y Nutrición 2006* [National Health and Nutrition Survey 2006]. Instituto Nacional de Salud Pública.
- O'Neill, D., O. Sweetman, and D. Van de gaer. 2000. "Equality of Opportunity and Kernel Density Estimation: An Application to Intergenerational Mobility." In T. Fomby, and R. C. Hill, eds., *Advances in Econometrics*, Vol. 14. Stanford: JAI Press.
- Parker, S.W., L. Rubalcava, and G. Teruel. 2008. "Evaluating Conditional Schooling and Health Programs." In T. Schultz, and J. Strauss, eds., *Handbook of Development Economics*, Vol. 4. Elsevier.
- Paes de Barros, R., F.H.G. Ferreira, J.R. Molinas Vega, and J. Saavedra Chanduvi. 2009. *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: The World Bank.

- Peragine, V., and L. Serlenga. 2008. "Higher Education and Equality of Opportunity in Italy." In J. Bishop, and B. Zheng, eds., *Inequality of Opportunity: Papers from the Second ECINEQ Society Meeting, Research on Economic Inequality*, Vol. 16. Bingley: Emerald Group Publishing.
- Psacharopoulos, G., and H.A. Patrinos. 1994. *Indigenous People and Poverty in Latin America: An Empirical Analysis*. Washington DC: The World Bank.
- Ramos, X., and D. Van de gaer. 2012. *Empirical Approaches to Inequality of Opportunity: Principles, Measures and Evidence*. FEB Working Paper 12/792. Ghent: Faculty of Economics and Business Administration, Ghent University.
- Reilly, J.J., E. Methven, Z.C. McDowell, B. Hacking, D. Alexander, L. Stewart, and C.J.H. Kelnar. 2003. "Health Consequences of Obesity." *Archives of Disease in Childhood* 88 (9): 748–52.
- Rivera, J.A., E. Monterrubio, T. González-Cossío, R. García-Feregrino, A. García-Guerra, and J. Sepúlveda. 2003. "Nutritional Status of Indigenous Children Younger than Five Years of Age in Mexico: Results of a National Probabilistic Survey." *Salud Pública de México* 45: S466–76.
- Rivera, J.A., and J. Sepúlveda. 2003. "Conclusions from the Mexican National Nutrition Survey 1999: Translating Results into Nutrition Policy." *Salud Pública de México* 45: S565–75.
- Rivera, J.A., D. Sotres-Alvarez, J.-P. Habicht, T. Shamah, and S. Villalpando. 2004. "Impact of the Mexican Program for Education, Health, and Nutrition (PROGRESA) on Rates of Growth and Anemia in Infants and Young Children." *The Journal of the American Medical Association* 291 (21): 2563–70.
- Roemer, J. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian Planner." *Philosophy & Public Affairs* 22 (2): 146–66.
- . 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
- Rosa Dias, P. 2009. "Inequality of Opportunity in Health: Evidence from a UK Cohort Study." *Health Economics* 18 (9): 1057–74.
- Schick, A., and R.H. Steckel. 2010. *Height as a Proxy for Cognitive and Non-Cognitive Ability*. NBER Working Paper N 16570.
- Schultz, P.T. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics* 74 (1): 199–250.
- SEDESOL (Ministry of Social Development of Mexico). 2008. *Evaluación externa del Programa Oportunidades 2008. A diez años de intervención en zonas rurales (1997-2007)*. Mexico.
- Serdula, M.K., D. Ivery, R.J. Coates, D.S. Freedman, D.F. Williamson, and T. Byers. 1993. "Do Obese Children Become Obese Adults? A Review of the Literature." *Preventive Medicine* 22: 167–77.
- Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall/CRH.
- Swift, A. 2005. "Justice, Luck, and the Family: The Intergenerational Transmission of Economic Advantage from a Normative Perspective." In S. Bowles, H. Gintis, and M. Osborne Groves, eds., *Unequal Chances: Family Background and Economic Success*. Princeton University Press.
- Todd, P.E. 2004. *Design of the Evaluation and Method Used to Select Comparison Group Localities for the Six Year Follow-Up Evaluation of Oportunidades in Rural Areas*. Technical report, International Food Policy Research Institute.
- Trannoy, A., S. Tubeuf, F. Jusot, and M. Devaux. 2010. "Inequality of Opportunities in Health in France: A First Pass." *Health Economics* 19 (8): 921–38.
- Verme, P. 2010. "Stochastic Dominance, Poverty and the Treatment Effect Curve." *Economics Bulletin* 30 (1): 365–73.

SYMPOSIUM ON CONFLICT AND GENDER

Armed Conflict, Gender, and Schooling

Mayra Buvinić, Monica Das Gupta, and Olga N. Shemyakina

The impact of armed conflict on gender differentials in schooling appears to be highly context-specific, as the review of the literature and the findings from the three studies in this symposium reveal. In some settings boys' schooling is more negatively affected than that of girls. In others, the reverse is the case. Effects are largely shaped by events surrounding a conflict, pre-war gender differences in educational attainments, and education and labor market opportunities in the absence of war. Rigorous evaluations of post-conflict policies and aid projects can provide useful information to address educational needs and gender differentials in these environments. JEL codes: I20, J10, O12

This symposium focuses on the link between schooling and armed conflict, and gender differentials in these links. This overview briefly reviews the recent literature on these links, and then discusses the contributions made by the three studies in this symposium. The studies indicate that the impact varies greatly depending on the context in which the conflict takes place.

Armed conflict can cause long-lasting damage that goes well beyond the immediate death and destruction. Several reviews document how armed conflict damages physical, human, and social capital, constrains economic growth, and affects the well-being of populations.¹ During armed conflicts, households adopt a variety of coping strategies that are also commonly used in response to economic shocks.² However, these strategies may be less effective in settings affected by

Mayra Buvinić is a Senior Fellow at the United Nations Foundation, Washington, DC 20036; her email address is mayra.buvinic@gmail.com. Monica Das Gupta is a Research Professor at the University of Maryland, College Park, MD 20742; her email address is mdasgupta@gmail.com. Olga N. Shemyakina (corresponding author) is an Assistant Professor at the School of Economics, Georgia Institute of Technology, Atlanta, GA 30332; her email address is olga.shemyakina@econ.gatech.edu. This work was funded by a grant from the Government of Norway to the World Bank. The views expressed in the article belong to the authors' and should not be attributed to the World Bank or any affiliated organization or member country.

1. See for example the reviews by [Blattman and Miguel 2010](#); [Buvinić et al. 2013](#); [Collier et al. 2003](#); and [Justino, Leone, and Salardi \(2014\)](#).

2. For some of the literature on the responses to economic shocks, see [Kochar 1999](#); [Morduch 1995](#); and [Rose 1999](#); 2001.

THE WORLD BANK ECONOMIC REVIEW, VOL. 28, NO. 2, pp. 311–319
Advance Access Publication October 14, 2013

doi:10.1093/wber/lht032

© The Author 2013. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

armed conflict since households are exposed not only to economic shocks but also to damaged physical and service infrastructure, and high risks to personal safety. Shocks induced by natural disasters may have similar impacts on households' well-being, but such shocks are less disruptive of trust and social networks than armed conflict (e.g., [Cassar et al. 2011](#); [Rohner et al. 2011](#)).

WHAT DO WE KNOW ABOUT THE IMPACT OF ARMED CONFLICT ON SCHOOLING?

A large body of literature indicates that exposure to economic shocks in early childhood can have a negative and long-lasting impact on the health and educational attainment of individuals, in turn affecting their long-term earnings and well-being in adulthood.³

Similar to economic shocks, many studies also find that armed conflict negatively affects children's educational attainment. The magnitude of the loss and the relative impact on girls versus boys depends on contextual factors. In some settings, armed conflict is found to reduce boys' educational attainment more than that of girls. One reason is that boys may be enlisted in the military, argues a study on Bosnia and Herzegovina ([Swee 2011](#)); another is that they may be sent out to work to help the household cope with the conflict-induced shock. [Rodriguez and Sanchez \(2012\)](#) find that municipality-level exposure to conflict in Colombia increases the chances both of children aged 6-17 dropping out of school and of adolescent boys being employed.

Boys' schooling attainment may also decline more relative to girls' in settings where girls had low schooling during the pre-conflict period (and therefore had less to lose). For example, [Akresh and de Walque \(2010\)](#) use two Demographic and Health Surveys (DHS) collected before and after the 1994 genocide in Rwanda and find that boys from non-poor families who were of school-going age during the war suffered greater declines in their education compared to girls or boys from poorer households. They relate this result to violent targeting of wealthy Tutsi households and communities during the genocide. Similarly, the long-term effect of the collapse of the educational system during the Khmer Rouge period in Cambodia reduced secondary school attainment more for boys, since girls were less likely to be in school during the pre-conflict period ([de Walque 2006](#): Figure 9). In a related study, [Merrouche \(2011\)](#) finds that individuals from Cambodian regions with high levels of landmine contamination lost an average 0.5 years of education with a similar effect for boys and girls.

In other settings, girls' schooling is more affected by armed conflict, possibly because parents seek to protect their girls from rape and other threats to their honor, as suggested by studies on the civil war in Tajikistan ([Shemyakina 2011](#)) and the insurgency (a low intensity conflict) in Punjab, India ([Singh and Shemyakina 2013](#)). Economic effects also came in play, as households in rural

3. See for example the reviews by [Almond and Currie 2011](#); and [Strauss and Thomas 2008](#).

Punjab that had both boys and girls may have substituted the educational expenditure towards boys at the expense of girls in the face of uncertainty (Singh and Shemyakina 2013).

Small losses in schooling during armed conflict can translate into large losses in lifetime earnings. Ichino and Walter-Ebmer (2004) examine the educational attainment of children born in 1920-1949 in four European countries and find that the cohorts born in 1930-1939 (these who reached age 10 during or soon after the war) completed less schooling than other cohorts under consideration. This negative effect was observed only for children living in Austria and Germany – which were heavily engaged in World War II – but not for children from Sweden and Switzerland which were neutral during the war. The authors suggest that military service of fathers and/or their death in the war was one of the primary channels through which conflict had an impact on schooling. Further, the authors find that the war was associated only with a small loss schooling, but this small difference reduced the likelihood of obtaining higher education degrees by the affected cohorts and made a substantial impact on their lifetime income. Disaggregating the data by the intensity of the conflict, Akbulut-Yuksel (2009) finds that boys and girls of school-going age during the war who lived in German cities heavily targeted by allied bombings achieved a significantly lower education level than those living in areas exposed to fewer bombings.

The importance of the intensity of violence is also illustrated by Chamarbagwala and Morán (2011), who explore the effect of differing levels of violence on primary education attainment by children who were of school-going age during the 30 year long war in Guatemala. Rural Mayan children who were of school age during the period of worst violence (1979-84) and who lived in the departments with high intensity of human right violations lost the most schooling.

By contrast, in settings with long-running low intensity conflict where access to schooling is likely to be little disrupted, education level can increase. De Groot and Goksel (2011) observe that in the Basque country, the demand for higher education by citizens from the middle level education group went up.

Two studies explore the effect of random abduction by the Lord Resistance Army (LRA) in Uganda on former child soldiers. Former male child soldiers were found to have lower education and earnings, and were less likely to be employed in skill or capital intensive occupations than those who were never abducted (Blattman and Annan 2010). The story was different for females. Educated females were more likely to be abducted, due to their perceived value as midwives, note-takers and radio-operators (Annan et al. 2011). These young women were less likely to be released, and during their abduction they were forced into marriages faster than those with less education. Abduction alone did not have a significant effect on women's schooling, which Annan et al. explain by the lack of opportunities for women in Uganda in general. However, female abductees who had children during their abduction (“forced mothers”) were

eight times less likely to re-enroll in school than female returnees without children. Foltz and Opoku-Agyemang (2011) focus on the effect of low intensity conflict in 2000-2005 on general population in Uganda and find little gender differential in school attainment.

In contrast to most studies that focus on children of school-going age, León (2012) also investigates the effect of exposure to political violence in Peru in *utero* and till age 6 on schooling attainment. Children's exposure to violent events during these early years of life is negatively and statistically significantly associated with children's long-run school attainment. By contrast, regional exposure to violence during one's *primary* schooling years has a negative impact only in the short but not the long run. León attributes the observed long-run effects to factors such as the deaths of teachers⁴ on the one hand, and on the other, to the poorer health of individuals affected early in life (as measured by their weight-for-height and height-for-age z-scores), and potentially poorer health of mothers due to the conflict. He does not differentiate between the impacts of violent events on schooling of boys versus girls.

Some studies have focused also on the effect of conflict on the *quality* of schooling. Brück et al. (2013) study the relationship between student test scores and district-level fatalities during the 2nd Intifada in Palestine. In both the West Bank and the Gaza Strip, the chance to pass the secondary school final exam was lower in school districts with higher fatality levels. In the West Bank (but not the Gaza Strip) the observed effect is stronger for boys than girls. The authors cite as a possible reason that boys in this region are more exposed to injuries, serious accidents, and post-traumatic stress disorder. Additionally, using various measures of school quality, the authors establish that the decline in test scores could be attributed to the negative effect of conflict on the quality of schooling.

In Turkey, Kibris (2013) finds substantially lower test scores in the regions affected by the low intensity Kurdish-Turkish conflict, and suggests that an inability to attract and retain good teachers by these regions is one of the channels through which this conflict has affected educational outcomes. Monteiro and Rocha (2012) find a negative impact of gang violence in Brazilian favelas on students' math test scores. The closer the school's location is to violent events, the lower are the test scores. Violence within three to four months before the Prova Brazilia exam (but not after) has a negative impact on the exam test scores. Local violence by drug gangs has a large negative impact on school resources (turnover of principals, teacher absenteeism and school closures). The authors find no substantial effect of violence in the previous academic year on this year's student test scores, indicating that schools are resilient in reconstituting themselves.

4. León (2012) notes though that both the insurgents and the government forces, left schools intact but often destroyed other infrastructure.

THE CONTRIBUTION OF THE PAPERS IN THIS SYMPOSIUM

The studies in this issue evaluate the effect of armed conflict on primary school attainment by gender in Timor-Leste, Nepal and Burundi. The authors draw on a variety of datasets collected by international and local organizations and use rigorous data analysis methods to study these effects. These studies help us understand how different institutional environments surrounding the conflict and post-conflict reconstruction shape gender differences in schooling outcomes. The studies by Valente and by Verwimp and Van Bavel focus on the short-term effects of the conflicts in Nepal and Burundi respectively, while Justino, Leone, and Salardi explore both, the short and long-run effects of the anti-colonial struggle of Timor Leste. All studies focus on the educational experiences of the cohort that was of school age during the conflict.

Valente (2014) focuses on the 1996-2006 Maoist insurgency in Nepal. The conflict was driven by the insurgents' strong ideological agenda to improve the social and economic standing of disadvantaged groups, and eliminate caste, gender and ethnicity driven inequalities in Nepal. The author uses the 2001 and 2006 DHS data and several district level conflict-intensity variables to identify the impact of armed conflict on educational outcomes of children of school-going age during this war. The intensity of Maoist activity in a district as measured by the number of deaths is associated with increased girls' schooling attainment, while the effect for boys is smaller and not so robust. However, conditional on the number of deaths, a rise in Maoist abductions – that often targeted school-age children – decreased educational attainments of girls in more affected districts.

The conflict escalated in 2001, so in addition to the standard difference in difference regression framework, Valente augments the analysis by exploring the effect of change in violence between 2001 and 2006 on the educational attainment of children who were of primary schooling age in these survey years. For instance, the 10-year olds in 2001 were exposed to less conflict than the 10-year olds in 2006, who experienced an intensified conflict for most of their schooling years. Examining education of children of comparable age but exposed to differing degrees of conflict exploits more fully the available data, and the results show that the duration of exposure to more intense conflict dramatically changes outcomes with children surveyed in 2006 obtaining more education than children surveyed in 2001. This innovative technique is possible due to the timing of the surveys. The study also credibly shows that different aspects of conflict may have varying impact on outcomes.

Valente suggests that an increase in female educational attainment during conflict was primarily linked to the Maoists' efforts to remove educational barriers for the disadvantaged groups of population and in particular, women. Women were actively present within structures of Maoist organizations; participated in politics and were viewed as important to the spread of Maoist propaganda as they had access to other members of the household (Ariño 2008). Thus, the

engagement of women and the goals of the Maoists, argues Valente, may have contributed to an increased educational attainment during the conflict, and especially by girls in the Maoist controlled districts.

Justino, Leone, and Salardi (2014) focus on the long-term struggle against occupation and colonization of Timor Leste by Indonesia. The intensity of violence during this long conflict varied over time and space. Post-conflict development in Timor Leste was characterized by a strong role of the international organizations, and a reconstruction process that benefited girls in particular. Under the East Timor Transitional Authority (ETTA), school fees were substantially reduced which increased enrollment of girls, children from poor households and rural areas (World Bank 2003). Also in Timor-Leste, the displaced tended to quickly move back to their places of residences and to re-establish community-run schooling programs (Justino, Leone, and Salardi 2014).

The authors study the impact of the conflict on school *attendance* in 2001, immediately following the 1999 wave of violence, a short-term outcome, and on primary school *attainment* in 2007, a longer-term result. In the analysis, they use the same cohort of individuals surveyed in different years. To examine the effect of the conflict on school attendance, the authors construct a panel-like dataset using retrospective data on school attendance for three consecutive years from a cross-sectional survey (i.e., before, during and after the 1999 wave of violence). The short-term effect of household level conflict exposure (as measured by displacement and damage to dwelling) is a reduced primary school attendance for boys and girls.

The identification of conflict exposure to violence over a longer time period is performed by exploiting variation in peaks of violence and in different forms of warfare that took place in Timor Leste at different times. In the long-run, the authors observe a substantial loss in educational attainment of boys, but not girls, who grew during different periods of violence. They explain this result by a higher workforce participation rates among boys, and the substitution by households of future gains from investment in education for current consumption.

Burundi experienced an armed conflict since 1965, with conflicts erupting almost every 10-15 years. The conflict was sustained by the economic differences and long-standing grievances between Hutu and Tutsi. Verwimp and Van Bavel evaluate the impact of the 1993-2000 war in Burundi and the 1993-1994 massacres on the educational attainment of boys and girls who were of primary school age at the time, using individual and household survey data collected in 2002. They use several province- and child-level conflict intensity measures. To differentiate households by their economic status, the authors use self-reported retrospective data on the household *pre-war* (1993) economic characteristics. This differentiation of households by their pre-war economic status is important, as post-war wealth status is likely to be affected by the conflict. Similar to Akresh and de Walque (2010) study for Rwanda, they find that the negative effect of conflict on education is significantly more pronounced for boys (but not girls) from non-poor households – reducing the gender gap in schooling. They

attribute this finding to the predominance of non-poor boys amongst those sent to school in the pre-conflict period, so they had the most to lose from schooling disruption during the conflict. In pre-war poor-households, educational attainment of both, boys and girls declined. Further, the authors note that the negative effect of displacement on education is particularly strong for children who had to move multiple times, especially these from poor households.

DISCUSSION

The three studies in this issue illustrate how the context shapes the impact of armed conflict on children's schooling, including gender differentials in schooling. In several settings, boys' schooling is more negatively affected than that of girls – sometimes because boys are withdrawn by their families to contribute to household income, as in Timor Leste (Justino, Leone, and Salardi 2014), or because they are drafted into serving in the conflict. Where boys have much greater access to schooling than girls in the pre-conflict situation, as in Burundi (Verwimp and Van Bavel 2014), boys are more affected by the conflict-induced reduction of access to schooling, which reduces the gender gap in schooling. Elsewhere, as in Nepal (Valente 2014), where insurgents were ideologically motivated to reduce inequalities in access to schooling, a greater presence of insurgents (as measured by killings) helped increase girls' access to schooling – except when the insurgents stepped up abductions. In cultures which emphasize protecting girls' reputations and consequently their marriageability, conflict's disruptions can reduce their access to schooling. All these differences prevail despite the fact that conflict typically disrupts schooling infrastructure and teachers' ability to fully perform their duties that should affect education of boys and girls in a similar manner.

Rigorous evaluation research is needed on post-conflict societies and policies, including the effectiveness of post-conflict reconstruction efforts in addressing gaps and gender inequalities in schooling created by the conflict, as well as the effectiveness of aid projects in improving children's school outcomes.

REFERENCES

- Akbulut-Yuksel, M. 2009. "Children of War: The Long-Run Effects of Large-Scale Physical Destruction and Warfare on Children." IZA Discussion Paper. IZA-Institute for the Study of Labor, Bonn, Germany.
- Akresh, R., and D. de Walque. 2010. "Armed Conflict and Schooling: Evidence from the 1994 Rwandan Genocide." IZA Discussion Paper. IZA-Institute for the Study of Labor, Bonn, Germany.
- Almond, D., and J. Currie. 2011. "Human Capital Development Before Age Five." In D. Card and O. Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Part B, pp. 1315–486. Amsterdam, The Netherlands: Elsevier.
- Annan, J., C. Blattman, D. Mazurana, and K. Carlson. 2011. "Civil War, Reintegration, and Gender in Northern Uganda." *Journal of Conflict Resolution* 55(6): 877–908.

- Ariño, M.V. 2008. "Nepal: A Gender View of the Armed Conflict and Peace Process." *Quaderns de Construcció de Pau* No.4. Escola de Cultura de Pau, Barcelona, Spain.
- Blattman, C., and J. Annan. 2010. "The Consequences of Child Soldiering." *Review of Economics and Statistics* 92(4): 882–98.
- Blattman, C., and E. Miguel. 2010. "Civil War." *Journal of Economic Literature* 48 (1): 3–57.
- Brück, T., M. Di Maio, and S. Miaari. 2013. "Exposure to Violence and Student Achievement in Palestine: Evidence from the Second Intifada." DiW Berlin Working Paper. Berlin, Germany.
- Buvinić, M., M. Das Gupta, U. Casabonne, and P. Verwimp. 2013. "Violent Conflict and Gender Inequality: An Overview." *The World Bank Research Observer* 28(1): 110–38.
- Cassar, A., P. Grosjean, and S. Whitt. 2011. "Civil War, Social Capital and Market Development: Experimental and Survey Evidence on the Negative Consequences of Violence." University of San Francisco, San Francisco, CA.
- Chamarbagwala, R., and H.E. Morán. 2011. "The Human Capital Consequences of Civil War: Evidence from Guatemala." *Journal of Development Economics* 94(1): 41–61.
- Collier, P., V.L. Elliot, H. Hegre, A. Hoeffler, M. Reynal-Querol, and N. Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy*. Washington, DC, and New York: World Bank and Oxford University Press.
- de Groot, O.J., and I. Göksel. 2011. "Conflict and Education Demand in the Basque Region." *Journal of Conflict Resolution* August 55(4): 652–77.
- de Walque, D. 2006. "The Socio-Demographic Legacy of the Khmer Rouge Period in Cambodia, in Population Studies." *Population Studies* 60(2): 223–31.
- Foltz, J.D., and K.A. Opoku-Agyemang. 2011. "Low-Intensity Conflict and Schooling Outcomes: Evidence from Uganda." Unpublished Working Paper. University of Wisconsin, Madison, WI.
- Ichino, A., and R. Winter-Ebmer. 2004. "The Long-Run Educational Cost of World War II." *Journal of Labor Economics* 22(1): 57–86.
- Justino, P., M. Leone, and P. Salardi. 2014. "Short- and Long-Term Impact of Violence on Education: The Case of Timor Leste." *The World Bank Economic Review* 28(2): 320–53.
- Kochar, A. 1999. "Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India." *Review of Economics and Statistics* 81(1): 50–61.
- Kibris, A. 2013. "The Conflict Trap Revisited: Civil Conflict and Educational Achievement." Working Paper. Sabanci University, Istanbul, Turkey.
- León, G. 2012. "Civil Conflict and Human Capital Accumulation: The Long Term Effects of Political Violence in Perú." *Journal of Human Resources* 47(4): 991–1023.
- Merrouche, Q. 2011. "The Long Term Educational Cost of War: Evidence from Landmine Contamination in Cambodia." *Journal of Development Studies* 47 (3): 399–416.
- Monteiro, J., and R. Rocha. 2012. "Drug Battles and School Achievement: Evidence from Rio de Janeiro's Favelas." NEUDC 2012 Presentation. Dartmouth, NH. http://www.dartmouth.edu/~neudc2012/docs/paper_260.pdf. Last accessed: August 8, 2013.
- Morduch, J. 1995. "Income Smoothing and Consumption Smoothing." *Journal of Economic Perspectives* 9(3): 103–14.
- Rodriguez, C., and F. Sanchez. 2012. "Armed Conflict Exposure, Human Capital Investments and Child Labor: Evidence from Colombia." *Defense and Peace Economics* 23(2): 161–84.
- Rohner, D., M. Thoenig, and F. Zilibotti. 2011. "Seeds of Distrust: Conflict in Uganda." Working Paper No. 417. Department of Economics, Center for Institutions, Policy and Culture in the Development Process, University of Zurich, Zurich, Germany.
- Rose, E. 1999. "Consumption Smoothing and Excess Female Mortality in Rural India." *Review of Economics and Statistics* 81(1): 41–49.
- . 2001. "Ex Ante and Ex Post Labor Supply Response to Risk in a Low-Income Area." *Journal of Development Economics* 64: 371–88.

- Shemyakina, O. 2011. "The Effect of Armed Conflict on Accumulation of Schooling: Results from Tajikistan." *Journal of Development Economics* 95(2): 186–200.
- Singh, P., and O. Shemyakina. 2013. "Gender-Differential Effects of Conflict on Education: The Case of the 1981-1993 Punjab Insurgency." HiCN Working Paper N 143, Households in Conflict Network.
- Strauss, J., and D. Thomas. 2008. "Health over the Life Course." In P. Schultz and J. Strauss, eds., *Handbook of Development Economics*, Vol.4. Amsterdam, The Netherlands: North-Holland.
- Swee, E.L. 2011. "On War and Schooling Attainment: The Case of Bosnia and Herzegovina." Unpublished manuscript, University of Melbourne, Melbourne, Australia.
- Valente, C. 2014. "Education and Conflict in Nepal." *World Bank Economic Review* 28(2): 354–83.
- Verwimp, P., and J. Van Bavel. 2014. "Schooling, Violent Conflict, and Gender in Burundi." *World Bank Economic Review* 28(2): 384–411.
- World Bank. (2003). "Timor-Leste Poverty Assessment: Poverty in a New Nation: Analysis for Action. Volume II." Joint Report of the Government of the Democratic Republic of Timor-Leste, ADB, JICA, UNDP, UNICEF, UNMISSET and the World Bank. <http://purl.pt/915/1/cd1/ta100/ta140%20All-Vol2-Final.pdf>. Last accessed: March 4, 2012.

Short- and Long-Term Impact of Violence on Education: The Case of Timor Leste

Patricia Justino, Marinella Leone, and Paola Salardi

This paper analyzes the impact of the wave of violence that occurred in Timor Leste in 1999 on education outcomes. We examine the short-term impact of the violence on school attendance in 2001 and its longer-term impact on primary school completion of the same cohorts of children observed again in 2007. We compare the educational impact of the 1999 violence with the impact of other periods of high-intensity violence during the 25 years of Indonesian occupation. The short-term effects of the conflict are mixed. In the longer term, we find evidence of a substantial loss of human capital among boys in Timor Leste who were exposed to peaks of violence during the 25-year long conflict. The evidence suggests that this result may be due to household trade offs between education and economic welfare. JEL Codes: I20, J13, O12, O15

The developmental consequences of violence and conflict are far reaching, affecting millions of men, women, and children (World Bank 2011). The objective of this paper is to examine one important channel linking violent conflict and development outcomes: the education of children living in contexts of conflict and violence. The paper focuses on the case of Timor Leste, particularly the last wave of violence in 1999 during the withdrawal of Indonesian troops from the territory. We analyze the short-term impact of the 1999 violence on primary school attendance in 2001 and its longer-term impact on primary school completion in 2007. In addition, we separately examine the impact of early periods of high-intensity violence (HVI) during the 25 years of Indonesian occupation and the effects of the entire conflict on primary school completion in 2007 to compare the average impact of the overall conflict period with the educational impact of singular peaks of violence. This is a

Patricia Justino (corresponding author) is a Senior Research Fellow at the Institute of Development Studies in Brighton, United Kingdom, the Director of MICROCON (www.microconflict.eu) and the Co-founder and Co-director of Households in Conflict Network (www.hicn.org). Her email address is p.justino@ids.ac.uk. Marinella Leone is a PhD candidate at the Department of Economics, University of Sussex, United Kingdom. Her email address is m.a.leone@sussex.ac.uk. Paola Salardi is a PhD candidate at the Department of Economics, University of Sussex, United Kingdom and a Researcher for MICROCON. Her email address is p.salardi@sussex.ac.uk.

THE WORLD BANK ECONOMIC REVIEW, VOL. 28, NO. 2, pp. 320–353
Advance Access Publication February 28, 2013

doi:10.1093/wber/lht007

© The Author 2013. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

unique and important feature of this paper because long conflicts are not characterized by constant levels of violence. Although armed conflict has considerable effects on people's lives, there is an important theoretical distinction between the conflict process and the violence that occurs at different times and in different places (Kalyvas 2006).

From a theoretical perspective, the long-term developmental effects of violent conflict are ambiguous. Standard neoclassical growth models predict that the temporary destruction of capital can be overcome in the long run by higher investments in affected areas.¹ However, the long-term destructive effects of violent conflict may remain entrenched in certain regions and among some population groups even if economic growth converges at the aggregate level. Recent research on the microlevel effects of violent conflict has shown that the negative impact of conflict on educational outcomes, labor market participation, and the health status of individuals and households may be observed decades after the conflict.²

Children may be particularly affected by conflict because many human capital investments are age specific. The destruction of human capital during childhood is a well-documented mechanism explaining long-term trends in household welfare (Alderman, Hoddinott, and Kinsey 2006; Case and Paxson 2008; Maccini and Young 2009).

The educational effects of violent conflict are particularly substantial. The existing literature shows that violent conflict almost always results in reductions in educational access and attainment (Akresh and de Walque 2011; Alderman, Hoddinott, and Kinsey 2006; Chamarbagwala and Morán 2010; Shemyakina 2011). Relatively minor shocks to educational access during childhood can lead to significant and long-lasting detrimental effects on individual human capital accumulation (Akbulut-Yuksel 2009; Ichino and Winter-Ebner 2004; Leòn 2012).

We analyze the short- and long-term impacts of violence on primary school attendance and completion in Timor Leste using data from two nationally representative household surveys collected in 2001 and 2007. We focus on primary school outcomes because most individuals in Timor Leste (approximately 65 percent) have, at most, only primary school education (TLSS 2007b). Our identification strategy exploits both individual-level violence measures and temporal and geographical variation in the incidence of the conflict using data from the East Timor Human Rights Violations Database (CAVR 2006).

Our results show mixed evidence for the impact of violent conflict on educational outcomes. Mirroring the findings of Bellows and Miguel and others, we find evidence for a rapid recovery in educational outcomes among girls in

1. See the discussion in Blattman and Miguel (2010) and the evidence in Bellows and Miguel (2006), Davis and Weinstein (2002), and Miguel and Roland (2011).

2. See reviews in Blattman and Miguel (2010) and Justino (2009, 2012a).

Timor Leste. However, we find that the 1999 wave of violence in Timor Leste, as well as the peaks of violence in the 1970s and 1980s, resulted in persistent negative effects on primary school attendance and completion among boys. We present evidence suggesting that boys were less able to benefit from postconflict recovery as a result of household trade offs between education and economic survival that may have led to the removal of boys from school.

The paper is structured as follows. Section I provides a descriptive background of the conflict in Timor Leste and the country's education sector. In section II, we describe the datasets, discuss our identification strategy, and present some descriptive results. Section III discusses our empirical results as well as a range of robustness checks. Section IV concludes the paper.

I. VIOLENT CONFLICT AND THE EDUCATION SECTOR IN TIMOR LESTE

Timor Leste was under Portuguese colonial rule from 1500 to 1974. After the Portuguese left, Indonesia forcefully annexed the territory, leading to a guerrilla war spurred by the Revolutionary Front for an Independent East Timor and its armed wing (the Armed Forces for the National Liberation of East Timor). Several thousand individuals were forcibly displaced during the Indonesian occupation and forced to live in extreme conditions without adequate food, shelter, or health facilities (Felgueiras and Martins 2006; Gusmão 2004). Approximately 60,000 people lost their lives in the early years of the occupation. The number of deaths reached 200,000 by the end of the occupation (UNDP 2002).

The situation in Timor Leste received little international attention until the Santa Cruz massacre in November 1991, in which Indonesian forces killed 200 protesters. The massacre was broadcast by the international media and raised considerable awareness of human rights violations during the Indonesian occupation. The independence movement received support from the Portuguese government and international organizations, including the UN. These events, in addition to the 1997 financial crisis, resulted in Indonesia agreeing to a referendum on the independence of Timor Leste. On August 30, 1999, 79 percent of the population of Timor Leste voted in favor of independence.

The aftermath of the referendum generated a wave of destruction, violence, and human rights violations by Indonesian forces and militias (Alonso and Brugha 2006). The number of killings during this wave of violence has been estimated at between 1,000 and 2,000 people, approximately 0.2 percent of the Timorese population (Robinson 2003; UNDP 2002). This wave of violence was characterized by massive displacement and the destruction of private dwellings and public infrastructure following the "scorched-earth" tactics employed by the Indonesian troops and pro-Indonesia militia groups (CAVR 2006; UNDP 2002). Approximately 80 percent of the country's infrastructure and buildings were destroyed during the withdrawal of Indonesian troops and

militias (UNDP 2002). In October 1999, a United Nations Transitional Administration was established in Timor Leste.

Variation in the Conflict across Time and Space

The conflict in Timor Leste has evolved in different ways over time and across space. The Timor Leste Commission for Reception, Truth, and Reconciliation, established in 2001, has identified three distinct phases of the conflict during the period between December 1975 and September 1999 (CAVR 2005). The first phase, from 1975 to 1984, was related to the initial Indonesian invasion and occupation of Timor Leste. The first few years, from 1975 to 1979, were the most intense in terms of killings and destruction. The second phase, from 1985 to 1998, was characterized by the consolidation and normalization of the occupation. Although people were killed in this phase (for instance, during the Santa Cruz massacre), the violence during this period was of relatively low intensity. The third phase of the conflict was identified with the 1999 withdrawal of Indonesian troops and the accompanying wave of violence. The main peaks of violence across these three periods were 1975–79, 1983, and 1999, coinciding with more intense fighting between the two factions (CAVR 2005). There were two main types of victims during this last wave of violence. The first was urban households, some (but not all) of which were supporters of the independence movement among or related to the Timorese intelligentsia. Some of these individuals were targeted and killed, whereas others fled from their areas of residence, fearing attacks by the Indonesian troops and militias in Dili and other urban areas (CAVR 2006; Robinson 2003). The second set of victims was mostly poor farmers who fled to safer areas or fell victim to the scorched-earth tactics employed by Indonesian forces withdrawing from Timor Leste (CAVR 2006).

The conflict was also characterized by significant variation at the geographical level, which we exploit in our empirical analysis. The violence was primarily concentrated in specific areas, and its geographic variation generally followed the movement of the Indonesian military forces. The occupation was more intense initially in the western region of Timor Leste because of the proximity to the West Timor border. It then spread to the central and eastern regions. The last wave of violence in 1999 was particularly intense in the western region and the urban areas of the central regions (CAVR 2005). The concentration of violence in 1999 in the western districts was also due to a long-established network of pro-Indonesian groups since before 1999. In contrast, the eastern and central regions were important areas for the resistance forces (Robinson 2003). We will explore this variation in violence across time and space in the empirical analysis below.

The levels of violence experienced in Timor Leste declined considerably after independence. In 2006, Timor Leste faced renewed civil strife as a result of fighting between different factions of the independence movement (Muggah et al. 2010; Scambray 2009). Although fighting and violence have become less

pronounced, some areas of Timor Leste continue to face serious challenges in terms of insecurity, youth unemployment, and violence (Muggah et al. 2010). This paper specifically focuses on the effects of the 1999 wave of violence and the previous years of the Indonesian occupation, but we also discuss the potential implications of the 2006 civil strife on our results in section III.

The Education Sector in Timor Leste

Beginning in 1999, substantial funds from bilateral and multilateral donors flowed into Timor Leste to support the reconstruction and rehabilitation of the country. Although Timor Leste was severely devastated during the 1999 wave of violence, the reconstruction of state institutions, school systems, infrastructure, and markets was relatively successful and rapid (World Bank 2003b). The main development indicators for the country in 2001 were close to the pre-1999 values. However, Timor Leste was (and is) one of the world's least developed countries (UNDP 2002).

Under Portuguese colonial rule, the Catholic Church was the major provider of education, with schooling primarily available for the elite in urban areas. The literacy rate was approximately 5 percent in 1975, and gender disparities were large (UNDP 2002). The Indonesian government expanded educational access to the entire population of Timor Leste, primarily as a means of controlling the population (Nicolai 2004). Enrollment rates increased over those years, and gender gaps began to close (UNDP 2002). Despite this progress, educational performance under the Indonesian occupation was characterized by delayed school entry, high repetition rates, and high dropout rates owing to the low quality of schools and teaching and high fees. Some Timorese were also unwilling to send their children to school because this was perceived as a sign of participation in the repressive Indonesian system (UNDP 2002). In 1995, less than half of individuals aged between 15 and 19 had completed primary school education (UNDP 2002).

The school system was almost totally destroyed in the immediate aftermath of the 1999 violence, and schools did not reopen until October 2000. However, children were still able to attend classes taught in the open air in makeshift camps (Rohland and Cliffe 2002), and substantial effort was applied to the reconstruction of the education system in Timor Leste (World Bank 2003a). In particular, the Trust Fund for East Timor included substantial funding for the renovation of damaged schools and the construction of new ones (USD 27.8 million over three years). Within a few months, many schools had been rebuilt, thousands of books had been replaced, and teachers had been recruited (Rohland and Cliffe 2002; World Bank 2003a).

During this rapid reconstruction process, primary school enrollment rates improved significantly. This increase was aided by the elimination of school fees and the reintroduction of Portuguese as the primary language of instruction. As a result, a large number of over-age students enrolled in primary school for the first time, and net primary school enrollment in Timor Leste

rose from 65 to 74 percent between 1999 and 2001. Gender differentials decreased significantly as a result of a large increase in female literacy rates (World Bank 2003a). However, the reconstruction of the school system in Timor Leste faced numerous challenges owing to the shortage of teachers and schools (UNDP 2006). Makeshift open-air schools were not ideal means of teaching children, and emergency funds were only available for a limited period of time. In 2007, most of the Timorese population continued to have little or no education.

II. DATA DESCRIPTION AND IDENTIFICATION STRATEGY

Our empirical study is based on two cross-sectional household surveys: the Timor Leste Living Standard Measurement Surveys (TLSS), which were jointly conducted by the National Statistics Directorate in Timor Leste and the World Bank in 2001 and 2007, including a broad range of individual- and household-level indicators. The TLSS 2001 surveyed 1,800 households from 100 *sucos* (villages), covering nearly 1 percent of the population (TLSS 2001). The survey included direct questions on the exposure of individuals and households to the violence in 1999. The TLSS 2007 covered a sample of 4,477 households from all 498 *sucos* in Timor Leste (TLSS 2007a). The TLSS 2007 was conducted over a period of 12 months between December 2007 and January 2008.³ The TLSS 2007 did not contain direct information on exposure to violence. To identify individuals and households affected by violence, we exploit data on the number of killings across time and space collected in the Human Rights Violations Database to identify districts and years that experienced HVI at the beginning of and during the occupation and following the withdrawal of Indonesian troops in 1999.⁴

Identification Strategy I: The Impact of Violence on School Attendance in 2001

We first investigate the short-term impact of the 1999 violence on the school attendance of boys and girls observed in 2001.⁵ We consider two different channels of exposure to violence. The first identifies individuals belonging to households that were displaced as a result of the 1999 wave of violence (all

3. The survey was launched in March 2006 but had to be suspended due to the outbreak of internal violence in the country (mostly in Dili). The survey was resumed in January 2007 and conducted over one year. All households interviewed in 2006 (351 households) were revisited and reinterviewed in 2007. Those not found at the time of the new interview (34 households) were replaced with new households (TLSS 2007a).

4. These data were compiled by the Commission for Reception, Truth, and Reconciliation from voluntary statements made by people (victims, perpetrators, and others) affected by violence.

5. We do not analyze primary school completion in 2001 because most children who were of school age in 1999 were still in school in 2001.

members displaced). The second identifies individuals in households that report having their homes completely destroyed by the violent attacks in 1999.⁶

The TLSS 2001 contains useful retrospective information on school attendance in three different school years: 1998/99, 1999/00, and 2000/01. We are interested in the year of recovery (2000/01). Because the 1999 violence primarily occurred in the summer and fall of 1999, we can assume with a high degree of confidence that the 1998/99 school year was not affected by conflict, whereas the 1999/00 school year began during the wave of violence. Note that many children continued to be able to attend school in 1999. However, these were generally makeshift open-air schools in internally displaced person camps established by the international community (Nicolai 2004; Richter 2009).

To employ the retrospective information on school attendance provided in the dataset, we exploit the time variation in school attendance status. We construct a panel dataset in which each individual is observed over three school years, and attendance status is time variant. We focus our analysis on individuals who were of primary school age over the 1998–2001 period, ensuring that all children had a minimum age of seven in 1998/99 and a maximum age of 12 in 2000/01.⁷ We estimate the following equation using a linear probability model:

$$E_{it} = \alpha + \beta_1 T_2 + \beta_2 T_3 + \beta_3 V_i^k * T_2 + \beta_4 V_i^k * T_3 + \alpha_i + e_{it} \quad [1]$$

where E_{it} is a binary variable for school attendance for individual i at time t . T_2 and T_3 are year dummies for the 1999 violence and for the first year of the postviolence period (school year 2000/01), respectively. The reference year is the previolence year, 1998/99. The model includes individual fixed effects, α_i . e_{it} is the random error term. All standard errors are clustered at the village level.

Violence-affected individuals are identified using two different measures, V_i^k , with $k = 1, 2$ depending on whether displacement or the destruction of a home is included in the specification. Of the children in the sample, 16 percent and 25 percent live in households that were displaced or had their homes destroyed, respectively. We allow the violence measure to interact with both year

6. The 2001 TLSS also contains self-reported information on the number of household members who have died as a result of violence. In our sample, 148 individuals (living in 27 households) reported the violent death of a household member. Of these individuals, 88 percent were also affected by displacement and/or dwelling destruction, and only 13 of those 148 individuals were children between the ages of 7 and 12 during the violence. We have reestimated our model in table 5 excluding these 13 children. The results remain unchanged. These estimates are not reported because of space constraints but are available upon request.

7. We have analyzed a larger sample that includes children of primary school age in the year of the violence (i.e., between 7 and 12 years old in 1999). This includes individuals aged 6 in 1998 and aged 13 in 2000. The inclusion of these individuals may generate “spurious” results because they are not all of primary school age. We have estimated the model using both samples. The results (not shown) are very similar; therefore, we opted to concentrate on the most restrictive sample.

dummies. The estimation of our specification employs a difference-in-difference methodology. $V_i^k * T_2$ represents the difference-in-difference term between the prewar year and the year of conflict, whereas $V_i^k * T_3$ represents the difference-in-difference term between the prewar year and the postwar year. We focus our attention on the coefficient β_4 because we are primarily interested in understanding the effects of violence on postwar outcomes. We also explore both the separate and joint impacts of each channel of violence by adding a triple interaction between the two violence dummies and the time dummies. This specification allows us to isolate the impact of only being displaced, only having the home destroyed and being affected by both shocks.⁸ This specification ensures that the control group does not include individuals affected by violence.

In table 1, we present average school attendance rates, disaggregated by gender and age groups, for the same cohort of children aged 7–10 years in 1998, 8–11 years in 1999, and 9–12 years in 2000. In general, the attendance rates for the whole sample increase over time and are higher for girls. There are, however, considerable differences in attendance rates between children affected by violence and those who do not report victimization. These differences are reported in figure 1, where we disaggregate school attendance averages between violence-affected and unaffected individuals. As expected, we observe a decline in school attendance in 1999 for children affected by violence.

We present the individual and household characteristics of children affected by the 1999 violence in table 2. The table demonstrates that children from displaced households are better off overall than those from households that were not displaced. Many of these were urban households that fled their areas of residence because they feared being targeted by the Indonesian troops stationed in Dili and other urban areas in the central regions (CAVR 2005; Robinson 2003). Households that report having their homes destroyed by violent attacks or affected by both shocks are generally poor farmers living in rural areas. These households are likely to be indiscriminate victims of the scorched-earth tactics employed by the Indonesian troops withdrawing to West Timor (CAVR 2005). Interestingly, we find that boys (aged 10–12) affected by displacement work more hours than unaffected individuals, whereas the opposite is true for girls.

We exploit the panel nature of the data to estimate the causal effects of the 1999 wave of violence on education outcomes. We estimated a fixed effects model,⁹ which allows us to eliminate time-invariant unobserved individual characteristics that may be correlated with the conflict measure and our

8. Of the children in our sample, 67.1 percent were not affected by any shock. Moreover, 7.5 percent of all children were only displaced, and 17.4 percent only had their homes destroyed. Finally, 8 percent of the sample was affected by both shocks.

9. The fixed effects model is more appropriate than a random effects model because we would have to assume that the unobserved component of the individual fixed effects and the other covariates specified in the equation are uncorrelated. This assumption is likely to be violated in our case. This choice is also supported by Hausman test results.

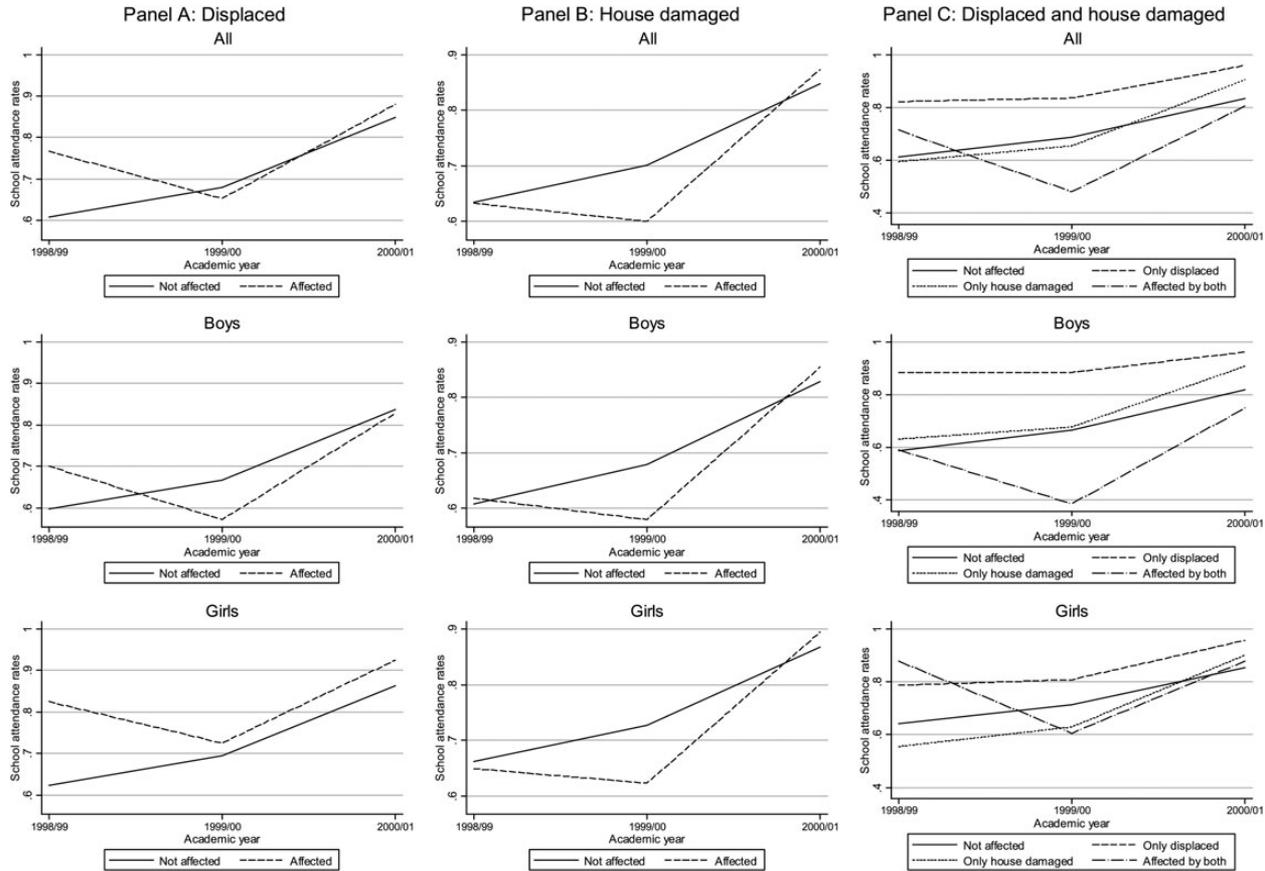
TABLE 1. Attendance Rates of Children Aged 7–12 Years Between 1998 and 2001

	All: 8–11 years old				Younger cohort: 8–9 years old				Older cohort: 10–11 years old			
	All	Boys	Girls	<i>t</i> test	All	Boys	Girls	<i>t</i> test	All	Boys	Girls	<i>t</i> test
1998/99	0.634 (0.012)	0.611 (0.018)	0.659 (0.017)	n.s.	0.509 (0.020)	0.498 (0.027)	0.521 (0.028)	n.s.	0.750 (0.018)	0.720 (0.026)	0.782 (0.025)	n.s.
1999/00	0.676 (0.012)	0.654 (0.017)	0.700 (0.017)	n.s.	0.622 (0.019)	0.602 (0.027)	0.647 (0.028)	n.s.	0.726 (0.018)	0.705 (0.026)	0.749 (0.026)	n.s.
2000/01	0.854 (0.014)	0.836 (0.019)	0.874 (0.019)	*	0.822 (0.022)	0.789 (0.030)	0.860 (0.031)	**	0.884 (0.021)	0.881 (0.029)	0.887 (0.029)	n.s.
N	966	512	454		466	251	215		500	261	239	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n.s. = not statistically significant. We consider the same cohort over time: the sample is aged 7–10 years in 1998, 8–11 years in 1999, and 9–12 years in 2000.

Source: Authors' computations using TLSS 2001.

FIGURE 1. School Attendance Rates by Channel of Violence Exposure



Source: Authors' own computations using TLSS 2001.

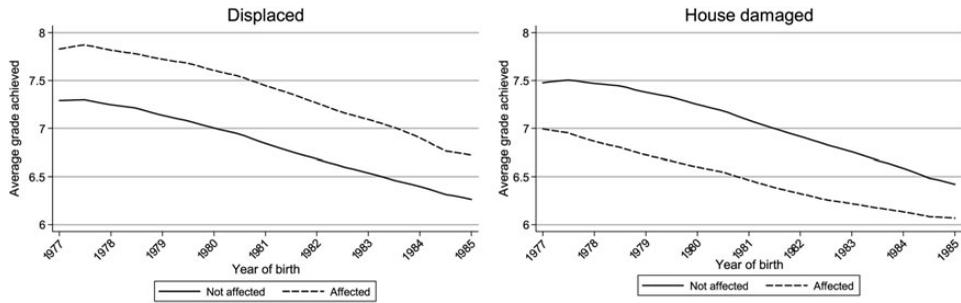
TABLE 2. Individual and Household Characteristics by Channel of Violence Exposure in 2001

	All children 7–12						Boys 7–12						Girls 7–12					
	Displaced			House damaged			Displaced			House damaged			Displaced			House damaged		
	0	1	<i>t</i> test	0	1	<i>t</i> test	0	1	<i>t</i> test	0	1	<i>t</i> test	0	1	<i>t</i> test	0	1	<i>t</i> test
Panel A – All children (7–12 years old)																		
Being female	0.472	0.500		0.484	0.454	n.s.												
Speaking Indonesian	0.575	0.720	***	0.584	0.632	n.s.	0.577	0.701	**	0.580	0.634	n.s.	0.574	0.738	***	0.589	0.631	
Speaking Portuguese	0.028	0.033	n.s.	0.030	0.027	n.s.	0.032	0.028	n.s.	0.031	0.035	n.s.	0.024	0.037	n.s.	0.029	0.018	
HH head is a farmer	0.646	0.556	**	0.623	0.659	n.s.	0.649	0.570	n.s.	0.639	0.634	n.s.	0.642	0.542	*	0.606	0.690	**
Education grade of HH head	3.114	3.651	n.s.	3.332	2.783	**	2.954	3.210	n.s.	3.099	2.685	n.s.	3.293	4.084	n.s.	3.581	2.899	*
Education grade of the mother	1.870	2.785	***	2.200	1.435	***	1.827	2.252	n.s.	2.061	1.406	**	1.918	3.318	***	2.348	1.470	***
Education grade of the father	2.824	3.495	*	3.037	2.589	**	2.802	2.869	n.s.	2.944	2.446	n.s.	2.848	4.121	**	3.136	2.762	
Living in urban areas	0.402	0.533	***	0.419	0.427	n.s.	0.396	0.551	***	0.413	0.431	n.s.	0.409	0.514	**	0.426	0.423	
Per capita monthly HH expenditure	238,963	262,113	n.s.	244,940	234,904	n.s.	250,901	249,806	n.s.	245,971	263,914	n.s.	225,635	274,421	*	243,843	200,024	***
N	1236	214		1080	370		652	107		557	202		584	107		523	168	
Panel B – Children aged 10–12 (labor market characteristics)																		
Has worked in the past seven days	0.063	0.088	n.s.	0.070	0.056	n.s.	0.054	0.167	**	0.071	0.063	n.s.	0.073	0.019	**	0.070	0.048	n.s.
Working hours	1.468	2.098	n.s.	1.722	1.084	n.s.	1.117	4.292	**	1.582	1.406	n.s.	1.857	0.148	***	1.868	0.711	n.s.
Has performed domestic chores	0.902	0.912	n.s.	0.897	0.922	n.s.	0.889	0.854	n.s.	0.888	0.875	n.s.	0.916	0.963	n.s.	0.907	0.976	n.s.
N	602	102		525	179		316	48		268	96		286	54		257	83	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n.s. = not statistically significant. In the title, 0 refers to not affected individuals while 1 refers to affected individuals. HH indicates household. Per capita monthly HH expenditure is expressed in Rupiah in real values using CPI of 2001.

Source: Authors' computations using TLSS 2001.

FIGURE 2. Preconflict Trends in Education Levels



Note: The plots are based on the estimation of separate kernel-weighted local polynomial regressions of the last completed grade of schooling against age using an Epanechnikov kernel.

Source: Authors' own computations using TLSS 2001.

dependent variable. Our specification also includes year dummies that allow us to control for unobserved time-variant heterogeneity.

To ensure that our key identifying assumption is not violated, we checked whether trends in education before the 1999 violence were parallel between groups affected by the violence and those unaffected by the violent events. We examined the average school grades of affected individuals and unaffected individuals who were not exposed to the 1999 violence during their primary school years and who were old enough in 1999 to have at least completed primary school.¹⁰ The results indicate that it is unlikely that preexisting differences in education trends drive our postconflict outcomes (see figure 2). This evidence, combined with the association of the violence with the Indonesian troop movements described in section I, strongly indicates that the effects of the violence that occurred in Timor Leste in 1999 on individual educational outcomes are unlikely to be driven by a systematic correlation between the determinants of individual educational attainment levels before 1999 and the incidence of the 1999 violence at the individual level.

Despite the evidence discussed above, there is a small possibility that this strategy may be unable to control for all of the unobservable individual characteristics that may be correlated with both conflict incidence and educational outcomes. In particular, there are two common omitted variables in empirical analyses of conflict that may affect our results (see Kalyvas 2006). The first variable is a household's level of support for armed groups. Supporters of proindependence groups in Timor Leste were likely to be targets of violent attacks by Indonesian forces. In that case, the correlation between the conflict variables and this potentially omitted variable would be positive. If supporters were also more educated and hence more likely to send their children to school, our estimated effect would be

10. We do not include cohorts born after 1985 because the educational attainment of these individuals might be censored.

biased upward. The use of a fixed effects model controls for these effects as a component of time-invariant individual heterogeneity. The results of this paper may nevertheless indicate a lower negative impact of the conflict on education than if we were able to account for this potential unobservable in the case that levels of support changed during the conflict. This is unlikely to have been the case in Timor Leste in light of the discussion in section I. Another common omitted variable is the level of control of different armed factions. In the case of Timor Leste, the level of control of the Revolutionary Front for an Independent East Timor and the Indonesian troops is likely to vary with the geographical characteristics of each area as well as their proximity to West Timor. We control for this by including individual fixed effects in our specifications.

Identification Strategy II: The Impact of Violence on Primary School Completion in 2007

In this section, we investigate the longer-term consequences of the violence experienced in 1999 in Timor Leste on primary school completion in 2007 among the cohorts of children analyzed above. We then compare these results to the educational impact of the peaks of violence that occurred in earlier years of the conflict.

To construct a measure of exposure to violence, we matched information on the number of killings (provided in the Human Rights Violations Database dataset)—which varies over time and across districts—to information on the year and district of birth of each individual (provided in the [TLSS 2007a](#) dataset). We focus on the number of killings as our main conflict variable because we find that it serves as a good proxy for the intensity of the conflict across time and space. The occurrence of killings largely tracked the movements of the Indonesian military operations ([Silva and Ball 2006](#)). The number of killings also proxies for the destruction of homes and infrastructure and the displacement of people during the 1999 wave of violence, given the manner in which it occurred (i.e., the scorched-earth technique employed by Indonesian troops as they moved toward West Timor).

Matching this measure of violence to the year and district of birth of each individual allows us to identify whether and for how long each individual was exposed to the conflict during his or her primary school years. Our violence measure is defined as $V_{jt} = \sum_{a=7}^{12} v_{t+a}^j$, where each v_{t+a}^j takes a value of one if the individual was of primary school age in districts and years affected by the conflict. Specifically, j is the district of birth, t is the year of birth, and a is the primary school age (from 7 to 12). This measure ranges from zero to six if, from none to all six of a child's primary school years, respectively, were classified as exhibiting HVI. Because we only have information on the years in which individuals were supposed to have attended primary school,¹¹ we

11. These are not the years in which the individuals actually attended school because we do not have access to this information. The existence of a delay in school means that the “supposed” years of attendance might not coincide with the “actual” years of school attendance. However, given the way in which we identify our control and treatment groups, we do not expect this difference to affect our results.

assume that the district of birth is the district where the child attended school at the time of the violent events.

We define districts and years of HVI as those in which the number of killings in that year and district are above a given threshold, defined as the mean of the number of killings plus one standard deviation. The years in which the conflict was the most intense, as defined by our threshold, are 1975–1979, 1983, and 1999. This observation coincides with the history of the conflict discussed in section I (see [CAVR 2005](#)).¹²

The definition of HVI districts and years as a binary variable instead of a continuous one is primarily justified by our interest in capturing the incidence of violent conflict rather than its scale and magnitude.¹³ In addition, the distribution of killings is highly right skewed, further justifying the use of a binary variable. A Kernel density plot of the number of killings (not shown) demonstrates that where and when the conflict events occurred, we observe a considerably higher number of violent events; otherwise, we observe a low to negligible number of events. Finally, and more important, the use of a discrete variable allows us to minimize potential biases deriving from the potential underreporting of violent events. The Human Rights Violations Database dataset was compiled from voluntary statements, which may have resulted in biased reports. For instance, individuals living in remote areas or sick and disabled people may have not been able to report abuses, whereas victims of sexual abuse or traumatized people may not have reported their true levels of exposure to violence. In contrast, socially active individuals may have been more likely to volunteer information ([Silva and Ball 2006](#)). Under these circumstances, the use of a continuous measure may lead to biased estimates reflecting potential self-selection into reporting violence ([Leòn 2012](#)). The direction of this bias is impossible to predict a priori and depends on how unobservable characteristics related to underreporting may be correlated with conflict exposure and the dependent variable.¹⁴

To estimate the effect of the 1999 violence on school completion in 2007, we include individuals born between 1977 and 1992 in our sample. The treatment group includes individuals who were between 7 and 12 years old in 1999 in HVI districts (born between 1987 and 1992). We do not include individuals born after 1992 because they may have not completed primary school by 2007. The control group includes individuals who were not of primary school age in 1999 (born between 1977 and 1986).

12. The districts most affected by violence in the earlier years of the conflict are Baucau, Lautem, Viqueque, Ainaro, Manufahi, Manatuto, Aileu, Dili, Ermera, and Bobonaro. Those most affected by the 1999 violence are Dili, Ermera, Bobonaro, Covalima, Liquica, and Oecussi.

13. We have checked the robustness of all results to the use of a continuous variable and two different thresholds of violence intensity defined as the number of killings in each district and year (i) above the mean plus half of a standard deviation and (ii) above the mean plus two standard deviations. The results obtained are largely similar to those reported in the paper and are available from the authors upon request.

14. We thank an anonymous referee for noting this issue.

To analyze the impact of earlier peaks of violence, we focus our analysis on a sample of individuals born between 1968 and 1984. The treatment group includes individuals who were of primary school age between 1975 and 1979 and in 1983 (born between 1968 and 1976) in HVI districts. We exclude those born before 1968 because the schooling system was very different before the Indonesian troops invaded Dili in 1975. We also do not include individuals born between 1985 and 1986 as they may have been affected by the 1999 violence although placebo tests presented later indicate that they have not been affected. One interesting aspect of this analysis is that the treatment term informs us not only about the effects of exposure to HVI but also about the number of years of primary school affected by this exposure to violence.

Finally, we analyze the effect of the whole conflict on school attainment in 2007. For this purpose, we consider the full sample of individuals born between 1968 and 1992, where the treatment groups are those identified above and the control group includes individuals born between 1977 and 1986. This allows us to calculate the average educational effect of exposure to any period of the conflict for boys and girls in different age groups.

To analyze the effect of the conflict on primary school completion in 2007, we estimate the following equation:

$$G_{ibjt} = \beta V_{jt} + \alpha_j + \alpha_t + \alpha_j t + X'_b \gamma + \varepsilon_{ijt} \quad [2]$$

where G_{ibjt} refers to primary school completion for individual i of household b born in district j in year t , defined as a binary variable equal to one if the individual has completed at least primary school and zero otherwise. The adoption of a binary variable as the dependent variable in place of a continuous one is motivated by our interest in primary school completion rather than school attainment in general. The education sector in Timor Leste is extremely underdeveloped, and most of the population is illiterate. Primary education is therefore a major concern in the country.¹⁵

In the regression above, all standard errors are clustered at the year and district of birth levels. The term X_b is a vector of household characteristics (education of the household head and whether the household head is a farmer). The term V_{jt} is defined as above and identifies individuals exposed to HVI. β is our parameter of interest, indicating whether an additional year of primary school

15. We examined whether our results are robust to the use of alternative definitions of the educational outcome measure. To investigate the robustness of the results to the use of a continuous rather than a binary variable, we used a maximum likelihood estimated ordered probit model for school grade attainment allowing for the censorship of those still in school. This estimation follows the methods proposed in Glewwe and Jacoby (1994), Holmes (2003), and Zhao and Glewwe (2010). None of the key findings on the impact of the intensity of violence on educational outcomes reported in this paper are materially altered under this alternative approach. The results of these exercises are available on request from the authors. We are grateful to an anonymous referee for encouraging us to investigate this issue further.

exposure to the conflict affects the probability of primary school completion after the conflict ended compared to an individual who was not affected by HVI during her primary school years. The two parameters α_j and α_t are fixed effects for the districts of birth and the years of birth, respectively, and the term $\alpha_j t$ represents district-specific linear trends.¹⁶

In table 3, we report the differences in average primary school completion in 2007 between individuals exposed to high- or low-intensity violence in each of the three samples analyzed. These descriptive statistics show that boys exposed to HVI in 1999 (1977–1992 sample) exhibit a lower attainment rate than those who are less exposed to violence. The opposite is true for girls. Children exposed to earlier peaks of high intensity violence (1968–1984 sample) exhibit a lower completion rate than those living in districts and years in which the violence was not as intense.

The empirical strategy discussed above assumes that no systematic relationship exists between the intensity of the violence across districts and preconflict education levels at the district level. The existence of time-varying unobservables that are correlated with both the outcome and the conflict variables would bias our results. We have discussed this issue in the section above. We show here that the assumption also holds for the medium- and long-term analysis. The inclusion of district fixed effects in equation [2] allows us to account for time-invariant differences in education levels across districts. By including district-specific time trends, we account for any difference in trends across districts and hence for any time-varying characteristics in a given district. However, this identification strategy still relies on the assumption that there is no correlation between preconflict trends in education and violence in specific districts. To test for this, we conducted placebo tests on cohorts that supposedly were not exposed to the conflict during their primary school years (table 4).

Because the geographical variation of the conflict differs between the early years and 1999, we estimate two separate models by defining different violence-affected districts and “placebo” cohorts. We construct two violence-affected district dummies equal to one if the individual’s district of birth is located in one of the HVI districts as defined above, during the early years of the conflict or during the 1999 violence, and zero otherwise.

The first placebo test concentrates on the early years of the conflict. We are unable to analyze preconflict cohorts because, as explained above, the cohort born before 1968 would have attended a different school system. Therefore, we define those born between 1977 and 1980 as exposed “placebo” cohorts and

16. We reestimated the equation including a cubic district trend and a square root district trend to account for possible nonlinear trends across districts. We do not find any difference in the estimates, and we therefore only show the results that include a linear district trend. Results are available upon request.

TABLE 3. Average Primary School Completion in 2007

	All			Boys			Girls		
	Low-intensity violence	High-intensity violence	<i>t</i> test	Low-intensity violence	High-intensity violence	<i>t</i> test	Low-intensity violence	High-intensity violence	<i>t</i> test
Panel A – All primary school age children									
1977–1992 sample	0.725 (0.006)	0.724 (0.013)	n.s.	0.752 (0.008)	0.709 (0.020)	**	0.698 (0.008)	0.739 (0.017)	**
1968–1984 sample	0.624 (0.007)	0.572 (0.023)	***	0.680 (0.010)	0.658 (0.032)	n.s.	0.569 (0.009)	0.472 (0.034)	***
1968–1992 sample	0.679 (0.005)	0.674 (0.009)	n.s.	0.720 (0.007)	0.692 (0.013)	**	0.636 (0.006)	0.654 (0.013)	n.s.
Panel B – Children of grade 1–3 age									
1977–1992 sample	0.731 (0.005)	0.692 (0.019)	**	0.751 (0.008)	0.673 (0.029)	***	0.709 (0.007)	0.711 (0.025)	n.s.
1968–1984 sample	0.623 (0.007)	0.571 (0.024)	***	0.679 (0.010)	0.659 (0.032)	n.s.	0.567 (0.010)	0.471 (0.035)	***
1968–1992 sample	0.687 (0.004)	0.637 (0.012)	***	0.723 (0.006)	0.666 (0.016)	***	0.650 (0.006)	0.605 (0.016)	**
Panel C – Children of grade 4–6 age									
1977–1992 sample	0.719 (0.005)	0.759 (0.018)	**	0.738 (0.008)	0.749 (0.028)	n.s.	0.700 (0.007)	0.769 (0.023)	***
1968–1984 sample	0.622 (0.007)	0.524 (0.037)	***	0.680 (0.010)	0.634 (0.051)	n.s.	0.564 (0.009)	0.385 (0.054)	***
1968–1992 sample	0.675 (0.004)	0.690 (0.014)	n.s.	0.711 (0.006)	0.713 (0.021)	n.s.	0.638 (0.006)	0.664 (0.019)	n.s.

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n.s. = not statistically significant.

Source: Authors' computations using TLSS 2007a.

compare them to those born between 1981 and 1984.¹⁷ As a further check, we also analyze violence exposure for cohorts born between 1977 and 1981 and compare them to those born between 1982 and 1986. The treatment term is the interaction between the placebo cohort and the HVI dummies. We expect to find no effect of “exposure” for cohorts who were not of primary school age but were born in districts with HVI. We repeat the analysis with a focus on the 1999 violence. Individuals born between 1982 and 1986 were not of primary school age in 1999. We define this latter cohort as the “placebo” cohort and compare their exposure to that of those born between 1977 and 1981 in districts with high- and low-intensity violence. The results in table 4 show that cohorts who were not supposed to be of primary school age during the most violent years, but who were born in HVI districts, do not show significant differences in primary school completion rates relative to the same cohorts born in districts of low-intensity violence. This result supports our identification assumptions.

III. EMPIRICAL RESULTS

In this section, we discuss the results of the short- and long-term analyses.

School Attendance in 2001

The results in table 5 report the impact of the two channels of exposure to violence in 1999 on school attendance in the 1999/00 and 2000/01 school years. We are primarily interested in the differential effects of the violence on school attendance in the postviolence period, T_3 (2000/01), relative to the previolence year, T_1 (1998/99).

The results show a negative and significant impact of displacement on school attendance in 2000/01 for the overall sample. We find that being affected by displacement alone (panel C, table 5) decreases school attendance by 8.5 percentage points on average, with stronger effects for boys. Individuals affected by both shocks experience a reduction in school attendance of 13.3 percentage points on average, with girls being more severely affected.¹⁸ The effects are stronger for younger children.

These results suggest that different violence channels affect school attendance in heterogeneous ways. School attendance is most severely disrupted for children, particularly girls, who are affected by both types of violence. Considering the channels separately, we observe that displacement is the most disruptive channel in terms of consequences on children’s school attendance because all household assets are likely to have been lost. *Ibáñez and Moya*

17. The cohorts truly exposed to the early years of the conflict are those born between 1968 and 1976. In our placebo test, we examine the cohorts immediately following these.

18. We estimated a pooled model with interactions of the violence measures with the female dummy. The results reported in panel C are statistically different between girls and boys, as in table 5.

TABLE 4. Placebo Test for Differences in Trends in Education Levels

	1977–1984 sample			1977–1986 sample			1977–1986 sample		
	(1) All	(2) Boys	(3) Girls	(4) All	(5) Boys	(6) Girls	(7) All	(8) Boys	(9) Girls
HVI district*Cohort 1977–80	0.062 (0.047)	0.070 (0.054)	0.052 (0.061)						
HVI district*Cohort 1977–81				0.021 (0.044)	0.059 (0.048)	–0.018 (0.058)			
HVI district(a)	0.014 (0.031)	0.031 (0.032)	–0.003 (0.043)	0.040 (0.030)	0.014 (0.029)	0.063 (0.043)			
HVI district*Cohort 1982–86							–0.015 (0.043)	0.044 (0.050)	–0.078 (0.054)
HVI district(b)							–0.051 (0.031)	–0.037 (0.040)	–0.060 (0.038)
N	2,542	1,255	1,287	3,402	1,699	1,703	3,402	1,699	1,703
R-squared	0.158	0.156	0.153	0.140	0.131	0.141	0.141	0.130	0.151

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses clustered at the year of birth * district level. Regressions include year and district fixed effects and controls (whether the household head is a farmer and the household head's level of education). HVI district(a) equals one if the individual's district of birth is found to be a conflict-affected district during the early years of conflict (1975–1979 and 1983), as defined by our violence measure. HVI district(b) equals one if the individual's district of birth is found to be a conflict-affected district during the 1999 violence as defined by our violence measure.

Source: Authors' computations using TLSS 2007a.

TABLE 5. Impact of 1999 Violence on School Attendance in 2001

	(1) All 8–11	(2) Boys 8–11	(3) Girls 8–11	(4) All 8–9	(5) Boys 8–9	(6) Girls 8–9	(7) All 10–11	(8) Boys 10–11	(9) Girls 10–11
Panel A – Impact of displacement									
D*T ₂	-0.184*** (0.045)	-0.199*** (0.056)	-0.172*** (0.060)	-0.198*** (0.047)	-0.212*** (0.065)	-0.185** (0.074)	-0.188*** (0.061)	-0.209** (0.088)	-0.170** (0.077)
D*T ₃	-0.127*** (0.037)	-0.111** (0.049)	-0.141*** (0.048)	-0.182*** (0.053)	-0.138* (0.077)	-0.233*** (0.071)	-0.089** (0.038)	-0.106* (0.055)	-0.066 (0.055)
N	2,898	1,536	1,362	1,398	753	645	1,500	783	717
R-squared	0.151	0.155	0.146	0.217	0.199	0.241	0.110	0.130	0.091
Panel B – Impact of house damage									
H*T ₂	-0.101** (0.042)	-0.109** (0.054)	-0.091* (0.054)	-0.075 (0.051)	-0.137* (0.070)	-0.012 (0.070)	-0.129*** (0.048)	-0.077 (0.067)	-0.196*** (0.059)
H*T ₃	0.027 (0.041)	0.016 (0.051)	0.040 (0.052)	0.046 (0.059)	0.027 (0.077)	0.061 (0.078)	0.004 (0.037)	0.014 (0.052)	-0.016 (0.056)
N	2,898	1,536	1,362	1,398	753	645	1,500	783	717
R-squared	0.147	0.153	0.142	0.209	0.198	0.226	0.110	0.122	0.106

(Continued)

TABLE 5. Continued

	(1) All 8–11	(2) Boys 8–11	(3) Girls 8–11	(4) All 8–9	(5) Boys 8–9	(6) Girls 8–9	(7) All 10–11	(8) Boys 10–11	(9) Girls 10–11
Panel C – Impact of displacement and house damage									
D*T ₂	–0.060 (0.049)	–0.076 (0.060)	–0.050 (0.062)	–0.070 (0.057)	–0.086 (0.068)	–0.058 (0.087)	–0.079 (0.069)	–0.122 (0.109)	–0.064 (0.085)
D*T ₃	–0.085* (0.047)	–0.154*** (0.059)	–0.041 (0.065)	–0.153** (0.064)	–0.248*** (0.070)	–0.098 (0.100)	–0.039 (0.055)	–0.048 (0.110)	–0.020 (0.065)
H*T ₂	–0.015 (0.050)	–0.030 (0.068)	0.002 (0.057)	0.027 (0.063)	–0.039 (0.091)	0.084 (0.077)	–0.062 (0.053)	–0.011 (0.076)	–0.130** (0.058)
H*T ₃	0.087* (0.050)	0.045 (0.064)	0.134** (0.058)	0.114 (0.072)	0.036 (0.096)	0.173** (0.086)	0.052 (0.044)	0.065 (0.063)	0.028 (0.057)
D*H*T ₂	–0.233** (0.094)	–0.174 (0.114)	–0.296** (0.123)	–0.282*** (0.104)	–0.193 (0.136)	–0.364** (0.152)	–0.156 (0.122)	–0.116 (0.171)	–0.158 (0.151)
D*H*T ₃	–0.133* (0.077)	0.037 (0.103)	–0.304*** (0.087)	–0.122 (0.114)	0.166 (0.158)	–0.413*** (0.121)	–0.123 (0.075)	–0.129 (0.128)	–0.119 (0.098)
N	2,898	1,536	1,362	1,398	753	645	1,500	783	717
R-squared	0.162	0.164	0.163	0.228	0.217	0.257	0.122	0.134	0.117

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports fixed effect estimates. Robust standard errors in parentheses are clustered at the village level. Regressions include time effects (T₂ refers to 1999/00; T₃ refers to 2000/01). D, H, and D*H are dummies, respectively defined as one if individual was displaced with the whole household, whether the house was completely damaged, and whether the individual was affected by both violent shocks.

Source: Authors' computations using TLSS 2001.

(2010) show similar evidence for Colombia. The destruction of a home affects household wealth, but perhaps less so if the household was able to retain other assets or to live with friends, neighbors, or relatives.

School Completion in 2007

In table 6, panel A, we report the estimates of our analysis of the effect of the 1999 violence on primary school completion in 2007.

The coefficient for the violence measure is negative but not statistically significant. However, once we split the sample into boys and girls (columns 2 and 3),¹⁹ the results show that boys exposed to violence during their primary school years are 18.3 percentage points less likely to have completed primary school eight years after the 1999 violence relative to boys not exposed to violence. This represents a 25 percent decrease in the probability of primary school completion. In contrast, we observe that girls exposed to the 1999 violence are 10.4 percentage points more likely to have completed primary school in 2007. This represents a 14 percent increase in the probability of girls completing primary school. We do not find any statistically significant differences across age groups.

We now turn to the effect of the peaks of violence in the earlier years of the conflict on primary school completion in 2007. In table 6, panel B, we report the results for the sample of individuals born between 1968 and 1984. We find that an additional year of exposure decreases school completion in 2007 for all individuals by 2.6 percentage points and by 3 percentage points for boys. Therefore, the likelihood of primary school completion for boys was reduced (for an average exposure of one year and 10 months) by 5.6 percentage points. The effect is particularly strong for boys attending the last three years of primary school (grades four to six) (column 5). We do not find a significant effect for girls.

The results in table 6, panel C, report the effect of the overall conflict on primary school completion in 2007. The sample includes individuals born between 1968 and 1992. The results indicate the average effect of exposure to both the first years of the conflict and the 1999 violence. Because we examine the effects of both periods of high intensity violence and because only one year of primary school could have been affected during the 1999 violence, we have transformed our treatment term into a binary variable (exposed or not exposed during primary school) to ensure that we do not confound the results. These results indicate that boys exposed to the conflict in any period are, on average, 7.4 percentage points less likely to complete primary school in 2007 than those less exposed to violence. This effect represents a 10 percent decrease in the probability of primary school completion for boys. This effect is stronger

19. Similar to 2001, we estimated a pooled model for 2007 including an interaction term with the female dummy. The results show that the effects are statistically different between boys and girls in panels A and C. We report separate estimates for clarity in the exposition.

TABLE 6. Impact of Conflict on Primary School Completion in 2007

	(1) All	(2) Boys	(3) Girls	(4) All	(5) Boys	(6) Girls
Panel A – Impact of the 1999 violence (1977–1992 sample)						
Years of prim. school in HVI	–0.041 (0.029)	–0.183*** (0.044)	0.104** (0.047)			
Years of grade 1–3 in HVI				–0.069* (0.038)	–0.210*** (0.056)	0.080 (0.064)
Years of grade 4–6 in HVI				–0.040 (0.029)	–0.183*** (0.044)	0.105** (0.048)
N	6,676	3,383	3,293	6,676	3,383	3,293
R-squared	0.150	0.144	0.180	0.150	0.144	0.180
Panel B – Impact of early years of conflict (1968–1984 sample)						
Years of prim. school in HVI	–0.026** (0.011)	–0.030** (0.014)	–0.021 (0.017)			
Years of grade 1–3 in HVI				–0.021* (0.012)	–0.022 (0.016)	–0.018 (0.020)
Years of grade 4–6 in HVI				–0.040** (0.018)	–0.054** (0.021)	–0.031 (0.028)
N	5,195	2,625	2,570	5,195	2,625	2,570
R-squared	0.338	0.358	0.318	0.338	0.358	0.318

Panel C – Impact of entire conflict (1968–1992 sample)

HVI in primary school	−0.012 (0.020)	−0.074*** (0.027)	0.055* (0.030)			
HVI in grade 1–3				−0.006 (0.019)	−0.035 (0.023)	0.031 (0.029)
HVI in grade 4–6				−0.019 (0.019)	−0.075*** (0.027)	0.044 (0.029)
N	9,329	4,753	4,576	9,329	4,753	4,576
R-squared	0.241	0.231	0.266	0.241	0.231	0.266

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses clustered at the year of birth * district level. Regressions include year and district fixed effects, district time trend (defined as the interaction between the district category and birth year), and controls (whether the household head is a farmer and the household head’s level of education).

Source: Authors’ computations using [TLSS 2007a](#).

among boys attending the last three years of primary school. The overall effect on girls is positive (most likely driven by the 1999 effects), corresponding to an 8.5 percent increase in the likelihood of primary school completion.

Discussion

The results above indicate that violent conflict in Timor Leste had mixed effects on education. On average, the wave of violence in 1999 resulted in immediate hardships for the education of boys and girls. Girls, however, recovered from the negative consequences of the 1999 violence in the medium term. When the same cohort was observed in 2007, girls affected by the conflict had a higher and statistically significant positive chance of completing primary school than girls who were not exposed to the violence. We find no effect of the earlier peaks of violence on girls' primary school completion, but we find a positive and statistically significant effect (at 10 percent) of the entire conflict on girls' primary school completion.

In contrast, boys exposed to the wave of violence in 1999 had a much lower probability of having completed primary school by 2007 relative to boys unaffected by the violent events. Earlier peaks of violence as well as the entire conflict have similar negative effects on the education of boys in Timor Leste, particularly among boys attending the last grades of primary school. The effect of earlier peaks of violence is smaller than the impact of the 1999 violence, although we observe the persistence of significant negative educational effects of the earlier years of the conflict in the longer term. The difference in the magnitudes of the impacts of different peaks of violence may be because individuals affected by the earlier violence may have had the opportunity to complete primary school or may reflect the particularly violent nature of the 1999 events. In all cases, boys were rather severely affected by violence over the 25 years of the Timor Leste conflict.

The 1999 wave of violence in Timor Leste was brutal, but the recovery was rapid, as discussed in section I. Although problems remain, violence-affected areas have clearly benefited from this reconstruction effort. This finding is in line with results reported for other conflict-affected countries in [Bellows and Miguel \(2006\)](#) and elsewhere. Our results show, however, that only girls affected by the violence seem to have been able to recover (and even improve) their educational outcomes. The results for boys are persistently negative.²⁰

Given the discussion in section I, it is unlikely that this result is explained by supply-side factors, such as the destruction of schools or the absence of teachers. The postconflict reconstruction process had clear, positive impacts on the educational outcomes of girls exposed to violence, possibly because of a strong consideration of gender concerns in the UN interventions in Timor Leste

20. An uneven negative impact of violent conflict on boys' educational outcomes is also reported in [Akresh and de Walque \(2011\)](#) for Rwanda and [Verwimp and Van Bavel \(2011\)](#) for Burundi. For a review of the literature on the impact of violent conflict on education, see [Justino \(2012b\)](#).

(Olsson 2009). However, it is highly unlikely that these programs would have been biased against educating boys. A more likely explanation is that the negative impact of the conflict on boys' education in Timor Leste is related to the different roles that boys and girls play within the household.

As mentioned in section II, boys who were affected by the violence in 1999, on average, tended to work more and longer hours in 2001 (table 2). We estimated a reduced form regression of the incidence of conflict on child labor in the aftermath of the 1999 violence. The results (not shown) indicate a positive correlation between conflict exposure (displacement) and the probability of boys working: boys affected by displacement are 11 percentage points more likely to work than boys unaffected by violence. Affected girls, however, are 3.6 percentage points less likely to work.²¹ Other studies have shown that child labor is a key factor in explaining low school enrolment rates in Timor Leste, particularly among boys. For instance, as Pedersen and Arneberg (1999) report, "Poverty is the main reason why some 20 percent of children never get the chance to go to school. [. . .] Children, especially boys, work when their parents do not have jobs or their families are headed by single mothers" (p. 83). This argument is in line with findings in the literature regarding household coping strategies in the face of adverse shocks, which have widely documented the use of children as an economic security mechanism (see Dasgupta 1993; Duryea, Lam, and Levinson 2007; Nugent and Gillaspay 1983). In areas experiencing violent conflicts, households may decide to replace dead, injured, absent, or disabled adult workers with children (if they have not also become fighters). Rodriguez and Sanchez (2009) analyze the effect of war on child labor and find that violent attacks by armed groups in Columbian municipalities significantly increased the probability of school dropout and the presence of children, particularly boys, in the labor market.

The above results suggest that household economic needs in Timor Leste may also have resulted in boys dropping out of school, a mechanism that may, in turn, explain the negative impact of the conflict on boys' education. This mechanism is not conclusive, and it is possible that school dropout may have occurred if boys joined armed groups as fighters or occupied other supporting roles. Data to test this alternative hypothesis are unavailable, but there are some indications that children joined both the proindependence troops and paramilitary groups and militias. UNICEF (2001) states that "[b]oth the proindependence and prointegration forces in East Timor used children as armed combatants during period of the Indonesian occupation and its violent resolution after the 1999 referendum. On both sides of the conflict the age of child

21. These results were obtained from a probit regression estimated for a sample of children aged 10–14 years where the dependent variable was whether the individual worked in the week prior to the 2001 survey. Controls include household per capita expenditures, whether they speak Indonesian, grades completed by the mother and father, the occupation of the head of household, gender of the head of household, household size, and region of residence. These estimates are not shown in the paper because of space constraints but are available upon request from the authors.

soldiers ranged from 10 to 18 years old, although most were between the age of 15 and 18 years old” (p. 18). Given this age range, it is unlikely that our results are strongly driven by increases in the number of child soldiers. However, we cannot completely exclude this channel given the lack of sufficiently rigorous empirical evidence.

Taken together, the various pieces of evidence discussed above point to school dropout—most likely owing to economic necessity, but potentially for other reasons—as an important channel through which the conflict may have negatively affected educational outcomes among boys in Timor Leste. These effects may have considerable consequences for the country’s future economic and political stability given the accumulation of negative education shocks among boys over the 25-year conflict, which may have trapped a significant number of individuals in cycles of low human capital and low productivity. In particular, recent studies have reported that large numbers of young men who dropped out of school during the conflict in Timor Leste are currently members of gangs and martial arts groups in Dili, which are responsible for increases in insecurity and violence in Timor Leste (Kostner and Clark 2007; Muggah et al. 2010; Scambary 2009).

Robustness Checks

We performed several robustness checks to address some important issues that may affect the results discussed above. In addition to the various validity checks reported in previous sections, we separately address the possible exposure of the 2007 sample to the civil violence that erupted in Timor Leste in 2006 and potential biases in the 2001 and 2007 results due to nonrandom migration patterns. Supporting tables are presented in the appendix.

As mentioned in section I, in 2006, Timor Leste experienced substantial internal civil strife owing to fighting between different factions of the independence forces. The violence in 2006 resulted in 37 killings, 2,000 severely damaged houses, 3,000 completely destroyed houses, and 150,000 displaced people (Muggah et al. 2010; Scambary 2009). Most displaced people were located in the vicinity of Dili (where 65 internally displaced person camps were located) and were still displaced in 2007. Despite the decision to restart the 2007 survey once the violence had subsided (see footnote 3), it is possible that some of the results discussed in the section above are not due to exposure to the 1999 violence but due to exposure to the civil upheaval in 2006. To control for this potential exposure to the violence in 2006, we explore a variable in the 2007 dataset that captures whether an individual was absent from home in the past 12 months for security reasons (2.7 percent of the sample). Our calculations show that individuals who were absent from home for security reasons in 2006 all resided in Dili. Therefore, we believe that this dummy reliably captures the level of exposure to the 2006 violence. The results in table A.1 are nearly identical to those in table 6, indicating that our main conclusions are unlikely to be biased by the effects of the civil violence in 2006.

Another important concern is that some individuals migrated at some point in their lives.²² The 2001 and 2007 datasets provide information on their places of birth and their current places of residence. The data do not allow us to establish when this migration occurred or whether these individuals migrated for conflict-related reasons. Thus, the migration variable is potentially noisy and prone to misclassification error. The direction of the endogeneity bias is difficult to predict a priori.²³ If, for instance, individuals did not choose their new place of residence randomly (Kondylis 2010) and those who migrated went to areas in which economic conditions are typically better (for instance, urban areas), our results would likely be underestimated. Conversely, the effect of the violence would be overestimated if migrants relocated to places where they received inferior education. In addition, individuals who decided to migrate may differ from those who did not migrate. If this is the case and, for example, only wealthier and more educated households were able to migrate, then including these individuals in our estimates would underestimate the overall effect of the violence.

To assess whether the bias deriving from migration is a serious concern in our analysis, in tables A.2 and A.3, we present estimates from regressions that include a sample of individuals who never moved from their places of birth. These estimates test whether the results in section III hold when we restrict the sample to nonmigrants. The results are broadly comparable in terms of magnitude, signs, and significance to those obtained using the full sample. In addition, the proportions of individuals who migrated to a different place are 13 percent of the 2001 sample and 19 percent and 24 percent of the 2007 samples (the 1977–92 and 1968–84 cohorts, respectively). This finding suggests that even in the extreme scenario where the estimated effect was zero for migrants, the overall estimated effect would only be attenuated by approximately one-quarter of its value. We are therefore quite confident that our results are not biased as a result of migration choices.

IV. FINAL REMARKS

The aim of this paper was to examine the effects of the 25 years of conflict in Timor Leste on educational outcomes among boys and girls exposed to

22. The migration decision should be interpreted as distinct from the occurrence of displacement in 1999 in Timor Leste. Although it is relatively common in the conflict literature to treat displacement as a migration decision (see, for instance, Chamarbagwala and Morán [2010] for Guatemala, Kondylis [2010] for Bosnia-Herzegovina, and Ibañez and Moya [2010] for Colombia), this is not an appropriate means of addressing displacement in the case of Timor Leste because our displacement variable is based on the respondents' reported displacement experience rather than a migration outcome. We have also estimated the determinants of migration and found that displacement does not play a significant role in migration decisions.

23. It is important to note that issues regarding the potential endogeneity of the migration decision need to be considered as distinct in this context from potential endogeneity concerns regarding the displacement measure, which have been discussed in section II.

violence. We began by analyzing the impact of the wave of violence that occurred during the withdrawal of Indonesian troops in 1999. We first analyzed the short-term impact of the 1999 violence on primary school attendance in 2001 and its longer-term impact on school completion for the same cohorts of children observed again in 2007. We compared these latter results to the impacts of the peaks of violence in the 1970s and 1980s on schooling outcomes observed in 2007 (among those who were of primary school age at the time of the various violent events) and to the overall average educational impact of the conflict. This approach enabled us to compare the impact of a long-duration conflict on educational outcomes during the overall conflict and during peaks of violence.

In line with the existing literature on the effects of violent conflict on educational outcomes, we find that the conflict in Timor Leste led to considerable adverse impacts on educational outcomes, particularly among boys exposed to the violence. We find, however, that the impact of the conflict on girls' education, although negative in the short term in terms of school attendance, did not hinder their school attainment in the longer term because they were able to benefit from the rapid reconstruction of the education system in violence-affected areas. In contrast, the 25 years of violent conflict had a clearly negative impact on the education of boys in Timor Leste that persisted across generations. This result is consistent for different peaks of violence throughout the conflict in Timor Leste. Generations of young Timorese boys have experienced considerable reductions in their accumulation of human capital, which may now be reflected in increases in insecurity, unemployment, and violence in the country since 2006.

We have discussed evidence suggesting that the negative impact of violence on boys' education is due to boys dropping out of school. This is likely to be caused by household investment trade offs between education and economic survival, where boys would have been removed from school to participate in household economic activities. It is also possible that a small number of young boys may have dropped out of school to join armed groups.

These results have important policy implications. One implication is the importance of educational recovery in areas affected by violent conflict. The Timor Leste case suggests that early recovery may have positive results for the lives of children (girls, in this case). Another key implication is that reconstruction policies must pay greater attention to their redistributive impacts across genders and different population characteristics. Although girls recovered quickly from the conflict, boys did not, despite the large investment in the early recovery of the education system in Timor Leste. The evidence for Timor Leste suggests that boys were very vulnerable to both the direct effects of violence on education outcomes and indirect effects through household welfare mechanisms. This result implies that much more attention must be paid to understanding how children are affected by violent conflict and the different roles girls and boys assume during and after the conflict because these are likely to

perpetuate the risks associated with renewed conflict and persistent vulnerabilities across generations.

ACKNOWLEDGMENTS

This work was supported by the World Bank's Gender and Development Unit, Poverty Reduction and Economic Management Network, with generous funding from the Government of Norway. While writing the paper, Marinella Leone was co-funded by the project "Actors, Markets, and Institutions in Developing Countries: A Micro-Empirical Approach (AMID)," a Marie Curie Initial Training Network funded by the European Commission under its 7th Framework Programme (contract Number 214705 PITN-GA-2008-214705). Patricia Justino and Paola Salardi were co-funded by the Integrated Project on the Micro-Level Analysis of Violent Conflict – MICROCON, which was funded by the European Commission's 6th Framework Programme (project no. 028730).

We are very grateful to Toan Do and Monica Das Gupta at the World Bank, three anonymous referees, and the journal's editor for very useful comments and discussions. We would also like to thank Denis Cogneau, Marc Gurgand, Ana María Ibáñez, Sylvie Lambert, Edoardo Masset, Barry Reilly, Ricardo Santos, Olga Shemyakina, Philip Verwimp, participants in two World Bank workshops in Washington D.C. and Oslo, participants at the EUDN PhD workshop at Tinbergen Institute in Amsterdam, at the AEL Conference on Development Economics at DIW in Berlin and the CRED Workshop at FUNDP in Namur for valuable discussions and comments. All remaining errors are ours.

APPENDIX – ROBUSTNESS CHECKS

TABLE A1. Robustness Check: Impact of Conflict on Primary School Completion in 2007, Controlling for 2007 Civil Violence

	Impact of 1999 violence (1977–1992 sample)			Impact of early years of conflict (1968–1984 sample)		
	All	Boys	Girls	All	Boys	Girls
Years of prim. school in HVI	-0.042 (0.029)	-0.186*** (0.044)	0.105** (0.047)	-0.026** (0.011)	-0.032** (0.014)	-0.021 (0.017)
Absent home past 12 months	0.083*** (0.023)	0.066 (0.043)	0.104*** (0.032)	0.086*** (0.030)	0.095** (0.040)	0.079* (0.045)
N	6,676	3,383	3,293	5,195	2,625	2,570
R-squared	0.151	0.144	0.181	0.339	0.359	0.318

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses clustered at the year of birth * district level. Regressions include year and district fixed effects, district time trend (defined as the interaction between the district category and birth year), and controls (whether the household head is a farmer and the household head's level of education).

Source: Authors' computations using TLSS 2007a.

TABLE A2. Robustness Check: Impact of 1999 Violence on School Attendance in 2001, Nonmigrant Sample

	(1) All	(2) Boys	(3) Girls	(4) All	(5) Boys	(6) Girls	(7) All	(8) Boys	(9) Girls
D*T ₂	-0.200*** (0.052)	-0.213*** (0.063)	-0.188** (0.072)				-0.053 (0.058)	-0.071 (0.067)	-0.042 (0.076)
D*T ₃	-0.144*** (0.038)	-0.130*** (0.050)	-0.157*** (0.053)				-0.082 (0.053)	-0.150** (0.066)	-0.037 (0.074)
H*T ₂				-0.124*** (0.047)	-0.119* (0.061)	-0.129** (0.062)	-0.027 (0.057)	-0.031 (0.077)	-0.022 (0.067)
H*T ₃				0.005 (0.044)	0.003 (0.054)	0.007 (0.056)	0.077 (0.055)	0.043 (0.070)	0.118* (0.065)
D*H*T ₂							-0.263** (0.108)	-0.206 (0.129)	-0.326** (0.143)
D*H*T ₃							-0.167* (0.088)	0.002 (0.113)	-0.343*** (0.096)
N	2,553	1,383	1,170	2,553	1,383	1,170	2,553	1,383	1,170
R-squared	0.157	0.161	0.154	0.154	0.157	0.150	0.169	0.169	0.173

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports fixed effect estimates. Robust standard errors in parentheses clustered at the village level. Regressions include time effects (T₂ refers to the 1999/00 school year, and T₃ refers to the 2000/01 school year); D, H, and D*H are dummies, respectively, defined as 1 if the individual was displaced with the whole household, whether the house was completely damaged, and whether the individual was affected by both violent shocks.

Source: Authors' computations using TLSS 2001.

TABLE A3. Robustness Check: Impact of Conflict on Primary School Completion in 2007, Nonmigrant Sample

	Impact of 1999 violence (1977–1992 sample)			Impact of early years of conflict (1968–1984 sample)		
	All	Boys	Girls	All	Boys	Girls
Years of prim. school in HVI	-0.033 (0.036)	-0.166*** (0.054)	0.103* (0.057)	-0.015 (0.013)	-0.018 (0.016)	-0.014 (0.021)
N	5,446	2,803	2,643	3,963	2,041	1,922
R-squared	0.151	0.150	0.178	0.330	0.374	0.292

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses clustered at the year of birth * district level. Regressions include year and district fixed effects, district time trend (defined as the interaction between the district category and birth year), and controls (whether the household head is a farmer and the household head's level of education).

Source: Authors' computations using TLSS 2007a.

REFERENCES

- Akbulut-Yuksel, M. 2009. "Children of War: The Long-Run Effects of Large-Scale Physical Destruction and Warfare on Children." HiCN Working Paper 62. Households in Conflict Network. www.hicn.org.
- Akresh, R., and D. de Walque. 2011. "Armed Conflict and Schooling: Evidence from the 1994 Rwandan Genocide." HiCN Working Paper 47. Households in Conflict Network. www.hicn.org.
- Alderman, H., J. Hoddinott, and B. Kinsey. 2006. "Long Term Consequences of Early Childhood Malnutrition." *Oxford Economic Papers* 58 (3): 450–74.
- Alonso, A., and R. Brugha. 2006. "Rehabilitating the Health System after Conflict in East Timor: A Shift from NGO to Government Leadership." *Health Policy and Planning* 21 (3): 206–16.
- Bellows, J., and E. Miguel. 2006. "War and Institutions: New Evidence from Sierra Leone." *American Economic Review* 96 (2): 394–99.
- Blattman, C., and E. Miguel. 2010. "Civil War." *Journal of Economic Literature* 48 (1): 3–57.
- Case, A., and C. Paxson. 2008. "Stature and Status: Height, Ability, and Labor Market Outcomes." *Journal of Political Economy* 116 (3): 499–532.
- CAVR. 2005. "Chega!, The CAVR Report." The Commission for Reception, Truth and Reconciliation in Timor-Leste. <http://www.cavr-timorleste.org/en/cheгаReport.htm>.
- . 2006. "Human Rights Violations Database." Hrdag. Available at: http://www.hrdag.org/resources/timor-leste_data.shtml.
- Chamarbagwala, R., and H. E. Morán. 2010. "The Human Capital Consequences of Civil War: Evidence from Guatemala." *Journal of Development Economics*. 94 (1): 41–61.
- Dasgupta, P. 1993. *An Inquiry into Well-Being and Destitution*. Oxford: Clarendon Press.
- Davis, D. R., and D. E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review* 92 (5): 1269–89.
- Duryea, S., D. Lam, and D. Levinson. 2007. "Effects of Economic Shocks on Children's Employment and Schooling in Brazil." *Journal of Development Economics* 84: 118–214.
- Felgueiras, J., and J. Alves Martins. 2006. *Nossas Memórias de Vida em Timor*. Braga, Portugal: Editorial A.O.
- Glewwe, P., and H. G. Jacoby. 1994. "Student Achievement and Schooling Choice in Low-Income Countries." *Journal of Human Resource* 28 (3): 843–64.

- Gusmão, X. 2004. *A Construção da Nação Timorese: Desafios e Oportunidades*. Lisboa, Portugal: Editora Lidel.
- Holmes, J. 2003. "Measuring the Determinants of School Completion in Pakistan: Analysis of Censoring and Selection Bias." *Economics of Education Review* 22 (3): 249–64.
- Ibañez, A.-M., and A. Moya 2010. "Do Conflicts Create Poverty Traps? Asset Losses and Recovery for Displaced Households." In R. Di Tella, S. Edwards, and E. Schargrofsky, eds., *The Economics of Crime: Lessons for and from Latin America*, 137–74. Chicago, IL: University of Chicago Press.
- Ichino, A., and R. Winter-Ebmer 2004. "The Long-Run Educational Cost of World War II." *Journal of Labor Economics* 22 (1): 57–86.
- Justino, P. 2009. "Poverty and Violent Conflict: A Micro-Level Perspective on the Causes and Duration of Warfare." *Journal of Peace Research* 46 (3): 315–33.
- 2012a. "War and Poverty." In M. Garfinkel, and S. Skaperdas, eds., *The Oxford Handbook of the Economics of Peace and Conflict*. Oxford: Oxford University Press.
- 2012b. "Violent Conflict and Human Capital Accumulation." In G. Brown and A. Langer, eds. *Elgar Companion to Civil War and Fragile States*, Cheltenham, UK: Edward Elgar Publishing.
- Kalyvas, S. N. 2006. *The Logic of Violence in Civil War*. Cambridge, MA: Cambridge University Press.
- Kondylis, F. 2010. "Conflict Displacement and Labor Market Outcomes in Post-War Bosnia & Herzegovina." *Journal of Development Economics*. 93 (2): 235–48.
- Kostner, M., and S. Clark 2007. *Timor-Leste Youth in Crisis: Situational Analysis and Policy Options*. Washington, DC: The World Bank.
- Leôn, G. 2012. "Civil Conflict and Human Capital Accumulation: The Long-term Effects of Political Violence in Peru." *Journal of Human Resources* 47 (4): 991–1023.
- Maccini, S. L., and D. Yang 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review* 99 (3): 1006–26.
- Miguel, E., and G. Roland 2011. "The Long-Run Impact of Bombing Vietnam." *Journal of Development Economics* 96 (1): 1–15.
- Muggah, R., O. Jutersonke, R. Murray, E. Rees, and J. Scambary 2010. "Urban Violence in an Urban Village. A Case Study of Dili, Timor-Leste." Geneva Declaration Working Paper. Geneva, Switzerland: Geneva Declaration Secretariat.
- Nicolai, S. 2004. "Learning Independence: Education in Emergency and Transition in Timor-Leste since 1999." UNESCO International Institute for Educational Planning, Paris.
- Nugent, J. B., and T. Gillaspay 1983. "Old Age Pensions and Fertility in Rural Areas of Less Developed Countries: Some Evidence from Mexico." *Economic Development and Cultural Change* 31 (4): 809–29.
- Olsson, L. 2009. *Gender Equality and United Nations Peace Operations in Timor Leste*. Leiden, The Netherlands: Brill Publisher.
- Pedersen, J., and M. Arneberg 1999. "Social and Economic Conditions in East Timor." International Conflict Resolution Program, Columbia University, NY.
- Richter, K. 2009. "Changes in Subjective Well-being in Timor-Leste on the Path to Independence." *World Development* 37 (29): 371–84.
- Robinson, G. 2003. "East Timor 1999. Crimes against Humanity." Report Commissioned by the United Nations Office of the High Commissioner of Human Rights, OHCHR, Geneva, Switzerland.
- Rodriguez, C., and F. Sanchez 2009. "Armed Conflict Exposure, Human Capital Investments and Child Labour: Evidence from Colombia." HiCN Working Papers 68. Households in Conflict Network. www.hicn.org.
- Rohland, K., and S. Cliffe 2002. "The East Timor Reconstruction Program: Successes, Problems and Tradeoffs." World Bank Conflict Prevention and Reconstruction, Working Paper 2. Washington, DC.
- Scambary, J. 2009. "Anatomy of a Conflict: The 2006–2007 Communal Violence in East Timor." *Conflict, Security and Development* 9 (2): 265–288.

- Shemyakina, O. 2011. "The Effect of Armed Conflict on Accumulation of Schooling: Results from Tajikistan." *Journal of Development Economics* 95(2): 186–200.
- Silva, R., and P. Ball 2006. "Profile of Human Rights Violations in Timor-Leste, 1974-1999." Report by the Benetech Human Rights Data Analysis Group to the Commission on Reception, Truth and Reconciliation. <http://www.hrdag.org/timor>.
- TLSS. 2001. *Timor-Leste Household Survey 2001*. Dili, Timor-Leste: Direcção Nacional de Estatística, Ministério de Finanças.
- . 2007a. *Timor-Leste Household Survey 2007*. Dili, Timor-Leste: Direcção Nacional de Estatística, Ministério de Finanças.
- . 2007b. *Final Statistical Abstract: Timor-Leste Survey of Living Standards 2007*. Direcção Nacional de Estatística, Ministério de Finanças, Dili, Timor-Leste.
- UNDP. 2002. *East Timor Human Development Report 2002: The Way Ahead*. New York: UNDP.
- . 2006. *East Timor Human Development Report 2006: The Path Out of Poverty. Integrated Rural Development*. New York: UNDP.
- UNICEF. 2001. *East Timorese Children Involved in Armed Conflict: Case Studies Report, October 2000-February 2001*. Dili, East Timor: UNICEF.
- Verwimp, P., and J. Van Bavel 2011. "Schooling, Violent Conflict and Gender in Burundi." HiCN Working Paper 101. Households in Conflict Network. www.hicn.org.
- World Bank. 2003a. *Timor Leste Education: The Way Forward*. Washington, DC: The World Bank.
- . 2003b. *Timor Leste. Poverty in a New Nation: Analysis for Action*. Washington, DC: The World Bank.
- . 2011. *World Development Report 2011: Conflict, Security and Development*. Washington, DC: The World Bank.
- Zhao, M., and P. Glewwe 2010. "What Determines Basic School Attainment in Developing Countries? Evidence from Rural China." *Economics of Education Review* 29 (3): 451–54.

Education and Civil Conflict in Nepal

Christine Valente

Between 1996 and 2006, Nepal experienced violent civil conflict as a consequence of a Maoist insurgency, which many argue also brought about an increase in female empowerment. This paper exploits variations in exposure to conflict by birth cohort, survey date, and district to estimate the impact of the insurgency on education outcomes. Overall conflict intensity, measured by conflict casualties, is associated with an increase in female educational attainment, whereas abductions by Maoists, which often targeted school children, have the reverse effect. Male schooling tended to increase more rapidly in areas where the fighting was more intense, but the estimates are smaller in magnitude and more sensitive to specification than estimates for females. Similar results are obtained across different specifications, and robustness checks indicate that these findings are not due to selective migration. JEL codes: I20, J12, O12

Between 1996 and 2006, Nepal experienced violent civil conflict as a consequence of a Maoist insurgency. This paper investigates the impact of being exposed to this insurgency at a young age on education outcomes.

This study makes both an empirical and a methodological contribution to the growing literature on the impact of civil conflict on human capital formation. First, this paper extends our understanding of the impact of civil conflict on education to include a conflict of moderate intensity. With just over 13,000 casualties and less than 1 percent of the population forcibly displaced, the level of violence considered in this study was much lower than in conflict episodes considered in previous research.¹ Second, two alternative identification strategies

Christine Valente is a Lecturer (Assistant Professor) at the University of Bristol, United Kingdom; her email address is christine.valente@bristol.ac.uk. The author thanks INSEC for sharing their conflict data and Martha Ainsworth, Quy-Toan Do, Helge Holterman, Steve McIntosh, Gudrun Østby, Kati Schindler, Olga Shemyakina, Helen Simpson, Sarah Smith, Frank Windmeijer, Hassan Zaman, the World Bank's Nepal Country Director's office, and participants at two Gender and Conflict Research Workshops at the World Bank (Washington) and Peace Research Institute Oslo for their useful comments. The author also thanks three anonymous referees for their valuable comments and suggestions. This work was supported by the World Bank-Norway Trust Fund. The views expressed in this paper are those of the author alone and do not necessarily reflect those of the funding agencies. A supplemental appendix to this article is available at <http://wber.oxfordjournals.org/>.

1. The estimated number of individuals forcibly displaced in Nepal is approximately 200,000 (USAID, 2007). Studies reviewed in this paper have considered the impact of conflict on education in Tajikistan (between 50,000 and 100,000 deaths and 10 percent of the country's population internally displaced in the two worst years of conflict, Shemyakina, 2011b), Guatemala (200,000 deaths during the worst conflict period, Chamarbagwala and Morán, 2011), Rwanda (nearly 10 percent of the population killed, Akresh and de Walque, 2008), and Peru (just under 70,000 deaths, León 2012).

are employed to increase confidence in the reliability of the estimates: the first relies on variation in conflict exposure across birth cohorts and geographic areas in a single survey, as is standard in the literature (e.g., Akresh and de Walque 2008; Shemyakina 2011a; Chamarbagwala and Morán 2011), and the second relies on variation in exposure to conflict among school-aged individuals between household surveys and geographic areas.

Educational attainment is generally expected to be adversely affected by exposure to armed civil conflict. Direct youth enrollment in the military, limited mobility, and the destruction of schools may all negatively affect the ability of children to attend school. Increased poverty may drive parents to remove children from school (to avoid direct costs) and put them to work (to avoid opportunity costs). Political instability and reduced life expectancy may decrease expected returns to education and, in turn, reduce investments in human capital. Moreover, the schooling of girls is often more sensitive to worsening economic conditions than that of boys. A conflict environment may also hinder the functioning of education programs by weakening government institutions and imposing logistical and staff security challenges on local and international NGOs.

However, the general expectation that schooling is disrupted in conflict areas may not be well founded in the particular case of Nepal. National trends do not indicate an increase in poverty coinciding with the conflict but rather a steady decline in poverty (World Bank 2005). Despite difficulties with public service provision, basic health and education services have been maintained (Armon et al. 2004). In addition, the insurgency may have had a positive effect on schooling outcomes despite the fighting through both intended and unintended consequences of the Maoist presence. The insurgents have been reported to police teacher absenteeism (Hart 2001; see also Collins 2006 and Devkota and van Teijlingen 2010, for a similar argument regarding health workers) and have explicitly opposed caste- and ethnicity-biased traditions; these actions may have directly benefited both male and female education. In addition, the insurgents have publicly opposed gender inequality, including gender inequality in access to schooling. For instance, it has been reported that “the Maoists have taken a strong stand on this issue – insisting that girls of school age attend the local facilities, even to the point of holding parents accountable and liable to punishment for the non-attendance of their daughters” (Hart 2001, p. 32). Although the egalitarian rhetoric has not been followed completely in practice, a number of women were directly involved in combat, and there is anecdotal evidence of improved conditions for women in areas controlled by the Maoists, such as decreases in polygamy, domestic violence, and alcoholism (Lama-Tamang et al. 2003; Manchanda 2004; Geiser 2005; Aguirre and Pietropaoli 2008; Ariño 2008). The diffusion of the egalitarian Maoist ideology may also have increased the aspirations of young girls for their own education and the aspirations of parents for their daughters’ education. One unintended aspect of the insurgency may have contributed to improving schooling outcomes, especially for girls: female labor

force participation increased (Menon and Rodgers 2011), and there is evidence that when women have more control over household expenditure investments in children increase, especially for girls (e.g., Thomas 1990; Duflo 2003). Thus, in the case of Nepal, contrary to most episodes of violent conflict, the direction of the effect of the Maoist insurgency on schooling outcomes seems unclear a priori.

Nonetheless, one particular aspect of the Nepalese conflict is likely to have been unambiguously detrimental to education: the common insurgent practice of abducting civilians. Parents may have been deterred from sending their children to school out of fear that they would be abducted by the insurgents (Human Rights Watch 2004). Quoting figures from the Informal Sector Service Center (INSEC), UNESCO (2010) reports that between 2002 and 2006, the Maoists abducted 10,621 teachers and 21,998 students (p.8). According to additional data provided by INSEC, the total number of abductions by Maoist forces during the conflict amounted to more than 85,000. Although most abductees were seemingly returned unharmed after a few days of intensive indoctrination (Becker 2009; Macours 2011), a number of youths joined the Maoist fighters (in 2003, an estimated 30 percent of Maoist forces were aged 14–18 years). The indoctrination sessions held during abductions are likely to have played a part in their recruitment.

In this paper, I exploit differences in the intensity of violence experienced by individuals born at different times, surveyed at different times, and in different districts to shed light on the ways in which experiencing the insurgency at a young age affected educational outcomes.

Individual data from the 2001 and 2006 Demographic and Health Surveys (DHS) of Nepal are merged with detailed conflict data collected by INSEC, namely, the number of conflict fatalities, school destructions, and abductions by Maoists at the district level.

I find that overall conflict intensity, as measured by conflict casualties, was associated with an increase in female educational attainment, whereas abductions by Maoists, which often targeted school children, had the reverse effect. Male schooling also tended to increase more rapidly in areas where the fighting was more intense, but the estimates are smaller in magnitude and more sensitive to specification. Similar results are obtained across different identification strategies, and robustness checks indicate that these findings are not due to selective migration.

In the next section, I review the existing evidence on the impact of armed conflict on education outside Nepal with an emphasis on male-female differences. I then present the Nepalese conflict in section II, the data in section III, the empirical strategy in section IV, and the estimation results in section V. Section VI concludes.

I. LITERATURE REVIEW

A number of cross-country analyses suggest that political instability has large negative effects on growth but that recovery to equilibrium levels tends to be rapid (see [Blattman and Miguel 2010](#), for a review). At the microeconomic level, the results of an emerging body of literature on the impact of war-related destruction or civil conflict on educational attainment show that violent conflict often leads to worse educational outcomes, but estimates vary substantially by conflict, gender, and educational level. Overall, girls in postprimary education appear to experience the worst effects.

In war-torn Germany and Austria, school-aged individuals exposed to war received fewer years of education ([Ichino and Winter-Ebmer 2004](#); [Akbulut-Yuksel 2009](#)). In Guatemala, where the worst period of the Guatemalan civil war saw nearly 200,000 deaths, [Chamarbagwala and Morán \(2011\)](#) find that individuals who were of schooling age in departments that were more affected by the war completed fewer years of schooling and that this effect was much more marked for girls. In Bosnia and Herzegovina, [Swee \(2009\)](#) estimates that cohorts of children exposed to greater conflict intensity at the municipal level were less likely to complete secondary schooling, but primary schooling attainment was unaffected. [Shemyakina \(2011a\)](#) finds that girls (but not boys) who were of schooling age during the Tajik civil war were less likely to complete mandatory schooling in areas severely affected by conflict events. [Rodriguez and Sanchez \(2009\)](#) find that in Colombia, children aged 12 years and older who were exposed to more violence at the municipal level were more likely to drop out of school and enter the labor market. [León \(2012\)](#) finds that individuals who were born and raised in Peruvian districts that were more affected by conflict-related violence completed fewer years of education. Three recent papers, one by [Akresh and de Walque \(2008\)](#) and two by [Annan, Blattman, and colleagues \(Annan et al. 2009; Blattman and Annan 2010\)](#), illustrate the marked heterogeneity in findings on the impact of civil conflict across demographic groups and across conflicts. [Akresh and de Walque \(2008\)](#) estimate that cohorts of children exposed to the extremely violent Rwandan genocide, which killed 10 percent of the country's population, completed 18.3 percent fewer years of education. However, contrary to results from Guatemala, for example ([Chamarbagwala and Morán 2011](#)), they find that due to the nature of the conflict, nonpoor, male individuals were more negatively affected. Studying the effect of forced recruitment into the Ugandan Lord's Liberation Army, [Annan et al. \(2009\)](#) and [Blattman and Annan \(2010\)](#) find dramatically different effects for men and for women in the opposite direction of those obtained by, for example, [Shemyakina \(2011a\)](#) and [Chamarbagwala and Morán \(2011\)](#). The abducted men in their sample, who were abducted, on average, for just over 15 months, experienced much worse educational attainment and labor market outcomes as well as poorer psychological health ([Blattman and Annan 2010](#)). However, these authors find no such effects for female abductees, which they attribute to the lack of opportunities for women in general ([Annan et al. 2009](#)).

II. CONFLICT IN NEPAL

Nepal was an absolute monarchy until 1990. Despite multiparty democratic elections in 1991, a Maoist insurgency broke out in February 1996 in the Rolpa district and ended in 2006. The insurgency was initially concentrated in a few Communist strongholds in Western Nepal, but by the end of the war, conflict-related casualties were recorded in 73 of the 75 Nepalese districts. The Maoist presence varied from sporadic attacks to the organization of local governments and law courts. Over the course of the conflict, Maoists attacked government targets, such as army barracks, police posts, and local government buildings (Do and Iyer 2010). They were also reported to terrorize, loot, abduct, and physically assault civilians (Bohara et al. 2006). However, government security forces also killed civilians and were accused of using children for spying, torturing, displacing, and summarily convicting civilians (Bohara et al. 2006).

The principal objective of the insurgents was the creation of a constituent assembly to draft a new constitution. Other important stated aims were land redistribution and equality for all castes, language groups, and women.

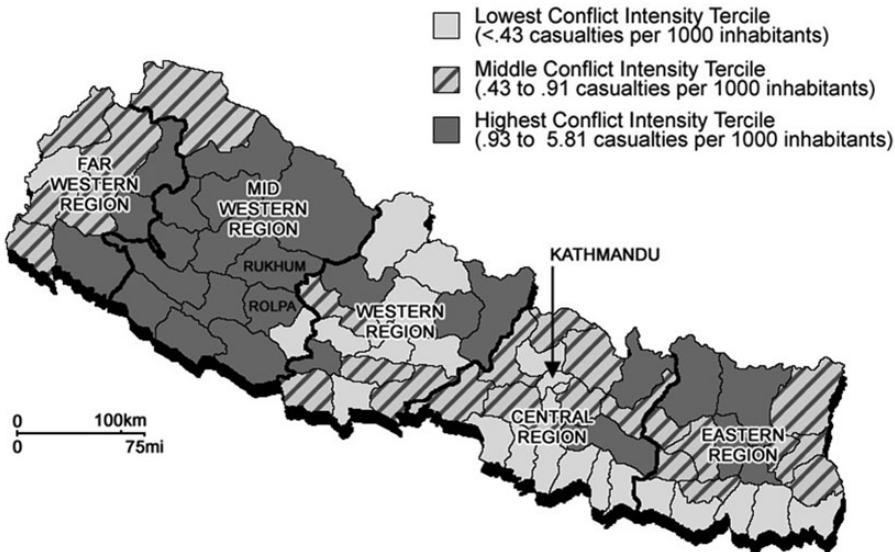
A crucial moment in the conflict was the Maoists' abandonment of a short-lived cease fire in November 2001. From that point, the government's response intensified dramatically, involving the Royal Nepal Army and leading to an escalation of violence (see figure S1 in the supplemental appendix, available at <http://wber.oxfordjournals.org/>). Building on opposition to King Gyanendra's authoritative reaction to the prolonged conflict, the Maoists joined forces with some of the country's major political parties, leading to the signing of a peace agreement in November 2006 and the creation of an interim government led by a power-sharing coalition that included the Maoists.

The intensity of conflict varied substantially across the districts of Nepal, as illustrated in figure 1, which depicts the distribution of districts between the three terciles of conflict deaths per 1,000 inhabitants. One specific characteristic of the Nepalese conflict that is likely to be particularly relevant for an analysis of educational outcomes is the insurgents' practice of abducting civilians, particularly school children, en masse for short periods of intensive indoctrination. As illustrated by figure S2, there is a positive correlation between the number of abductions by Maoists and the intensity of fighting as measured by conflict-related casualties, but the relationship is not systematic. Hence, it is possible to consider the effect of abductions over and above that of overall conflict intensity.² Districts with the highest proportion of abductees among the population are found in the middle tercile, which may be due to a lesser need for indoctrination in Maoist strongholds.³ Districts in the top quartile of the distribution of

2. The correlation coefficient between total conflict deaths and abductions per 1,000 inhabitants is 0.14.

3. Hutt (2005) also links abductions to the weakness of support for the Maoists: "The Maoists know that much of their support is hollow and based on fear. Maoist cadres have taken to mounting temporary abductions of large numbers of school teachers and students, who are taken to remote locations and subjected to political indoctrination sessions" (Hutt, 2005, p.86).

FIGURE 1. Conflict Intensity across Districts of Nepal



Notes: Author's calculations are based on casualties recorded in INSEC (2009) and district population figures from the 1991 population census (Central Bureau of Statistics, 2009). District terciles are defined by the distribution of total district casualties per 1,000 inhabitants.

abductions per 1,000 inhabitants that are not also in the top quartile of the distribution of casualties per 1,000 inhabitants tend to be found at the far western or eastern borders, close to districts characterized by intense fighting (i.e., numerous casualties).

Several arguments have been advanced to explain the district variation in the intensity of the insurgency, including geography (Murshed and Gates 2005; Bohara et al. 2006; Do and Iyer 2010), poverty (Murshed and Gates 2004; Do and Iyer 2010), a lack of political participation (Bohara et al. 2006), and inter-group inequality (Murshed and Gates 2005; Macours 2011). Determinants of district conflict intensity are therefore likely to be correlated with the explained variables of interest, which could give rise to omitted variable bias. As long as the omitted variables in question are constant over time, the inclusion of district fixed effects will suffice to remove any bias. If there are time-varying omitted variables correlated with both conflict intensity and the explained variables of interest, the inclusion of district fixed effects will not remove all potential biases, and additional steps must be taken to shed light on the causal impact of the insurgency. In section IV, I test for the presence of such time-varying omitted variables and discuss how I address potential threats to identification.

Despite the civil conflict, Nepal has experienced steady growth in real gross GDP (5 percent per year between 1995/96 and 2003/04), an additional increase

in disposable income due to substantial flows of remittances from abroad (representing 12.4 percent of the GDP), a steady decrease in poverty over the period (from 42 percent in 1995/96 to 31 percent in 2003/04), and an improvement in human development indicators, such as primary school enrollment (up from 57 percent to 73 percent) and child mortality, which decreased by 5 percent per year (World Bank 2005; Macours 2011).

However, the positive outlook for Nepal as a whole may mask unequal progress due to heterogeneous conflict intensity across districts. Indeed, national trends may hide a slower decrease in poverty, or even an increase in poverty, in more conflict-affected areas. In this paper, I exploit variation in the intensity of exposure to violent conflict by birth cohort, survey year, and district to investigate differential changes in primary educational attainment and completed years of education across districts that experienced varying degrees of violence.

III. DATA

DHS have been conducted in a number of developing countries as part of the Measure DHS project, a reputable USAID-funded project. The second and third DHS in Nepal were conducted in 2001 and 2006, respectively, and are nationally representative repeated cross-sections. The timing of these surveys is particularly useful because the surveys either preceded or followed the bulk of the fighting.⁴

For each DHS, a household survey collected the usual individual demographic and education data as well as household-level socioeconomic information. More detailed information was then collected from all women and a subset of men of reproductive age (if ever married, in the case of the 2001 survey). The data used in this analysis come from the household survey as well as migration information from the detailed interviews with women in the 2006 DHS.⁵

Summary statistics by district conflict intensity and specification subsample can be found in table S1.

The data used to measure conflict intensity are taken from electronic files provided by INSEC, an independent, well-regarded human rights NGO based in Kathmandu with reporters in each of the 75 Nepalese districts who monitor human rights violations. The INSEC data files contain the number of conflict-related deaths per month per district of Nepal between February 1996 and December 2006 as well as the total number of school destructions and abductions by Maoists at the district level, which are used to construct most measures of exposure to conflict used in this paper. Data from INSEC have been extensively used in the media, international agencies, and government reports and in a number of academic studies, including those by Bohara et al. (2006) and Do and Iyer (2010). However, conflict deaths and school destructions are easier

4. In 2001, six out of 257 sampling units had to be dropped from the sample for security reasons (Ministry of Health et al., 2002, p.6).

5. In the 2001 survey, children listed on the household roster cannot be matched to their mothers.

to monitor than abductions, and there are some surprising figures in the abduction data provided by INSEC, such as only 284 abductions by Maoists in Rolpa during the entire conflict. A degree of measurement error is likely to affect any conflict event data. If uncorrelated with the actual number of conflict events, this classical measurement error would lead to attenuation bias. However, the measurement error would have to be both inversely related to the true number of conflict events and very severe for it to lead to a reversal of the sign of the estimated effect of conflict. Such a result appears implausible given the degree of consensus on INSEC conflict data. Furthermore, findings using the number of casualties are consistent with the simple difference-in-difference calculations in table 1, in which the conflict event counts are collapsed into binary indicators. These are more blunt indicators of conflict intensity, but they are also less prone to measurement error.

In addition, this study uses two indicators of Maoist control over a given district by 2003 based on classifications reported by Hattlebak (2007). I consider, in turn, two alternative definitions of Maoist control. I first categorize as under Maoist control any district that is categorized as such by both Maoists and the government (Definition 1). I then apply what Hattlebak (2007) considers a more reliable classification, the government classification (Definition 2).

IV. EMPIRICAL STRATEGY

I exploit the fact that surveyed individuals have been exposed to varying degrees of conflict intensity according to their district of residence, year of birth, and whether they were surveyed in 2001 or in 2006.

The baseline estimation strategy is similar to that in much of the literature estimating the impact of civil conflict on individual outcomes reviewed in section I. The strategy exploits differences in exposure to conflict by birth year cohort and district of residence for individuals surveyed at the end of the conflict in 2006.

To check the robustness of the baseline results, I use a second identification strategy in which the source of identification is the change in the intensity of conflict within the district between 2001 and 2006, just before and just after the escalation of the conflict. By 2006, individuals born in 1991 and 1996, for example, would have been exposed to the same total amount of conflict before or during their schooling careers (albeit at different times in their lives). Hence, a comparison of the schooling outcomes of individuals born in 1991 and 1996 in the 2006 DHS would not be particularly informative. However, individuals born in 1991 who were observed at age 10 years in the 2001 survey experienced much less conflict by the time their education data were collected in 2001 than individuals born in 1996 and observed at age 10 years in the 2006 survey. Therefore, comparing their education outcomes at age 10 years is informative.⁶ This second

6. The variation in exposure between surveys also varies substantially between districts. For instance, in Mahottari, there were 0.14 additional deaths per 1,000 inhabitants between the two DHS, whereas in Jumla, there were close to three additional deaths per 1,000 inhabitants during the same period.

approach allows me to use variation in conflict intensity over time that would be discarded in the traditional approach. In addition, the second approach provides the opportunity for useful checks of the robustness of my findings to potential migration and mortality biases.

I consider two outcome variables: a binary indicator for primary schooling completion and the number of years of education completed. Less than 45 percent (20 percent) of the male (female) adult population surveyed in the 2006 Nepal DHS had completed primary education, so primary schooling completion is a relevant cutoff in the present context. Given the recent occurrence of the conflict, a focus on primary education also has an advantage in that many individuals whose primary schooling careers coincided with the conflict period are old enough to have completed their primary education; hence, their long-term primary schooling outcomes are observed. Finally, given the high prevalence of voluntary migration in Nepal, it is important to test the robustness of my findings to migration bias. I do this by comparing the schooling attainment of children under 15 years of age surveyed in 2001 and 2006 in a given district. This age group is appropriate as long as I focus on primary education. As of 2004, 97 percent of Nepali migrants were men aged 15–44 years who typically left their wives and children behind (Lokshin and Glinskaya 2009). By focusing on children under 15 at the time of the survey, the individual is thus both unlikely to have migrated himself and unlikely to have accompanied a migrant parent. The DHS did not collect detailed migration data, but it does provide data on the date of arrival at the current location for women of reproductive age. Before the age of 15 years, the overwhelming majority of children are still living with their mothers; hence, I can further test the robustness of my findings for migration bias by restricting the sample to children whose mothers had not migrated since the beginning of the conflict.

Specification 1: Exploiting Differences in Exposure to Conflict between Birth Year Cohorts within Districts

Similar to previous studies on the impact of conflict on educational attainment, I first use data from the postconflict DHS (2006) and exploit variations in exposure to conflict by birth year cohort and district. In its simplest form, the estimating equation can be written as follows:

$$EDUC_{ij}^t = \delta_j + \alpha_t + \beta TOTCONF_j^t + \varepsilon_{ij}^t \quad (1)$$

where $EDUC_{ij}^t$ is, in turn, a dummy variable equal to one if individual i in district j born in year t has completed primary education and zero otherwise or the number of years of education completed by this individual; δ_j represents district fixed effects; α_t represents birth year dummies; and $TOTCONF_j^t$ is the interaction between a dummy equal to one when the individual belongs to the treated cohort and the number of conflict casualties (per 1,000 inhabitants) in district j

during 1996–2006. In the baseline regressions, I define the treated cohort as those aged 5 to 9 years at the beginning of the conflict in 1996, whereas the control cohort includes individuals aged 16 to 19 years at the beginning of the conflict. This choice is discussed in the preliminary analysis at the end of this section.

Under the assumption that there is no correlation between the number of district casualties and unobserved factors varying with district and birth cohort, β is the causal effect of a one-unit increase in $TOTCONF_j^t$ on the primary completion rate or on the number of years of education completed by exposed cohorts. A one-unit increase in $TOTCONF_j^t$ roughly corresponds to one standard deviation in the district-level distribution of casualties (0.98). Another way of appraising the magnitude of β is to consider a one-unit increase in $TOTCONF_j^t$ as a move from the district with the least conflict to the 53rd district (out of 75) in order of conflict intensity or from the 53rd district to the 69th district in order of conflict intensity—a very large increase in conflict intensity.

There are five “developmental regions” in Nepal, which are relatively homogeneous in terms of their level of development (see figure 1). In equation (1), I implicitly restrict birth cohort effects (α_t) to be identical across development regions. To reduce the potential for unobserved cohort-district varying factors to bias the estimate of the effect of conflict exposure, in the main set of results, I report estimates of equation (1R) in which the birth year intercepts are allowed to vary by development region (α_t^R).⁷ Here, the effect of conflict is identified by using the difference in exposure to conflict by district and birth year cohort, net of birth year trends common to all districts within a given development region:

$$EDUC_{ij}^t = \delta_j + \alpha_t^R + \beta TOTCONF_j^t + \varepsilon_{ij}^t. \quad (1R)$$

In addition, I estimate variants of equation (1) in which I add regressors capturing specific aspects of the conflict that are likely to have affected schooling outcomes, namely, the number of school destructions and abductions by Maoist forces (per 1,000 inhabitants) during the conflict. I also consider whether Maoist control over the district had an effect on primary attainment by estimating variants of equation (1) in which I replace $TOTCONF_j^t$ with an indicator variable that switches on when the district was under Maoist control at the height of the conflict.

It would be desirable to control for the socioeconomic status of the household in which an individual was raised. However, household characteristics are not included as regressors in this specification because these are only observed at the time of the survey in 2006, 10 years after the beginning of exposure to conflict, and may therefore be caused by the conflict. In addition, at the time of the survey, individuals in the control group were 25–28 years old, and members of

7. Note that region dummies are subsumed under the district fixed effects, but the interactions between regions and birth year dummies are not.

the treated group were 14–18 years old. Therefore, it is difficult to imagine household characteristics that would not depend on the individual's education level at the time of the survey. In the second identification strategy described below, I observe school-aged children in their household; hence, I can control for household characteristics.

Specification 2: Comparison of Outcomes in 2001 and 2006 for a Given Age at Interview

In equation (1), I only use data from the postconflict DHS (2006) and exploit variations in exposure to conflict according to birth year cohort and district. However, there is a comparable survey for 2001, just before the conflict escalated, which allows me to estimate the impact of the conflict using an alternative identification strategy based on variation in conflict exposure by survey date and district. The idea is to exploit the fact that a child aged, for example, 10 years in 2001 will have experienced much less conflict during his lifetime than another child aged 10 years in 2006 in the same district, and the difference in conflict exposure between these two children will also vary across districts. Finding results similar to those obtained using the traditional identification strategy in equation (1) would bolster confidence in the reliability of my estimates. More specifically, I estimate the following:

$$\begin{aligned}
 EDUC_{ij}^s &= \theta_j + \lambda_s + \gamma CONFEXP_{ij}^s + X_{ij}^s \varphi + \mu_{ij}^s \\
 s &= 2001, 2006 \\
 CONFEXP_{ij}^s &= \sum_{y=birth_{ij}}^s CONF_{jy}
 \end{aligned}
 \tag{2}$$

where $EDUC_{ij}^s$ is the primary education completion dummy or the number of completed years of education of individual i in district j observed in survey year s , θ_j represents district fixed effects, λ_s is a survey dummy equal to one for DHS 2006 and zero for DHS 2001, $CONF_{jy}$ is the number of conflict deaths per 1,000 inhabitants that occurred in district j in year y , and $CONFEXP_{ij}^s$ is the number of conflict deaths per 1,000 inhabitants in district j that occurred between the individual's birth year and survey year s in which the individual is interviewed, which I calculate from yearly district death counts. When $EDUC_{ij}^s$ is the primary education indicator, the sample comprises children aged 10–18 years, who may have completed primary education at the time of the survey. When $EDUC_{ij}^s$ is the number of years of education completed, the sample comprises children aged 5–14 years, who are of primary school age at the time of the survey. Under the assumption that there is no correlation between the cumulative number of district casualties between 1996 and year s and unobserved district-survey-varying factors, γ is the causal effect of a one-unit increase in $CONFEXP_{ij}^s$ on the rate of primary schooling completion (the number of years of education completed) by

the 10- to 18-year-old (5- to 14-year old) group. The magnitude of γ can be directly compared to that of β because both $TOTCONF_j^t$ and $CONFEXP_{ij}^s$ are expressed in district casualties per 1,000 inhabitants.

X_{ij}^s is a set of controls that contains age at interview dummies and their interaction with the survey dummy in all specifications, thus allowing the educational attainment of each age group to vary independently between the two surveys.⁸ These covariates are included to control for potential differences in the 2001–2006 change in the district-level age composition of the 10- to 18-year-old or 5- to 14-year-old group that, if correlated with conflict intensity, could bias the results. In some variants, X_{ij}^s also includes household characteristics at the time of the survey (rural location and education of the household head). These characteristics are only measured at the time of the survey and could potentially be caused by past conflict. However, finding similar results when these observable household characteristics are included would suggest that potential changes in the composition of households due to mortality or migration do not drive my findings.

To further reduce the potential for unobserved time-varying factors to bias the estimate of the effect of conflict exposure, in the main set of results, I report estimates of equation (2R) in which the coefficients of the survey dummy, the age at survey intercepts, and their interactions are allowed to vary by development region. Here, the effect of conflict exposure is identified using the within-district change in conflict exposure at age x between 2001 and 2006, net of 2001–2006 changes in educational attainment at age x common to all districts in a given development region:

$$EDUC_{ij}^s = \theta_j + \lambda_s^R + \gamma CONFEXP_{ij}^s + X_{ij}^{sR'} \varphi + \mu_{ij}^s. \quad (2R)$$

This second identification strategy has three important advantages compared to the traditional approach based on variation in exposure by birth cohort and district for individuals observed at a single point in time. First, it has the advantage of comparing cohorts that are born only five years apart but have experienced very different degrees of conflict (i.e., a 10-year-old in 2001 in district j was exposed to much less conflict than a 10-year-old in 2006 in district j but was born only five years earlier), which reduces concerns regarding potential confounders, including differential migration patterns. Second, when the sample is restricted to children aged 14 years and under, the concern regarding selection bias due to voluntary migration decreases because most migrants are men aged 15–44 years, who typically leave their wives and children behind (Lokshin and Glinskaya 2009). Third, I can further test the robustness of my findings to

8. When estimating equation (2) without any of the controls included, γ is positive and statistically significant for both females and males in the completed years of education regression and positive and statistically significant for females only in the primary education completion regression (full results are available upon request).

migration bias by excluding from the sample children surveyed in 2006 whose mothers moved to their current location after 1996.⁹

All specifications are estimated using linear district fixed-effects panel data models. All models allow for error terms to be correlated in an arbitrary fashion within a district to avoid overrejection of the null hypothesis of zero treatment effect due to serial correlation, following [Bertrand et al. \(2004\)](#).

Preliminary Analysis

An inspection of the data shows that although the legally mandated age for beginning schooling is six years old and there are five years of primary schooling, a sizeable proportion of children are enrolled in primary school before age six (70.1 percent at five years) and until age 14 (16.1 percent) in the 2006 DHS, with numbers subsequently decreasing sharply (7.9 percent at 15 years and 2.85 percent at 16 years).¹⁰ Therefore, an analysis of the long-term effect of conflict on primary schooling completion should consider children aged at least 14 years at the time of the survey, and control cohorts should have been at least 15 years at the beginning of the conflict (and preferably slightly older). When estimating variants of equation (1), I therefore define the treated cohort as comprising individuals aged five to nine years at the beginning of the conflict in 1996, such that all treated cohorts are exposed to the conflict during most of their potential primary schooling careers and the youngest exposed cohort is observed at age 14 in the 2006 DHS. In the main regressions, the control cohorts include individuals aged 16 to 19 years at the beginning of the conflict—that is, individuals who were born too early to have their primary education affected by the conflict but are as close, and therefore as comparable, to the treated cohorts as possible. I also provide a robustness check in which the control group comprises individuals aged 18–25 years at the start of the conflict. In the baseline specification, I exclude cohorts aged 10 to 15 years in 1996 because the treatment status of these cohorts is less clear. Many of these individuals could have been enrolled in primary schooling during the conflict, but they were not exposed to conflict during most of their primary schooling careers (see figures S3a and S3b).

Panel A of table 1 illustrates the basic identification strategy used in the baseline specification. This panel shows the difference in the increase in primary schooling between cohorts exposed (row (1)) and not exposed (row (2)) to conflict during their primary schooling careers in districts experiencing above-median conflict casualties (columns (1) and (4)) compared to districts experiencing below-median conflict casualties (columns (2) and (5)). Women born too early to be affected by the conflict during their primary schooling years have a

9. The education data used in this paper come from the DHS household datasets. Information on date of arrival at the present location was only collected in individual interviews with women aged 15–49 years. The same exclusion could not be implemented for the 2001 DHS because individuals listed on the household roster cannot be matched to their mothers.

10. In the 2001 DHS, 44.4 percent of five-year-olds, 24.6 percent of 14-year-olds, and 13.6 percent of 15-year-olds were enrolled in primary schooling.

TABLE 1. Preliminary Difference-in-Difference Calculations, Completion of Primary Schooling

		Female			Male		
		Primary Education Rate by					
		Number of Casualties in District			Number of Casualties in District		
		(1)	(2)	(3)	(4)	(5)	(6)
		High	Low	Difference	High	Low	Difference
Panel A: Binary DiD Experiment							
(1)	Age 5 to 9 in 1996	0.67	0.64	0.03	0.80	0.79	0.01
(2)	Age 16 to 19 in 1996	0.27	0.43	-0.16	0.63	0.70	-0.07
	Difference	0.40	0.21	0.19 (0.045)***	0.17	0.09	0.08 (0.038)*
Panel B1: Placebo DiD Experiment 1							
(1)	Age 16 to 19 in 1996	0.27	0.43	-0.16	0.63	0.70	-0.07
(2)	Age 26 to 29 in 1996	0.13	0.20	-0.07	0.50	0.53	-0.03
	Difference	0.14	0.23	-0.09 (0.045)**	0.13	0.17	-0.04 (0.047)
Panel B2: Placebo DiD Experiment 2							
(1)	Age 16 to 19 in 1996	0.27	0.43	-0.16	0.63	0.70	-0.07
(2)	Age 20 to 23 in 1996	0.15	0.32	-0.17	0.58	0.62	-0.04
	Difference	0.12	0.11	0.01 (0.040)	0.05	0.08	-0.03 (0.047)

(Continued)

TABLE 1. Continued

		Female			Male		
		Primary Education Rate by					
		Number of Abductions in District			Number of Abductions in District		
		(1)	(2)	(3)	(4)	(5)	(6)
		High	Low	Difference	High	Low	Difference
Panel C: Binary DiD Experiment							
(1)	Age 5 to 9 in 1996	0.70	0.62	0.08	0.82	0.77	0.05
(2)	Age 16 to 19 in 1996	0.33	0.37	-0.04	0.61	0.70	-0.09
	Difference	0.37	0.25	0.12 (0.053)**	0.21	0.07	0.14 (.038)***
Panel D1: Placebo DiD Experiment 1							
(1)	Age 16 to 19 in 1996	0.33	0.37	-0.04	0.61	0.70	-0.09
(2)	Age 26 to 29 in 1996	0.15	0.19	-0.04	0.53	0.52	0.01
	Difference	0.18	0.18	0.00 (0.051)	0.08	0.18	-0.10 (0.049)**
Panel D2: Placebo DiD Experiment 2							
(1)	Age 16 to 19 in 1996	0.33	0.37	-0.04	0.61	0.70	-0.09
(2)	Age 20 to 23 in 1996	0.22	0.27	-0.05	0.62	0.59	0.03
	Difference	0.11	0.10	0.01 (0.041)	-0.01	0.11	-0.12 (0.045)***

Notes: District casualties are expressed per 1,000 inhabitants. “High” and “Low” refer to above-median or below-median district totals per 1,000 inhabitants. Standard errors clustered at the district level are in parentheses. All first differences (i.e., row (1) – row (2) for a given conflict category) are statistically significant, except for values marked with . DiD indicates difference-in-difference. *p < 0.10, **p < 0.05, ***p < 0.01.

Source: INSEC 2009, Central Bureau of Statistics 2009. Education data are based on Nepal DHS 2006.

much lower rate of primary schooling in high-conflict districts compared to women in low-conflict districts. In contrast, primary schooling completion is slightly higher in high-conflict districts compared to low-conflict districts among the cohort of women who were entering primary school around the beginning of the conflict, resulting in an additional increase in female primary schooling of 19 percentage points between the two cohorts in high-conflict areas compared to low-conflict areas. A qualitatively similar but less dramatic effect is observed among men.

To shed light on the direction of the potential biases due to differential preconflict trends, I conduct several control or “placebo” experiments in which conflict exposure is artificially assigned to cohorts who were too old to be affected by the conflict. In panel B1, I compare the change in primary schooling attainment between cohorts aged 16–19 years and cohorts aged 26–29 years at the beginning of the conflict in above-median versus below-median conflict intensity districts. The difference-in-difference is negative, and for females, it is statistically significant. This finding indicates that rates of primary schooling completion were improving more slowly in areas where more conflict occurred in 1996–2006 when comparing cohorts that were potentially enrolled in primary school (i.e., aged 5–14 years) in the 1972–1984 period and cohorts that were potentially enrolled in the 1982–1994 period. If this trend had continued during the conflict period, the estimates presented in this paper would be a lower bound of the true effect of conflict; that is, the positive coefficient of the conflict variable would be an overly conservative estimate, especially for females.

The ideal placebo experiment would be based on the actual cohorts involved in the experiment of interest in the absence of conflict. Such a test is clearly not feasible. However, it is possible to conduct an additional placebo experiment based on cohorts born immediately before the period relevant to the experiment of interest to check for differences in trends as close as possible to the period of interest. The results of this additional placebo experiment comparing cohorts aged 16–19 years at the beginning of the conflict with those aged 20–23 years are shown in panel B2. During this period immediately preceding the conflict, I cannot reject the hypothesis that the evolution of primary schooling was parallel in districts with above- and below-median conflict casualties for both males and females.

In panel C, the experiment is conducted by replacing the below- and above-median casualty categories with below- and above-median Maoist abduction categories. The results in panel C show that primary schooling has progressed more rapidly in districts with above-median Maoist abductions. However, two-thirds of districts classified as high (low) conflict based on the median number of casualties are also classified as high (low) conflict based on the median number of abductions. Therefore, these simple two-by-two calculations may capture the effect of overall conflict intensity, the effect of Maoist abductions, or both. In the regressions that follow, I disentangle the effect of overall conflict intensity and Maoist abductions by including both conflict variables.

Panels D1 and D2 show results for tests that replicate the placebo tests for panels B1 and B2, where the definition of high- and low-conflict districts based on above- and below-median conflict casualties is replaced with the definition based on above- and below-median district abductions. Females experienced similar preconflict primary education trends in high- and low-abduction districts. Male cohorts found in high-abduction districts experienced slower progress in preconflict primary education relative to low-abduction districts. If the same trends continued for cohorts considered in the experiment of interest, then the effect of exposure to Maoist abductions would tend to be biased downward for males (i.e., to be more negative) but not for females. On the contrary, in the regression analysis, I find that after controlling for district casualties, the education of females suffered from Maoist abductions, but that the education of males did not. Therefore, the difference in male trends observed in panels D1 and D2 does not drive my conclusions.

V. RESULTS

The preliminary analysis in section IV suggested that primary schooling completion rates tended to increase more rapidly during armed conflict in areas that experienced a high intensity of conflict, especially for girls. In tables 2 to 4, I present estimates of the impact of exposure to conflict on educational outcomes to determine whether this striking conclusion of the preliminary analysis is confirmed when using more detailed information on the intensity of conflict, controlling for unobserved heterogeneity between individual districts and between regions over time and using different identification strategies.

Table 2 reports findings on the impact of conflict exposure on primary schooling completion. The first two columns present estimates of the long-term effect of conflict intensity on primary schooling completion using the baseline specification (equation (1R)). The last four columns indicate the robustness of these findings through comparison of the change in primary completion rates for the 10- to 18-year-old group for districts with varying degrees of conflict intensification between the 2001 and 2006 DHS (equation (2R)). The last two columns include controls for rural location and the educational attainment of the household head.

The results in the first column indicate that areas with more fighting witnessed a larger increase in female primary education attainment. Casting this result in terms of the distribution of conflict violence, an increase in violence of one standard deviation of the district-level distribution of casualties during the conflict (0.98 casualties per 1,000 inhabitants) increases female primary schooling attainment by 5.6 percentage points. This is roughly the effect of a move from the 5th to the 75th percentile of the district-level conflict distribution of total casualties. The sign and order of magnitude of this effect is confirmed when comparing cohorts born only five years apart but exposed to very different levels of conflict using equation (2R). Across all specifications in table 2, conflict exposure does

TABLE 2. Impact of Conflict Intensity Measured by Casualties on Primary Schooling Completion

Explained Variable and Sample	(1)	(2)	(3)	(4)	(5)	(6)
	Primary Education - Female	Primary Education - Male	Primary Education - Female 10–18	Primary Education - Male 10–18	Primary Education - Female 10–18	Primary Education - Male 10–18
Specification	Eq. (1R)	Eq. (1R)	Eq. (2R)	Eq. (2R)	Eq. (2R)	Eq. (2R)
= 1 if 5–9 in 1996 × District casualties during 1996–2006 ($TOTCONF_{ij}^t$)	0.0555** (0.0272)	0.0241 (0.0251)				
District casualties before survey ($CONFEXP_{ij}^s$)			0.0764* (0.0431)	0.0036 (0.0368)	0.0811** (0.0368)	0.0094 (0.0338)
= 1 if rural					-0.1094*** (0.0205)	-0.0622*** (0.0154)
= 1 if head has primary education					0.0543*** (0.0133)	0.0684*** (0.0115)
= 1 if head has secondary education					0.2127*** (0.0120)	0.2046*** (0.0139)
= 1 if head has higher education					0.3249*** (0.0251)	0.2597*** (0.0253)
Panel variable	District	District	District	District	District	District
<u>Included dummies:</u>						
Year of birth	Yes	Yes	No	No	No	No
Region × Year of birth	Yes	Yes	No	No	No	No
DHS 2006	No	No	Yes	Yes	Yes	Yes

(Continued)

TABLE 2. Continued

Explained Variable and Sample	(1)	(2)	(3)	(4)	(5)	(6)
	Primary Education - Female	Primary Education - Male	Primary Education - Female 10–18	Primary Education - Male 10–18	Primary Education - Female 10–18	Primary Education - Male 10–18
Specification	Eq. (1R)	Eq. (1R)	Eq. (2R)	Eq. (2R)	Eq. (2R)	Eq. (2R)
Age at interview	No	No	Yes	Yes	Yes	Yes
Region × DHS 2006	No	No	Yes	Yes	Yes	Yes
DHS 2006 × Age at interview	No	No	Yes	Yes	Yes	Yes
Region × Age at interview	No	No	Yes	Yes	Yes	Yes
DHS 2006 × Region × Age at interview	No	No	Yes	Yes	Yes	Yes
Observations	3,823	3,055	9,595	9,267	9,584	9,255
No. of clusters	75	75	69 ^a	69 ^a	69 ^a	69 ^a
R-squared	0.1106	0.0368	0.2077	0.3021	0.2602	0.3372
p value male vs. female ^b	0.345		0.074		0.044	

Notes: All specifications are estimated using the panel fixed-effects estimator and include a constant. District casualties are expressed per 1,000 inhabitants. Columns (1) and (2): Sample only includes individuals surveyed in the Nepal DHS 2006 and aged 5–9 years (treatment group) or 16–19 years (control group) at the beginning of the conflict in 1996. Columns (3) to (6): Sample only includes individuals surveyed in Nepal DHS 2001 or 2006 aged 10 to 18 years at the time of the survey. Standard errors clustered at the district level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

^aDHS data collection was somewhat affected by the conflict in 2001. Hence, contrary to DHS 2006, four districts were not covered: Dolpa, Jajarkot, Rolpa, and Rukhum. The small districts of Manang and Mustang were not surveyed, but these districts did not experience any casualties during the conflict.

^bp value of an F test of equality between the reported treatment effects for males and females.

Source: INSEC 2009, Central Bureau of Statistics 2009, Nepal DHS 2001, and Nepal DHS 2006.

TABLE 3. Impact of Alternative Conflict Variables on Primary Schooling Completion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	School Destructions Female	School Destructions Male	Maoist Abductions Female	Maoist Abductions Male	Maoist Control Female – Definition 1	Maoist Control Male – Definition 1	Maoist Control Female – Definition 2	Maoist Control Male – Definition 2
Explained Variable	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education
Specification	Eq. (1R)	Eq. (1 R)	Eq. (1 R)	Eq. (1R)	Eq. (1R)	Eq. (1R)	Eq. (1R)	Eq. (1R)
= 1 if 5–9 in 1996 × District casualties during 1996–2006 (TOTCONF _i)	0.0539* (0.0302)	0.0301 (0.0291)	0.0638** (0.0275)	0.0231 (0.0259)				
= 1 if 5–9 in 1996 × District schools destroyed 2002–2006	0.4176 (3.1298)	–1.7848 (2.3395)						
= 1 if 5–9 in 1996 × Maoist Abductions during 1996–2006			–0.0022*** (0.0005)	0.0002 (0.0010)				
= 1 if 5–9 in 1996 × District controlled by Maoists (Definition 1)					0.0916 (0.0590)	0.2009*** (0.0529)		

(Continued)

TABLE 3. Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	School Destructions Female	School Destructions Male	Maoist Abductions Female	Maoist Abductions Male	Maoist Control Female – Definition 1	Maoist Control Male – Definition 1	Maoist Control Female – Definition 2	Maoist Control Male – Definition 2
Explained Variable	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education	Primary Education
Specification	Eq. (1R)	Eq. (1 R)	Eq. (1 R)	Eq. (1R)	Eq. (1R)	Eq. (1R)	Eq. (1R)	Eq. (1R)
= 1 if 5–9 in 1996 × District controlled by Maoists (Definition 2)							0.1456*** (0.0457)	0.0923* (0.0506)
Panel variable	District	District	District	District	District	District	District	District
<u>Included dummies:</u>								
Year of birth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region × Year of birth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,823	3,055	3,823	3,055	3,823	3,055	3,823	3,055
No. of clusters	75	75	75	75	75	75	75	75
R-squared	0.1106	0.0370	0.1118	0.0368	0.1105	0.0415	0.1130	0.0382
p value male vs. female ^a		0.466		0.0134		0.208		0.456

Notes: All specifications are estimated using the panel fixed-effects estimator and include a constant. School destructions and abductions by Maoists are expressed per 1,000 inhabitants. Sample only includes individuals surveyed in the Nepal DHS 2006 and aged 5–9 years (treatment group) or 16–19 years (control group) at the beginning of the conflict in 1996. Definition of a district controlled by Maoists based on matches between People’s Army and government classifications as of 2003 (Definition 1) or government classification (Definition 2), according to Hattlebak (2007). Standard errors clustered at the district level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

^ap value of an F test of equality between the reported treatment effects for males and females.

Source: INSEC 2009, Central Bureau of Statistics 2009, Nepal DHS 2006,

not appear to significantly affect male primary schooling completion, although the gender difference in the conflict effect is only statistically significant in columns (3) to (6).

These results are robust to including controls for household characteristics, suggesting that the results are not driven by a change in household composition due to, for example, selective mortality or migration (columns (5) and (6)). Table S2 presents specifications similar to those in table 2 but replaces the indicator for primary schooling completion with years of education completed. Similar results are obtained, indicating that an increase in violence of one standard deviation increases female educational attainment by 0.6 years.

Table S3 presents three different specifications to further check the robustness of the baseline results in the first two columns of table 2 to the following changes in specification: restricting birth year fixed effects to be identical across the five development regions of Nepal, changing the control cohort, and replacing the number of casualties with its natural logarithm. The results in table S3 confirm that primary education progressed more rapidly during the conflict in districts experiencing more casualties and that this effect was more robust across specifications for females.

Next, I investigate whether specific aspects of the conflict had different effects on primary schooling completion (table 3). First, I use INSEC data on the total number of school destructions per district to test whether these destructions had a negative effect on primary schooling completion despite the overall positive impact of the insurgency (columns (1) and (2)). For both genders, I find a statistically insignificant effect, which is likely because a district-level analysis lacks the power to identify the effect of school destructions. School destructions were a rare and isolated aspect of the conflict¹¹ that could be expected to have had a large effect on schooling at a disaggregated level but not at the district level. Second, I use INSEC data on the total number of abductions by Maoists per district to test whether a larger number of abductions, often targeting school children, had an adverse effect on schooling. The results in columns (3) and (4) indicate that abductions had a negative effect on female primary schooling. An increase in the number of abductions (per 1,000 inhabitants) by one standard deviation of the district-level distribution (16.82) decreases female primary schooling attainment by 3.7 percentage points. In other words, the effect of a move from the 5th to the 75th percentile of the district-level distribution of total abductions yields a 1.6 percentage point decline in female primary completion. Third, I test whether primary schooling completion improved more in districts controlled by Maoists where the insurgents were better able to affect schooling provision according to their ideology (columns (5) to (8)). There is no clear-cut definition of insurgent control, with discrepancies between the classifications used by the People's Army and the government (Hatlebakk 2007). Therefore, I use two alternative classifications. The choice of definition affects the magnitude

11. According to the data provided by INSEC, 76 schools were destroyed.

TABLE 4. Impact of Conflict Intensity on Completed Years of Education, 5- to 14-year-olds

Explained Variable and Sample	(1)	(2)	(3)	(4)
	Years of Education Female 5–14	Years of Education Male 5–14	Years of Education Female 5–14 – Mother here since 1996	Years of Education Male 5–14 – Mother here since 1996
Specification	Eq. (2R)	Eq. (2R)	Eq. (2R)	Eq. (2R)
District casualties before survey ($CONFEXP_{it}^s$)	0.2795** (0.1147)	0.1224 (0.1207)	0.2633** (0.1172)	0.0755 (0.1391)
= 1 if rural	–0.3916*** (0.0788)	–0.1559** (0.0664)	–0.4703*** (0.0914)	–0.1770** (0.0703)
= 1 if head has primary education	0.2210*** (0.0428)	0.2301*** (0.0361)	0.2051*** (0.0453)	0.2381*** (0.0374)
= 1 if head has secondary education	0.8263*** (0.0431)	0.7064*** (0.0386)	0.8619*** (0.0509)	0.7242*** (0.0388)
= 1 if head has higher education	1.1291*** (0.1035)	1.0828*** (0.0846)	1.3829*** (0.1137)	1.2016*** (0.1124)
Panel variable	District	District	District	District
<u>Included dummies:</u>				
DHS 2006	Yes	Yes	Yes	Yes
Age at interview	Yes	Yes	Yes	Yes
Region × DHS 2006	Yes	Yes	Yes	Yes
DHS 2006 × Age at interview	Yes	Yes	Yes	Yes

Region × Age at interview	Yes	Yes	Yes	Yes
DHS 2006 × Region × Age at interview	Yes	Yes	Yes	Yes
Observations	11,793	12,116	9,772	9,959
No. of clusters	69 ^a	69 ^a	69 ^a	69 ^a
R-Squared	0.5062	0.6077	0.4909	0.5996
p value male vs. female ^b		0.079		0.065

Notes: All specifications are estimated using the panel fixed-effects estimator and include a constant. District casualties are expressed per 1,000 inhabitants. Sample only includes individuals surveyed in the Nepal DHS 2001 and 2006 and aged 5–14 years at the time of the survey. In columns (3) and (4), the 2006 sample is restricted to individuals whose mothers were interviewed individually and whose mothers reported having lived in their current place of residence as of 1996. The same exclusion could not be implemented for the 2001 DHS because individuals listed on the household roster cannot be matched to their mothers. Standard errors clustered at the district level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

^aDHS data collection was somewhat affected by the conflict in 2001. Hence, contrary to DHS 2006, four districts were not covered: Dolpa, Jajarkot, Rolpa, and Rukhum. The small districts of Manang and Mustang were not surveyed, but these districts did not experience any casualties during the conflict.

^b p value of an F test of equality between the reported treatment effects for males and females.

Source: INSEC 2009, Central Bureau of Statistics 2009, Nepal DHS 2001, and Nepal DHS 2006.

and significance of estimates, but the overall message is that primary schooling has tended to become more prevalent over time for both genders in areas controlled by the Maoists.

Tables S4 and S5 replicate the analysis in table 3 with birth year fixed effects restricted to be identical across development regions of Nepal (tables S4 and S5) and the control cohort replaced with individuals aged 18–24 years at the beginning of the conflict (table S5). The same conclusions apply as those drawn from the set of preferred results in table 3.

In table 4, I turn to the estimated effect of an increase in conflict intensity between 2001 and 2006 on completed years of education of children of primary schooling age (5–14 years) at the time they were surveyed (as per equation (2R)).

Column (1) of table 4 indicates that an increase in casualties since birth by one standard deviation increases the completed years of education by just over one-quarter of a year for girls aged 5 to 14 years in 2006 compared to girls from the same district who were the same age when surveyed in 2001, before the conflict escalated. For boys, the coefficient of interest is less than half the magnitude of that for girls and is significantly different from the estimated conflict effect for girls (at the 10 percent significance level). The estimates are very similar when restricting the 2006 sample to children whose mothers had not moved since 1996 (columns (3) and (4)), which confirms that changes in composition due to migration patterns are not driving these findings. In table S6, I repeat the analysis in table 4 but restrict the age intercepts and survey year dummy to be identical across Nepal's development regions. The results for the female sample are almost identical, but estimates for the male sample are now nearly as large as those for the female sample and are statistically significant. Echoing the findings for primary schooling completion, overall, these results confirm that primary education progressed more rapidly during the conflict in districts that experienced more casualties, and this effect is more robust across specifications for females.

In conclusion, the results presented in this section provide no support for the hypothesis that the Nepalese civil conflict had a negative effect on schooling overall. There is a robust positive effect of the intensity of the insurgency on female educational attainment, but there is less of an effect for male educational attainment. There is also evidence of a decrease in female primary schooling completion where insurgents were more prone to abductions, holding overall conflict intensity constant.

VI. CONCLUDING REMARKS

Despite experiencing a substantial civil conflict between 1996 and 2006, Nepal has surprisingly enjoyed one of the best periods in its history in terms of economic growth and poverty reduction. At present, however, little is known about whether this period of development at the aggregate level hides disparities at a more disaggregated level due to the wide variation in conflict intensity across the country.

In this paper, I exploit variation in exposure to conflict by birth cohort, survey date, and district to estimate the impact of conflict intensity on schooling outcomes.

I find no support for the hypothesis that civil-conflict-related violence, as measured by the number of conflict casualties, had a negative effect on the quantity of schooling attained by children of either gender. On the contrary, there is robust evidence that female primary schooling attainment increased in districts that experienced more conflict deaths relative to districts with fewer conflict deaths. This result holds irrespective of whether one compares (within a given district) (i) the completion of primary education for cohorts exposed and not exposed to the conflict and observed at the end of the conflict or (ii) years of education completed by a given age for cohorts observed before (2001) and after (2006) a sharp escalation of the conflict and that were therefore exposed to very different degrees of conflict. It is also robust to a number of changes in specifications. In particular, robustness checks indicate that changes in household composition due to conflict-induced migration patterns do not drive this finding.

However, one aspect of the Nepalese civil conflict that is particularly relevant to schooling outcomes had adverse consequences on female primary schooling: the widespread insurgent practice of abducting civilians, many of whom were school children.

The findings reported in this paper echo the positive changes observed for Nepal as a whole during the conflict period in terms of economic growth, education, and child health. The present analysis shows that the progress in education observed at the country level does not hide a slower increase in districts where more fighting occurred, but the insurgent practice of abducting civilians adversely affected female educational outcomes. The estimates presented in this paper are in line with the existing qualitative literature on the Nepalese civil conflict, which consistently reports mixed conclusions with respect to the impact of the conflict on education and female empowerment (e.g., [Hart 2001](#); [Lama-Tamang 2003](#); [Manchanda 2004](#); [Pettigrew and Shneiderman 2004](#); [Geiser 2005](#); [Aguirre and Pietropaoli 2008](#); [Ariño 2008](#); [Falch 2010](#)).

Education, particularly female educational attainment, appears to have benefited from the societal changes induced directly or indirectly by the insurgency more than it was adversely affected by the loss of income and other disruptions caused by the conflict. Data limitations prevent a more detailed analysis of the channels through which the conflict affected education beyond the distinction between the effect of conflict as a whole and that of abductions. However, potential mechanisms suggested by the existing anthropological and peace studies literature include Maoist efforts to remove barriers to schooling for all children from the lower castes and to reduce teacher absenteeism (e.g., [Hart 2001](#); [Lama-Tamang 2003](#)), which could have benefited both male and female education; the Maoist influence in encouraging or coercing parents to send girls to school ([Hart 2001](#)); and the Maoists' effect on female empowerment. Although the exact figure is contested, a substantial share of the guerillas in the Maoist

ranks was female. Many more females were involved in the Maoist movement without direct participation in combat, such as by disseminating propaganda (Lama-Tamang et al. 2003; Pettigrew and Shneiderman 2004), and even larger numbers may have been influenced by the Maoist discourse on gender equality. In addition, there is anecdotal evidence of an improvement in the condition of women in areas controlled by the Maoists, such as decreases in polygamy, domestic violence, and alcoholism, as well as greater support for women to divorce their husbands (Lama-Tamang et al. 2003; Manchanda 2004; Geiser 2005; Ariño 2008). Although the insurgents' rhetoric was often in contrast with their actual practice (Pettigrew and Shneiderman 2004), the presence of females in their ranks and the propaganda promoting female autonomy may have increased female bargaining power within the household as well as female aspirations. According to Hart (2001), "girls and women are strongly encouraged to gain an education and to participate in society generally and in activities connected to the 'People's War' in particular. This directly challenges their traditional role and apparently stimulates girls to consider leading lives beyond marriage and the home (Hart 2001, p.35)". Furthermore, an unintended consequence of the conflict has been that women have adopted roles typically reserved for men. Women's involvement in the labor market increased as a consequence (Menon and Rodgers 2011). The rise in female labor market participation may have increased returns to female schooling and motivated girls to obtain more education and parents to invest more in their daughters' education. Increased female earnings are also likely to improve the ability of mothers to influence the way household resources are spent. Moreover, there is evidence that when women have more control over household expenditures, such as because their own earnings make up a larger share of the household's income, investments in children increase; this is especially the case for girls (e.g., Thomas 1990; Duflo 2003), although this may not be the case in all contexts (Quisumbing and Maluccio 2003; Gitter and Barham 2008). Finally, the nature of the occupations of women outside the home also changed. In many areas, women were reported to take on leadership roles in local institutions, including schools (Pettigrew and Shneiderman 2004). This improvement in female representation in local institutions may have contributed to increased education, especially for girls.

Data limitations prevent rigorous tests of the role played by these different potential channels in explaining the finding that education, particularly female education, increased more in areas where the fighting was more intense.¹² Future research aiming to disentangle the role of each of the channels through which the insurgency may have improved educational outcomes would be valuable.

An issue beyond the scope of this paper is the important question of the effect of civil conflict on the quality of education, which is potentially large (for a review, see Shemyakina and Valente 2011). Data limitations have thus far

12. See the appendix for some insights based on self-reported measures of female empowerment available in the DHS.

precluded quantitative research on the impact of conflict on the quality of schooling, but there is growing evidence that cognitive skills, rather than completed years of education, matter for individual earnings and economic growth (e.g., Hanushek and Woessmann 2008). Therefore, even where the number of years of education completed is not adversely affected by civil conflict, such conflict may have deleterious effects on human capital if the quality of learning deteriorates.

From an international perspective, this paper contributes to unpacking the complexity that lies behind the generic term civil conflict. The idiosyncrasies of each conflict highlight the need for additional research on the impacts of different conflicts to shed light on the range of potential effects rather than a focus on extreme, but thankfully rare, instances.

From a policy perspective, the present findings call for measures that aim to protect school children and teachers from being directly targeted by combatants. As shown in this paper, even where primary education systems appear very resilient to surrounding violence, direct targeting of schools, however mild (e.g., brief abductions of pupils and teachers for indoctrination purposes), has adverse effects on schooling, especially for girls.

REFERENCES

- Aguirre, D., and I. Pietropaoli. 2008. "Gender Equality, Development and Transitional Justice: The Case of Nepal." *The International Journal of Transitional Justice* 2: 356–77.
- Akbulut-Yuksel, M. 2009. "Children of War: The Long-run Effects of Large-scale Physical Destruction and Warfare on Children." IZA Working Paper No.4407. Institute for the Study of Labor, Bonn, Germany.
- Akresh, R., and D. de Walque. 2008. "Armed Conflict and Schooling: Evidence from the 1994 Rwandan Genocide." World Bank Policy Research Paper No.4606. World Bank, Washington, DC.
- Annan, J., C. Blattman, D. Mazurana, and K. Carlson. 2009. "Women and Girls at War: 'Wives', Mothers, and Fighters in the Lord's Resistance Army." HICN Working paper 63. Households in Conflict Network, Brighton, UK.
- Ariño, M. V. 2008. "Nepal: A Gender View of the Armed Conflict and Peace Process." Quaderns de Construcció de Pau No.4. Escola de Cultura de Pau, Barcelona, Spain.
- Armon, J., C. Berry, D. Duncan, R. Calder, S. Clapham, and M. Harvey. 2004. "Service Delivery in Difficult Environments: The Case of Nepal." Department for International Development, Asia Policy Division.
- Becker, J. 2009. "Child Recruitment in Burma, Sri Lanka, and Nepal." In S. Gates, and S. Reich eds., *Child Soldiers in the Age of Fractured States*, 108–20. Pittsburg, PA: University of Pittsburg Press.
- Bertrand, M., E. Duflo, and S. Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119(1): 249–75.
- Blattman, C., and J. Annan. 2010. "The Consequences of Child Soldiering." *Review of Economics and Statistics* 92 (4): 882–98.
- Blattman, C., and E. Miguel. 2010. "Civil War." *Journal of Economic Literature* 48 (1): 3–57.
- Bohara, A., N. Mitchell, and M. Nepal. 2006. "Opportunity, Democracy, and the Exchange of Political Violence: A Subnational Analysis of Conflict in Nepal." *Journal of Conflict Resolution* 50 (1): 108–28.

- Central Bureau of Statistics [Nepal]. 2009. *Nepal Statistical Year book 2009*. Central Bureau of Statistics, Kathmandu. http://cbs.gov.np/?page_id=1233, last accessed May 22, 2013.
- Chamarbagwala, R., and H. E. Morán. 2011. "The Human Capital Consequences of Civil War: Evidence from Guatemala." *Journal of Development Economics* 94: 41–61.
- Collins, S. 2006. "Assessing the Health Implications of Nepal's Ceasefire." *Lancet* 368: 907–8.
- Devkota, B., and E. R. van Teijlingen. 2010. "Understand Effects of Armed Conflict on Health Outcomes: The case of Nepal." *Conflict and Health* 4: 20.
- Do, Q.-T., and L. Iyer. 2010. "Geography, Poverty and Conflict in Nepal." *Journal of Peace Research* 47 (6): 735–48.
- Duflo, E. 2003. "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa." *World Bank Economic Review* 17 (1): 1–25.
- Falch, A. 2010. "Women's Political Participation and Influence in Post-Conflict Burundi and Nepal." Peace Research Institute Oslo Working Paper. Oslo, Norway.
- Geiser, A. 2005. "Social Exclusion and Conflict Transformation in Nepal: Women, Dalit and Ethnic Groups." Working Paper 5/2005. Swisspeace, Bern, Switzerland.
- Gitter, S. R., and B. L. Barham 2008. "Women's Power, Conditional Cash Transfers, and Schooling in Nicaragua." *World Bank Economic Review* 22 (2): 271–90.
- Hanushek, E. A., and L. Woessmann 2008. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature* 46: 607–68.
- Hart, J. 2001. "Conflict in Nepal and its Impact on Children." A Discussion Document Prepared for UNICEF Regional Office South Asia, Refugee Studies Centre, Oxford.
- Hatlebakk, M. 2007. "LSMS Data Quality in Maoist Influenced Areas of Nepal." No 6, CMI Working Papers. CMI (Chr. Michelsen Institute), Bergen, Norway.
- Human Rights Watch. 2004. "Between a Rock and a Hard Place: Civilians Struggle to Survive in Nepal's Civil War." <http://www.hrw.org/en/reports/2004/10/06/between-rock-and-hard-place?print>, last accessed May 20, 2013.
- Hutt, M. 2004. "Nepal and Bhutan in 2004: Two Kings, Two Futures." *Asian Survey* 45 (1): 83–87.
- Ichino, A., and R. Winter-Ebmer 2004. "The Long-Run Educational Cost of World War II." *Journal of Labor Economics* 22 (1): 57–86.
- INSEC. 2009. *Electronic data files*. Informal Sector Service, Kathmandu.
- Lama-Tamang, M. S., S. M. Gurung, D. Swarnakar, and S. R. Magar 2003. "Social Change in Conflict Areas: Assessment Report." Prepared for UK Department for International Development (DFID) Nepal, Kathmandu.
- León, G. 2012. "Civil Conflict and Human Capital Accumulation: The Long-Term Effects of Political Violence in Perú." *Journal of Human Resources* 47 (4): 992–1022.
- Lokshin, M. M., and E. Glinskaya 2009. "The Effect of Male Migration on Employment Patterns of Women in Nepal." *World Bank Economic Review* 23 (3): 481–507.
- Macours, K. 2011. "Increasing Inequality and Civil Conflict in Nepal." *Oxford Economic Papers* 63 (1): 1–26.
- Manchanda, R. 2004. "Maoist Insurgency in Nepal. Radicalizing Gendered Narratives." *Cultural Dynamics* 162 (3): 237–58.
- Menon, N., and Y. Rodgers 2011. "War and Women's Work: Evidence from the Conflict in Nepal." World Bank Policy Research Working Paper No. 5745. World Bank, Washington, DC.
- Ministry of Health [Nepal], New ERA, and ORC Macro. 2002. *Nepal Demographic and Health Survey 2001*. Calverton, Maryland: Family Health Division, Ministry of Health; New ERA; and ORC Macro.
- Murshed, S. M., and S. Gates 2005. "Spatial-Horizontal Inequality and the Maoist Insurgency in Nepal." *Review of Development Economics* 9(1): 121–34.
- Pettigrew, J., and S. Shneiderman 2004. "Women and the Maobaadi: Ideology and Agency in Nepal's Maoist Movement." *Himal South Asia* 17 (1): 19–29.

- Quisumbing, A. R., and J. A. Maluccio 2003. "Resources at Marriage and Intrahousehold Allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa." *Oxford Bulletin of Economics and Statistics* 65: 283–327.
- Rodríguez, C., and F. Sánchez 2009. "Armed Conflict Exposure, Human Capital Investments and Child Labor: Evidence from Colombia." HiCN Working Paper No. 68. Households in Conflict Network, Brighton, UK.
- Shemyakina, O. 2011a. "The Effect of Armed Conflict on Accumulation of Schooling: Results from Tajikistan." *Journal of Development Economics* 95 (2): 186–200.
- 2011b. "Labor Market, Education and Armed Conflict in Tajikistan." World Bank Policy Research Working Paper No. 5738. World Bank, Washington, DC.
- Shemyakina, O., and C. Valente 2011. "Education and Violent Conflict in South Asia: Challenges and Solutions." Background paper prepared for the World Bank Regional Study on "Improving the Quality of Learning in South Asia." Mimeograph, University of Bristol, UK.
- Swee, E. L. 2009. "On War and Schooling Attainment: The Case of Bosnia and Herzegovina." HiCN Working Paper No. 57. Households in Conflict Network, Brighton, UK.
- Thomas, D. 1990. "Intra-Household Resource Allocation: An Inferential Approach." *Journal of Human Resources* 25 (4): 635–64.
- UNESCO. 2010. *Education under Attack*. Paris, France: United Nations Educational, Scientific and Cultural Organization.
- USAID. 2007. Nepal – Humanitarian Assistance Fact Sheet No. 1, June. http://pdf.usaid.gov/pdf_docs/PNADJ147.pdf, last accessed May 20, 2013.
- Wallenstein, P., and M. Sollenberg 2001. "Armed Conflict 1989–2000." *Journal of Peace Research* 38 (5): 629–44.
- World Bank. 2005. "Nepal - Resilience amidst Conflict: An Assessment of Poverty in Nepal, 1995–96 and 2003–04." Report No. 34834-NP. World Bank, Washington, DC.

Schooling, Violent Conflict, and Gender in Burundi

Philip Verwimp and Jan Van Bavel

We investigate the effect of exposure to violent conflict on human capital accumulation in Burundi. We combine a nationwide household survey with secondary sources on the location and timing of the conflict. Only 20 percent of the birth cohorts studied (1971–1986) completed primary education. Depending on the specification, we find that the probability of completing primary schooling for a boy exposed to violent conflict declined by 7 to 17 percentage points compared to a nonexposed boy, with a decline of 11 percentage points in our preferred specification. We also find that exposure to violent conflict reduces the gender gap in schooling, but only for girls from nonpoor households. Forced displacement is one of the channels through which conflict affects schooling. Our results are robust to various specifications and estimation methods. JEL codes: 012, I21, J16

During the past 30 years, civil conflict has affected almost three-fourths of all countries in sub-Saharan Africa (Gleditsch et al. 2002). Economists have studied the causes of war and their role in reducing growth and development (Collier and Hoeffler 1998; Miguel, Satyanath, and Sergenti 2004; Guidolin and La Ferrara 2007). The long-term economic consequences are particularly debated in the literature. Authors who find rapid economic recovery after war include Davis and Weinstein (2002) for Japan, Brakman, Garretsen, and Schramm (2004) for Germany, Miguel and Roland (2006) for Vietnam, and Bellows and Miguel (2009) for Sierra Leone. Convergence toward the country's long-term growth

Philip Verwimp (corresponding author) is associate professor in development economics and holder of the Marie and Alain Philippon Chair in Sustainable Human Development at the European Centre for Advanced Research in Economics and Statistics and the Solvay Brussels School of Economics and Statistics, Université Libre de Bruxelles, philip.verwimp@ulb.ac.be. Jan Van Bavel is a research professor in demography at the Centre for Sociological Research, Faculty of Social Sciences, University of Leuven, jan.vanbavel@soc.kuleuven.be. The research for this article was financed by the Norwegian Trust Fund to the World Bank. The authors thank seminar discussants and participants in the World Bank workshops on Gender and Conflict in Washington DC (June 2010) as well as in Oslo (December 2010), particularly Mayra Buvinic, Monica Dasgupta, Damien De Walque, Michael Lokshin Christine Valente, and Quentin Wodon. We thank Judson Caskey for his help with the implementation of the CGM command in STATA. We have also benefited from the comments by three anonymous reviewers as well as by the editor of the journal. Errors are the responsibility of the authors only. A supplementary appendix to this article is available online at <http://wber.oxfordjournals.org/>.

THE WORLD BANK ECONOMIC REVIEW, VOL. 28, NO. 2, pp. 384–411
Advance Access Publication May 14, 2013

doi:10.1093/wber/lht010

© The Author 2013. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

path is reached relatively rapidly, often within 15 years, as predicted by a neo-classical growth model.

The relatively rapid recovery of economic growth and other macro-level indicators do not tell us much about the distribution of long-term consequences at the micro level. This paper considers the consequences of civil war for human capital accumulation at the individual level. Gender differences are a critical source of heterogeneity in this respect; however, the direction of the gender effect is an empirical question. If, for example, conflicting parties engage child soldiers, it is likely that boys will be more affected than girls. Existing gender inequalities may be exacerbated during violent conflict; however, these inequalities may also be attenuated. For example, if a country needs the brains and labor of young women to work in the military industry during a dispute with a neighbor, the labor market position of women may benefit from the conflict. There is no universal rule to predict what the gender-specific impact will be. The gender-specific impact may be exacerbated in one domain (e.g., sexual violence), or the conflict may offer new opportunities (e.g., in paid labor or business). The direction of the effects as well as their magnitude differ from country to country and context to context depending on preexisting gender inequalities, the type of conflict, the duration of the conflict, and the institutional particularities of the war-affected country.

This paper focuses on the effect of civil war on schooling in Burundi. We attempt to determine the direction and magnitude of this effect in terms of schooling for both boys and girls. If schooling is negatively affected, this may, in turn, affect subsequent choices and opportunities for both men and women, including access to paid labor, age at marriage, number of children, and the socioeconomic characteristics of spouses. The level of schooling attained as a child and young adult is a fundamental driver of welfare throughout one's life.

We work with the *Enquête Démographique et de Santé* collected by UNFPA in 2002. This survey contains detailed information on each member of the interviewed households, including all births and deaths, schooling, wealth, and the history of migration during the civil war. We combine these surveys with event data on the location and timing of the conflict. The empirical identification strategy exploits variation in the onset and duration of the conflict across Burundi's provinces and related variation to determine which cohorts of children were exposed to the massacres and the civil war during their school-aged years.

We find that the completion of primary schooling in Burundi is affected by the massacres and the subsequent civil war. For every year that a school-aged boy was exposed to conflict in his province of residence, his probability of completing primary schooling decreased by 6 percentage points in our preferred specification. Boys from both poor and nonpoor households suffer from war. Girls suffer a general schooling disadvantage in Burundi; however, we find that violent conflict reduces the gender gap, although only for girls from nonpoor households. We show that forced displacement is one of the channels through which violent conflict affects schooling.

I. REVIEW OF THE LITERATURE ON SCHOOLING, GENDER, AND CONFLICT

There is a body of research analyzing how households cope with economic or agricultural shocks, such as crop failures, famines, or droughts (Fafchamps, Udry, and Czukas 1998; Dercon 2004), but there is little work on the microeconomic consequences of violent or nonviolent political shocks. Although few households have formal insurance against economic shocks, many have informal insurance mechanisms that they can use, such as self-insurance (portfolio spread, income diversification, temporary migration), village-level solidarity mechanisms, or even outside insurance against weather calamities (Dercon 2004). Such insurance mechanisms appear not to be available for political risks; at least, the scholarly community has largely failed to study potential coping mechanisms for political shocks. One of the findings of the coping literature in development economics is that nonpoor households are better able to cope with negative economic shocks compared to poor households. Using assets, savings, or their social capital, nonpoor households are more successful at cushioning the negative impact of weather, disease, or price shocks. The nascent literature on political shocks suggests that this poor versus nonpoor divide in terms of coping is nonexistent or much smaller than in the case of economic shocks. In the event of antiurban, Marxist, or cultural revolution-type conflicts, the nonpoor, educated part of the population may be hit harder than the poor uneducated part, with completely different effects on their welfare in comparison to economic or agricultural shocks.

Shemyakina's (2006) empirical work on violent conflict in Tajikistan finds that girls suffer a greater loss in education compared to boys. She attributes this finding to concerns over safety and low returns on girls' education. In contrast, Akresh and de Walque (2008) find that male Rwandan children in nonpoor households incur the strongest effect. Evans and Miguel (2004) find that young children in rural Kenya are more likely to drop out of school after their parents' death and that this effect is particularly strong for children who lose their mothers. Although Kenya was not the scene of violent conflict during the observed period, the finding is relevant because violent conflict produces many orphans, which may have a similar effect on children's schooling.

Combining a household panel with detailed data on allied bombings of German cities during WWII, Akbulut-Yuksel (2009) finds significant, long-lasting, detrimental effects of bombing on the human capital and labor market outcomes of individuals who were school aged during WWII. These individuals had 0.4 fewer years of schooling, on average, in adulthood in comparison to individuals not affected by the bombings. On average, affected children experienced a reduction of 6 percent in labor market earnings in relation to unaffected children.

Alderman, Hoddinott, and Kinsey (2006) find that Zimbabwean children affected by the civil war in the 1970s completed fewer grades of schooling or

started school later than those not affected by the shocks. Similar results are found by León (2011) for Peru, by Angrist and Kugler (2008) and Rodriguez and Sanchez (2009) for Colombia, by Chamarbagwala and Morán (2010) for Guatemala, and by de Walque (2006) for Cambodia.

The reasons that education during the war may be affected negatively include school closure, migration and displacement, the quality and availability of school facilities, and shocks to income and security (Justino 2011). Chamarbagwala and Morán (2010) find that individuals who were school aged in areas that were more affected by the war (1979–1984) in Guatemala completed fewer years of schooling and that this effect was stronger for girls. Their study suggests that loss of property and massive displacement led households to reallocate limited resources to providing young boys and, to a lesser extent, young girls with at least some primary education. Although both boys and girls received less secondary education as a result of the civil war, the effects were more pronounced for girls.

Justino (2011) observes that children who are needed to replace labor may be removed from school, which may deplete households of their stock of human capital for future generations. Akresh and de Walque (2009) and Shemyakina (2006) point to this mechanism as an explanation for the reduction in educational attainment and enrollment observed in contexts of civil war. In a recent paper, Rodriguez and Sanchez (2009) find that violent attacks in Colombian municipalities by armed groups have significantly increased the probability of school drop-outs and the inclusion of children in the labor market. We add that not only is the young generation is prevented from acquiring human capital, but educated members of older cohorts may be disproportionately killed, thereby depriving the country of its human capital stock.

II. CONFLICT, THE ECONOMY, AND EDUCATION IN BURUNDI

The 1990s were a particularly violent decade in Central Africa's history. Burundi and Rwanda experienced several episodes of mass murder and genocide, and the regional civil war in the Democratic Republic of Congo created an enormous loss of life and socioeconomic destruction. Most of the recent work on Burundi focuses on the causes of the latest episode of civil conflict (Nkurunziza and Ngaruko 2000), the progression of the crisis (Chrétien and Mukuri 2000), the year-by-year political dimensions of the conflict (Reyntjens and Vandeginste 1997; Reyntjens 1998), and possible conflict solutions (Ndikumana 2000). The proportion of people living below the nationally defined poverty line increased during this period from 35 to 68 percent, and the conflict led to double-digit inflation rates, which peaked at over 30 percent in 1997 (all figures from IMF 2007).

Civil conflict in Burundi began in 1965, three years after independence from the Belgian colonial administration, when a group of Hutu officers unsuccessfully attempted to seize power and overthrow the monarchy. This failed coup led to

a purge of Hutu from the army and government and marked the beginning of the political exclusion of the Hutu majority by the Tutsi minority. Power became the sole monopoly of the Tutsi, who effectively seized power in 1966 and proclaimed the First Republic. In 1972, a Hutu insurgency began in southwestern Burundi, resulting in considerable loss of life among the Tutsi residents. The subsequent Tutsi army repression eliminated all educated Hutu (Lemarchand 1994).

The next major confrontation was in 1988, when a Hutu insurgency began in the north. As in 1972, army repression was swift and took a heavy toll on local Hutus. However, unlike in 1972, the international community condemned the massacres and pressured the Buyoya regime to liberalize its political system. In June 1993, this situation led to the first free and fair elections in postindependence Burundi. The democratic transition did not last long. In October 1993, Melchior Ndadaye, the first democratically elected president and a Hutu, was assassinated by Tutsi army elements in a failed coup attempt, marking the start of another civil war. As the news spread to the rural provinces, Hutu peasants committed large-scale massacres of Tutsi and Hutu-supporting Union for National Progress (*Union pour le Progrès national*). Chrétien (1997) writes that districts in certain provinces were “almost completely ‘cleansed’ of all Tutsi elements.” The Tutsi army retaliated against the Hutu, continuing what would become the most severe civil war in Burundi’s history in terms of both human lives and socio-economic destruction (Ndikumana 2000).

Spatial and Temporal Intensity of the Conflict

In this paper, we use the term “violent conflict” to describe the massacres that occurred in the 1993–1994 period as well as the subsequent civil war. Because the exact timing and location of the massacres and the civil war plays an important role in our identification strategy (see section 4), we describe the evolution of the massacres and the civil war through time and space as follows:

- In 1993 and 1994, massacres occurred in many parts of the country with different intensities.
- From the end of 1994 to July 1996, civil war spread throughout the country.
- From July 1996 to August 2000, Major Buyoya returned to power after a bloodless coup. The civil war intensity was lower in most provinces, and the Arusha Peace and Reconciliation Agreement was signed in 2000.

The massacres were particularly intense in central and northern Burundi. Bundervoet (2009) estimates that in half of the provinces, more than 7 percent of all individuals lost their fathers in 1993. Table 1 provides the data per province and sketches the evolution of the civil war based on Chrétien and Mukuri (2000). Fighting began in October 1994 in the northwestern provinces of Cibitoke, Bubanza, Bujumbura Rural, and Ngozi. By early 1995, the violence spread to the bordering Kayanza province, and by April 1995, massacres of

TABLE 1. Primary Education Completed, by Province of Residence and Exposure to Violent Conflict

Province of residence in 1993 (1)	Death rate 1993 (2)	Poverty head count in 1990 (3)	Timing of the civil war 1995–1998 (4)	Primary education (% completed)		<i>t</i> test on the means (6)–(5) (7)
				Not exposed to violent conflict (5)	Exposed to violent conflict (6)	
Bubanza	4.2	22.4	1995–1998	15.90 [3.9]	2.4 [2.4]	–13.47**
Bujumbura Rurale	5.4	25.7	1995–1998	26.20 [2.7]	28.64 [3.3]	2.43
Bururi	3.8	37.7	1995/1996	25.04 [1.7]	18.97 [2.4]	–6.06**
Cankuzo	2.5	25.1	not affected	16.36 [2.0]	–	
Cibitoke	4.9	19.6	1995–1998	8.60 [2.3]	6.94 [3.0]	–1.66
Gitega	21.9	35.2	1996/1997	32.81 [3.4]	28.50 [2.4]	–4.03
Karuzi	26.7	66.8	1995/1996	23.20 [3.8]	9.60 [2.6]	–13.60***
Kayanza	35.4	44.9	1995/1996	27.01 [3.0]	20.70 [2.4]	–6.30*
Mwaro	12.8	24.0	1996/1997	20.85 [3.2]	10.81 [2.6]	–10.04***
Makamba	1.7	39.7	1996–1998	9.70 [1.1]	8.38 [1.5]	–1.32
Kirundo	12.1	34.0	1996/1997	22.00 [3.4]	16.23 [3.0]	–5.76*
Muyinga	16.0	27.8	1995/1996	21.17 [3.5]	11.29 [2.4]	–9.86***
Muramvya	7.8	24.0	1996/1997	39.43 [5.8]	25.97 [2.4]	–13.46**
Ngozi	25.7	42.5	1995/1996	16.81 [2.5]	9.44 [1.9]	–7.37***
Rutana	5.3	58.0	not affected	9.9 [2.8]	–	
Ruyigi	6.7	41.0	1995/1996	19.05 [3.8]	25.00 [8.3]	5.95
Rural Burundi	7% median	36.2	2.28 years average	19.79 [0.7]	15.98 [0.8]	–3.81***

Notes: Column (2) shows the percentage of survey respondents whose fathers were killed in the 1993 massacres; column (3) shows the Poverty Head Count; and column (4) shows the spread of the civil war over time and space. We only consider the “relevant” duration, which is the period during which school-aged children from the 1981–1986 birth cohorts could have been exposed to violence. Column (5) shows the birth cohorts not exposed to violent conflict (neither massacres nor the civil war) when they were between 7 and 12 years of age; column (6) shows the birth cohorts exposed to violent conflict (massacres, civil war, or both) when they were between 7 and 12 years of age. Standard errors are presented between brackets. *** significant at 1 percent, ** at 5 percent, and * at 10 percent.

Source: Bundervoet (2009), Republic of Burundi and World Bank (1995), Chrétien and Mukuri (2000), United Nations (1996), Bundervoet et al. (2009), UNFPA Enquête Démographique et de Santé (2002).

civilians and confrontations between army and rebel forces occurred in Karuzi, Bururi, Ruyigi, and Muyinga. By late 1995, fighting occurred in the central provinces of Gitega and Muramvya and the northern province of Kirundo. By then, the conflict had spread to almost all of the provinces of Burundi, with the exception of Cankuzo (in the east of the country) and Rutana and Makamba (in the south of the country). In July 1996, former president Buyoya again seized power in a bloodless *coup d'état* backed by the army. During late 1996 and early 1997, the armed conflict continued in Kayanza, Muramvya, Kirundo, and Gitega. In April 1997, the Arusha peace talks between the principal conflict parties began. In late 1997, insecurity increased again in Cibitoke, Bubanza, and Bujumbura Rural, provinces that remained unsafe until 1999.

The various conflict accounts provide no definitive explanation for why the massacres and the civil war affected some provinces earlier than others. However, the conflict's spatial spread was influenced by geography and natural endowments: (i) the proximity of the Democratic Republic of Congo's North Kivu region where the rebels had a base, explaining the early onset of war in the provinces of Cibitoke, Bubanza, and Bujumbura Rural; (ii) the presence of the Kibira forest in the north, which also served as a rebel base, explaining the spread of war to Kayanza and Ngozi provinces; and (iii) the Tanganyika Lake, which allowed the use of boats to bring the war to the southern province of Makamba.

Civilian Impacts of the Conflict

Between 1994 and 2001, an estimated 200,000 people lost their lives, a majority of them civilians (UNFPA 2002). To understand the micro-level impact of the war, we focus on displacement, looting of household assets, and the theft and burning of crops.¹

First, in its 2002 Demographic and Health Survey, UNFPA found that over 50 percent of the rural population had been displaced from their homes at least once between 1993 and 2000 as a result of violence. The average displacement duration for the entire sample was just over one year, meaning three agricultural seasons were missed because households could not cultivate or harvest their fields while they were displaced (UNFPA 2004). Displacement also meant that individuals were more likely to contract water- and vector-borne diseases while hiding in the forest. Because people could not carry significant amounts of food when fleeing their villages, adequate nutrition was a problem. Displacement also implied a lack of access to markets, health clinics, and schools because roads were unsafe or these structures had been damaged. Later in the war, civilians were forced into local resettlement camps by the government, and the camp conditions were poor (Human Rights Watch 2000). The displacement's impact on aggregate production from 1993 to 1998 showed production declines of 15

1. For an analysis of the health consequences of the civil war in Burundi, we refer to [Bundervoet et al. \(2009\)](#), Health and Civil War in Rural Burundi, *Journal of Human Resources*, 44, 2, p. 536–563.

percent in cereals, 11 percent in roots and tubers, and 14 percent in fruits and vegetables, with particularly dramatic declines in 1994 and 1995. Later in this paper, we will test the impact of displacement on schooling as a potential channel by which exposure to violent conflict can affect child schooling.

When the conflict ended in a given province, displaced households returned to their homes and fields. However, humanitarian interventions by either the government or nongovernmental organizations after the fighting ended were practically nonexistent because of the continued insecurity on all roads linking the capital, Bujumbura, to the countryside. By early 1995, rebel groups had begun to target and kill foreign nongovernmental organization workers and journalists who left Bujumbura to visit war regions. International development assistance dropped sharply during the crisis, from almost \$320 million before 1993 to below \$100 million in 1999 (IMF 2007).

Second, in addition to the displacement and killing of civilians, both rebel and government forces engaged in the looting of civilian property, particularly livestock, causing an unprecedented drop in household capital stock. Aggregate national figures show that the number of tropical livestock units in the country declined by 23 percent from 1990 to 1998, a decline that was predominantly due to theft and pillaging.

Third, [Human Rights Watch reports \(1998\)](#) document the theft and burning of household crops. Crops were stolen from fields and granaries, and coffee trees were particularly targeted for burning. Because coffee is the government's main source of tax revenue, rebels frequently burned coffee plantations to reduce government revenue, although we cannot quantify the magnitude of this destruction. Coffee is also an important source of income for small farmers who had less income to pay for other expenditures, including purchasing food crops, school fees, or health care.

Fourth, the conflict in Burundi is notorious for its adverse impact on women and girls. Rape was widespread, and there were many instances of brutality, even against children. Gender roles became more entrenched as boys and men were drafted by the army or recruited by rebel movements.

Education and Conflict in Burundi

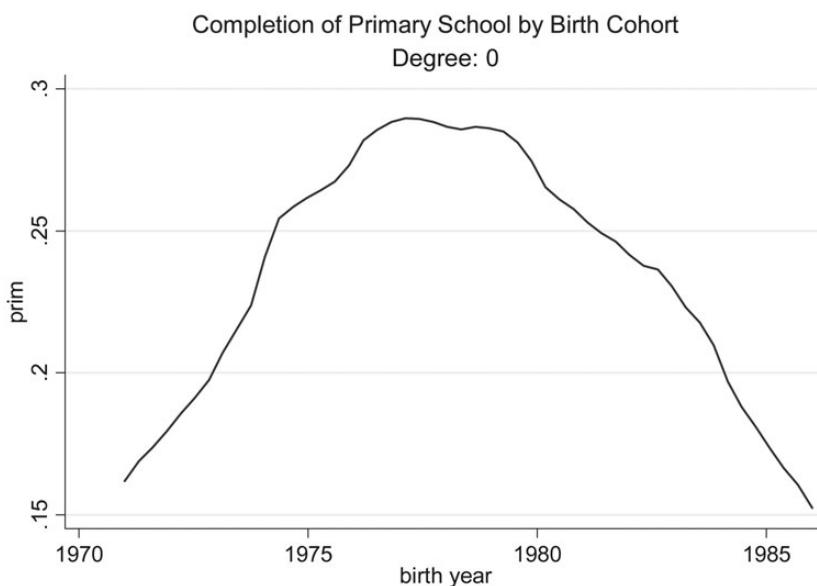
Access to education has been a long-standing source of inequality, tension, and conflict. In the cohorts under study, only 20 percent completed primary schooling. Education is directly related to jobs in the public sector, for which degree holders have the monopoly. The education system and jobs in the administration were dominated by Tutsi from the southern region of Bururi. Nkurunziza and Ngaruko (2002) write that in 1972, almost all educated Hutu were killed by the Tutsi army. Education was clearly a liability at that time.

In a new report on education and violent conflict, [UNESCO \(2011, p.51\)](#) calculates that the onset of conflict in Burundi marked an abrupt change in school enrollment. The decade before the conflict (1981–1991) saw an expansion of

enrollment for each new cohort, male and female. The gross enrollment rate increased from 33.2 to 70 percent in that decade (Ministère de l'Éducation 1999). The conflict-induced trend reversal can be observed in figure 1, which we computed with UNFPA 2002 data. The birth cohorts that could finish primary schooling before the beginning of the conflict show an upward trend in primary school completion, from a rate of less than 20 to almost 30 percent. The cohorts born between 1975 and 1980 show the highest primary school completion rates in the history of Burundi (up to 2005). This high completion rate was due to the expansion of primary education, which doubled the enrollment rate in five years (1982–1987) (UNESCO 1999; UNICEF 2008). Progress stops for the birth cohorts born at the end of the 1970s and is reversed from the 1980 birth cohorts onward, just when the first birth cohorts are confronted with the start of the violence. Children in Burundi officially attend primary school from age 7 to age 12, when they finish sixth grade (UNESCO 2011). Some children may start schooling later and complete primary schooling at a later age.

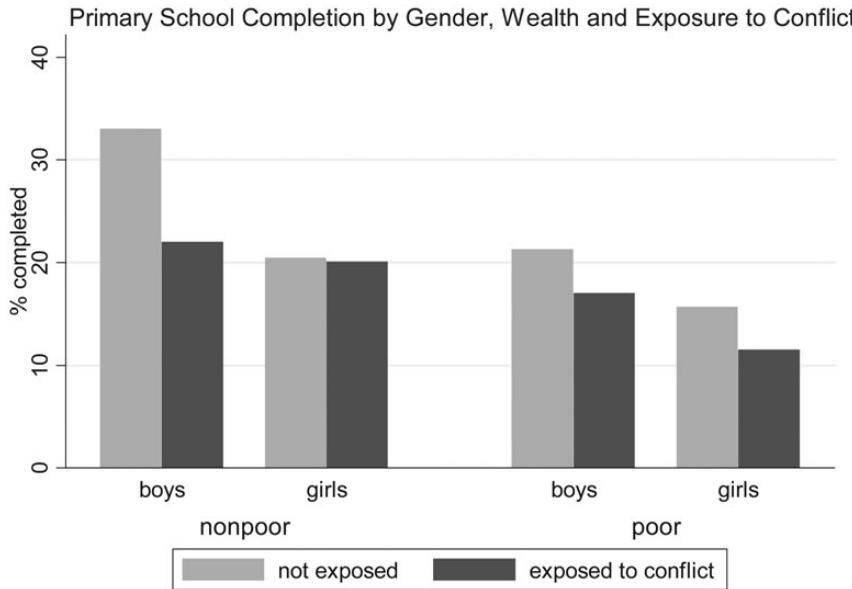
Figure 2 provides the key variables of our empirical approach (see below). Girls fare worse than boys, children from poor households fare worse compared to children from nonpoor households, and exposure to conflict negatively affects completion rates. Poverty at the household level is defined by livestock ownership before the start of the massacres and the civil war. This variable is the only pre-conflict wealth indicator available in the UNFPA survey (see below) and was registered through a recall question. Livestock ownership is one of the most important manifestations of wealth in rural Burundi.

FIGURE 1. Primary School Completion by Birth Cohort



Source: UNFPA Enquête Démographique et de Santé (2002).

FIGURE 2. Primary School Completion by Gender, Wealth, and Exposure to Conflict



Source: UNFPA Enquête Démographique et de Santé (2002).

The interaction of gender, poverty, and exposure in figure 2 offers surprising insights: the completion rate for girls from nonpoor households exposed to conflict is almost the same as for boys exposed to conflict, but it is significantly different for nonexposed boys and girls from nonpoor households. Moreover, we do not find this gender-gap-reducing effect of conflict on schooling for children in poor households.

In an extensive review of the damage to the education sector during the conflict in Burundi, Obura (2008, 94-96 and 99) observes that schools were destroyed or looted and teachers and children killed or displaced. Importantly, although the gross enrollment rate decreased, the Gender Parity Index did not decline during the conflict and even improved slightly, from 0.80 to 0.83. Obura also notes that a church-led education initiative, called Yaga Mukama, which existed before the war and provided two days of primary-school-level education per week to the rural poor, became very popular during the war and even acted as a substitute for formal education in affected areas.

III DATA AND IDENTIFICATION STRATEGY

In 2002, UNFPA collected demographic and health data (through the *Enquête Démographique et de Santé*) on almost 7,000 households. At the time of the survey, many Burundesi lived in camps for internally displaced persons.

A particular feature of this survey is that it is stratified over urban, rural, and camp locations in each province of the country. Weights are assigned to each observation in the survey representing the inverse of the probability of that observation being drawn in each sampled location. The sampling is based on a total population count by the National Institute of Statistics (*L'Institut de Statistique et d'Etudes Economiques du Burundi*) the year before the survey.² The focus of the survey is on household composition, schooling, and health, with significant attention to the potential impact of the conflict through displacement. Descriptive data are presented in tables 1 and 2. In our analysis, we dropped the observations from the capital of Bujumbura because it is nearly impossible to determine exactly when the population of Bujumbura was affected by the conflict. Arguments can be made for a very short as well as a very long time.

Conflict Variables

We construct four conflict exposure variables. The first two variables are general indicators of exposure to violent conflict, and the last two variables represent potential impact channels of conflict on education. First, we construct a binary variable to indicate whether a child resided in a province characterized by violent conflict. To determine which provinces were affected by the massacres (in 1993–1994) and the civil war (1995–1998), we use two sources.³ For the massacres, we rely on [Bundervoet \(2009\)](#), who computed the percentage of people whose fathers were killed. He applied the method proposed by [Gakidou and King \(2006\)](#) to correct for selection bias resulting from the absence from the survey of households where everyone was killed. Using that estimate (reproduced in table 2), we distinguish between eight provinces with a death rate higher than the median death rate (7 percent) and eight provinces with a lower death rate. The eight provinces with a lower than median death rate are defined as nonaffected provinces; the eight other provinces are defined as affected provinces. For exposure to the civil war, we rely on [Chrétien and Mukuri \(2000\)](#), who describe the spread of the violence over space and time. A child residing in a province engulfed by civil war during the child's primary school age is defined as exposed to civil war. We combine the exposure to massacres and to the subsequent civil war in one's province of residence during school age into one exposure variable.

Second, we determine the number of years that a child was exposed to violent conflict during the child's primary school ages. This variable is based on the combination of year of birth and province of residence at the onset of conflict. We cannot exclude the possibility that children moved to a more peaceful province after the onset of war in their province of residence. In that case, we overestimate

2. For details on the sampling method, we refer to annex B of the survey report "Situation démographique et sociale du Burundi: Résultats de l'enquête socio-démographique et de santé de la reproduction ESDR Burundi 2002," published by UNFPA Burundi.

3. Although the civil war lasted longer (until 2005), we only consider the relevant part of the war for our identification strategy: the potential exposure of children of primary school age from the birth cohorts under consideration (1971–1986)

TABLE 2. Individual and Household Characteristics, by Exposure to Violent Conflict (N = 5,856)

Name of the variable	Values	Not exposed to violent conflict (n = 3,586) (+) (1)	Exposed to violent conflict (n = 2,266) (+) (2)	t test on the means (2)–(1) (3)
<i>At the individual level</i>				
Age	16–31	25.1 [0.06] (3.79)	17.8 [0.03] (1.51)	–7.3***
Sex (% female)	0–1	60.6 [0.81] (0.48)	56.5 [1.04] (0.49)	–4.1***
Completed primary Education	0–1	19.8 [0.66] (0.40)	16.1 [0.77] (0.38)	–3.7***
No. of years exposed to violent conflict	0–4	0 [0.00] (0.00)	2.28 [0.02] (0.94)	2.28***
Number of times moved residence	0–4	0.087 [0.01] (0.39)	1.00 [0.02] (1.03)	0.91***
Years spent in a displacement camp	0–8	0.015 [0.01] (0.18)	0.89 [0.03] (1.66)	0.87***
<i>At the household level</i>				
Prewar wealth (Livestock in 1993)	0–20	1.45 [0.55] (4.76)	2.02 [0.84] (5.76)	0.57***

Notes: We only consider the “relevant” duration, which is the period that school-aged children from the 1981–1986 birth cohorts could have been exposed to the violence. (+) represents proportions in the case of binary variables and averages in the case of continuous variables. Standard errors are presented in brackets; standard deviations are presented in parentheses. *** significant at 1 percent, ** at 5 percent, and * at 10 percent.

Source: UNFPA Enquête Démographique et de Santé (2002), Chrétien and Mukuri (2000), United Nations (1996), and Bundervoet et al. (2009).

the duration of exposure, yielding a conservative estimate of the effect of conflict on schooling. However, UNFPA (2012, 141) notes that most conflict-induced migration occurred locally, within the same province. In addition, we test for individual exposure channels, such as the frequency of forced displacement and the time spent in a displacement camp, which are not measured at the level of the province of residence.

Many children born in the 1981–1986 period experienced at least one year of conflict during their primary school career. The oldest ones, born in 1981, were about to graduate from primary school when the conflict began. Depending on the province of residence, younger children experienced no, some, or many violent conflict during their school ages. Not all provinces were affected at the same time. In principle, the maximum duration of exposure to conflict during primary school age is six years. In practice, we do not find children exposed for more than four years in our sample.

The two other exposure-to-conflict variables are constructed directly from the UNFPA survey to index channels of influence of conflict on education. One variable is the number of times the child had to move residence forcibly during the

massacres and the civil war. The other variable is the number of years the child spent in a displacement camp, again during primary school age.

Identification Strategy

Our basic approach is a difference-in-differences strategy. We use the spatial and temporal variation of violent conflict in Burundi to infer the effect of exposure on child schooling. We compare children who were exposed to several years of conflict in their province of residence during their school-aged years with children of the same age residing in provinces that were not significantly affected and with children who were old enough to complete their schooling before the conflict started in both affected and nonaffected provinces. Building on figure 2 and previous tabulations, our baseline specification is a linear probability model of the following form:

$$Schooling_{ijt} = \alpha_j + \delta_t + \beta_1(Exposure_{jt}) + \gamma Z_i + \epsilon_{ijt} \quad (1)$$

where *Schooling* is our binary education variable for having completed primary school or not, measured for a child *i* residing in province *j* and born at time *t*. We denote with α_j the province fixed effects, with δ_t the birth cohort fixed effects, and with ϵ_{ijt} a random error term. The last term has an individual- and a household-level component. First, we calculate the $Exposure_{jt}$ variable as a binary measure to indicate a child residing in a province *j* that experienced violent conflict at the time when birth cohort *t* was of primary school age. Second, we use our estimate of the duration in number of years of exposure for a child residing in an affected province. In the latter case, β_1 , the coefficient of interest, measures the impact on schooling of an additional year of exposure to violent conflict. Including all provinces allows us to use variation in onset as well as in the duration of conflict across provinces to identify the war's causal impact on children's schooling. The provincial fixed effects control for any unobserved effect that does not change over time. To capture potential trends at the province level, we estimate a specification with province-level time trends. In section 6, we will test whether our variable of interest captures prewar province-level trends correlated with the duration of conflict.

In some specifications, we control for characteristics that are specific to the household in which the child lives. Importantly, and to avoid endogeneity, these household-level characteristics are measured in 1993, *before* the start of the conflict. *Z* is a vector of child-specific characteristics, such as the age, sex, level of education of the head of the household, and the wealth of the household. We do not include the *Z* variables in all specifications because by 1993, the older cohorts had already left their parental households; thus, these variables cannot affect their school completion.

We cluster our standard errors at the province level to control for intraprovince correlations (Bertrand et al. 2004). Clustering should occur at the province level because our shock is coded at this level. We face a problem of low numbers

of clusters, leading to larger standard errors and coefficients that are imprecisely estimated. We therefore use the bootstrapping method proposed by Cameron et al. (2008) to improve the estimation.

As stated in section 3.1, the spatial onset and subsequent spread of the war was determined by the proximity of a province to the border with the Democratic Republic of the Congo, the Kibira forest, or the Tanganyika Lake. These factors are exogenous to the level of education or other household characteristics. Voors et al. (2012) do not find evidence of the endogeneity of education (and other household characteristics) and exposure to violence at the household and village level. Although the authors cannot exclude occurrences of targeted violence, they note that “the probability of incorrectly maintaining the null of non-targeted violence is acceptably small” (950).

Furthermore, because the impact of conflict may differ according to the age at which the impact is felt, we also account for the age-specific onset of conflict in a separate set of regressions. In this case, the coefficient of the variable of interest indicates the effect of the onset of conflict in the province of residence at a given age on the probability of completing primary schooling.

Because we are also interested in a potential gendered effect of the impact of the civil war on human capital accumulation, we estimate the following specification:

$$Schooling_{ijt} = \alpha_j + \delta_t + \beta_1(Exposure_{jt}) + \beta_2 S_i + \beta_3(S_i * Exposure_{jt}) + \gamma Z_i + \varepsilon_{ijt} \quad (2)$$

where S_i is the sex of the child ($S_i = 1$ for girls), and the other variables are as in specification (1). In this specification, β_1 gives the effect of violent conflict on schooling for boys. The interaction effect between gender and conflict tells us whether there is an additional effect for girls, and the sum of β_1 and β_3 gives the total effect of conflict on schooling for girls. In addition, the sum of β_2 and β_3 gives the total effect of gender on schooling.

The above specifications do not specify the mechanism through which the impact of the conflict is channeled; the specification only provides a generic “exposure to civil war variable” in binary form, in number of years of exposure, or in age-specific onset. To analyze particular channels in more detail, we develop other specifications that use alternative measures of conflict to indicate a specific mechanism:

$$Schooling_{ijt} = \alpha_j + \delta_t + \beta_1(Channel_i) + \gamma Z_i + \varepsilon_{ijt}. \quad (3)$$

The channels are the time spent in a displacement camp during school age and the number of times the child moved residence during school age.

IV. FINDINGS

In table 3, we use the binary shock exposure variable that takes the value of one for children exposed to violent conflict in their province of residence during their school age years and zero for nonexposed individuals. The regressions in columns 1–3 are linear probability models. In the first column, we control for province and year of birth fixed effects and find that the coefficient of our variable of interest (exposure to violent conflict) is -0.09 , which means that the probability of completing primary schooling is 9 percentage points lower for children exposed to violence. Girls have a lower probability of completing primary schooling; however, there is a small, positive coefficient for the female*conflict exposure interaction variable. Given the rather large standard error, we are not able to interpret the point estimate of this interaction effect in column 1, an effect that is also not statistically significantly different from zero. Preconflict wealth, measured as livestock holdings, increases the probability of completing primary school. Columns 2 and 3 repeat the analysis controlling for linear age effects and for province-specific time trends. This control increases the magnitude of the coefficient of the variable of interest but affects neither the gender nor the interaction effect.

Driven by our exposure variable, which is measured at the province level, we need to cluster at that level (Bertrand et al. 2004). However, we have a small number of clusters. Standard errors for our variable of interest in the linear probability specifications are typically one-half to one-quarter of the magnitude of the coefficient of interest. These coefficients are thus imprecisely estimated and do not allow an interpretation of the point estimates.⁴

We use a well-developed method to address a small number of clusters, the CGM bootstrapping method, named after Cameron, Gelbach, and Miller (2008). The key element of the CGM method is that it resamples entire clusters and provides asymptotic refinement, which leads to improved inference. A drawback of this method is that the accompanying variance-covariance matrix cannot be calculated in the presence of fixed effects at the level of the clusters.⁵ Therefore, we run the CGM regression without fixed effects. We keep in mind (referring to the difference between columns 2 and 1) that this may overestimate the effect of exposure to conflict. The magnitude of the resulting coefficient (0.11), however, remains on the order of the previous estimates.

Given the small standard errors in the CGM procedure and the statistical significance of the gender*conflict dummy interaction, we are able to interpret the point estimates resulting from the regression in column 4. The female*conflict

4. We also clustered at the province-year of birth level, resulting in 256 clusters, which reduced the standard errors while yielding the same coefficients. However, we do not prefer this method because the intraprovince year of birth clusters may be correlated with each other. Cameron, Gelbach, and Miller (2008, 414) argue against clustering at the region-year level.

5. This drawback was confirmed by Judson Caskey in an email to us. We used the `cgmreg` do-file on professor Caskey's website to install the command.

TABLE 3. Linear Probability and CGM Regressions of Schooling, Conflict, Gender, and Household Wealth, with Binary Conflict Exposure Variable

Dependent variable: Child completed six years of primary schooling	(1) All LPM	(2) All LPM	(3) All LPM	(4) All CGM	(5) All, no weights CGM	(6) Poor only CGM	(7) Nonpoor only CGM
Violent conflict shock	-0.09** [0.04]	-0.16*** [0.04]	-0.17*** [0.04]	-0.11*** [0.01]	-0.07* [0.04]	-0.11*** [0.03]	-0.14*** [0.04]
Child is female	-0.08*** [0.02]	-0.08*** [0.02]	-0.07*** [0.02]	-0.06*** [0.01]	-0.07** [0.03]	-0.03*** [0.04]	-0.15*** [0.02]
Violent conflict*female	0.04 [0.03]	0.04 [0.03]	0.04 [0.03]	0.03** [0.01]	0.03 [0.03]	0.003 [0.01]	0.13*** [0.04]
Age (in years)		-0.009*** [0.003]		-0.006*** [0.000]	-0.002 [0.000]	-0.006*** [0.001]	-0.006*** [0.000]
Prewar wealth	0.008*** [0.002]	0.008*** [0.003]	0.007*** [0.002]	0.008*** [0.003]	0.009*** [0.003]	0.04*** [0.01]	0.005*** [0.002]
Intercept	0.06* [0.03]	0.40*** [0.08]	0.28*** [0.06]	0.39*** [0.03]	0.27*** [0.03]	0.34*** [0.04]	0.46*** [0.47]
Province fixed effects	Yes	Yes	Yes	No	No	Yes	Yes
Birth cohort fixed effects	Yes	No	No	No	No	No	No
Province-specific time trend	No	No	Yes	No	No	No	No
Sample Size	5,706	5,706	5,706	5,706	5,706	3,398	1,708

Notes: LPM represents linear probability model. Standard errors are presented between brackets. All regressions, except those in column (5), are weighted with the sample weights provided in the data set, and all regressions are clustered at the province level. *** significant at 1 percent, ** at 5 percent, and * at 10 percent.

Source: UNFPA Enquête Démographique et de Santé (2002).

interaction yields a coefficient of $+0.03$ with a standard error of 0.01 and is statistically significantly different from zero at the 1 percent level, meaning that the conflict in Burundi diminished the gender-gap in schooling somewhat. However, the gap still exists ($-0.06 + 0.03 = -0.03$). This result in the case of the binary exposure variable is only obtained if we weigh the regression with the sample weights provided in the survey, as column 5 shows. Therefore, caution is required in the interpretation of the interaction effect.

Repeating the analysis for poor and nonpoor households separately, we find in columns 6 and 7 that violent conflict diminishes the female disadvantage in schooling only for nonpoor households. The linear combination of the female and the female*exposure variables yields a coefficient of -0.02 ($= -0.15 + 0.13$) with a standard error of 0.02 and is statistically not significantly different from zero at the usual thresholds. Thus, there is no longer a gender gap in schooling between girls and boys from nonpoor households who are exposed to violent conflict. This result remains the same when we remove the sample weights: a coefficient of -0.02 ($= -0.127 + 0.106$), a standard error of 0.03 , and not statistically significantly different from zero at the usual thresholds (result not shown, available upon request). For girls from poor households, the linear combination is -0.03 ($= -0.033 + 0.003$) with a standard error of 0.01 and remains statistically significantly different from zero at the 1 percent level.

The F-test for the equality of the coefficients of interest for girls from poor and nonpoor households (i.e., the interaction of exposure and gender) in both equations confirms the above finding: 26.15^{***} , with $p < 0.001$, meaning that the coefficients are not equal. This result corroborates the intuition behind figure 2 in which boys and girls from nonpoor households exposed to conflict have similar completion rates. In sum, in nonpoor households, school completion rates for boys decrease to the level of girls such that the gender gap almost disappears. In contrast, in poor households where school completion rates are lower, girls are as affected as boys such that the gender gap persists.

Proceeding to *years of exposure* as our variable of interest in table 4, we find that the magnitude of the coefficients is approximately half of the binary case. Every additional year of exposure to violent conflict reduces the probability of completing primary schooling by 5 to 6 percentage points overall (columns 1 and 2) and by 8 percentage points for boys from nonpoor households (column 5). As shown in table 3, exposure to conflict diminishes the gender gap in primary school completion rates, particularly for girls from nonpoor households. The longer boys and girls are exposed to violent conflict, the stronger the reduction of the gender gap in school completion rates is. Again, referring to column 3, we arrive at this particular result only when using the sample weights provided in the survey.

Because exposure to shocks may have a different impact according to the age at which the child was exposed, we regress our outcome variable on a series of age-specific shocks. In table 5, we determine for each child the age at which the

TABLE 4. Linear Probability and CGM Regressions of Schooling, Conflict, Gender, and Household Wealth, with Number of Years of Conflict Exposure Variable

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Child completed six years of primary schooling	All LPM	All CGM	All, no weights CGM	Poor only CGM	Nonpoor only CGM
Years of violent conflict exposure	-0.05** [0.02]	-0.06*** [0.01]	-0.04*** [0.01]	-0.05*** [0.01]	-0.08*** [0.01]
Child is female	-0.08*** [0.02]	-0.06*** [0.01]	-0.08*** [0.03]	-0.04*** [0.01]	-0.14*** [0.02]
Violent conflict*female	0.02* [0.01]	0.02*** [0.00]	0.02 [0.01]	0.01*** [0.00]	0.06*** [0.01]
Age in years		-0.01*** [0.00]	-0.01 [0.00]	-0.01*** [0.00]	-0.01*** [0.00]
Prewar wealth	0.01*** [0.008]	0.01*** [0.002]	0.01*** [0.003]	0.04*** [0.007]	0.01*** [0.001]
Intercept	0.06 [0.04]	0.45 [0.03]	0.33 [0.08]	0.38 [0.04]	0.55 [0.04]
Province fixed effects	Yes	No	No	No	No
Year of birth fixed effects	Yes	No	No	No	No
Sample size	5,706	5,706	5,706	3,998	1,708

Notes: LPM represents linear probability model. Standard errors are presented between brackets. All regressions, except those in column (3), are weighted and clustered at the province level. ***significant at 1 percent, ** at 5 percent, and * at 10 percent.

Source: UNFPA Enquête Démographique et de Santé (2002).

child experienced the onset of violent conflict in the province of residence. Furthermore, we interact these dummy variables with the female variable. Examining column 2, where the use of the CGM model yields small standard errors, we find that the first four years (ages 7 to 10) are crucial, in declining order of magnitude, rather than the last two years (ages 11 and 12). For boys of poor households, the magnitude of the coefficients is smaller compared to boys from nonpoor households at all ages. The probability of completion of primary schooling for boys from nonpoor households decreases by 28, 21, 14, and 9 percentage points for the onset of conflict at ages 7, 8, 9, and 10 years, respectively. Table 5 clearly indicates that a cohort of children in Burundi was particularly affected by exposure to violent conflict if the conflict started in the province of residence at the moment that the cohort should have started primary school. A child who had yet to start school or who had only completed one or two years

TABLE 5. Linear Probability and CGM Regressions of Schooling, Conflict, Gender, and Household Wealth, with Age-Specific Conflict Onset

Dependent variable:	(1)	(2)	(2)	(3)
Child completed six years of primary schooling	All LPM	All CGM	Poor only CGM	Nonpoor only CGM
Age at onset of conflict				
Seven	-0.19 [0.11]	-0.15*** [0.01]	-0.13*** [0.01]	-0.28*** [0.03]
Eight	-0.13 [0.09]	-0.12*** [0.01]	-0.09*** [0.01]	-0.21*** [0.02]
Nine	-0.13* [0.07]	-0.06*** [0.01]	-0.05*** [0.01]	-0.14*** [0.04]
Ten	-0.08 [0.07]	-0.08*** [0.01]	-0.09*** [0.01]	-0.09** [0.04]
Eleven	-0.04 [0.04]	0.01 [0.02]	0.01 [0.02]	-0.03 [0.02]
Twelve	-0.02 [0.06]	-0.02 [0.02]	-0.07*** [0.01]	0.07 [0.07]
Child is female	-0.08*** [0.02]	-0.06*** [0.01]	-0.03*** [0.01]	-0.16*** [0.02]
Female*age at onset of conflict				
Seven	0.09* [0.05]	0.09*** [0.01]	0.05*** [0.01]	0.24*** [0.04]
Eight	0.10** [0.03]	0.07*** [0.01]	0.04** [0.01]	0.17*** [0.04]
Nine	0.07 [0.06]	0.05** [0.02]	0.03 [0.02]	0.16*** [0.05]
Ten	0.05 [0.03]	0.04*** [0.01]	-0.01** [0.01]	0.18*** [0.06]
Eleven	-0.02 [0.04]	-0.03** [0.01]	-0.07*** [0.02]	0.12*** [0.03]
Twelve	-0.04 [0.07]	-0.03 [0.02]	-0.02* [0.01]	-0.01 [0.01]
Prewar wealth	0.01*** [0.003]	0.01** [0.003]	0.04*** [0.01]	0.005*** [0.001]
Intercept	0.06 [0.03]	0.23*** [0.01]	0.02*** [0.01]	0.32*** [0.03]
Province fixed effects	Yes	No	No	No
Year of birth fixed effects	Yes	No	No	No
Sample Size	5,706	5,706	3,998	1,708

Notes: LPM represents linear probability model. Standard errors are presented between brackets. All regression are clustered at the province level and weighted with sample weights provided in the data set. *** significant at 1 percent, ** at 5 percent, and * at 10 percent.

Source: UNFPA Enquête Démographique et de Santé (2002).

of schooling at the start of the conflict may have been compelled to give up school. The gender gap in primary school completion, however, is reduced if the conflict starts at a young age in the province of residence of the child. Onset at an early age (up to age nine) reduces the gender-gap for girls from poor households. Onset at ages 10 to 12 aggravates the school completion chances of girls from poor households, in contrast to girls from nonpoor households. Onset at the ages of 11 or 12 for boys from nonpoor households does not affect their completion chances. Most likely, these boys are able to make up for potential lost months or years of schooling because their parents believe that a degree is within reach. We also tested a linear interaction between age and gender to account for possible trends in female enrollment over time, separate from the effect of conflict. The results (not shown) indicate that the effect of onset of conflict for girls at a younger age is slightly smaller; the effect is larger for girls who are slightly older (11 and 12 years old).

Exposure to violent conflict remains a broad term that is defined at the province-birth cohort level. Based on this definition, we cannot derive the channel by which the education of children at school age is affected during conflict. Possible channels are the destruction of school buildings or insecurity that makes parents keep children at home. One possible channel that affected almost one out of three households in Burundi during the war was forced displacement. Our data allow us to test this channel in two ways. The survey registered the number of times that each household member had to move residence because of the fighting and the length of stay in a displacement camp. It appears plausible that these two events would have a negative effect on the probability of a school-aged child completing primary schooling. Columns 1–3 in table 6 test these channels.

We find that the frequency of forced displacement and the length of stay in a displacement camp matter for school completion. Being uprooted from one's village because of ongoing or imminent violence proves to be disruptive to a child's school career to the extent that it decreases the probability of completing primary schooling, particularly if it occurs several times. The magnitude of the length of stay in a displacement camp is smaller but remains statistically significantly different from zero. As in previous regressions, the coefficient of the interaction between gender and the two alternative conflict measures is positive and statistically significant, meaning that displacement reduces the gender gap in school completion. When we test the effect of the three channels of violence (exposure to battles, forced displacement, and duration of stay in a camp, in column 3), all three channels exercise a negative and statistically significant effect on the completion of primary schooling. The probability of completing primary schooling declines by 8 percentage points as a result of exposure to conflict in the province of residence, by 6 percentage points for every instance of forced displacement, and by 2 percentage points for every year spent in a displacement camp. Importantly, the interaction effects of the alternative conflict measures with gender aggravate the completion chances for girls from poor households.

TABLE 6. CGM Regressions of Schooling, Conflict, Gender, and Household Wealth with Alternative Measures of Conflict Exposure (Robustness)

Dependent variable: Child completed six years of primary schooling	(1) Moved residence CGM	(2) Displaced in camp CGM	(3) Three measures CGM	(4) Poor only CGM	(5) Nonpoor only CGM	(6) 1978–82 cohorts excluded CGM	(7) 1971–74 cohorts excluded CGM
Violent conflict shock			–0.08*** [0.02]	–0.08*** [0.01]	–0.10** [0.04]	–0.09*** [0.02]	–0.06*** [0.01]
Times moved residence	–0.08*** [0.01]		–0.06*** [0.01]	–0.06*** [0.01]	–0.07*** [0.01]		
Years in displacement camp		–0.06*** [0.01]	–0.02*** [0.00]	0.01 [0.00]	–0.04*** [0.00]		
Child is female	–0.05*** [0.01]	–0.05*** [0.01]	–0.06*** [0.01]	–0.04*** [0.01]	–0.15*** [0.03]	–0.08*** [0.01]	–0.03*** [0.01]
Violent conflict*female			0.01 [0.02]	0.01 [0.01]	0.10*** [0.03]	0.06* [0.01]	–0.01 [0.01]
Moved residence*female	0.02*** [0.00]		0.02** [0.01]	–0.02*** [0.01]	0.03** [0.00]		
Years in camp*female		0.02*** [0.00]	0.00 [0.00]	–0.01*** [0.00]	0.02*** [0.00]		
Age in years	–0.006*** [0.00]	0.001 [0.00]	–0.01*** [0.00]	–0.01*** [0.00]	–0.01*** [0.00]	–0.01*** [0.00]	0.01 [0.00]
Prewar wealth	0.01*** [0.003]	0.01*** [0.003]	0.01*** [0.003]	0.04*** [0.01]	0.01*** [0.001]	0.01*** [0.001]	0.01** [0.003]
Intercept	0.31*** [0.02]	0.20*** [0.02]	0.46*** [0.03]	0.41*** [0.03]	0.52*** [0.05]	0.27*** [0.03]	0.18*** [0.04]
Province fixed effects	No	No	No	No	No	No	No
Birth cohort fixed effects	No	No	No	No	No	No	No
Sample size	5,706	5,706	5,706	3,998	1,708	3,981	4,550

Notes: Standard errors are presented between brackets. All regressions are clustered at the province level and are weighted with the sample weights provided in the data set. *** significant at 1 percent, ** at 5 percent, and * at 10 percent.

Source: UNFPA Enquête Démographique et de Santé (2002).

This result in column 5 is worse than our baseline result, where the gender gap for girls from poor households was neither reduced nor aggravated as a result of conflict exposure. These alternative measures thus show that the channel through which the conflict operates affects girls from poor and nonpoor households (column 6) differently. Displacement reduces the gender gap for the latter but widens it for the former.⁶ We call for caution in the interpretation of the result because we cannot exclude the possibility of a third, unobserved intervening variable (or set of variables) that correlates both with the value of the channel variable and with the education variable.

In columns 6 and 7, we perform robustness checks in which we omit the 1978–1982 and 1971–1974 birth cohorts from the analysis. For the 1978–1982 cohorts, the argument can be made that we are not sure whether these birth cohorts are affected by the violence. Some pupils may still be in primary school when they are 13 to 16 years old, in which case these older birth cohorts would be affected by the massacres and the civil war toward the end of their primary school career and would not constitute an adequate control group. Valente's (2011) paper on the schooling consequences of the conflict in Nepal makes a similar argument for dropping a few birth cohorts from the analysis. In column 6 of table 6, we thus infer the effects of violent conflict on affected cohorts when we are certain that the control group never experienced violence during their school careers and the treated group did. The result, again computed with the CGM method, is very similar to the one obtained in our baseline regression. For the 1971–1974 cohorts, the argument can be made that these cohorts are rather old, which may lead to a bias in the estimation of the time trend if the slope changes significantly over time. However, omitting these cohorts from the estimation (in column 7) leads to similar results.

V. ISSUES OF CONCERN FOR THE IDENTIFICATION STRATEGY: POVERTY, PREWAR TRENDS, SELECTIVE SURVIVAL, AND SELECTIVE MIGRATION

A first issue of concern for our identification strategy is that we may measure the effect of something other than exposure to violent conflict. If, for example, massacres were more intense or the civil war lasted longer in poor provinces compared to nonpoor provinces, then we may be measuring the effect of poverty instead of exposure to violent conflict. Although we control for wealth (in the form of livestock) in our regression analysis, this variable is measured at the household level. Because our exposure variable is measured at the province level, we must ensure that we are not capturing another effect. To that end, we analyzed data on the death rate in 1993 and the duration of the civil war in poor and

6. The figures in table 6, as in all tables, are rounded off at two decimals for aesthetic reasons. The coefficient and the standard error of the gender*years in the camp variable in column 4 are, for example, -0.0125 [0.0032].

nonpoor provinces. Poverty is measured as the percentage of the population under the poverty line in 1990 (prior to the start of the massacres and the civil war). A province is defined as poor if the percentage of the population under the poverty line is higher than 36.2 percent, the poverty headcount in Burundi in 1990. The difference in the 1993 death rate between poor and nonpoor provinces was -5.3 percentage points; however, this difference is not statistically significant at the usual thresholds. Similarly, the difference in the duration of the civil war between poor and nonpoor provinces is six months, but it is not statistically significant. [Bundervoet et al. \(2009\)](#) find very few correlations between the timing of the conflict onset (no, early, or late) at the province level, the length of exposure to conflict at the individual level, and a range of household characteristics. We conclude from this finding that there appears to be no selection into violence of provinces or individuals based on prewar characteristics. Therefore, it appears unlikely that our exposure variable is capturing a wealth or other effect.

Second, although we include province fixed effects in our specifications to control for time-invariant province characteristics, there may be a problem of endogeneity with time-varying province characteristics. To test for that endogeneity, we analyze the potential correlation between a prewar province-level trend in primary school completion rates and the duration of conflict in that province. To this end, we compute a prewar trend, defined as the difference between the average school completion rate of the three oldest (1971–1973) and the three youngest (1978–1980) prewar cohorts, and we regress this trend on the duration of violent conflict in the province. We perform this test at the province and the individual level. The results are presented in the online appendix, tables A1 and A2. We do not find any statistically significant effects in various specifications using the usual thresholds. We note that the specification in table 3, column 2 also includes a province-specific time trend.

A third issue is potential bias caused by selective survival. Because the survey, by definition, only contains data on children who survived violence until the time of the survey, we must account for potential survivor bias in the sample. Particularly, if death during the conflict was not a random event, we may over- or underestimate the effects of the conflict on schooling depending on the direction of the bias. The debate on the selectivity of violence in Burundi is ongoing ([Bundervoet 2009](#); [Voors et al. 2012](#)). The findings in [Bundervoet \(2009\)](#) suggest that the violent conflict affected the schooling of not only children who were school aged during the conflict but also those who had completed their primary education. Education in times of conflict in Burundi has proved to be a liability. Thus, our nonaffected cohort (1971–1980, in our approach not affected during their primary school career) suffers survival bias; the most educated cohort members were killed in 1993. Thus, on average, this cohort was more educated than we infer from the survivors in the 2002 survey. Assuming that there is no such survival bias for the affected cohort (which is likely because in 1993, they were too young to be targeted), the negative effect that we find for the affected

cohort would then be an underestimate of the true effect. We investigate this claim together with the next issue.

Fourth, in addition to selective killing, we may face a problem of selective migration. If migrants have a profile that differs from stayers, then we may over- or underestimate the impact of violent conflict on stayers. We thus need to address two potential threats: (i) people killed in the 1993 massacres and the subsequent civil war may have had a different profile than survivors; (ii) people who have migrated since 1993 may have a different profile than those who did not migrate. The latter issue can be divided into three categories of migrants/refugees: (ii.a) those who were internally displaced, (ii.b) those who were refugees and who returned to Burundi before 2002, and (ii.c) those who went abroad but did not return before 2002.

Persons in categories (ii.a) and (ii.b) are included in the sample. As mentioned above, one of the strengths of the survey design in 2002 was that it surveyed people living in displacement camps at the time of the survey. People who fled abroad but returned before 2002 are included because they were part of the target population at the time of the design of the survey. Thus, only groups (i) and (ii.c) represent a potential selection problem.

The 2002 survey allows us to investigate the profiles of people who were killed as well as migrants. We compare the profiles of households with and without at least one child killed in the 1993–2002 period (available in the online appendix, table A3). We conduct this comparison for the loss of boys as well as girls. Furthermore, we investigate the profiles of widowed persons. We find that parents who lost at least one daughter in the violence were less educated compared to parents who did not lose a daughter. Inferring from this finding that the killed daughters were more educated than the surviving daughters is premature given that the siblings (above age 15) of the deceased girls have a higher probability of completing primary education. This finding remains inconclusive for two other reasons: the low number of girls killed and the fact that we only dispose of the education data for siblings who still live in the parental home at the time of the survey. We do not find significant differences between the profiles of parents and siblings with and without at least one son killed. Regarding the death of spouses, we find a difference in the prewar level of livestock ownership. Households in which the husband died in the 1993–2002 period had significantly more livestock than households in which the husband was alive at the time of the survey.

These findings do not conflict with those of [Bundervoet \(2009\)](#). First, he also finds a higher level of prewar livestock among households with members killed, and second, his finding was based on the observation that fathers who were killed had more educated children. However, we are interested in the level of education of the deceased children, not the deceased fathers. A large part of the latter (and thus of Bundervoet's assertion) are born before 1971, a cohort that is not relevant for this paper.

Given that we only computed the profiles of parents, siblings, or husband/wives of people who were born in the 1971–1986 period and given that most of the above findings are not very conclusive or do not point in one clear direction, we conclude that selection bias caused by nonrandom killings is unlikely to bias our estimates in a particular direction.

We draw the same conclusion for the case of the migrants/refugees. Because we do not have data on the people who did not return to Burundi at the time of the survey, we attempt to obtain a profile by proxy. The closest we can get to the long-term refugees not registered in the 2002 survey is to consider the profile of those refugees who were abroad for many years and then returned to Burundi. From the figures (table A3.3) and in comparison with the stayers, these long-term refugees were slightly older, had a lower share of women, and had more educated heads of households. Had they returned, it would have increased the percentage that completed primary schooling in the nonaffected cohort (born 1971–1980). In that case, the estimates we find for the cohorts affected by violence can be considered an underestimate of the true effect.

VI. CONCLUSION

There is no universal theory that allows us to predict the direction of the gender effect of violent conflict on schooling. In times of peace, girls in Burundi are less likely to complete primary schooling compared to boys. This negative gender effect, irrespective of violent conflict, is a robust finding in all our specifications. However, is there an additional gender effect on schooling as a result of violent conflict? We find that the schooling of boys is negatively affected by conflict. For girls, we find that exposure to violent conflict reduces the gender gap in schooling, but only for girls from nonpoor households. This finding is confirmed across specifications and is consistent with the observations in [Obura \(2008\)](#), which present declining gross enrollment rates during the civil war but a stable and even slightly increasing Gender Parity Index. Exposure to violent conflict did not affect the gender gap in schooling between boys and girls from poor households.

The losses that we find in terms of schooling as well as the narrowing of the gender gap do not necessarily apply to other settings. [Justino \(2011\)](#) observes that the micro-level effects depend on the type of conflict and the socioeconomic profile of the victims. The magnitude of the observed effect in Burundi, a decline in the probability of completing primary schooling by 4 to 6 percentage points per year of exposure, cannot be compared straightforwardly with findings in papers using other dependent or independent variables. In her overview, [Justino \(2011\)](#) mentions a range of 0.4 to 1.2 years of education lost because of violent conflict. The magnitude of the effect in [Shemyakina \(2006\)](#), between 4 and 7 percentage points lower probability of completing the mandatory nine years in Tajikistan, is somewhat lower than our range of estimates of 7 to 17 percentage points for the exposed versus the nonexposed cohorts.

Policymakers should consider that conflict shocks may have different distributional consequences than the well-known economic or climatic shocks. Whereas price fluctuations or rain-level variability are known to affect the poorest part of the population much more than the nonpoor part, this is not necessarily the case in the event of shocks of a political nature, such as massacres or civil war. This paper demonstrates that groups considered least vulnerable in the development economics literature—boys, in general, and boys from nonpoor households, in particular—are most severely affected by violent conflict in terms of educational attainment. Primary school completion rates for boys from nonpoor households decrease to the level of girls from nonpoor households in case of exposure to violent conflict. As a result, the gender gap in schooling narrowed considerably during the conflict in Burundi among nonpoor households.

REFERENCES

- Akbulut-Yuksel, M. 2009. "Children of War: The Long-Run Effects of Large Scale Physical Destruction and Warfare on Children." IZA Discussion Paper No. 4407. Institute for the Study of Labor, Bonn, Germany.
- Alderman, H., J. Hoddinott, and B. Kinsey. 2006. "Long Term Consequences of Early Childhood Malnutrition." *Oxford Economic Papers* 58 (3): 450–74.
- Angrist, J. D., and A. D. Kugler. 2008. "Rural Windfall or a New Resource Curse? Coca, Income and Civil Conflict in Colombia." *Review of Economics and Statistics* 90 (2): 191–215.
- Bellows, J., and E. Miguel. 2009. "War and Local Collective Action in Sierra Leone." *Journal of Public Economics* 93 (11-12): 1144–57.
- Bertrand, M., E. Duflo, and S. Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119 (1): 249–75.
- Brakman, S., H. Garretsen, and M. Schramm. 2004. "The Strategic Bombing of German Cities During World War II and Its Impact on City Growth." *Journal of Economic Geography* 4 (2): 201–18.
- Bundervoet, T. 2009. "Livestock, Land and Political Power: The 1993 Killings in Burundi." *Journal of Peace Research* (46) 3: 357–76.
- Bundervoet, T., P. Verwimp, and R. Akresh. 2009. "Health and Civil War in Rural Burundi." *Journal of Human Resources* 44 (2): 536–63.
- Cameron, C., J. Gelbach, and D. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90: 414–27.
- Caskey, J. 2012. website consulted to install and execute the CGM command in STATA.
- Chamarbagwala, R., and H. Moran. 2010. "The Human Capital Consequences of Civil War: Evidence from Guatemala." *Journal of Development Economics* 94 (1): 41–61.
- Chrétien, J.-P. 1997. *Le Défi de l'Ethnisme: Rwanda et Burundi: 1990–1996*. Paris: Karthala.
- Chrétien, J.-P., and M. Mukuri. 2000. *Burundi, la Fracture Identitaire. Logiques de Violence et Certitudes 'Ethniques', (1993–1996)*. Paris: Karthala.
- Collier, P., and A. Hoeffler. 1998. "On the Economic Causes of Civil War." *Oxford Economic Papers* 50 (4): 563–73.
- Dasgupta, M., and S. Li. 1999. "Gender-Bias in China, the Republic of Korea and India, 1920–1990: Effects of War, Famine and Fertility Decline." Policy Research Working Paper 2140. The World Bank, Washington, DC.
- Davis, D., and D. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review* 92 (5): 1269–89.

- Dercon, S. 2004. "Growth and Shocks: Evidence from Rural Ethiopia." *Journal of Development Economics* 74 (2): 309–29.
- . 2004. *Insurance against Poverty*. Oxford: Oxford University Press.
- Fafchamps, M., C. Udry, and K. Czukas. 1998. "Drought and Saving in West Africa: Are Livestock a Buffer Stock?" *Journal of Development Economics* 55 (2): 273–305.
- Gakidou, E., and G. King. 2006. "Death by Survey: Estimating Adult Mortality without Selection Bias from Sibling Survival Data." *Demography* 43 (3): 569–85.
- Ghobarah, H., P. Huth, and B. Russett. 2003. "Civil Wars Kill and Maim People Long After the Shooting Stops." *American Political Science Review* 97 (2): 189–202.
- Gleditsch, N., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand. 2002. "Armed Conflict 1946–2001: A New Dataset." *Journal of Peace Research* 39 (5): 615–37.
- Guidolin, M., and E. La Ferrara. 2007. "Diamonds Are Forever, Wars Are Not: Is Conflict Bad for Private Firms?" *American Economic Review* 97 (5): 1978–93.
- Human Rights Watch. 1998. *Proxy Targets: Civilians in the War in Burundi*. New York: Human Rights Watch.
- . 2000. *Emptying the Hills: Resettlement Camps in Burundi*. New York: Human Rights Watch.
- IMF. 2007. "Burundi: Poverty Reduction Strategy Paper." IMF Country Report No. 07/46. International Monetary Fund, Washington, DC.
- Justino, P. 2011 "Violent Conflict and Human Capital Accumulation." In G. Brown and A. Langer, eds., *Elgar Companion to Civil War and Fragile States*. Cheltenham, UK: Edward Elgar Publishing.
- Lemarchand, R. 1994. *Burundi: Ethnocide as Discourse and Practice*. Cambridge, UK: Cambridge University Press.
- León, G. 2011. "Civil Conflict and Human Capital Accumulation: The Long-term Effects of Political Violence in Peru." *Journal of Human Resources* 47(4):991–1022.
- Maccini, S., and D. Yang. 2006. "Under the Weather: Health, Schooling, and Socioeconomic Consequences of Early-Life Rainfall." University of Michigan, Ford School of Public Policy and Department of Economics.
- Miguel, E., and G. Roland. 2006. "The Long Run Impact of Bombing Vietnam." University of California, Berkeley, Department of Economics.
- Miguel, E., S. Satyanath, and E. Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* 114 (4): 725–53.
- Ministère de l'Éducation Nationale. 1999. *L'évaluation de m'éducation pour tous – bilan à l'an 2000, Rapport des Pays – Burundi, EFA 2000, Mid-Decade Assesments*. Bujumbura, Burundi: Ministère de l'Éducation.
- Moulton, B. 1986. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics* 32 (3): 385–97.
- Ndikumana, L. 2000 "Towards a Solution to Violence in Burundi: A Case for Political and Economic Liberalization." *Journal of Modern African Studies* 38 (3): 431–59.
- Nkurunziza, J., and F. Ngaruko. 2000. "An Economic Interpretation of Conflict in Burundi." *Journal of African Economies* 9 (3): 370–409.
- Obura, A. 2008. *Staying Power: Struggling to Reconstruct Education in Burundi since 1993*. Paris: International Institute for Educational Planning, UNESCO.
- Republic of Burundi. 1995. "Note de Stratégie Economique et Evaluation de la Pauvreté." Report No. 13592. Bujumbura, Burundi.
- . 1998. *Enquete Prioritaire: Etude Nationale Sur les Conditions de Vie des Populations*. Bujumbura, Burundi: Institut de Statistiques et d'Études Economiques du Burundi.
- . 2003. "Interim Strategic Framework for Accelerating Economic Growth and Reducing Poverty." Poverty Reduction Strategy Paper. Bujumbura, Burundi.

- Reyntjens, F. 1998. "Evolution Politique au Rwanda et au Burundi, 1997–1998." In M. Stefaan and F. Reyntjens, eds., *L'Afrique des Grands Lacs. Annuaire 1998–1999*, 71–96. Paris: L'Harmattan.
- Reyntjens, F., and S. Vandeginste. 1997. "Burundi. Evolution Politique en 1996–1997." In M. Stefaan and F. Reyntjens, eds., *L'Afrique des Grands Lacs. Annuaire 1997–1998*, 1–13. Paris: L'Harmattan.
- Rodriguez, C., and F. Sanchez. 2009. "Armed Conflict Exposure, Human Capital Investments and Child Labor: Evidence from Colombia." HiCN Working Papers, 68. Housholds in Conflict Network, Berlin, Germany.
- Shemyakina, O. 2006. "The Effect of Armed Conflict on Accumulation of Schooling: Results from Tajikistan." *Journal of Development Economics* 95 (2): 186–200.
- UNESCO. 2011. "Education for All Global Monitoring Report, March." UNESCO, Paris.
- UNFPA Burundi. 2004. "Situation démographique et sociale du Burundi: Résultats de l'enquête socio-démographique et de santé de la reproduction ESDR Burundi 2002."
- United Nations. 1996. "Report of the Secretary-General on the Situation in Burundi. Security Council, S/1996/660." United Nations, New York.
- Valente, C. 2011. "What Did the Maoists Ever Do for Us? Education and Marriage of Women Exposed to Civil Conflict in Nepal." World Bank Policy Research Working Paper 5741. World Bank, Washington, DC.
- Voors, M., E. Nillesen, P. Verwimp, E. Bulte, R. Lensink, and D. Vandersoest. 2012. "Violent Conflict and Behavior: A Field Experiment in Burundi." *American Economic Review* 102 (2): 941–64.
- World Health Organization. 2002. *World Report on Violence and Health*. World Health Organization, Geneva, Switzerland.

Forthcoming papers in

**THE WORLD BANK
ECONOMIC REVIEW**

- **Can We Trust Shoestring Evaluations?**
Martin Ravallion
- **Evaluation of Development Programs: Randomized Controlled Trials or Regressions?**
Chris Elbers and Jan Willem Gunning
- **Effects of Colombia's Social Protection System on Workers' Choice between Formal and Informal Employment**
Adriana Camacho, Emily Conover, and Alejandro Hoyos
- **Can Conditional Cash Transfers Compensate for a Father's Absence?**
Emla Fitzsimons and Alice Mesnard
- **Collective Action and Community Development: Evidence from Self-Help Groups in Rural India**
Raj M. Desai and Shareen Joshi
- **Excluding the Rural Population: the Impact of Public Expenditure on Child Malnutrition in Peru**
Gissele Gajate-Garrido
- **Is Small Better? A Comparison of the Effect of Large and Small Dams on Cropland Productivity in South Africa**
Elodie Blanc and Eric Strobl



THE WORLD BANK

1818 H Street, NW

Washington, DC 20433, USA

World Wide Web: <http://www.worldbank.org/>

E-mail: wber@worldbank.org

OXFORD  OPEN



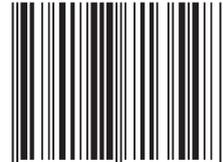
SFI[®]

Certified Sourcing

www.sfiprogram.org

SFI-00453

ISBN 978-0-19-871354-8



9 780198 713548