ISSN 0258-6770 (print) ISSN 1564-698X (online)

THE WORLD BANK ECONOMIC REVIEW

Volume 24 • 2010 • Number 1

www.wber.oxfordjournals.org

OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

THE WORLD BANK ECONOMIC REVIEW

The World Bank Economic Review is a professional journal for the dissemination of World Bank–sponsored and other research that may inform policy analysis and choice. 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. It is intended for readers familiar with economic theory and analysis but not necessarily proficient in advanced mathematical or econometric techniques. Articles illustrate how professional research can shed light on policy choices. Consistency with World Bank policy plays no role in the selection of articles.

Articles are drawn from work conducted by World Bank staff and consultants and by outside researchers. Non-Bank contributors are encouraged to submit their work.

Before being accepted for publication, articles are reviewed by three referees—one from the World Bank and two from outside the institution. Articles must also be endorsed by two members of the Editorial Board before final acceptance.

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.

THE WORLD BANK ECONOMIC REVIEW

Volume 24 · 2010 · Number 1

On Analyzing the World Distribution of Income

Anthony B. Atkinson and Andrea Brandolini

Consideration of world inequality should cause reexamination of the key concepts underlying the welfare approach to measuring income inequality and its relation to measuring poverty. This reexamination leads to exploration of a new measure that allows poverty and inequality to be considered in the same framework, incorporates different approaches to measuring inequality, and allows varied expressions of the cost of inequality. Applied to the world distribution of income for 1820–1992, the new measure provides different perspectives on the evolution of global inequality. JEL codes: D31, C80

There is currently a great deal of interest in the world distribution of income, as evidenced by the wide popular debate and by many academic articles (see the recent survey by Anand and Segal 2008). People are keen to know whether world inequality is growing or declining. They want to monitor progress toward eradicating world poverty, as in the UN Millennium Development Goals. The main argument of this article is that finding empirical answers to these questions requires first reconsidering the conceptual basis of the measurement of inequality and poverty. The move to a world canvas should be the

Anthony B. Atkinson (corresponding author, tony.atkinson@nuffield.ox.ac.uk) is a professor at Nuffield College, Oxford, and the London School of Economics, United Kingdom. Andrea Brandolini (andrea.brandolini@bancaditalia.it) is head of division at the Department for Structural Economic Analysis of the Bank of Italy. This article builds on Atkinson and Brandolini (2004), which analyzes international income inequality as well as world inequality and considers the intermediate class of inequality measures. That paper was presented at the 28th General Conference of the International Association for Research in Income and Wealth in Cork, Ireland; at the 4th International Conference on the Capability Approach: Enhancing Human Security in Pavia, Italy; and in a seminar at the University of Bari. The authors are most grateful to Conchita D'Ambrosio for her discussion of the paper at Cork; to seminar participants in Cork, Pavia, and Bari for helpful observations; to Stephen Jenkins for invaluable comments that led to a recasting of the paper; and to François Bourguignon and Peter Lambert for their excellent suggestions on this version. Further helpful remarks were made by Luigi Cannari, Fabrice Murtin, Alessandro Secchi, Paul Segal, and journal referees. This version was presented at the conference celebrating the 70th birthday of Sir James Mirrlees, and the authors are grateful to him and other participants for their most helpful comments. The authors thank Marco Chiurato and Federico Giorgi for excellent research assistance. The article was essentially completed during Atkinson's visit to the Economic Research Department of the Bank of Italy in 2006. The views expressed here are solely the authors' and do not necessarily reflect those of the Bank of Italy.

THE WORLD BANK ECONOMIC REVIEW, VOL. 24, NO. 1, pp. 1–37 doi:10.1093/wber/lhp020
Advance Access Publication January 8, 2010

occasion for a fundamental reexamination of underlying principles. While the issues raised apply at a national level as well, their heightened significance at a global level means that they can no longer be swept under the carpet. A critique of the standard inequality measures leads to an exploration of a new approach to measuring global inequality and poverty. This article is primarily about principles, but their application is illustrated by taking as a case study the data on the distribution of world income assembled by Bourguignon and Morrisson (2002).

There are three reasons why a reexamination is necessary. First, differences between incomes are much larger on a world scale than nationally. The Bourguignon and Morrisson data show the decile ratio (the ratio of the top to bottom decile groups) for all the world's inhabitants in 1992 as 24.7 (available at www.delta.ens.fr/XIX). The figure given by Gottschalk and Smeeding (1997, figure 2) was 5.8 in 1991 for the United States (for a different income concept) and 2.8 for Sweden, almost an order of a magnitude less than the global figure. Measuring world inequality thus requires evaluating a much wider range of incomes than that found in a typical advanced high-income country. (The move to a global scale is the focus here, but there are countries where the within-country income differences are much wider than in the United States, and the argument made here may also be seen as questioning the use of standard inequality measures within those countries.) As is discussed in section II, standard inequality measures impose too tight a straitjacket for applying them both to differences within countries and across the world. More flexibility is needed than can be accommodated with a single parameter, which is why the new measure explored in section III has several parameters.[1]

The second reason is the need to consider the relationship between measuring income inequality and measuring poverty. People are interested in both world inequality and world poverty, but the two literatures are separate (see Atkinson and Bourguignon 1999), with an uneasy relationship between them. The same criticism applies to studies at the national level, but it is easier to avoid a confrontation between the two concepts when they are moving in the same direction. At a global level, however, the proportion of the world population living on less than $1 a day is falling while the world Gini coefficient remains stubbornly high (see figure 1 later in the article). Do we give priority to one of the indicators? Some people have a lexicographic approach, giving total priority to poverty reduction, but others believe that there is some tradeoff between the two concerns. One possibility is to give both measures an independent role in a reduced-form social welfare function, as discussed by Fields (2006) and Kanbur (2008). The approach suggested here accommodates

1. A referee reasonably asked whether this argument is circular: that it suggests that the proper choice of inequality measure depends on how much inequality there is. It can be replied that the more flexible measure is appropriate in all circumstances but that, where income differences are sufficiently small, a single parameter measure may be a reasonable approximation.

differences in weighting of poverty and inequality in a social welfare function that can be tilted toward either concept by varying its parameters. More fundamentally, it goes to the heart of the difference between the two concepts by analyzing how society values an extra dollar at different places on the income distribution.

The third reason for a reexamination is that on a global scale, absolute as well as relative differences need to be considered. In 2005 the real per capita income of China was $4,091, or one-tenth the $41,674 of the United States (World Bank 2008, pp. 23–27). This means that China has to grow 10 times faster than the United States to achieve the same *absolute* increase in the production of goods and services per person. Even if China grows faster in relative terms, the absolute gap may widen. For example, with annual per capita growth rates of 5 percent in China and 2 percent in the United States, the absolute income gap between the two countries would widen for 49 years before starting to narrow, finally disappearing after 80 years. Concern for the absolute dimension of economic growth has far-reaching implications for assessing its distributive consequences, both between and within countries. As Livi Bacci (2001, p. 114) commented on Dollar and Kraay's (2002) conclusions on the "pro-poor" effect of economic growth, "it is not much of a relief for somebody living on $1 day to see that his income, up by 3 cents, is growing as much as the income of the richest quintile" (authors' translation).

At the empirical level, however, relative inequality measures predominate. Official publications do not report estimates of absolute inequality, and even academic studies are rare (one example is Del Río and Ruiz-Castillo 2001). Studies on global income inequality often take different routes, but they have in common a focus on relative measures of inequality (Chotikapanich, Rao, and Valenzuela 1997; Schultz 1998; Bhalla 2002; Bourguignon and Morrisson 2002; Milanovic 2002; Dowrick and Akmal 2005; and Sala-i-Martin 2006). Anand and Segal (2008) focus their survey on relative global inequality. Firebaugh (2003, pp. 72–3) briefly deals with the question to make it clear that "[i]nequality pertains to proportionate share of some item—not to size differences," and to avoid confusion, he introduces the terms "widening and narrowing gaps" to refer to changing absolute differences.

Only in two recent contributions has attention been drawn to the absolute/relative issue. Ravallion (2004, p. 19) notes that "[w]hile relative inequality has been the preferred concept in empirical work in development economics, perceptions that inequality is rising may well be based on absolute disparities in living standards." He shows how the "virtually zero correlation" between the relative Gini index and income growth becomes a "strong positive correlation" when an absolute Gini index is employed. Svedberg (2004, p. 28) highlights the importance of looking at the absolute distribution of income across countries and concludes that "[t]o pay more heed to the growing absolute

income gaps between rich and poor countries, and their consequences, seems an urgent task for future research into growth and distribution."[2]

Section I considers the application of the standard approaches to the world distribution of income and highlights the contrasting findings for trends in poverty, relative inequality, and the absolute cost of inequality. To understand this further, the "world social welfare function" underlying the exercises of measuring world income inequality and world poverty is made more explicit. The main tool in the analysis is the social marginal valuation of income, or the social value attached to an extra dollar received by people located at different points in the income distribution. Specifying how the social marginal valuation of income changes over the income scale is the first step in choosing an inequality measure, but expressing the cost of inequality relative to mean income is a second key step. These two steps underlie the construction of any inequality index.

Section II explains why the standard relative approaches to measuring inequality as well as the alternative, absolute approach proposed by Kolm (1976) fall short when applied over the whole range of world incomes. In effect, the existing measures excessively constrain how the social marginal valuation varies with income and provide no ready means to integrate the analyses of poverty and inequality. This leads to an exploration, in section III, of a new measure, grounded in an absolute approach but more flexible in form. The flexibility not only allows for a wider range of variation of income, as found on a global scale, but also shows how different measures can be obtained as limiting cases (and hence how the different approaches can be blended). The new measure, which differs in both of the key steps outlined above, is applied in section IV to the changes in the world distribution of income from 1820 to 1992. The data are not new—they are those of the Bourguignon and Morrisson (2002) dataset—but the new approach suggested here helps in understanding why people reach different conclusions about the evolution of world inequality and poverty. The main arguments are summarized in section V.

One important aspect should be clarified at the outset. Consideration of the world distribution as a whole, as in the studies cited above, assumes that there is a single world evaluation function. The main, but not the only, way in which inequality measures have been interpreted is in social welfare terms. In adopting such a welfarist perspective, this article posits the world social welfare function as a symmetric function $W(y_1, \ldots, y_n)$ of the real (purchasing power–adjusted) incomes, y_i, of the n people (households) in the world ranked by their income from lowest, y_1, to highest, y_n. There are assumed to be no other relevant differences between people apart from income,[3] which justifies the symmetry assumption. There is then a mapping from the properties of the

2. An important start has been made in studies of the global distribution combining average income and inequality measures; see Gruen and Klasen (2008).

3. The analysis is entirely static: it does not address the welfare evaluation of income changes, with which many people are concerned (see Bourguignon, Levin and Rosenblatt 2006).

world social welfare function to the properties of the inequality measure, and vice versa.

But there is an important difference between the world distribution and the distribution within a country. The people 1 to n are not all part of the same political entity. Redistributive mechanisms typically operate at the national level and are much more limited at the global level. The formulators of the social objective in a particular country may feel different degrees of responsibility for people who are citizens of that country and those who are citizens of other countries and so may treat them differently. This may, for example, lead to people with (real) income y being considered poor if they are citizens of country A but not if they are citizens of country B. Such differential treatment would, however, be inconsistent with there being a single symmetric world social welfare function. Some people would, for this reason, simply reject the idea of a world welfare function and hence any calculations of global inequality or global poverty (see, for example, Bhagwati 2004). Here, the aim is to make sense of such calculations, which implicitly assume a symmetric world social welfare function, treating as irrelevant the country of which a person is a citizen. It is on this assumption that the analysis is based.

Finally, while the article focuses on the social welfare approach to measuring inequality and poverty, that is not the only approach that should be considered. It would be possible to start from a set of axioms; it would be possible to consider other spaces, such as those of capabilities (see Sen 1992).

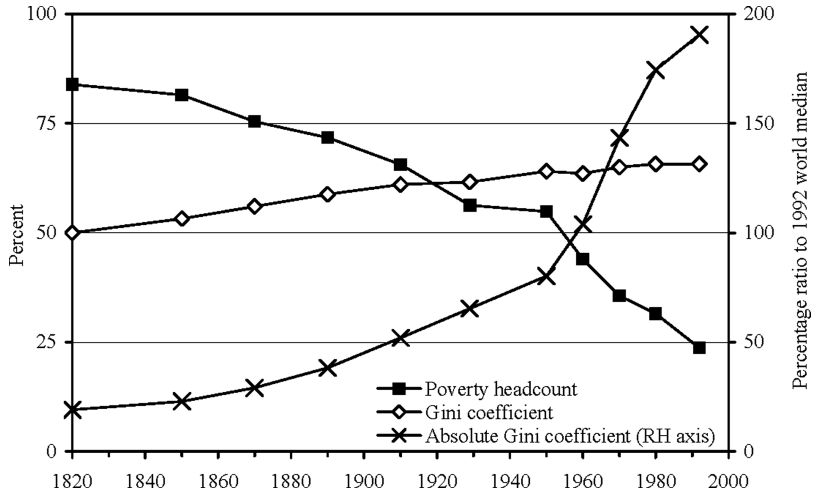
I. Applying Standard Indices to the World Income Distribution

The most popular index applied to measuring inequality is the Gini coefficient (half the mean difference divided by the mean). Figure 1 shows its value for the world income distribution for 1820–1992 using the Bourguignon and Morrisson data. Bourguignon and Morrisson's method is to use evidence on the national distribution (or the distribution for a group of countries) of the income shares of decile groups and of the top 5 percent. The groups are treated as homogeneous, which understates the degree of overall inequality. The distributional data are then combined with estimates of national GDP per capita, expressed in constant purchasing power parity (PPP) U.S. dollars at 1990 prices, which are derived from the historical time series constructed by Maddison (1995). The issues raised by this method and issues of data reliability are not considered here; the estimates are taken at face value.[4]

As Bourguignon and Morrisson (2002, p. 742) show, the Gini coefficient rose almost continuously from 1820 to 1950 and then more or less leveled off

4. See Deaton (2005) on the appropriateness of merging distributional measures from surveys with means from national accounts and Atkinson and Brandolini (2001) on the reliability of compilations of distribution statistics.

FIGURE 1. Evolution of World Inequality and Poverty, Standard Measures, 1820–1992



Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

between 1950 and 1992: "[T]the burst of world income inequality now seems to be over. There is comparatively little difference between the world distribution today and in 1950." If there is a Kuznets inverse-*U* curve for the world as a whole, then the world is slow to enter the downward phase: see the Gini coefficient in figure 1. On the other hand, measures of world poverty based on a constant purchasing power poverty standard show a steady—indeed an accelerating—downward trend. Figure 1 shows the world poverty headcount calculated by Bourguignon and Morrisson applying a standard comparable to that of the $1 a day standard used by the World Bank.

"Relative" and "Absolute" Approaches

The poverty measure in figure 1 represents an "absolute" approach, in that the poverty line is fixed in terms of purchasing power; a "relative" approach would make it proportional to the median or the mean of the distribution. However, an absolute approach does not imply that the line must be kept constant over time, as discussed below. This suggests a need for care in the use of the word "absolute," which may take on different meanings in the context of poverty measurement, as Foster (1998) shows.

A different use arises in measuring inequality. Following Kolm (1976), inequality measures are described as "relative" when they are invariant to proportional transformations (scale invariance) and "absolute" when they are invariant to additive transformations (translation invariance). The Gini coefficient described above is "relative." If all incomes are doubled in purchasing power, the Gini coefficient is unchanged: it is the *relative* mean difference.

There are good reasons for considering absolute income levels. With a doubling of real incomes from their 2005 values, per capita income in the United States remains 10 times that of China, but the absolute difference increases from \$37,583 to \$75,166. The world would be getting richer, but the differences between countries would be becoming larger in absolute terms. One way in which this can be reflected is by taking the absolute mean difference, or the absolute Gini coefficient (see figure 1), rather than the relative mean difference. The absolute mean difference has increased throughout the period, accelerating upward after 1950. This alternative—rather neglected—measure of inequality gives a different perspective on the evolution of world income distribution. If the \$1 a day poverty headcount is the optimistic view of recent decades of the world distribution, the absolute Gini is the pessimistic view.

Representing Different Social Values

Figure 1 helps explain why people may reach different conclusions about what is happening to world income distribution. People may look at poverty or inequality, and they may think of inequality in relative or absolute terms. This suggests that the functional form of the world social welfare function should reflect differences in social judgments. Indeed, Bourguignon and Morrisson (2002) show how alternative inequality indices may record different directions of change: the period 1980–92 saw the mean logarithmic deviation fall, the Theil index rise, and the Gini coefficient remain virtually unchanged. Different social values can be incorporated by using functional forms, such as those listed above, or by allowing a parameter to vary within a specific functional form. The analysis here uses the second approach, since it makes more transparent the underlying social values.

The constant elasticity index, I, introduced by Kolm (1969) and Atkinson (1970) allows users to choose different values of the elasticity, reflecting different views about the weights to be applied to changes at different points in the income distribution. The index, which is based on the mean of order $(1 - \varepsilon)$, is given by

$$(1) \qquad I = \begin{cases} 1 - \left[\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(\dfrac{y_i}{\mu}\right)^{1-\varepsilon}\right]^{1/(1-\varepsilon)}, & \varepsilon > 0,\ \varepsilon \neq 1 \\ 1 - \displaystyle\prod_{i=1}^{n}\left(\dfrac{y_i}{\mu}\right)^{1/n} & \end{cases}$$

where y_i denotes the income of person i in a population of n people with mean income μ. People are assumed to be ranked by increasing income, so that i indicates their position in the income distribution. Here, and throughout the article, income is assumed to be strictly positive. As ε rises, inequality receives more weight. Where $\varepsilon = 1$, the second version of the formula applies, and I is

equal to 1 minus the ratio of the geometric mean to the arithmetic mean. Where $\varepsilon = 2$, the value of I is higher since it is equal to 1 minus the ratio of the harmonic mean to the arithmetic mean.

The index I can be interpreted as expressing the cost of inequality in terms of the proportionate amount of income that could be subtracted from the mean without affecting the level of social welfare: $I = 1 - y_e^I/\mu$, where y_e^I is referred to as the equally distributed equivalent income, which can be written as $\mu(1 - I)$. This formulation involves two distinct steps, with choices to be made at each step, and this two-step distinction recurs throughout the article. The first step is to specify the function of individual incomes that is added across individuals. In effect, $y_i^{1-\varepsilon}/(1 - \varepsilon)$ is added across incomes, where division by $(1 - \varepsilon)$ ensures a nondecreasing function. (The degree of concavity of this function, captured by ε, embodies the chosen distributional values, as discussed further below.) This sum, divided by n, is denoted by Σ and referred to below as the additive element of the social welfare function.

The second key step in the measurement of inequality is to take a function of Σ and the mean income μ to arrive at an interpretable formulation. For the index I, the concave transformation is first reversed, to give $[(1 - \varepsilon)\Sigma]^{1/(1-\varepsilon)}$, and then divided by μ and subtracted from 1 to give I. The index I thus expresses the cost of inequality as the proportionate shortfall of the equally distributed income from the mean.

This is, however, a choice. The cost could be expressed differently, as discussed below. The two-step process has been described for the constant elasticity index, but it applies generally, including to nonadditive forms of Σ, such as that embodied in the Gini coefficient, G. In that case, $\mu(1 - G)$ gives the equally distributed equivalent income, or what Sen (1976) called "real national income": μ is a measure of aggregate economic performance, and $(1 - G)$ is the discount applied on account of the cost of inequality.

An increase in the income of person i raises social welfare, and the social marginal valuation of income can be defined as the value placed on an additional dollar received by a particular person. For the constant elasticity index, I, social welfare is defined by its ordinally equivalent representation constituted by the additive element Σ rather than by the equally distributed equivalent income $y_e^I = [(1 - \varepsilon)\Sigma]^{1/(1-\varepsilon)}$. The social marginal valuation of income, y_i, is hence equal to $y_i^{-\varepsilon}$.[5] The elasticity (defined positively) of the social marginal valuation of income is constant and equal to ε. For the index I, the marginal valuation tends to infinity as income goes to zero and to zero as income

5. Throughout the article, the social welfare function is defined in per capita terms rather than in its aggregate form, which implies that the social marginal valuation of income is divided through by n. Since what matters are the *relative* valuations of incomes i and j rather than their *absolute* values, the division by n is ignored in much of what follows—which affects all incomes equally—referring to the *individual* social marginal valuation of income. Note that the definition of social welfare in per capita terms has important implications for the interpretation of welfare changes when the population is growing. See footnote 15.

goes to infinity. For the Gini coefficient, G, the social marginal valuation of income is given by $[2 - (2i - 1)/n]$, where i is the person's rank in the income distribution and n is the total number of people.[6] For the poorest person, with $i = 1$, the social marginal valuation is $2 - 1/n$, which approaches 2 as n becomes large; for the median person (with n odd), it is 1; and for the richest person, it is $1/n$, which approaches zero as n becomes large. For both indices I and G, the social marginal valuation is nonnegative and nonincreasing.

The index I has been criticized, like the Gini coefficient, for being a relative measure: measured inequality is unchanged when all incomes increase (or decrease) in the same proportion. As discussed, it is a matter of concern at the global level that equal rates of growth in all countries imply widening absolute gaps. Kolm (1976) introduced the absolute index

$$K = \frac{1}{\kappa} \ln \left[\frac{1}{n} \sum_{i=1}^{n} e^{\kappa(\mu - y_i)} \right], \quad \kappa > 0. \tag{2}$$

The index K is absolute in the sense described earlier: inequality is unaffected by an equal addition to (or subtraction from) all incomes. With constant relative growth rates, inequality would increase.

As Kolm (1976, pp. 437–38) clearly recognized, the use of the index K involves two distinct departures, corresponding to the two key steps in the formulation described earlier. The first involves the different functional form in the additive element Σ: exponential rather than isoelastic. The second involves expressing the cost of inequality in absolute rather than relative terms. The index K represents the cost of inequality defined as the absolute amount of income that could be subtracted from the mean without affecting the level of social welfare: $K = \mu - y_e^K$, where y_e^K is the equally distributed equivalent income, equal to $\mu - K$ (see also Blackorby and Donaldson 1980). Inequality is said to cost \$X billion, rather than x percent of total income. In this respect, the index K is parallel to the absolute Gini coefficient. Equally, the measures I can be expressed in absolute terms ($\mu\ I$), and the measures K as a proportion of mean income. (The cost can be normalized in this way because an equally distributed equivalent distribution is being considered, and in this case absolute and proportional changes in the distribution are identical.)

The index K, like the index I, contains a free parameter κ that captures inequality aversion.[7] The larger κ, the higher is the weight attributed to the

6. This follows from writing the social welfare function as $\mu(1 - G)$ and G as $\Sigma_i(2i - n - 1)y_i/n^2\mu$. On the social welfare function implicit in the Gini coefficient, see Sen (1976) and Blackorby and Donaldson (1978).

7. The Kolm index, and more generally any nonrelative measure, is not *unit invariant*: a change in the unit of account of the incomes affects measured inequality, even if the underlying distribution is unaltered. Zheng (2007) proposes a new axiom of unit consistency requiring that income inequality rankings be preserved as the unit of account varies. The simpler approach adopted here takes account of the definition of units in the choice of κ.

lowest incomes; when κ tends to infinity, K tends to the difference $(\mu - y_1)$ between the mean income and the lowest income, y_1. The individual social marginal valuation of income, as computed from the additive element of the social welfare function, is given by $\exp(-\kappa y_i)$, and its elasticity with respect to income, defined positively, is equal to κy_i. The elasticity is increasing with income. Moreover, if the elasticity is specified at a particular value of income, then the value of κ can be deduced. If, for example, the elasticity is set equal to 1 at the mean income, then κ would equal the reciprocal of the mean.[8]

In empirical applications, the choice of the parameters ε and κ has to be considered. Researchers using the constant elasticity index I have chosen a range of values. Mirrlees (1978) straightforwardly proposed the "inverse square law," with a value of $\varepsilon = 2$. When used in official publications, however, the values tend to be lower. The study on high-income countries by Sawyer (1976) used values of 0.5 and 1.5. The U.S. Census Bureau (Jones, Weinberg, and U.S. Census Bureau 2000, p. 7, for example) publishes income distribution statistics taking values of 0.25, 0.5, and 0.75 (it also suggests that 1.0 is the maximum permissible value, although the expression for I indicates that this is not the case).

One way to pin down these values is by resorting to estimates of the social preferences implicit in tax systems. Christiansen and Jansen (1978) estimated the elasticity of the social marginal value of income implicit in the Norwegian system of indirect taxation in 1975 to be equal to 1.7 or to 0.9, depending on the model specification. Stern (1977) found an elasticity of around 2 for the British income tax system of the early 1970s.

Today, political preferences may be for less redistribution, so that lower values should also be considered. This has been suggested by experimental evidence, which provides a second source. Amiel, Creedy, and Hurn (1999) found broad support for median values of the elasticity of around 0.2. Such experiments typically ask people to think about the elasticity in terms of Okun's (1975) "leaky bucket." Suppose that a transfer costing $1 to a person with double the mean income is made to a person with half the mean income, with 50 cents being lost in the transfer, so that the recipient receives only 50 cents. Whether this "leaky" transfer increases social welfare depends on the relative valuation of marginal changes in income. An elasticity of 1 means that, compared to the $1 cost to the person with double mean income, four times the weight is attached to the 50 cents received by the person on half average income. So the transfer would raise social welfare. If the elasticity were 0.5, then the weight would only be twice, and the cost and the benefit would be equal. Put more generally, a loss ℓ is socially

8. The aim of this procedure is to fix the magnitude of κ. Once chosen, the value of κ is kept constant over time. This implies that, as real income grows, the *actual* elasticity of the social marginal value of income must also rise. To keep the elasticity constant over time, κ would have to be inversely proportional to the mean. However, this would change the nature of the index K, which would no longer be translation invariant.

acceptable up to the point at which $z^{\varepsilon}\ (1 - \ell) = 1$, where z is the ratio of the income of the donor to that of the recipient. This mental experiment is helpful in thinking about the implications of different values of the elasticity of the social marginal value of income, and it is considered again in the next section.

Applying Parameterized Measures to the World Income Distribution

In applying these measures to the world income distribution, values were taken for the elasticity in the interval [0.125, 2.0], which should cover a wide range of social preferences. As is clear from figure 2, adopting different values for ε gives very different measures of the cost of world inequality, varying in 1992 from 10 percent with $\varepsilon = 0.125$ to 74 percent with $\varepsilon = 2$. But the time trend does not differ much from that of the Gini coefficient, shown without markers. For the index K, figure 3 assumes that the values of the elasticity apply at the world median income in 1992, estimated from the Bourguignon and Morrisson data to be $1,712 at 1990 PPP. Here the cost of inequality is expressed absolutely, and the comparator is the absolute Gini coefficient, again shown without markers. The time path of the K index for elasticities of 1 and 2 is similar to that for the absolute Gini, and there is no great difference between the K index and the corresponding absolute version of the I index. The time paths for the elasticity of 0.125 show more difference.

These findings suggest that the major difference between the inequality indices I and K applied at a world scale lies in expressing the cost of inequality in absolute terms. Of the two key stages identified earlier, the expression of cost is crucial. The individual functional form plays less of a role.[9] But this is not necessarily the case when considering a wider range of functional forms, as examined next.

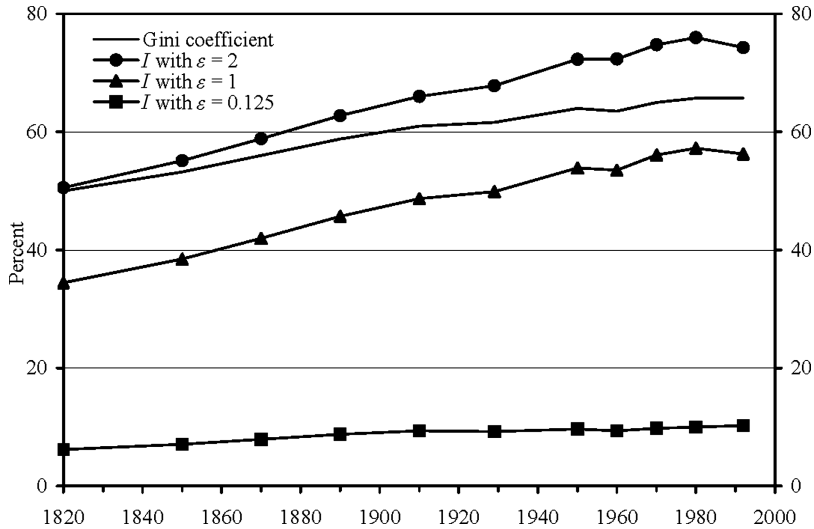
II. Sensitivity to Different Transfers

The functional forms considered so far do not allow sufficient flexibility when considering the world distribution. This may be seen by returning to the hypothetical leaky bucket experiment and the effect of transfers of income at different points in the world distribution. The essential problem is that of devising a path for the social marginal valuation of income that treats appropriately both transfers within a rich country, such as the United States, and transfers between people in rich countries and the poor in poor countries.

Table 1 shows the means for decile groups in a selection of countries (or groups of countries), according to the Bourguignon and Morrisson data for 1992, with income expressed relative to the 1992 world median ($1,712 at 1990 PPP). Thus, the first row in table 1 shows that the mean income for the

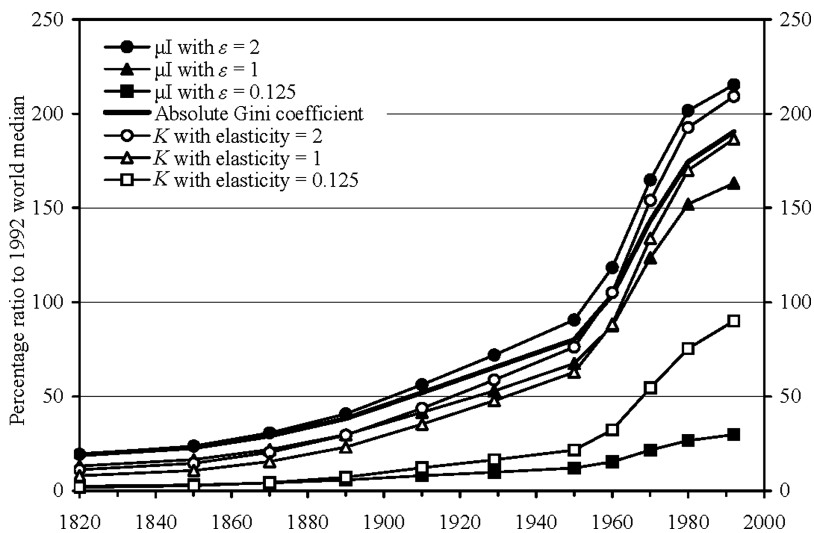
9. The same considerations apply to Kolm's (1976) "centrist" index and Bossert and Pfingsten's (1990) intermediate indices. These alternatives are discussed in Atkinson and Brandolini (2004).

FIGURE 2. Evolution of World Inequality, 1820–1992: Different Parameter Values



Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

FIGURE 3. Evolution of World Inequality, Absolute Measures, 1820–1992



Note: The elasticity of the *K* index is computed at the 1992 world median.
Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

first (lowest) decile group for 46 African countries (total population of 357 million) is 0.15 of the world median. The average income for the tenth (highest) decile group in the United States in 1992 is some 40 times the world median.

TABLE 1. World Incomes in 1992 Expressed Relative to the World Median and Social Marginal Valuation of Income

		Social marginal valuation of income								
Income relative to world median	Country and decile groups	Constant elasticity, $\varepsilon = 2$	Constant elasticity, $\varepsilon = 1$	Constant elasticity, $\varepsilon = 0.125$	Kolm index elasticity, $\kappa m = 0.125$ at world median	Gini coefficient[a]	Alternative 1: direction of poverty gap ($\lambda = 4$, $\beta = 12$, $\delta = \delta_0 = 0.5$)	Alternative 2: less angular ($\lambda = 4$, $\beta = 4$, $\delta = \delta_0 = 0.5$)	Alternative 3: direction of Kolm ($\lambda = 4$, $\beta = 2$, $\delta = -4$, $\delta_0 = 0$)	Alternative 4: Gini-like ($\lambda = 4$, $\beta = 3$, $\delta = \delta_0 = 2$)
0.15	46 African countries, decile group 1	44.444	6.667	1.268	1.112	1.988	4.893	2.850	2.571	1.036
0.20	Nigeria, decile group 2	25.000	5.000	1.223	1.105	1.968	4.846	2.759	2.388	1.036
0.28	India, decile group 1	12.755	3.571	1.172	1.094	1.892	4.687	2.592	2.131	1.035
0.34	Philippines-Thailand, decile group 1	8.651	2.941	1.144	1.086	1.760	4.445	2.451	1.963	1.034
0.40	Indonesia, decile group 1	6.250	2.500	1.121	1.078	1.711	4.034	2.299	1.815	1.033
0.48	Mexico, decile group 1	4.340	2.083	1.096	1.067	1.578	3.207	2.086	1.642	1.031
0.59	Philippines-Thailand, decile group 3	2.873	1.695	1.068	1.053	1.371	1.994	1.790	1.446	1.028
0.68	Russia, decile group 1	2.163	1.471	1.049	1.041	1.294	1.400	1.564	1.315	1.024
0.76	China, decile group 5	1.731	1.316	1.035	1.030	1.211	1.158	1.384	1.216	1.020
0.80	Indonesia, decile group 3	1.563	1.250	1.028	1.025	1.162	1.096	1.304	1.173	1.017
0.88	Egypt, decile group 4	1.291	1.136	1.016	1.015	1.059	1.031	1.163	1.095	1.012
1.01	North Africa, decile group 4	0.980	0.990	0.999	0.999	0.995	0.999	0.989	0.993	0.999
1.11	Turkey, decile group 4	0.812	0.901	0.987	0.986	0.962	0.993	0.894	0.931	0.986
1.27	37 Latin American countries, decile group 7	0.620	0.787	0.971	0.967	0.866	0.991	0.796	0.853	0.956
1.40	45 Asian countries, decile group 6	0.510	0.714	0.959	0.951	0.807	0.990	0.749	0.807	0.921
1.49	Mexico, decile group 5	0.450	0.671	0.951	0.941	0.759	0.990	0.728	0.781	0.891
1.57	Portugal-Spain, decile group 1	0.406	0.637	0.945	0.931	0.737	0.990	0.714	0.761	0.860
1.68	Poland, decile group 4	0.354	0.595	0.937	0.919	0.709	0.990	0.701	0.739	0.809
1.76	United States, decile group 1	0.323	0.568	0.932	0.909	0.681	0.990	0.695	0.726	0.767
2.00	Brazil, decile group 7	0.250	0.500	0.917	0.882	0.602	0.990	0.684	0.696	0.624
2.36	Germany, decile group 1	0.180	0.424	0.898	0.844	0.516	0.990	0.679	0.672	0.419
2.77	United States, decile group 2	0.130	0.361	0.880	0.802	0.472	0.990	0.677	0.659	0.283
3.03	Italy, decile group 2	0.109	0.330	0.871	0.776	0.425	0.990	0.677	0.655	0.244
3.44	Germany, decile group 2	0.085	0.291	0.857	0.737	0.402	0.990	0.677	0.651	0.219
7.02	Italy, decile group 5	0.020	0.142	0.784	0.471	0.238	0.990	0.677	0.649	0.208
9.19	United States, decile group 5	0.012	0.109	0.758	0.359	0.164	0.990	0.677	0.649	0.208
10.01	Germany, decile group 7	0.010	0.100	0.750	0.324	0.149	0.990	0.677	0.649	0.208

(Continued)

TABLE 1. Continued

Income relative to world median	Country and decile groups	Social marginal valuation of income								
		Constant elasticity, $\varepsilon = 2$	Constant elasticity, $\varepsilon = 1$	Constant elasticity, $\varepsilon = 0.125$	Kolm index elasticity, $\kappa m = 0.125$ at world median	Gini coefficient[a]	Alternative 1: direction of poverty gap ($\lambda = 4$, $\beta = 12$, $\delta = \delta_0 = 0.5$)	Alternative 2: less angular ($\lambda = 4$, $\beta = 4$, $\delta = \delta_0 = 0.5$)	Alternative 3: direction of Kolm ($\lambda = 4$, $\beta = 2$, $\delta = -4$, $\delta_0 = 0$)	Alternative 4: Gini-like ($\lambda = 4$, $\beta = 3$, $\delta = \delta_0 = 2$)
11.08	United States, decile group 6	0.008	0.090	0.740	0.284	0.131	0.990	0.677	0.649	0.208
14.79	France, decile group 9	0.005	0.068	0.714	0.178	0.069	0.990	0.677	0.649	0.208
20.66	United States, decile group 9	0.002	0.048	0.685	0.086	0.032	0.990	0.677	0.649	0.208
38.79	United States, decile group 10	0.001	0.026	0.633	0.009	0.005	0.990	0.677	0.649	0.208

Note: Decile group 1 is the lowest and decile group 10, the highest.

a. As income refers to the mean income of each decile group (as a ratio to the world median), in the expression for the social marginal valuation of income, the term $(2i - 1)/n$ represents the mean rank of all people in the decile group and is calculated as the sum of the cumulative share of all groups poorer than the one indicated and half the population share of the group itself.

Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

Now consider the individual social marginal valuation of income, expressed initially as an isoelastic function of income, $y^{-\varepsilon}$, so that the social valuation of an extra dollar accruing to a person with income y is 2^{ε} times that of an extra dollar accruing to a person with income $2y$. The implied social marginal valuations of income, expressed as a ratio to the social marginal valuation of the median income, are shown for three values of ε in table 1. As envisaged in the leaky bucket experiment, the value of ε determines the degree of loss that people are willing to accept when making a redistributive transfer. For domestic redistribution in the United States, the mean for decile group 6 is four times the mean for decile group 2, according to the Bourguignon and Morrisson data. Then $\varepsilon = 2$ implies that a transfer of \$1 from decile group 6 to decile group 2 would raise social welfare if all but $1/4^2 = 1/16$ leaked away before reaching decile group 2, or that a loss of up to almost 94 cents would be acceptable. This degree of leakage might appear too high. Put another way, the implied social marginal valuation for a person in decile group 2 in the United States would be 16 ($=4^2$) times that for a person in decile group 6, and the implied marginal valuation for a person in decile group 2 would be 196 ($=14^2$) times that of a person in decile group 10 (the mean income of decile group 10 being 14 times that of decile group 2). If $\varepsilon = 1$, then for a transfer of \$1 from decile group 6 to decile group 2, the maximum acceptable leakage is 75 cents, and the marginal valuation for a person in decile group 2 would be 14 times that for a person in decile group 10. If $\varepsilon = 0.5$, the central value used by the U.S. Census Bureau, the maximum acceptable leakage for a transfer of \$1 from decile group 6 to decile group 2 would fall to 50 cents, and the marginal valuation for a person in decile group 2 would be 3.75 times that for a person in decile group 10.

How does this extend to the world scale? Table 1 shows that the average income of the top 10 percent in the United States is some 140 times that of the bottom 10 percent in India. A value of $\varepsilon = 0.5$ implies that a transfer of \$1 from U.S. decile group 10 to India decile group 1 would be acceptable if the loss is 92 cents or less (if 8 cents are received). Would such a level of loss be acceptable?[10] The social marginal valuation of income accruing to decile group 1 in India is, at $\varepsilon = 0.5$, nearly 12 times that of a person in decile group 10 in the United States.

Some might believe that a lower value of ε should be applied. A value of $\varepsilon = 0.25$ implies that the social marginal valuation of income for a person in the bottom decile group in India is 3.44 times that of a person in the top decile group in the United States; a value of $\varepsilon = 0.125$ implies that the marginal valuation would be 1.85 times that of a person in decile group 10 in the United States and that a loss of up to 46 cents would be acceptable. However,

10. It should be noted that issues of agency are not considered here, in particular the fact that the United States has less control over the leakages with an international transfer than it has with a domestic transfer.

what are the implications of low values of ε for the evaluation of transfers from other countries to a person in decile group 1 in India? Table 1 shows that a relatively low-income person in Western Europe, say a person in decile group 2 in Germany, might have an income 12.5 times that of a person in decile group 1 in India. A value of $\varepsilon = 0.125$ implies that the marginal valuation of income for a person in decile group 1 in India is only 1.37 times that for a person in decile group 2 in Germany. This will strike many people as too low.

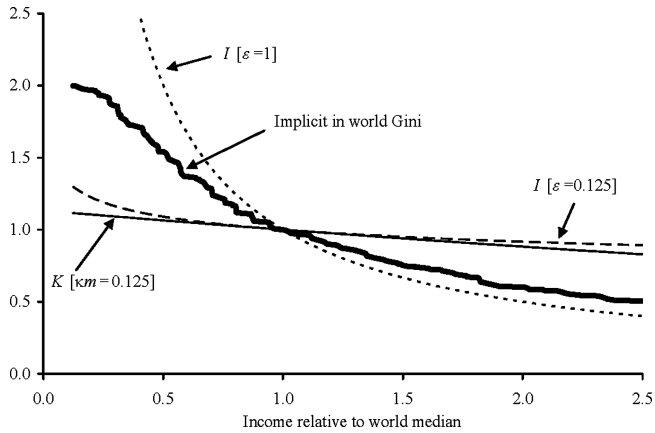
Moreover, reducing ε to such low values would have implications for transfers within the United States. With $\varepsilon = 0.125$, for example, a transfer would be made from decile group 10 to decile group 2 only if the leakage was less than 28 cents, which seems a limiting requirement. (A considerable fraction of those in decile group 2 are below the official U.S. poverty line.) The marginal value of $1 to a person in decile group 2 would be treated as worth only 1.4 times $1 to a person in decile group 10. Adjusting the parameter to fit the world distribution is, in effect, squeezing the range of distributional weights applied within the United States. Adopting values more appropriate to the within-country situation instead, however, implies a very wide range of marginal valuations on the global scale. With the inverse square law ($\varepsilon = 2$), for example, the marginal value of income to a person in the bottom decile group in India is almost 20,000 times that to a person in the top decile group in the United States.

These difficulties arise from the straitjacket imposed by the assumption of a constant elasticity. To quote Little and Mirrlees (1974, p. 240), "there is no particular reason why [the social marginal valuation] should fall at the same proportional rate at all consumption levels. Why should twice as much consumption deserve a quarter of the weight, whether consumption is low or high?"

Anand and Sen (2000) make a case for a variable elasticity function in which elasticity increases with income. As they note, this can be achieved by adopting the Kolm absolute index, K. Table 1 shows the marginal valuation of income implied by the Kolm index with an elasticity of 0.125 at the world median. This has a large effect on the marginal valuations within the United States: the marginal value of $1 to a person in decile group 2 rises to 90 times that to a person in decile group 10. But it would have little effect on the marginal valuations of income for the person in decile group 1 in India relative to that of a low-income person in Western Europe, rising from 1.37 to 1.48. The use of the Kolm index relaxes the constant elasticity assumption, but it does not reconcile both ends of the world distribution. The same consideration would apply if the social welfare function proposed by Anand and Sen (2000) were used, which combines the constant relative and constant absolute inequality versions.

The Gini coefficient, possibly the most used among inequality indices, provides an insightful alternative. As seen above, the social marginal valuation implicit in the Gini coefficient depends on the income rank order and is

FIGURE 4. Social Marginal Valuation of Income



Note: All values of the social marginal valuation of income are normalized by its value at the world median.

Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

bounded above by 2 and below by zero. (In Table 1, this is approximated by the mean rank of all people in each decile group, calculated as the sum of the cumulative share of all groups poorer than the one indicated and half the population share of the group itself.) The Gini coefficient has another appealing property, which may be seen in figure 4 (corresponding to table 1). With the Gini index, the social marginal valuation of income declines above the 1992 world median in a fashion similar to the constant elasticity $\varepsilon = 1$ but differs at lower values. Initially, the marginal valuation falls slowly with income, but then the decline accelerates up to the mode.[11] Finding a functional form that has this "slow, quick, slow" property would enable, at least in part, differentiating between incomes received *within* poor and rich countries, while also bounding the differentiation *between* poor and rich countries. The widespread use of the Gini coefficient in studies of the world distribution can be seen as an implicit revelation of preference for such a pattern.

At the same time, despite its popularity, the Gini coefficient has two features that are open to challenge. The first is that, unlike the I and K indices, it is not additively separable in incomes. It lacks the property that the ratio of the social marginal valuations of income for person i and person j depends only on their incomes. Consider an example. Suppose that the European Union is contemplating a switch from a policy transferring \$1 to a person in decile group 4 in Turkey (under a program for countries applying for EU membership) to a policy transferring \$1 to a person in decile group 1 in India (under its development program). With Gini weights, the social marginal valuation for decile

11. The kernel estimates of the world distribution of income by Bourguignon and Morrisson (2002, figure 1) have a secondary mode, but the broad shapes are consistent with the statement in the text.

group 1 in India is 1.97 times that for decile group 4 in Turkey (see table 1). Between these two groups lie the bottom six decile income groups in China. If rapid development in China were to shift these decile groups above decile group 4 in Turkey, the fall in the income ranking in the world population would cause the social marginal valuation for the Turkish decile group to rise from 0.962 to 1.218. As a result, the social marginal valuation for decile group 1 in India relative to decile group 4 in Turkey would fall by more than a fifth, to 1.55. Incomes in India and Turkey would have remained the same, but the attractiveness of the switch in policy would have been affected by development elsewhere. This is the argument for assuming additive separability (although there may be circumstances in which separability might not be an appropriate assumption).

The second problem with the Gini coefficient arises from its treatment of high incomes. It is going too far to say that it involves "spiteful egalitarianism" (Feldstein 2005, p. 12), but it is true that the Gini weights do tend to zero very fast at the top of the income scale, as can be seen from table 1. It is not clear that the social marginal valuation for a person in decile group 9 in France should be 2.14 times that for a person in decile group 9 in the United States. It might be desirable to allow for the possibility that the social marginal valuation remains strictly above a positive value as income tends to infinity.

III. Toward a New Approach

The previous discussion provides the rationale for exploring a new measure. The objective is to design a measure that combines the "slow, quick, slow" empirical property of the Gini coefficient with additive separability, while allowing for a strictly positive social marginal valuation of income at all income levels. The second motivation for devising a new measure goes back to the objective of measuring poverty and inequality within a common framework. This can be achieved by assigning the role of a poverty line to a particular income level, a feature not part of any of the measures considered so far. Identifying a poverty threshold within the social welfare function helps to show that concern about poverty may arise because incomes are unequally distributed and some people fall below the poverty line or because mean income is below the poverty line (or both). Put differently, poverty may occur even if everyone has the same income, if a society is globally poor. Clearly this depends on how the poverty line is defined. A society could not be globally poor if the poverty line were taken as some percentage (less than 100 percent) of the mean income.

Several approaches are considered here. That just described, often referred to as a "relative" poverty line, may be contrasted with "absolute" poverty lines that are independent of mean income, although it should be noted that "absolute" poverty lines are not necessarily constant over time. As Sen (1983) has

stressed, a standard fixed in one evaluative space, such as that of capabilities, might imply a poverty line in terms of income that varies over time.

Achieving the objectives of increasing flexibility and integrating poverty and inequality requires introducing several parameters governing the form of the social welfare function. Understandably, there may be resistance to being asked to consider a measure of inequality that requires thinking first about the values of different parameters. The popularity of the Gini coefficient is in part due to the fact that it does not require specifying any parameters. However, this does not mean that there are no implicit value judgments underlying the Gini coefficient; as already shown, its properties can be challenged. The virtue of parameterization, as argued in Atkinson (1970), is that it forces the user to make explicit choices about the instrument of evaluation and it allows readily for different views about the importance of redistribution. At the same time, guidance on the choice of parameter values may be welcomed. One aim of the numerical application in the next section is to give a flavor of the consequences of different choices.

Consider the following four-parameter measure of global social welfare:

$$\Sigma = \frac{1}{n}\sum_i W_i \equiv \frac{1}{n}\sum_i \left\{ y_i - \frac{\lambda}{\beta} e^{\beta(\delta_0 - \delta)} \ln\left[1 + e^{\beta(\delta - y_i)}\right] \right\} \tag{3}$$

where β is positive and has the dimension of 1/income, δ_o and δ have the dimension of income, and λ is a non-negative pure number. As a consequence, the expression Σ used to evaluate the total world welfare has the dimension of income.[12]

Specification (3) embodies the two key steps described earlier—the shape of the individual function and the calculation of the welfare cost—both of which involve assumptions. The first step is to adopt an exponential form that tilts the measure in the direction of index K rather than index I. Indeed, as shown below, the Kolm measure may be seen as a limiting case of specification (3). In this sense, Σ is an absolute measure.[13] The first element in specification (3) is mean income, from which the second term, which captures the unequal distribution of income weighted by the parameter λ, is subtracted. In this sense, too, Σ is an absolute measure. Note that Σ can be negative.

How can the different parameters be interpreted? It is useful to begin with the first derivative:

$$W'_i = \frac{1}{n}\left[1 + \lambda \frac{e^{\beta(\delta_0 - \delta)}}{1 + e^{\beta(y_i - \delta)}}\right] \tag{4}$$

12. The social welfare function is assumed to be defined over incomes, not individual utilities. This is not to assert that there exists a well-behaved utility function such that the private marginal valuation of income can be written in this form.

13. It would be an interesting extension to consider a version closer to the I measure. The authors are grateful to Peter Hammond for suggesting the derivation of the K or I indices as a limiting case.

As before, the divisor n is ignored, and the term in brackets is referred to as the individual social marginal valuation of income for person i. There are four parameters in specifications (3) and (4), but δ_0 plays only an instrumental role; unless explicitly signaled, it is assumed that $\delta_0 = \delta$, reducing to three the parameters that need to be chosen: λ, β, and δ.

The parameter λ captures the varying importance attached to distributional concerns. If no weight is attached to distribution, then one simply sets $\lambda = 0$, the social marginal valuation is everywhere 1, and that is the end of the story. The implications of different, nonzero, choices of λ may be seen from considering the fact that (with $\delta_0 = \delta$) the social marginal valuation falls monotonically from $[1 + \lambda/(1 + e^{-\beta\delta})]$ when y_i is 0 to 1 as y_i tends to infinity. The social marginal valuation of a person with zero income is at most $(1 + \lambda)$ times that for the richest person, so that $\lambda = 4$ corresponds to a maximum ratio of five, which implies a maximum socially acceptable loss of 80 percent from a transfer from the richest person to the poorest. This value of λ is applied in the illustrations below, although in the light of the large world income differences, this may be regarded as a conservative choice.

The two remaining parameters, β and δ, determine the nature of the concern for inequality and poverty. Specification (3) gives a special status to the income level δ, and one interpretation, taken up below, is that of a poverty line. Other interpretations can be given as well, however, and variations in δ allow the measure to adopt either a Kolm-like form or a Gini-like form. The parameter β determines the "angularity" of the measure, which has a natural interpretation in each of the cases, now discussed in turn. Because the discussion below focuses not on incomes but on their ratios to the median m, the actual values of the parameters in the income space are δm, $\delta_0 m$ and β/m.

The Poverty Gap

Some people believe that poverty is a concern, but not inequality. This position is exemplified by Feldstein (2005, p. 12): "I have no doubt about the appropriateness of transferring income to the very poor . . . the emphasis should be on eliminating poverty and not on the overall distribution of income or the general extent of inequality." This position has been called "charitable conservatism" (Atkinson 1990). An attraction of the measure explored here is that it encapsulates the poverty gap if δ is set as the poverty line, with $\delta_0 = \delta$, and β tends to infinity. Under these assumptions, the social welfare function (3) becomes:[14]

$$\Sigma_{\beta\to\infty,\delta_0=\delta} = \frac{1}{n}\sum_i y_i - \lambda\frac{1}{n}\sum_i \max[0, (\delta - y_i)] \tag{3a}$$

14. As β goes to infinity, if $y_i \geq \delta$ the term $(1/\beta)\ln[1 + \exp(\beta(\delta - y_i))]$ in equation (3) tends to zero; if $y_i < \delta$, application of L'Hôpital's Rule allows the limit to be calculated as $(\delta - y_i)$.

Thus, world welfare is evaluated as the mean minus λ times the aggregate poverty gap per person of the total population.[15] As may be seen from equation (4), as β tends to infinity, with $\delta_0 = \delta$, the social marginal valuation equals $(1 + \lambda)$ where income is below δ and 1 where it is above δ. Distributional concern is concentrated below the poverty line, to an extent that depends on λ. Where $\lambda = 4$, four times the poverty gap is subtracted from national income: multiplying by λ allows for the concerns of those who feel that the small size of the poverty gap, expressed per person of the total population, understates its significance.

A Less "Angular" Version

With the poverty gap, the social marginal valuation is constant as a function of income when income is below the poverty line, falls like a stone at $y = \delta$, and is again constant for all incomes above the poverty line. For some people, this is too abrupt. They might well want to taper the marginal valuation as income approaches the poverty line and to recognize that the needs of the "near-poor," just above the poverty line, are greater than those of people comfortably above. The 1991 modification to the Human Development Index (HDI) was based on the argument that "the idea of diminishing returns to income is now better captured by giving a progressively lower weight to income beyond the poverty cut-off point, rather than the zero weight previously given" (United Nations Development Programme 1991, p. 15).[16] The HDI modification took the form of a fractional weight above the poverty line, but such a less "angular" version can also be achieved using specification (3) by retaining δ $(= \delta_0)$ as the poverty line and taking a finite value of β. With β finite, the social marginal valuation changes more smoothly around δ. This may be seen from the second derivative:

$$W_i'' = -\frac{1}{n}\frac{\beta\lambda e^{\beta(\delta_0-\delta)}}{[1 + e^{-\beta(y_i-\delta)}][1 + e^{\beta(y_i-\delta)}]} \tag{5}$$

which has its minimum value (the steepest downward slope for the marginal

15. Formulation of the social welfare function in per capita terms implies that world welfare goes up, *ceteris paribus*, whenever the aggregate poverty gap grows less than the population. However, one could argue that what matters in assessing poverty is the amount of resources necessary to eliminate poverty—the absolute aggregate poverty gap, not its value per person. This corresponds to viewing the world poverty as measured by the absolute number of the poor rather than by their number relative to the total population. Which of these two conceptions of poverty is chosen has important consequences for interpreting the evolution of poverty and welfare, as the absolute and per person aggregate poverty gaps need not move in the same direction. Chakravarty, Kanbur, and Mukherjee (2006) attempt to unite these two conceptions of poverty by developing a family of poverty measures that avoid the population replication axiom.

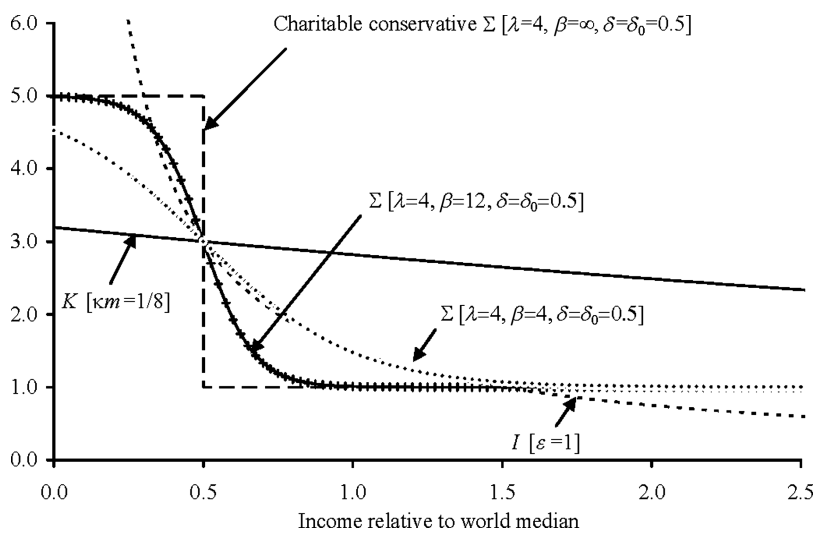
16. This quotation is drawn from Anand and Sen (2000), who present an extensive (and sympathetic) critique of the treatment of the social marginal valuation of income in successive versions of the HDI.

valuation) at $y_i = \delta$. Both before and after $y_i = \delta$ the slope is less steep. The value of β determines how sharply the social marginal valuation changes around the point of inflexion. This is illustrated in figure 5, where the poverty line is 0.5. The marginal valuation at the poverty line is $(1 + \lambda/2)$, independent of β. All the curves relating to the new measure in figure 5 pass through this point since λ has a common value of 4. (To ease comparison, the curves for the Kolm and the constant elasticity measures are rescaled to go through this point as well.) With $\beta = 12$, the function is a "smoothed" version of the poverty gap, giving some additional weight to people above the poverty line, but the weight falls rapidly away: at the world median, the social marginal valuation is indistinguishable from that with the poverty gap. With $\beta = 4$, on the other hand, less significance is attached to the poverty line. Those with incomes up to three times the poverty line receive a perceptible additional weight, which is similar to that assigned to them by the (rescaled) constant elasticity index I with $\varepsilon = 1$; for higher incomes, the social marginal valuation stabilizes at 1, the lower bound for the new measure, while it keeps declining for the index I. With $\beta = 4$, those below the poverty line get lower weight, relative not only to the poverty gap version and the function with higher values of β but also to the constant elasticity measure.

Toward an Inequality Measure

If δ is no longer regarded as the poverty line, the new measure can represent the views of people who are concerned with overall inequality. With $\delta = 0$

FIGURE 5. Social Marginal Valuation of Income with New Measure, Poverty Line Version



Source: Authors' elaboration.

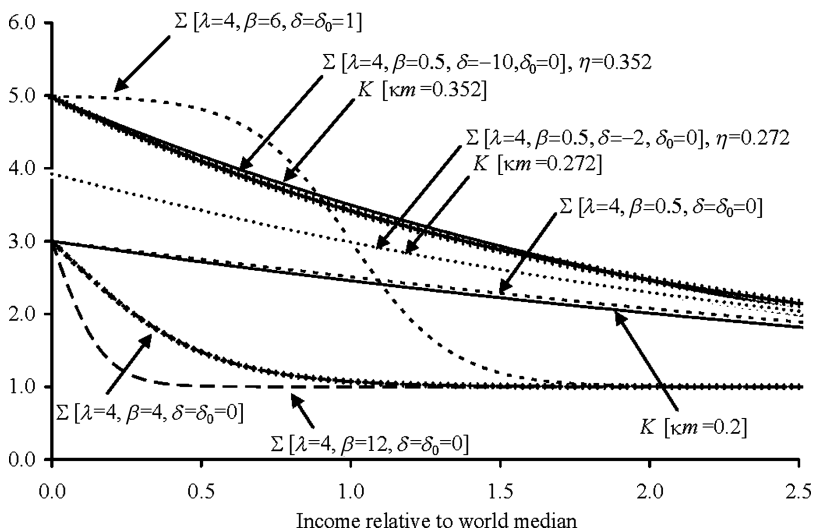
(and $\delta_0 = 0$), there is no interior point of inflexion, and (with $\lambda = 4$) the social marginal valuation of income has the form shown in figure 6 by the three curves starting from the same value 3 (the social valuation at zero income is $(1 + \lambda/2)$). The three curves are based on different values of β and illustrate different speeds of approach to the limiting value of 1: the greater the value of β, the more rapidly the weight attributed to higher incomes converges to 1. For the range of incomes shown in figure 6, the curve with the lowest value of β (0.5) has some similarity with the Kolm index with an elasticity of 0.2 at the median (after rescaling so that it also starts from 3 when income is nil).

There is indeed a close relationship with the Kolm index. If δ_0 is held to zero, but δ is allowed to tend to minus infinity, the individual social marginal valuation of income becomes $(1 + \lambda\ e^{-\beta y_i})$, which for large λ approaches the Kolm form with β corresponding to κ in equation (2). As δ tends to minus infinity, equation (3) becomes:

$$\Sigma_{\delta \to -\infty, \delta_0 = 0} = \frac{1}{n}\sum_i y_i - \frac{\lambda}{\beta}\frac{1}{n}\sum_i e^{-\beta y_i} \tag{3b}$$

The separation of δ_0 and δ is introduced to allow this limit to be taken. (The limit may be seen from equation (3) by regarding $\exp(\beta\delta)$ as the denominator and applying L'Hôpital's Rule.) Arrival at a form similar to the Kolm index underlines the absolute rather than relative nature of the generalization, but the difference remaining where λ is finite should be stressed: as income goes to

FIGURE 6. Social Marginal Valuation of Income with New Measure, Inequality Version



Source: Authors' elaboration.

infinity, the social marginal valuation of income goes to zero in the case of the Kolm index while it approaches one with specification (3b). Thus, the social evaluation of an extra dollar accruing to the poorest person relative to an extra dollar accruing to the richest person approaches infinity with the Kolm index, while it is at most $(1 + \lambda)$ with this formulation.

The similarity with the Kolm index is illustrated by the curves in the upper part of figure 6. The two curves virtually coincide within the shown income range, with the continuous curves that correspond to the Kolm indices having the same elasticity at the median, rescaled to start from the same value at zero income. (The two curves would, however, depart from their Kolm counterparts at some higher level of income.)

Slow, Quick, Slow

So far, δ has been allowed to vary downward. If δ is allowed to be positive, a measure is obtained with the "slow, quick, slow" property. This is illustrated in figure 6 by setting $\delta = 1$ (once again equal to δ_0) and $\beta = 6$. The key element is the behavior of the second derivative of the social welfare function. From equation (5), it may be seen that the third derivative of the social welfare function explored here is first negative (for $y_i < \delta$) and then positive (for $y_i > \delta$). The literature on transfer "sensitivity" (Atkinson 1973; Kolm 1976; Davies and Hoy 1985) shows that the assumption that the third derivative is positive is equivalent to the "principle of diminishing transfers," or third order stochastic dominance.

Suppose that the two people in the earlier leaky bucket experiment (person 1 poorer than person 2) are now each joined by a friend with income h above theirs, and that \$1 is simultaneously transferred from person 1's friend to person 1 and \$1 from person 2 to person 2's friend. In other words, there are two mean-preserving transfers of the same size, but in opposite directions. Then, the principle of diminishing transfers means that more weight is attached to the transfer affecting the poorer person and that social welfare increases (see Shorrocks and Foster 1987, for a more general treatment). With the social welfare function explored here, this principle is assumed to apply at middle and higher incomes, above the point of inflexion δ. In contrast, over the initial range of incomes, up to δ, there is increasing sensitivity to transfers.

As before, the parameters can be calibrated by considering the elasticity of the social marginal valuation. As shown by equation (6), below, this varies with income. With $\delta = \delta_0$, the elasticity at the point of inflexion $y_i = \delta$ is equal to $\lambda\beta\delta/[2(2 + \lambda)]$. In figure 6, $\lambda = 4$ and $\beta = 6$, so that the elasticity at $\delta = 1$ is 2. In the example below, a Gini-like measure is constructed by taking the point of inflexion at twice the world median income, $\delta = 2$, and setting $\beta = 3$; this leaves the elasticity at the point of inflexion unchanged at 2, but gives a much lower elasticity of 0.11 at the median income. With higher values of δ, the flatter, initial section applies over a wider range. Indeed, by letting δ go to infinity, distributional indifference becomes a limiting case.

On the Interpretation of the New Measure

The new measure explored here has been constructed to embody a desired pattern of change in the social marginal valuation of income. But how is the new measure to be interpreted? Its interpretation may be aided by re-arranging the expression for social welfare. By adding and subtracting from equation (3) the term $\frac{\lambda}{\beta}e^{\beta(\delta_0-\delta)}\ln\left[1+e^{\beta(\delta-\mu)}\right]$, social welfare can be treated as being made up of two components:

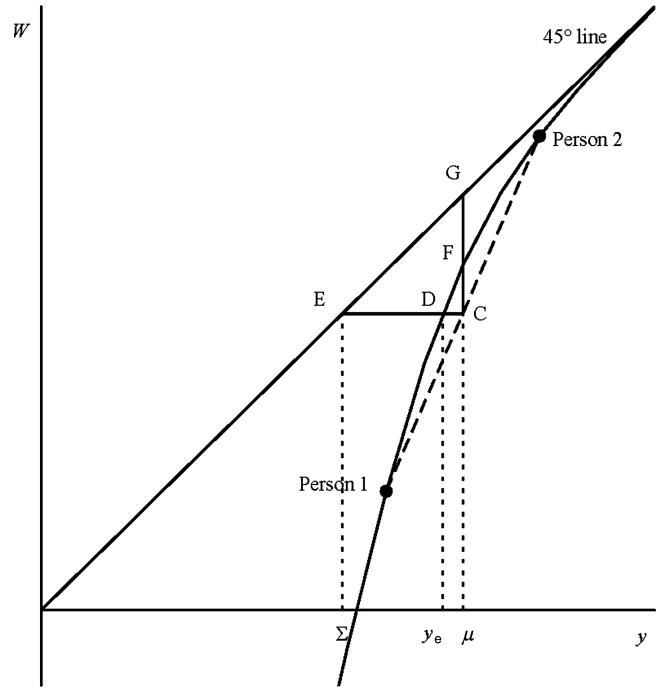
$$\Sigma = \Sigma(\mu) - \sigma = \Sigma(\mu) - \left\{\frac{\lambda}{\beta}e^{\beta(\delta_0-\delta)}\frac{1}{n}\sum_i \ln\left[\frac{1+e^{\beta(\delta-y_i)}}{1+e^{\beta(\delta-\mu)}}\right]\right\} \tag{6}$$

The first term on the right side of equation (6), $\Sigma(\mu)$, is the level of social welfare attained if everyone has an income equal to the mean, μ. In general, this level of social welfare is less than μ, although it approaches μ as the mean tends to infinity. (With the poverty gap, it is equal to μ once the mean passes the poverty line.) This reflects the fact that it is a welfare measure and that there are diminishing returns in the transformation of income into welfare. The second term, denoted by σ, represents the costs of income differences. The term reduces to zero if all incomes are equal to the mean.

How this new measure differs from earlier approaches can be illustrated in the simple example in figure 7. If there are two people with incomes (measured on the horizontal axis) as shown and mean μ, the achieved level of social welfare is given by point C (the midpoint). Welfare is measured on the vertical axis. The I and K measures proceed by calculating the equally distributed equivalent income, y_e (obtained by reading across horizontally from C to D), and the cost of inequality is the loss CD. Unlike the I and K social welfare functions, however, the new measure has the same dimension as income. This implies that the level of welfare, Σ, can be directly compared with the mean income, μ, and that there is no need to introduce the equally distributed equivalent income. GC, the overall difference between μ and Σ, is made up of two components, GF and FC: GF reflects the diminishing returns in the transformation of income into welfare and shrinks as income grows; FC measures the cost of inequality, the second term in equation (6).

In the case of the poverty gap, the curve in figure 7 becomes a kinked line, coinciding with the 45° line from the level δ of the poverty income onward. The distance GC is λ times the aggregate poverty gap per person. This gap consists, potentially, of two components, either of which may be zero. Where the mean income is above the poverty line, poverty is entirely due to the unequal distribution of income. If the mean income is below the poverty line, there can be both aggregate poverty (corresponding to the difference between μ and $\Sigma(\mu)$) and distributional poverty. Aggregate poverty can remain even if incomes are equalized. Indeed, if everyone has an income below the poverty line, then the component FC disappears even though income differences remain (since

FIGURE 7. Interpreting the Function W: A Two Person Example



Source: Authors' elaboration.

the poverty gap is unaffected by transfers of income among people below the poverty line). The problem of poverty can therefore be seen as a problem of distribution or a problem of the overall level of income.

These observations highlight the crucial role of δ when it is seen as the poverty line. The parameter δ (and δ_0) could be defined as a fraction of mean income, a purely relative approach that is not explored here. On an absolute approach, δ (and δ_0) is independent of mean income, but, as noted, this does not imply that it should be kept constant over a long period. Where the underlying concern relates to a more fundamental space, such as the achievement of a level of functionings, the necessary level of income may be changing as a result of economic and social progress. This issue is taken up again in the next section.

In the general case, the concern is not with the GF component but only with the FC component, the costs of inequality. It is instructive to see how the new measure departs from Kolm's absolute approach on the costs of inequality. The Kolm index K measures the costs of inequality as the absolute difference between the mean and the equally distributed equivalent income: $K = \mu - y_e^K$. Equation (6) expresses the cost of inequality as the difference between the social welfare at the mean, $\Sigma(\mu)$, and the social welfare for the actual income

distribution, Σ. As by definition social welfare at the equally distributed equivalent income equals Σ, the term $\Sigma(\mu) - \Sigma(y_e)$ rather than $\mu - y_e$ is being taken as the cost of the unequal distribution of income. For the Kolm-like measure defined by (3b), this term equals $e^{-\beta\mu}(e^{\beta K} - 1)\lambda/\beta$, where K is the Kolm index with $\kappa = \beta$. For given mean income, the two measures generate the same ranking, but the cost of inequality defined here is smoothed out by a rise in mean income. Raising all incomes by \$1 leaves the Kolm index unchanged by construction but reduces the costs of inequality with the Kolm-like measure, and it does so at a decreasing rate as mean income rises: the richer the economy, the less an extra \$1 is worth.

The new inequality measure σ is decomposable by population subgroups (see Cowell 1980; Shorrocks 1980):

$$\sigma = \sum_j w_j \sigma_j + \frac{\lambda}{\beta} e^{\beta(\delta_0 - \delta)} \sum_j w_j \ln\left[\frac{1 + e^{\beta(\delta - \mu_j)}}{1 + e^{\beta(\delta - \mu)}}\right], \tag{7}$$

where subscript j refers to the J mutually exclusive subgroups of the population, and w_j is the population share of subgroup j. The first term on the right side of equation (7) is the population-weighted average of within-group inequalities; the second term is between-group inequality, calculated after attributing the group mean income to each member in a group. For the poverty gap measure (equation 3a), the decomposition is:

$$\sigma_{\beta\to\infty,\delta_0=\delta} = \sum_j w_j \sigma_j + \lambda \sum_j w_j \max[0, (\delta - \mu_j)] - \lambda \max[0, (\delta - \mu)] \tag{8a}$$

When the overall mean is above the poverty line, the between-group component is λ times the weighted average of the aggregate poverty indicator, that is, the difference $(\delta - \mu_j)$ if positive; when the overall mean falls short of the poverty line, the aggregate poverty indicator must be subtracted from this sum.

IV. The New Approach Applied to World Inequality

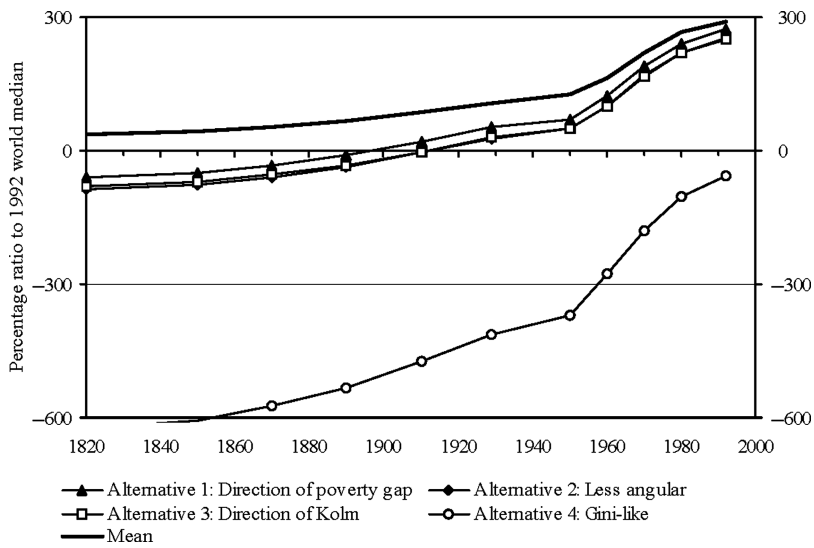
The alternative measure suggested above is now applied to the evidence on world inequality. The pattern of the social marginal valuation of income is illustrated in the final four columns of table 1 (alternatives 1–4), where the maximum acceptable leakage is taken to be 80 percent ($\lambda = 4$). The first two alternatives take δ as the poverty line (assumed to be half world median income in 1992), and set δ_0 equal to δ. With alternative 1, β has a high value, reflecting concern about poverty but not about inequality (in the direction of the poverty gap version). The social marginal valuation of income falls sharply once the poverty line is reached and is essentially constant above world median income. Alternative 2, with a smaller value of β (=4), corresponds to a less angular position. Below the poverty line, the social marginal valuation is lower

than with alternative 1, but it crosses at the poverty line. For incomes up to the world median, the weight attached to marginal income is at least 40 percent higher than that attached to marginal income in the United States.

In contrast, alternatives 3 and 4 break the link between δ and the poverty line and lean toward measures of inequality. Letting $\delta = -4$ (and $\delta_0 = 0$) moves in the direction of the Kolm index, K. Alternative 4 goes in the opposite direction, setting $\delta = \delta_0 = 2$ with $\beta = 3$, which generates a Gini-like inequality measure (but with additive separability and decomposability by population subgroups). The social marginal valuation first falls slowly and then more quickly, exhibiting increasing and then decreasing sensitivity to transfers. The difference in transfer sensitivity is particularly important when considering the world scale of incomes. Individual countries may lie largely within the increasing or the decreasing phase (see table 1). Even so, alternative 4 is consistent with substantial redistribution within the United States: the social marginal valuation for decile group 1 is more than three times that for the U.S. median.

These four alternative measures are applied to the world income estimates of the Bourguignon and Morrisson database. Figure 8 shows the evolution of world social welfare from 1820 to 1992, where welfare has the dimension of per capita income and is expressed as a percentage of world median income in 1992. As noted earlier, welfare may be negative, as was the case for all four alternatives until the beginning of the 20th century. For the poverty line measures, it is scarcely surprising that the earlier values are so low since a contemporary (1992) standard is being applied, but the inequality measures are

FIGURE 8. Evolution of World Social Welfare, Alternatives 1–4, 1820–1992



Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

also absolute in the sense described in the previous section. This applies not only to alternative 3, approaching the Kolm index, but also to the Gini-like alternative 4. Indeed the Gini-like measure is initially off the scale.

All four measures indicate a considerable improvement over the period, driven by the growth of mean income (the thick top line). However, while the upward tendency was similar, the rates of increase in social welfare differed from those in mean income. For example, the (absolute) annual increase in mean income between 1980 and 1992 was double that between 1890 and 1910, but the rise in social welfare using the Gini-like index was only a quarter higher. Distributionally adjusted income, as with the new social welfare measure, may give rather different pictures of different historical periods.

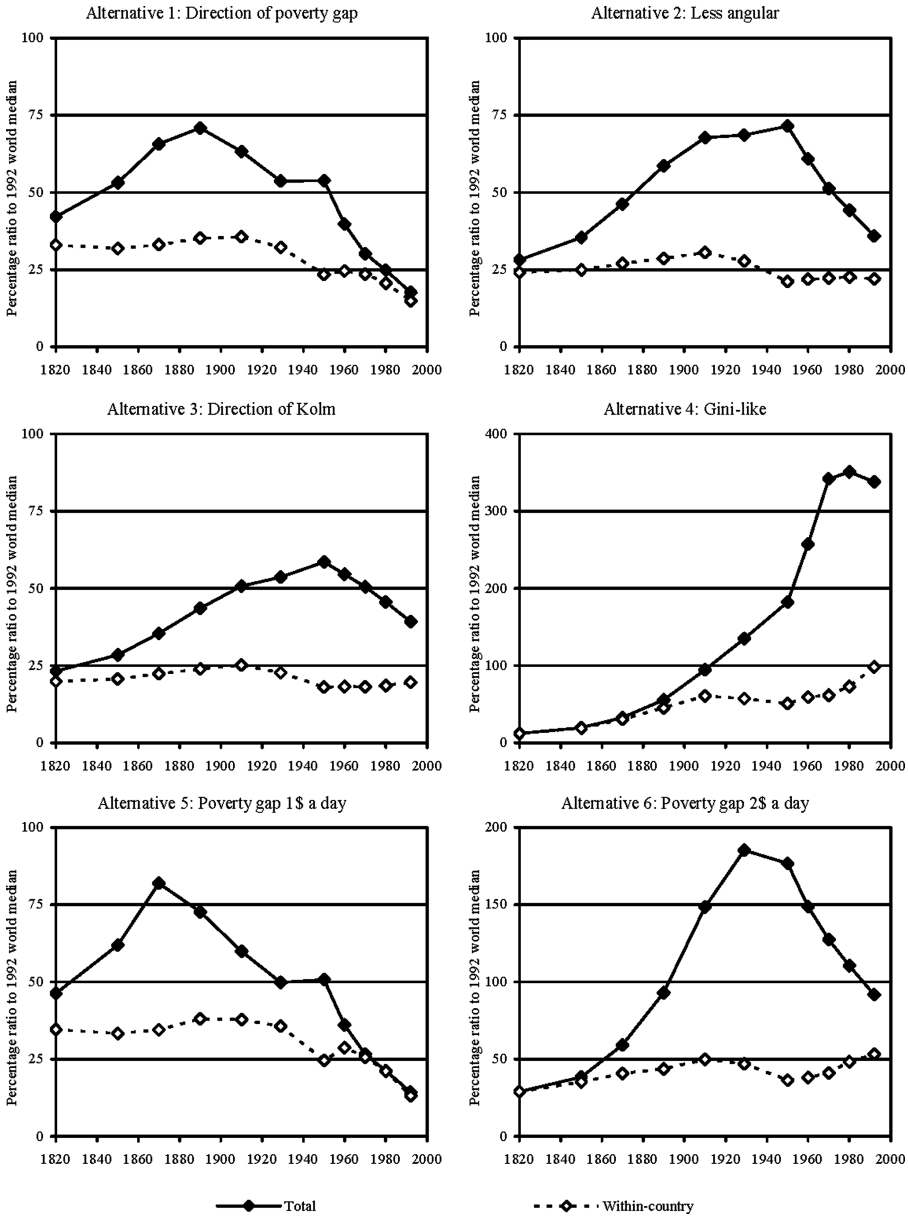
The absolute cost of inequality, σ, is given in figure 9, again expressed as a percentage of the 1992 world median, so that a value of 100 corresponds to a cost of U.S. $1,712 per person. Figure 9 has a panel for each of the four alternatives and one for the poverty gap measures defined in (3a) with δ set at 0.5 and one with it set at 1 (alternatives 5 and 6). These last two values roughly correspond to the $1 a day and $2 a day poverty lines, as defined by Bourguignon and Morrisson (2002). With alternative 1, whose parameters lead in the direction of the poverty gap, the cost due to inequality rises until 1890 and then declines, accelerating after the Second World War. The time path with the less angular version in alternative 2 and the Kolm-like version in alternative 3 also have an inverse-U shape, but the peak cost of inequality is reached much later, in 1950. In contrast, with the Gini-like measure, alternative 4, the cost due to inequality increases steadily, then very rapidly between 1950 and 1970 before reaching a peak in 1980. (Recall that the Gini-like measure is not the same as the Gini coefficient: the social marginal valuation of income received by one person does not depend on what is happening elsewhere in the distribution.)

Thus, the two inequality versions of the new measure, the Kolm-like and the Gini-like, move in opposite directions after 1950. If the two poverty gap measures represented by alternatives 5 and 6 are compared with alternatives 1 and 2, all are found to share the same inverse-U shape, in particular the steep downward trend after 1950, though they differ in the time of the turning point and in the size of the change. With the $1 a day poverty line, the turning point is 1870; with the $2 a day line, it is in the 20th century.

For all six alternatives, within-country inequality, the population-weighted average of inequality calculated within countries or groups of countries, is far more stable than the total, suggesting that the secular variation in the total cost is driven largely by changing income differences across countries. Notice, however, how a significant rise of within-country inequality from 1970 to 1992 offsets the international convergence in mean incomes with alternative 4. The less angular poverty measure and the Kolm-like inequality measure level off, and the $2 a day poverty gap and the Gini-like inequality measure show a rise after 1950.

These results assume that the cost is measured in absolute terms. Figure 10 shows that some differences arise if the cost is measured as a proportion of

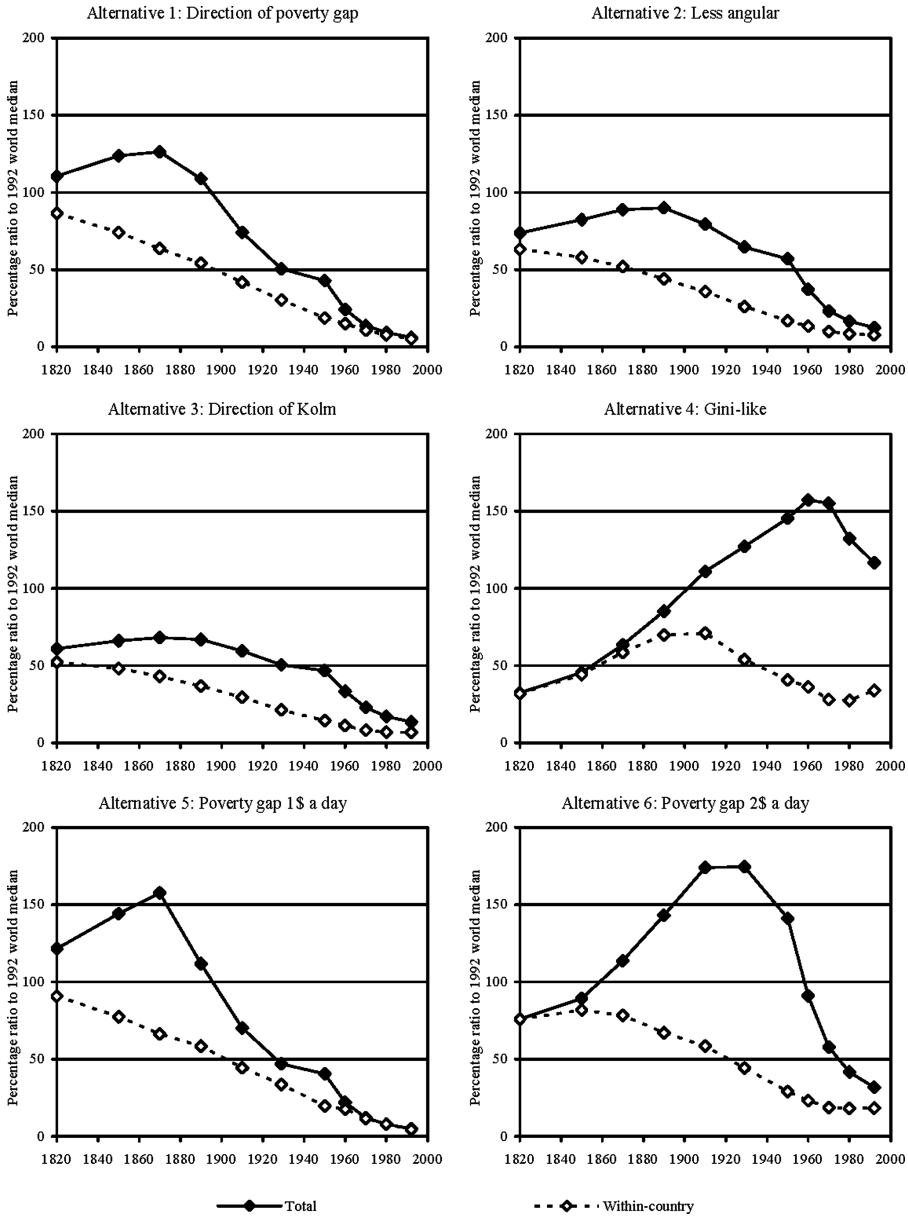
FIGURE 9. Evolution of the Absolute Cost of World Inequality, Alternatives 1–6, 1820–1992



Note: Within-country inequality is the population-weighted average of inequality calculated within the 33 countries or groups of countries included in the Bourguignon and Morrisson (2002) database.

Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

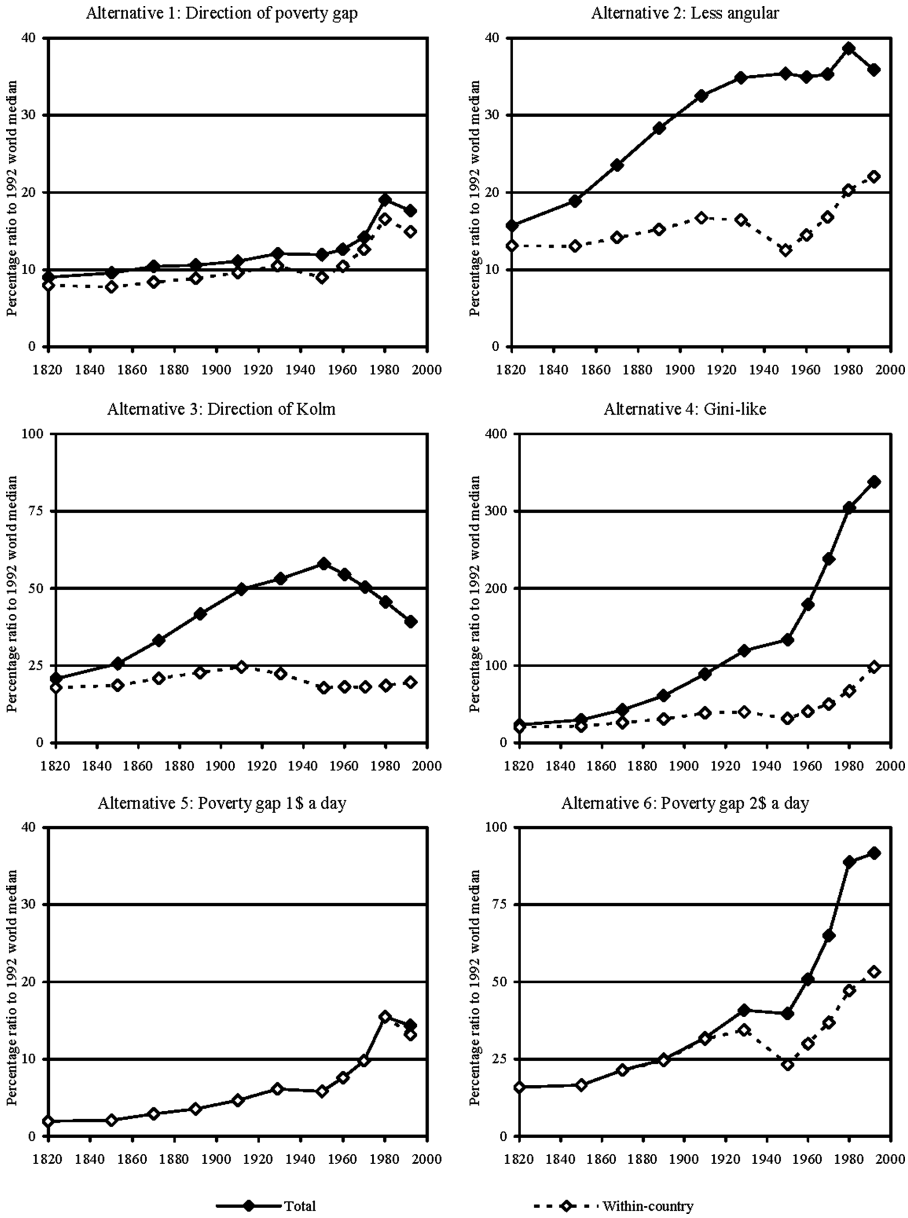
FIGURE 10. Evolution of the Relative Cost of World Inequality, Alternatives 1–6, 1820–1992



Note: Within-country inequality is the population-weighted average of inequality calculated within the 33 countries or groups of countries included in the Bourguignon and Morrisson (2002) database.

Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

FIGURE 11. Evolution of the Absolute Cost of World Inequality, Alternatives 1–6, Time-variable δ, 1820–1992



Note: Within-country inequality is the population-weighted average of inequality calculated within the 33 countries or groups of countries included in the Bourguignon and Morrisson (2002) database.

Source: Authors' elaboration on the Bourguignon and Morrisson (2002) database.

mean income. The peaks with the less angular version (alternative 2) and with the Kolm-like measure (alternative 3) come earlier—toward the end of the 19th century. The relative cost due to inequality with the Gini-like measure also peaks earlier, in 1960, and then falls thereafter. But the differences are not nearly as striking as those found for the standard measures presented at the opening of this article.

These evaluations are based on a value of δ that is kept constant across the period 1820–1992. For the two poverty lines, δ is assumed to be half the world median income in 1992. This value sets a very high standard: in 1820 only Western European countries, the United States, and Argentina-Chile enjoyed a mean income greater than δ. It is reasonable to wonder how the results would change if this extreme absolutist hypothesis were relaxed by varying the poverty standard over time in stepwith economic and social progress.

Figure 11 shows the consequences of recomputing the measure retaining the value of δ for 1992 but assuming that it grew over time along with median world income. Doing so amounts to applying the values of δ from table 1 to the median income in each year rather than to the median income in 1992, taking the increase in median income as a reference point. It should be stressed that this does *not* assume that the poverty line is proportionate to median, or mean, incomes. The (externally derived) time variation in δ may involve a faster or slower rate of growth. All other parameters are kept unchanged. As shown in figure 11, under a time-varying δ, the secular pattern of world income inequality looks considerably different from the one reported earlier for all poverty-type measures (alternatives 1 and 2) and alternatives 5 and 6: the inverse-U shape turns into a steadily ascending trend, which flattens out only after 1980. The impact is barely visible on the two inequality-type measures, except for the upward trend of the Gini-like measure, now continuing even after 1970. Assuming a time-varying δ also affects the within-country component, which tends to account for a much larger share of the overall inequality: in alternatives 1 and 5, it almost wipes out the between-country component.

To sum up, contrary to what is suggested by the earlier analysis using the standard inequality indices, the conclusions depend very much on distributional judgments.

V. Conclusions

The effects of globalization on world income inequality have been much debated in recent years. In the literature, as noted by Anand and Segal (2008, p. 61), "no consensus emerges concerning the direction of change in global inequality in the last 20–30 years." Some commentators have stressed the impressive growth performance of emerging economies such as China, India, and other countries in Southeast Asia and have concluded that world inequality

and poverty must have decreased. Others have countered that these impressive rates of growth have not yet translated into absolute increases in income comparable with those of developed economies, given the very different levels of GDP per capita. Thus, world income gaps must have risen. This article argued that—before such judgments can be made—the foundations of inequality measurement need to be reexamined. The sheer scale of global income differences means that the tools applied to inequality measurement at the domestic level cannot simply be carried over. In the discourse about global justice, both poverty and inequality and their interrelation have to be considered, as do the different meanings of "absolute" and "relative."

Differences of view about the evolution of world inequality and poverty stem in part from differences in how to measure them. In seeking to provide a framework for considering the cost of world inequality and poverty that encompasses different types of concern, this article adopts a welfarist approach (without endorsing it as the only possible approach). Its first findings, in section I, suggested that the differences in conclusions about changes in world inequality could be attributed largely to how the cost is expressed. However, section II argued that existing measures of inequality impose too tight constraints on how social marginal valuation varies with income and provide no ready means to integrate the analysis of poverty and inequality.

To encompass the global income differences and allow for concerns about poverty as well as inequality, section 3 explored a new parameterized measure of global social welfare. This measure has, in a sense explained in the article, an absolute structure (and it would be interesting to consider the parallel, relative measure), but it is sufficiently flexible to include different value systems and to incorporate a poverty line. Including several approaches within a single measure helps not only to better understand their interrelation but also to obtain measures that blend different concerns. People differ, for example, in the relative importance they attach to poverty and inequality. This difference appears fundamental, but it can be embedded within the new measure explored here. Letting one of the key parameters increase allows the measure to take on a poverty gap form, whereas lower values permit a less angular version of the poverty gap, tapering the measure for the near-poor. If other parameters are allowed to vary, more general concerns about inequality can be introduced. These may follow the pattern of standard welfare-based measures, with declining sensitivity to transfers with movement up the income scale. Or they may exhibit increasing and then decreasing sensitivity to transfers, mimicking the Gini coefficient but with the property of additive separability (and subgroup decomposability).

The new measure can accommodate a constant poverty line or one varying over time with economic and social development, an alternative with considerable consequences for interpreting the evolution of the world income distribution. Finally, the overall weight attached to distributional issues can be varied according to individual interests. For example, one person may be

concerned about poverty but not attach much weight to this consideration, relative to total income. Another person might feel that, in the context of world poverty, little weight should be attached to additional income received by those at the top of the distribution. Stated more pragmatically, the new measure can exhibit a willingness to redistribute within rich countries without magnifying to an implausible degree the willingness to make transfers across the whole spectrum of world incomes.

References

Amiel, Yoram, John Creedy, and Stan Hurn. 1999. "Measuring Attitudes Towards Inequality." *Scandinavian Journal of Economics* 101(1):83–96.

Anand, Sudhir, and Paul Segal. 2008. "What Do We Know about Global Income Inequality?" *Journal of Economic Literature* 46(1):57–94.

Anand, Sudhir, and Amartya K. Sen. 2000. "The Income Component of the Human Development Index." *Journal of Human Development* 1(1):83–106.

Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2(3):244–63.

———. 1973. "More on the Measurement of Inequality." Massachusetts Institute of Technology, Cambridge, Mass. Published in 2008 in *Journal of Economic Inequality* 6(3):277–83.

———. 1990. "Public Economics and the Economic Public." *European Economic Review* 34(2–3):225–48.

Atkinson, Anthony B., and François Bourguignon. 1999. "Poverty and Inclusion from a World Perspective." In *Governance, Equity, and Global Markets. Proceedings of the Annual Bank Conference on Development Economics in Europe, June 21–23 1999*, Conseil d'Analyse Économique and World Bank. Paris: La Documentation Française.

Atkinson, Anthony B., and Andrea Brandolini. 2001. "Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries as a Case Study." *Journal of Economic Literature* 39(3):771–99.

———. 2004. "Global World Income Inequality: Absolute, Relative or Intermediate?" Paper presented at the 28th General Conference of the International Association for Research in Income and Wealth, 22–28 August, Cork, Ireland, and at the 4th International Conference on the Capability Approach: Enhancing Human Security 5–7 September, Pavia, Italy (www.iariw.org/papers/2004/brand.pdf).

Bhagwati, Jagdish. 2004. *In Defense of Globalization*. Oxford: Oxford University Press.

Bhalla, Surjit. 2002. *Imagine There's No Country. Poverty, Inequality, and Growth in the Era of Globalization*. Washington, DC: Institute for International Economics.

Blackorby, Charles, and David Donaldson. 1978. "Measures of Relative Equality and Their Meaning in Terms of Social Welfare." *Journal of Economic Theory* 18(1):59–80.

———. 1980. "A Theoretical Treatment of Indices of Absolute Inequality." *International Economic Review* 21(1):107–36.

Bossert, Walter, and Andreas Pfingsten. 1990. "Intermediate Inequality: Concepts, Indices, and Welfare Implications." *Mathematical Social Sciences* 19(2):117–34.

Bourguignon, François, Victoria Levin, and David Rosenblatt. 2006. "Global Redistribution of Income." Policy Research Working Paper 3961. World Bank, Washington, DC.

Bourguignon, François, and Christian Morrisson. 2002. "Inequality Among World Citizens: 1820–1992." *American Economic Review* 92(4):727–44. (www.delta.ens.fr/XIX).

Chakravarty, Satya R., Ravi Kanbur, and Diganta Mukherjee. 2006. "Population Growth and Poverty Measurement." *Social Choice and Welfare* 26(3):471–83.

Chotikapanich, Duangkamon, D.S. Prasada Rao, and Rebecca Valenzuela. 1997. "Global and Regional Inequality in the Distribution of Income: Estimation with Limited and Incomplete Data." *Empirical Economics* 22(4):533–46.

Christiansen, Vidar, and Eilev S. Jansen. 1978. "Implicit Social Preferences in the Norwegian System of Indirect Taxation." *Journal of Public Economics* 10(2):217–45.

Cowell, Frank A. 1980. "On the Structure of Additive Inequality Measures." *Review of Economic Studies* 47(3):521–31.

Davies, James B., and Michael Hoy. 1985. "Comparing Income Distributions under Aversion to Downside Inequality." CSIER Working Paper. Ontario, Canada: University of Western Ontario, Centre for the Study of International Economic Relations.

Deaton, Angus. 2005. "Measuring Poverty in a Growing World (Or Measuring Growth in a Poor World)." *Review of Economics and Statistics* 87(1):1–19.

Del Río, Coral, and Javier Ruiz-Castillo. 2001. "Intermediate Inequality and Welfare: The Case of Spain, 1980–81 to 1990–91." *Review of Income and Wealth* 47(2):221–37.

Dollar, David, and Aart Kraay. 2002. "Growth is Good for the Poor." *Journal of Economic Growth* 7(3):195–225.

Dowrick, Steve, and Muhammad Akmal. 2005. "Contradictory Trends in Global Income Inequality: A Tale of Two Biases." *Review of Income and Wealth* 51(2):201–29.

Feldstein, Martin S. 2005. "Rethinking Social Insurance." *American Economic Review* 95(1):1–24.

Fields, Gary. 2006. "Should Poverty and Inequality Measures Be Combined?" In Alain de Janvry, and Ravi Kanbur, eds., *Poverty, Inequality, and Development. Essays in Honor of Erik Thorbecke*. New York: Springer.

Firebaugh, Glenn. 2003. *The New Geography of Global Income Inequality*. Cambridge, Mass.: Harvard University Press.

Foster, James E. 1998. "Absolute versus Relative Poverty." *American Economic Review Papers & Proceedings* 88(2):335–41.

Gottschalk, Peter, and Timothy M. Smeeding. 1997. "Cross-National Comparisons of Earnings and Income Inequality." *Journal of Economic Literature* 35(2):633–87.

Gruen, Carola, and Stephan Klasen. 2008. "Growth, Inequality, and Welfare: Comparisons across Space and Time." *Oxford Economic Papers* 60(2):212–36.

Jones, Arthur F., Daniel H. Weinberg, and U.S. Census Bureau. 2000. *The Changing Shape of the Nation's Income Distribution, 1947–1998*. Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration, and U.S. Census Bureau.

Kanbur, Ravi. 2008. "Globalization, Growth, and Distribution: Framing the Questions." Working Paper 5. Commission on Growth and Development, Washington, DC.

Kolm, Serge-Christophe. 1969. "The Optimal Production of Social Justice." In Julius Margolis, and Henri Guitton, eds., *Public Economics. An Analysis of Public Production and Consumption and Their Relations to the Private Sectors*. London: Macmillan.

———. 1976. "Unequal Inequalities I." *Journal of Economic Theory* 12(3):416–42.

Little, Ian M. D., and James A. Mirrlees. 1974. *Project Appraisal and Planning for Developing Countries*. London: Heinemann.

Livi Bacci, Massimo. 2001. "Remarks in the Roundtable: 'Lotta alla povertà nell'economia globale: un confronto sul. *World Development Report 2000/2001, Attacking Poverty*, della Banca Mondiale.'" *QA - La questione agraria* 2:113–16.

Maddison, Angus. 1995. *Monitoring the World Economy*. Paris: Organisation for Economic Co-operation and Development.

Milanovic, Branko. 2002. "True World Income Distribution, 1988 and 1993: First Calculation based on Household Surveys Alone." *Economic Journal* 112(476):51–92.

Mirrlees, James A. 1978. "Social Benefit–Cost Analysis and the Distribution of Income." *World Development* 6(2):131–38.

Okun, Arthur M. 1975. *Equality and Efficiency*. Washington, DC: Brookings Institution.

Ravallion, Martin. 2004. "Competing Concepts of Inequality in the Globalization Debate." In Susan M. Collins, and Carol Graham, eds., *Brookings Trade Forum 2004. Globalization, Poverty, and Inequality*. Washington, DC: Brookings Institution Press.

Sala-i-Martin, Xavier. 2006. "The World Distribution of Income: Falling Poverty and . . . Convergence, Period." *Quarterly Journal of Economics* 121(2):351–97.

Sawyer, Malcolm C. 1976. "Income Distribution in OECD Countries." In *OECD Economic Outlook. Occasional Studies*. Paris: Organisation for Economic Co-operation and Development.

Schultz, T. Paul. 1998. "Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why." *Journal of Population Economics* 11(3):307–44.

Sen, Amartya K. 1976. "Real National Income." *Review of Economic Studies* 43(1):19–39.

———. 1983. "Poor, Relatively Speaking." *Oxford Economic Papers* 35(2):153–69.

———. 1992. *Inequality Reexamined*. Oxford: Clarendon Press.

Shorrocks, Anthony F. 1980. "The Class of Additively Decomposable Inequality Measures." *Econometrica* 48(3):613–25.

Shorrocks, Anthony F., and James E. Foster. 1987. "Transfer Sensitive Inequality Measures." *Review of Economic Studies* 54(3):485–97.

Stern, Nicholas H. 1977. "Welfare Weights and the Elasticity of the Marginal Valuation of Income." In Michael J. Artis and A. Robert Nobay, eds., *Studies in Modern Economic Analysis: The Proceedings of the Association of University Teachers of Economics, Edinburgh, 1976*. Oxford: Blackwell.

Svedberg, Peter. 2004. "World Income Distribution: Which Way?" *Journal of Development Studies* 40(5):1–32.

United Nations Development Programme. 1991. *Human Development Report 1991*. Oxford: Oxford University Press.

World Bank. 2008. *Global Purchasing Power Parities and Real Expenditures*. 2005 International Comparison Program. World Bank, Washington, DC.

Zheng, Buhong. 2007. "Unit-Consistent Decomposable Income Measures." *Economica* 74(293): 97–111.

Do Autocratic States Trade Less?

Toke S. Aidt and Martin Gassebner

Does the political regime of a country influence its involvement in international trade? A theoretical model that predicts that autocracies trade less than democracies is developed, and the predictions of the model are tested empirically using a panel of more than 130 countries for 1962–2000. In contrast to the existing literature, data on the regime type of individual countries are used rather than information about the congruence of the regime type of pairs of trading countries. In line with the model, autocracies are found to import substantially less than democracies, even after controlling for official trade policies. This finding is very stable and does not depend on a particular setup or estimation technique. JEL codes: F13, F14, O24, P45, P51

Is there a systematic relationship between economic and political liberalization? Does a country's political regime systematically affect its involvement in international trade? The first question has received much attention recently, with studies of the determinants of democracy (for example, Barro 1999; Acemoglu and others 2008) and economic freedom (for example, Boockmann and Dreher 2003; Dreher and Rupprecht 2007) as well as studies of the relationship between democracy and economic freedom (for example, Sturm and de Haan 2003; Giavazzi and Tabellini 2005).[1] The second, more specific question is much less well researched, and this article aims to provide new answers.

Knowledge of how political regimes influence international trade comes primarily from the political science literature. Two seminal works find that democracy encourages trade. Mansfield, Milner, and Rosendorff (2000) stress the congruence between the political regime of pairs of trading countries. They show that pairs of democratic countries trade more than pairs consisting of a

Toke S. Aidt (toke.aidt@econ.cam.ac.uk) is a senior lecturer at the Faculty of Economics, Jesus College, University of Cambridge. Martin Gassebner (gassebner@kof.ethz.ch, corresponding author) is a researcher at the KOF Swiss Economic Institute, ETH Zurich. The authors thank Axel Dreher, Heiner Mikosch, Jan-Egbert Sturm, the journal editor, three anonymous referees, and participants in a seminar at the University of Wijgmegen, the Silvaplana Workshop of Political Economy, the European Trade Study Group as well as the Workshop International Relations Goettingen for helpful comments on earlier drafts.

1. Convergence and contagion trends of the two phenomena have recently been studied by Nieswiadomy and Strazicich (2004) and Gassebner, Gaston, and Lamla (forthcoming).

THE WORLD BANK ECONOMIC REVIEW, VOL. 24, NO. 1, pp. 38–76 doi:10.1093/wber/lhp022
Advance Access Publication January 15, 2010

democracy and an autocracy.[2] Milner and Kubota (2005) study the relationship between political regime type and trade policy in a sample of developing countries and show that democratic political institutions are associated with liberal trade policy.

This article adds to the literature in two related ways. First, it argues that the theoretical foundations of the previous studies (discussed in more detail in section I) overlook the importance of regime differences in political accountability and external monitoring and how these differences induce societies to build more or less effective incentives for customs officials. The lack of political accountability makes it possible for political leaders to extract rents by imposing restrictions on international trade. Moreover, within a hierarchical government structure, the lack of effective external monitoring (for example, due, to the absence of free media) makes it less likely that political leaders choose to strengthen institutions to reduce trade-distorting red tape and other unofficial trade barriers. This effect is strengthened when autocratic leaders share the rents generated by red tape through a structured, hierarchical chain of rent sharing. The theoretical contribution of this article is thus the prediction that autocracies—societies with weak political accountability and underdeveloped external monitoring and rent-sharing chains—trade less with the rest of the world than do democracies—societies with strong political accountability and well-developed external monitoring—for two reasons. Democracy limits the scope for rent extraction through trade restrictions. And democracy encourages institutional reforms that reduce bureaucratic inefficiencies (and trade-distorting red tape).

The second way this article adds to the literature is by looking beyond the existing empirical work's focus on the congruence of the political regime of pairs of trading countries, which does not cast light on how the political regime of individual countries affects trade volumes. Do autocracies trade less than democracies? This article answers in the affirmative. Furthermore, it uses a much larger data set, with a longer time horizon and deeper country coverage than previous studies, and estimates dyadic as well as single-country panel models. And its empirical design demonstrates that regime differences in trade policy, while playing a role, cannot fully account for the observed differences in trade flows. Both the observation that autocracies trade less and the observation that they trade less conditional on trade policy are consistent with the theoretical model. But evidence is also presented that countries with a free press and effective political accountability trade more, suggesting that the particular transmission channels of the model are important.

Some researchers have argued that international trade encourages democratization (for example, Li and Reuveny 2003; Rigobon and Rodrik

2. Using a less rich setup, Morrow, Siverson, and Tabares (1998) also find that democracies trade more with each other. And Daumal (2008) finds that federalist systems increase international trade.

2005; López-Córdova and Meisner 2008). This possibility is obviously a concern when estimating the impact of regime type on trade flows: countries that are not involved in international trade could be autocracies for that reason. This issue is addressed partly by allowing for unobserved country- and time-fixed effects in the empirical specification, partly by lagging the empirical indicators used to capture regime differences between countries, and partly by using instrumental variables. In particular, two new instruments are introduced for the political regime type.

By addressing whether systematic differences in trade integration exist between democracies and autocracies, this article contributes to the broader debate about trade and development and the role of "good governance" in fostering economic progress. First, trade integration is often seen as an engine of economic development. That autocracies trade less may therefore be one reason why so many remain underdeveloped. Second, trade integration and the underlying trade distortions are argued to be endogenous outcomes generated by the quality of political institutions and the type of bureaucracy that governments build. The natural policy implication is that better institutions will lead to better policies and less inefficiency, ultimately enhancing trade integration. This highlights the importance of the World Bank's and other international institutions' recent emphasis on a good governance agenda around the world.

The article is organized as follows. Section I presents the model, compares it with existing models, and develops the two hypotheses that govern the empirical investigation. Section II provides evidence on the main assumptions of the model and on its two transmission channels, using partial-regression leverage plots. Section III develops the empirical strategy in detail. Section IV presents the main result. Section V summarizes an extensive set of robustness checks. Section VI introduces the instruments and reports the results from the instrumental variable estimation. Section VII summarizes the findings and offers some concluding remarks.

I. A Model of Political Regimes and Trade Flows

This section presents a model that illuminates two new channels through which regime types can affect trade flows. One channel is the political accountability channel. It is harder for citizens in autocratic countries to hold their rulers accountable, so rulers are more free to use trade taxes to extract rents. The other channel is the external monitoring channel. Lack of a free press, for example, weakens external monitoring in autocratic societies. Rulers can compensate for this by building internal control mechanisms that weed out red tape and other unofficial trade obstructions introduced by the customs services. However, complementarity between external monitoring and internal control mechanisms, modeled as an efficiency wage, implies that autocratic rulers have less incentive to build or strengthen such internal control mechanisms, so the

customs service in an autocracy is freer to introduce and maintain red tape. Both channels suggest that, all else being equal, autocracies trade less than democracies, even conditional on similar official trade policies.

The theoretical work on the link between political regime type and trade flows or policy has focused on the role of international agreements or the effect of an extension of the voting franchise rather than on accountability and external monitoring.[3] The first approach is taken by Mansfield, Milner, and Rosendorff (2000), who study how the incentives to enter a trade agreement differ between pairs of countries with different political regimes. The presumed difference between democracy and autocracy is that the executive in a democracy is constrained by the requirement that the legislature must ratify any trade agreement, while the executive in an autocracy is free of such constraints. With the additional assumptions that the legislature is more protectionist than the executive and that trade negotiations take place sequentially, as suggested by Putnam (1988), Mansfield, Milner, and Rosendorff (2000) show that pairs of democracies agree on a less protectionist trade policy than mixed pairs of autocracies and democracies. The reason for this somewhat surprising result is that a trade war is more costly for a pair of democracies than for other pairs. As a consequence, pairs of democracies face worse outside options than other pairs and hence agree to more concessions than mixed pairs do. While this prediction is robust to a range of bargaining structures, the model is silent on how much pairs of autocracies trade relative to pairs of democracies.[4]

The model in this article shares the presumption that the critical difference between autocracies and democracies is the lack of effective constraints on the executive in autocracies but departs in three important ways. First, it focuses on individual countries and thus on unilateral trade policy, allowing predictions about how the regime type affects trade flows and trade policy. Second, it focuses explicitly on the incentives that the threat of replacement provides for rulers and politicians in different types of political regimes, allowing democracy and autocracy to be conceptualized along a continuum within the same analytic structure. Third, it combines an explicit economic structure with a stylized political structure.

The other approach in theoretical work is taken by Milner and Kubota (2005), who maintain that the link between democratization and free trade is

3. There is also a large literature on the political economy of trade protection (for example, Hillman 1982; Mayer 1984; Grossman and Helpman 1994; Aidt 1997). The aim of this literature is to explain trade protection within the context of competitive political systems often embodied in some form of democratic institution rather than to explain differences between broad regimes types such as autocracy and democracy.

4. Dai (2002) criticizes the theoretical findings of Mansfield, Milner, and Rosendorff (2000) and argues that their main proposition depends on the preferences of the executive and that it is, therefore, not generally true that democratic pairs trade more than mixed pairs. However, as pointed out by Mansfield, Milner, and Rosendorff (2002), this critique is valid only if the two-level game structure of international negotiations is replaced by a structure in which the legislature of a democracy negotiates directly with its counterpart or with the dictator if paired with an autocracy.

an enlarged constituency of government that changes the identity of the median voter. Under autocracy, the constituency of government is typically a small group of individuals who are well endowed with capital. Under democracy with universal suffrage, the median voter is a worker with a small capital endowment. In countries with a comparative advantage in producing labor-intensive goods (such as developing countries), the Stolper-Samuelson Theorem implies that the median voter benefits from trade liberalization as both a consumer and a laborer. The model here complements this. It leaves aside the effects of political transition on the constituency of government and the role of special interests in an autocracy and in a democracy. Instead, it highlights the degree to which rulers can be held accountable for their actions and to which their incentives to invest in "good" institutions vary systematically across regime types within a specific-factors model of international trade.

The focus of the model is more closely related to that of Adserà, Boix, and Payne (2003), who also stress the role of political accountability with free and competitive elections and the role of the media in creating incentives for politicians to reduce rent extraction. They do not, however, trace out the implications for trade policy, nor do they study a hierarchical structure of government or the implications of political failures.

The Economy

The model considers a small open economy that produces two goods and has an infinite time horizon. The stage model is similar to the specific-factors model of trade employed by Grossman and Helpman (1994) and many others. Good 0 is a numeraire good produced with constant returns to scale with labor as the only input and with an input–output coefficient of 1. Good 1 is produced by labor and sector-specific capital.[5] The profit function is $\pi(p)$, where p is the domestic price of the good and p^* is the international price. Domestic supply is $\partial\pi/\partial p = y(p)$. Labor can move freely between sectors, so the wage rate in the private sector is $w^{\mathrm{P}} = 1$.

The economy is populated by a continuum of agents with measure 1, which are called workers. Each worker supplies one unit of labor inelastically to the labor market in return for wage income w^{P}. Each worker also owns an equal share of the specific factor used in the production of good 1, receiving income $\pi(p)$ from this source each period. Workers consume both goods and spend their entire income each period. Their utility function is $x_0 + u(x_1)$. Optimization subject to the budget constraint yields individual demands, $x_1 = d(p)$ and $x_0 = w^{\mathrm{P}} + \pi(p) - pd(p)$, and the indirect utility $v(p) = w^{\mathrm{P}} + \pi(p) + s(p)$, where $s(p) = u(d(p)) - pd(p)$. All utilities are discounted with the factor $\beta \in (0,1)$.

Good 1 is traded internationally, and net imports are $m\ (.) = d(p) - y(p)$.[6] Workers care about the domestic price of good 1 for two reasons. First, it

5. The claims to the specific factor cannot be traded.
6. Individual and aggregate demand for good 1 are identical.

affects them as consumers; second, it affects their profit income. Taking the derivative of the indirect utility function with respect to p yields

$$\frac{\partial v}{\partial p} = -m(p). \tag{1}$$

Accordingly, if good 1 is imported (m (.) > 0), workers want the domestic price to be as low as possible, while if good 1 is exported (m (.) < 0), they want the domestic price to be as high as possible, to boost their profit income. Trade flows are distorted by two types of policy interventions. First, the ruler of the country (the government) can levy a trade tax τ on good 1. If $\tau > 0$ and good 1 is imported (exported), τ is a tariff (export subsidy); if $\tau < 0$ and good 1 is imported (exported), τ is an import subsidy (export tax). Second, the customs officials in charge of regulating international trade can introduce various unofficial trade barriers, refered to as red tape. The per unit cost of red tape is denoted by θ. To be concrete, it is assumed that good 1 is imported and thus that τ is a tariff.[7] Effective trade distortion, $\tau + \theta$, can thus be defined as the difference between the domestic and the foreign price—or $\tau + \theta = p - p^*$.[8] The revenues from the trade tax are

$$r(\tau, \theta) = \tau m(\tau, \theta), \tag{2}$$

where $(\partial m(\tau, \theta)/\partial\tau) < 0$—that is, an increase in τ pushes up the domestic price, which reduces domestic demand and increases domestic production. Red tape reduces the tax revenues raised for each value of τ because $\partial m(\tau, \theta)/\partial\theta) < 0$.[9] Assuming that $2(\partial m(\tau,\theta)/\partial\tau) + \tau\ (\partial^2 m(\tau, \theta)/\partial\tau^2 < 0$, this means that $r(\tau,\theta)$ is a Laffer curve. Finally, it is assumed that $(\partial m(\tau, \theta)/\partial\theta) + \tau(\partial^2 m(\tau, \theta)/\partial\tau\partial\theta) < 0$, such that the revenue-maximizing tariff falls with θ. Taken together, these assumptions imply that the revenue-maximizing effective trade tax $(\tau + \theta)$ is increasing in θ.[10]

7. It will be clear from the objective function of the ruler that imports or exports are never subsidized. If good 1 is exported, the ruler will impose an export tax, and if it is imported, the ruler will protect domestic production with a tariff. Both are equally bad from the point of view of workers. In both cases, workers want the domestic price to be as close as possible to the world market price (and thus import and export taxes to be zero). Thus the focus on tariffs here does not lead to any loss of generality. A similar analysis can be conducted for export taxes.

8. If good 1 is exported, the wedge would be $p^* - p = \tau + \theta$.

9. Since $p = p* + \tau + \theta, (\partial m(\tau, \theta)/\partial p) = \partial m(\tau, \theta)/\partial\theta = \partial m(\tau, \theta)/\partial\tau < 0$.

10. Let the effective trade tax be denoted $\tau^E = \tau + \theta$; then $\partial\tau^E/\partial\theta = \partial\tau/\partial\theta + 1$. Using the first-order condition for revenue maximization and noting that $(\partial^2 m(\tau, \theta)/\partial\tau^2) = (\partial^2 m(\tau, \theta)/\partial\tau\partial\theta)$,

$$\begin{aligned}\frac{\partial\tau^E}{\partial\theta} &= -\frac{(\partial m(\tau, \theta)/\partial\tau) + \tau\left(\partial^2 m(\tau, \theta)/\partial\tau^2\right)}{2(\partial m(\tau, \theta)/\partial\tau) + \tau(\partial^2 m(\tau, \theta)/\partial\tau^2)} + 1 \\ &= -1 + \frac{(\partial m(\tau, \theta)/\partial\tau)}{2(\partial m(\tau, \theta)/\partial\tau) + \tau(\partial^2 m(\tau, \theta)/\partial\tau^2)} + 1 > 0.\end{aligned}$$

Politics

The country is governed by a ruler whose objective is assumed to be extracting rents from the economy. These rents are spent on the numeraire good[11] and come from two sources: rents from official trade taxes and rents generated by red tape. The idea is that customs officials in corrupt countries often are part of a sophisticated, institutionalized form of rent seeking, with the rents from red tape passing from the customs officials up a chain that stops with ruler. The ruler's utility at time t is $u_R = r(\tau, \theta) + \theta b$, where b is the value to the ruler of the red tape introduced by the customs officials.[12] The ruler's capacity to extract rents depends on the quality of the underlying institutions, measured by $Q \in [0,1]$. At one end of the spectrum is a fully functional democracy with a free press, competitive elections, and respect for civil rights ($Q = 1$), and at the other end is a dictatorship with no effective way for citizens to hold the ruler accountable and no free press to monitor events ($Q = 0$). Depending on Q, the ruler can, therefore, be thought of as an unconstrained dictator, a democratically elected politician, or someone in between.[13] The intrinsic objectives of rulers in autocracies and democracies are not assumed to be different. Instead, the quality of institutions forces democratically elected rulers to behave differently from dictators.

The ruler must employ a bureaucracy to run the customs services,[14] collect tariffs, and hand them over to the ruler. The bureaucrats may also create red tape θ. To capture the idea of a rent-seeking chain, it is assumed that the ruler as well as the bureaucrats benefits from red tape—for example, because bribes can be collected and shared. Red tape can either be low (absent) or high, that is, $\theta \in \{0, \overline{\theta}\}$, where $\overline{\theta} > 0$. The rent that the representative bureaucrat gets from introducing red tape is θB, where B is a positive constant. For simplicity, it is assumed that the bureaucrat holds

11. This assumption can be relaxed. A ruler who consumes good 1 has a direct interest in the domestic price of that good and would then effectively be trading off the revenue implications of policy choices, with the desire to minimize the distortion of the domestic price. The conflict of interest between the ruler and citizens would still be present, and the main results of the analysis would be unaffected qualitatively.

12. The implicit motivation for delegating decision-making power to a ruler is that a government is needed to secure private contracts and ensure that markets can operate. The model could be extended to include a public good. In this case, the ruler can keep only the difference between what is spent on public goods and total tax revenues. All the results are essentially unaffected, so for simplicity the Leviathan assumption is used, as in Brennan and Buchanan (1980).

13. Political accountability is simply assumed to be positively related to a free press. However, as shown by Adserà, Boix, and Payne (2003), the relationship can be endogenized by noting that a free press can provide information to the electorate. This information allows voters to reduce the rents extracted by their rulers.

14. This implicitly assumes that, in addition to workers, there is a pool of potential rulers and bureaucrats in the society that can be called on to serve.

office for only one period and consumes good 0 only.[15] The ruler benefits directly from red tape but also realizes that it distorts trade flows and reduces tariff revenues. The ruler might therefore want to design a control system of incentives for bureaucrats to refrain from introducing red tape, which is modeled as an efficiency wage. This is just one example of a costly institution that a ruler might build to discipline the bureaucracy; most important are that the ruler's choice to pay (or not to pay) the efficiency wage is endogenous and that it is costly to pay the efficiency wage. Rulers can also use penalties to control customs officials, with the maximum penalty assumed to be $\eta \geq 0$.

The incentive to pay the efficiency wage depends critically on the effectiveness of external monitoring. The media play an important part in these activities. A free press, for example, can report on malfeasance, and the ruler can take appropriate action. It is assumed that the external monitoring technology discovers malfeasance of a particular customs official with probability $1 - z(.)$ and that the ruler cannot use the rents from red tape to deduce who introduced it. The effectiveness of external monitoring is exogenous but systematically related to the quality of institutions, Q. It is assumed that $z'(Q) < 0$ and that $z(1) = 0$ and $z(0) = 1$—that is, external monitoring is more effective in societies with high-quality institutions. This reflects systematic differences in media freedom between broad regimes types such as autocracies and democracies. Within-regime variation in media freedom is not modeled, even though in practice it may play a role. Gehlbach and Sonin (2008), for example, show that government control over media in nondemocratic countries is stronger when the government's need to use the media to mobilize the population is larger or when the advertisement market is large.

A bureaucrat who introduces red tape and is discovered, which happens with probability $1 - z(Q)$, is immediately fired and bears the cost of the penalty and thus loses wage income from the public sector and rents from red tape and returns to the private sector. In the private sector, the former bureaucrat receives w^p starting from the next period onward. A bureaucrat who is not discovered, which happens with probability $z(Q)$, keeps the public sector wage for the current period (w_t) and any rent from creating red tape, and returns to the private sector in the subsequent period. The expected utility of a bureaucrat who introduces red tape in period t can be written as

$$z(Q)[w_t + \overline{\theta}B] - (1 - z(Q))\eta + \frac{\beta w^p}{1 - \beta} \tag{3}$$

15. This is not important for the results. It is straightforward to extend the model to allow bureaucrats to hold office forever. Moreover, a bureaucrat who consumes good 1 might be concerned about the impact that red tape will have on the domestic price of the good. In reality, however, that bureaucrat would be only one among many, and it is unlikely that this effect will be taken into account, but even if it were, the main results are still valid as long as the value of red tape is sufficiently high.

and that of a bureaucrat who refrains from doing so as $w_t + (\beta w^p/1 - \beta)$, where w_t is the public sector wage.[16] To strengthen the incentives of the bureaucracy, the ruler may, as suggested by Becker and Stigler (1974), offer an efficiency wage. Its cost is financed from tariff revenues. The efficiency wage that ensures that no red tape is introduced is given by

$$w^e = \max\left\{\frac{z(Q)}{1 - z(Q)}\overline{\theta}B - \eta, 0\right\}. \tag{4}$$

Effective monitoring and harsh penalties reduce the efficiency wage. Faced with the public sector wage w_t, the optimal choice of the bureaucrat in any period t can then be summarized as

$$\theta_t(w_t) = \begin{cases} 0 & \text{if } w_t \geq w^e \\ \overline{\theta} & \text{if } w_t < w^e \end{cases} \tag{5}$$

Citizens attempt to hold the ruler accountable for actions while in office. For simplicity, it is assumed that only workers have political voice. This can be thought of as a situation in which the ruler needs to please the masses, an assumption that makes sense in a democracy but also in many autocracies.[17] Workers try to replace rulers judged to extract too much rent, and the extent to which they can do so depends on the quality of institutions, Q. In a fully democratic society, elections and a free press provide accountability (Ferejohn 1986; Persson and Tabellini 2000; Besley and Prat 2006), but even in autocracies and dictatorships, rulers may be constrained by the threat of a coup or a popular revolt (Acemoglu and Robinson 2001). At the beginning of each period, workers announce a performance standard that the ruler has to satisfy to be "reappointed" at the end of the period. Workers base their performance standard on the utility they get from the policies implemented by the ruler and the bureaucrat within the period. The performance standard announced at the beginning of period t is denoted by $\widehat{v}_t$. The standard requires the ruler to introduce a policy package (τ_t, w_t) that yields at least the utility level $\widehat{v}_t$ in order to be considered for reappointment.

In a well-functioning democracy with a free press (high Q), a ruler (politician) who complies with the standard is guaranteed reappointment while a ruler who does not comply is certain of dismissal. Accountability is, however, seriously weakened in societies with dysfunctional institutions (low Q). Absence of regular and fair elections, intimidation of the opposition, electoral fraud, suppression of the press, and the like can significantly reduce the degree of accountability. This article focuses on a particular type of governance failure that directly impacts the degree of accountability rulers are subjected to.

16. To ensure a positive supply of bureaucrats, it is assumed that $\overline{\theta}B > 1$.
17. The model could be extended to allow for lobbying.

Definition 1 (*q-failure*) *Workers can promise to dismiss only a ruler who does not satisfy* $\widehat{v}_t$ *in period t with probability* $1 - q(Q) \in [0,1]$ *with* $q'(Q) < 0$, $q(1) = 0$ *and* $q(0) = 1$.

A q-failure arises when citizens cannot in all cases dismiss underperforming rulers, and a society with $q(Q)$ close to 1 can be interpreted as a dictatorship in which the ruler can rule unchallenged. A society with poor institutions (Q close to 0) suffers from significant q-failures and lacks an effective external control mechanism (z is close to 1). By contrast, in a society with strong institutions (Q close to 1), q-failures are mostly absent, and external monitoring—for example, through free media—is effective (z is close to 0).

The interaction among rulers, bureaucrats, and workers can be summarized as follows. At the beginning of each period, a new bureaucrat enters office, and workers announce a performance standard. Next, the ruler decides on the tariff and the public wage for the period. Then the bureaucrat decides how much red tape to introduce, and the monitoring technology determines whether to fire the bureaucrat prematurely. At the end of the period, workers observe their utility levels, judge the performance of the ruler against the utility standard, and decide whether to reappoint the incumbent ruler. This, together with random events, as captured by q, determines whether the incumbent is replaced by another ruler. After this, the sequence of events is repeated.

Analysis and Results

Given a sequence of standards $\{\widehat{v}_t\}_{t=0}^{\infty}$, the ruler faces the choice between complying and hoping to stay in power (which allows the ruler to collect rents in the future) or not complying and collecting the maximum rent now.

A ruler who decides not to comply at time t (that is, to deviate, D) sets

$$\left\{\tau_t^D, w_t^D\right\} = \arg\max_{\tau_t, w_t} r(\tau_t, \theta(w_t)) - E(w_t) + \theta(w_t)b. \tag{6}$$

In doing so, the ruler anticipates how the public wage affects the choices of the bureaucrat. It is costly to provide wage incentives, and the expected wage bill is

$$E(w_t) = \begin{cases} zw_t & \text{if } w_t < w^e \\ w_t & \text{if } w_t \geq w^e \end{cases} \tag{7}$$

The ruler knows that the wage has to be paid only if the bureaucrat is not discovered adding red tape. Clearly, either $w_t^D = 0$ or $w_t^D = w^e$ is optimal. In the former case, the optimal tariff is

$$\tau^{D1} = \arg\max_{\tau_t} r(\tau_t, \overline{\theta}) + \overline{\theta}b \tag{8}$$

and the rent is $r\,(\tau^{D1}, \overline{\theta}) + \overline{\theta}b$ for all t. In the latter case, it is

$$\tau^{D2} = \arg\max_{\tau_t} r(\tau_t, 0) - w^e \tag{9}$$

and the rent is $r(\tau^{D2}, 0) - w^e$ for all t. In both cases, the workers attempt to replace the ruler at time $t+1$ but fail to do so with probability $q(Q)$. The ruler's expected payoff is

$$V_t(D) = \max\{r(\tau^{D1}, \overline{\theta}) + \overline{\theta}b, r(\tau^{D2}, 0) - w^e\} + \beta q(Q)V_{t+1}*, \tag{10}$$

where V^*_{t+1} is the continuation value of holding office at the beginning of period $t+1$. The optimal deviation policy depends on the quality of the monitoring institutions as described by lemma 1.

Lemma 1 (*the optimal deviation policy*) *Let* $Z(Q) = (z(Q)/1 - z(Q))$ *with* $Z' < 0$. *Moreover*,

$$\Delta R^D \equiv \frac{r(\tau^{D2}, 0) - r(\tau^{D1}, \overline{\theta}B) - \overline{\theta}b + \eta}{\overline{\theta}B.}$$

Then

(1) *If* $Z(Q) \geq \Delta R^D$, *then* $(\tau^{D1}, 0)$ *is optimal.*
(2) *If* $Z(Q) < \Delta R^D$, *then* (τ^{D2}, w^e) *is optimal.*

Proof. The lemma follows from a straightforward comparison of the net rents collected by the ruler, in each case using equation (4). ■

The quality of institutions Q effectively determines whether it is in the ruler's interest to maintain strong wage incentives for the bureaucrat. If institutional quality is high (Q close to 1), external monitoring is effective (z is low), it is cheap to pay the efficiency wage, and it is optimal to weed out red tape, even for a ruler who has decided to disregard citizen demands. If, by contrast, institutions are weak (Q close to 0) and the monitoring technology is ineffective (z is high), the ruler has no incentive to build (costly) incentives for the bureaucrat. The ruler simply focuses on maximizing tariff revenues subject to red tape. The intuition behind these results is that external monitoring and the efficiency wage are complements rather than substitutes. It is simply cheaper to pay the efficiency wage if the institutional framework allows for effective external monitoring.

An increase in the value derived from red tape (b) makes it more likely that the ruler decides not to pay the efficiency wage. Thus, insofar as institutionalized rent-seeking chains are more common in an autocracy than in a democracy, an autocratic ruler has an additional reason to pay low wages: to encourage customs officials to generate red tape whose rents directly benefit the ruler. But if an autocratic ruler can punish harder (for example, by not having to worry about the rule of law), it is cheaper to pay the efficiency wage, all else being equal, making this choice more likely. Whether to pay the efficiency wage then

depends partly on whether the ruler can punish hard (η is larger) relative to the value of the rent earned from red tape ($b\overline{\theta}$). In situations where the value of the rent is very large and the capacity to punish is limited, paying the efficiency wage is too costly, while in situations where the value of the rent is low and the capacity to punish is larger, paying the efficiency wage becomes the preference.

A ruler who decides to comply (C) in period t selects the policy package

$$\left\{\tau_t^C, w_t^C\right\} = \arg\max_{\tau,w} r(\tau_t, \theta(w_t)) - E(w_t) + \theta(w_t)b \tag{11}$$

subject to $v(\tau_t, \theta_t) \geq \widehat{v}_t$. Again, the ruler either sets $w_t^C = 0$ or $w_t^C = w^e$, and

$$\tau^{C1}(\widehat{v}_t) = \arg\max_{\tau_t} r(\tau_t, \overline{\theta}) + \overline{\theta}b \tag{12}$$

subject to $v(\tau_t, \overline{\theta}) \geq \widehat{v}_t$ is optimal for $w_t^C = 0$ and

$$\tau^{C2}(\widehat{v}_t) = \arg\max_{\tau_t} r(\tau_t, 0) - w^e \tag{13}$$

subject to $v(\tau_t, 0) \geq \widehat{v}_t$ is optimal for $w_t^C = w^e$. Since $v(\tau, \theta)$ is decreasing in τ, the ruler must reduce the tariff below the respective rent-maximizing levels to satisfy the constraints. The expected payoff is

$$V_t(C) = \max\left\{r(\tau^{C1}(\widehat{v}_t), \overline{\theta}) + \overline{\theta}b, r(\tau^{C2}(\widehat{v}_t), 0) - w^e\right\} + \beta V_{t+1}*. \tag{14}$$

As shown by lemma 2, the quality of institutions, through its impact on external monitoring, also plays a key role for the choice between the two compliance strategies.

Lemma 2 (*within-period optimal compliance*). *Let* $Z(Q) = (z(Q)/1 - z(Q))$ *with* $Z' < 0$. *Moreover, suppose that* $\widehat{v}_t \geq \max\left\{v\left(\tau^{D1}, \overline{\theta}\right), v(\tau^{D2}, 0)\right\}$ *and let*

$$\Delta R_t^C \equiv \frac{r\left(\tau^{C2}(\widehat{v}_t), 0\right) - r\left(\tau^{C1}(\widehat{v}_t), \overline{\theta}\right) - \overline{\theta}b + \eta}{\overline{\theta}B}.$$

Then

(1) *If* $Z(Q) \geq \Delta R_t^C$, *then the optimal compliance policy is* $\left(\tau^{C1}(\widehat{v}_t), 0\right)$.
(2) *If* $Z(Q) < \Delta R_t^C$, *then the optimal compliance policy is* $\left(\tau^{C2}(\widehat{v}_t), w^e\right)$.

Proof. The lemma follows from a straightforward comparison of the net rents collected by the ruler in each case using equation (4). ■

The intuition behind lemma 2 is similar to that behind lemma 1. In societies with strong institutions, the ruler has an incentive to strengthen institutions and pay an efficiency wage; in a society with weak institutions, this incentive is absent.

The sequence of performance standards is incentive compatible if at all t

$$V_t(C) \geq V_t(D). \tag{15}$$

Workers select the sequence of standards that yields the highest lifetime utility subject to incentive compatibility. The structure of the model implies that the optimal choice is stationary; that is, $\widehat{v}_t = \widehat{v}^*$ for all t, where $\widehat{v}^*$ is defined by

$$\begin{aligned} &\max\{r(\tau^{C1}(\widehat{v}^*), \overline{\theta}) + \overline{\theta}b, r(\tau^{C2}(\widehat{v}^*), 0) - w^e\} \\ &= \frac{1-\beta}{1-\beta q(Q)} \max\{r(\tau^{D1}, \overline{\theta}) + \overline{\theta}b, r(\tau^{D2}, 0) - w^e\}. \end{aligned} \tag{16}$$

Incentive compatibility requires that $q(Q) < 1$; otherwise, institutions are so bad that no ruler would ever comply with any standard other than the rent-maximizing one. It is also clear from equation (16) that workers' welfare is increasing in the quality of institutions—that is, $(\partial\widehat{v}*/\partial Q) > 0$.

Why is the volume of international trade different in autocracies and democracies? To study this, two extremes are compared. At one extreme is a society with well-functioning democratic institutions and a free press (a democracy): $Q \to 1^-$. At the other end is a society with seriously dysfunctional institutions (an autocracy): $Q \to 0^+$. In the real world, most societies fall between these extremes. The following proposition states the main implications of the model.

Proposition 1 (*regime type and the volume of trade*). *Assume that min* $\{\Delta R_t^C, \Delta R_t^D\} > 0$.[18]

(1) *The effective trade distortion is higher in autocracies than in democracies and, as a consequence, autocracies trade less with the rest of the world than do democracies.*
(2) *For given official trade policy* (τ), *autocracies trade less with the rest of the world than do democracies because of differences in red tape and other unofficial trade distortions.*

Proof. *Part 1.* Consider an autocracy with $Q = 0$. This implies that $q(0) = 1$ and $z(0) = 1$. Lemma 1 implies that the optimal deviation entails $w^D = 0$ and $\tau^D = \tau^{D1}$, while lemma 2 implies that the optimal compliance policy is $w^C = 0$ and $\tau^C = \tau^{C1}$ for all t. However, since $q(0) = 1$, equation (16) implies that the ruler implements $\tau = \tau^{D1} = \tau^{C1}$ $(v(\tau^{D1} + \overline{\theta}))$ and $w = 0$ each period until replaced by a new ruler who behaves likewise. Workers get $v(\tau^{D1} + \overline{\theta})$, and the

18. A sufficient condition for this to hold is that $\eta \geq b\overline{\theta}$. However, what really matters is that democratic leaders choose to pay the efficiency wage while autocratic rulers do not. Insofar as the punishment cost and the value of red tape vary across regimes, the value of red tape to democratic leaders needs only to be not too high relative to the cost to bureaucrats of being caught creating red tape. For autocratic rulers, no restrictions need to be imposed on the relative size of η and $b\overline{\theta}$ to expect that they are both finite.

effective trade distortion is $\tau^{D1} + \overline{\theta}$. Consider next a democracy with $Q = 1$. This implies that $q(1) = 0$ and $z(1) = 0$. Under the assumption that min $\{\Delta R_t^C, \Delta R_t^D\} > 0$, lemmas 1 and 2 imply that $w^D = w^C = w^e$ and that $\tau^D = \tau^{D2}$ and $\tau^C = \tau^{C2}$ at all t. The effective trade distortion is τ^{C2}. Let v^{**} denote equilibrium utility of a worker, where v^{**} is defined by equation (16):

$$r(\tau^{C2}(v^{**}), 0) - w^e = (1 - \beta)(r(\tau^{D2}, 0) - w^e). \tag{17}$$

Suppose that $\beta = 0$. Then $r(\tau^{C2}\ (v^{**}),0) = r(\tau^{D2}, 0)$. The assumptions listed at the end of section I imply that $\tau^{D2} < \tau^{D1} + \overline{\theta}$. Since indirect utility is decreasing in the domestic price of good 1, it follows that $v\ (\tau^{D2}) > v(\tau^{D1} + \overline{\theta})$. Consequently, for $\beta = 0$, it follows that $v^{**} = v(\tau^{D2})$ is larger than $v(\tau^{D1} + \overline{\theta})$. Since from equation (17), v^{**} is increasing in β, and $v(\tau^{D1} + \overline{\theta})$ is independent of β, it follows that the best incentive-compatible standard under democracy v^{**} entails higher utility than what is obtained under autocracy, $v(\tau^{D1} + \overline{\theta})$. Since indirect utility is decreasing in the effective tariff rate, it follows that $\tau^{D1} + \overline{\theta} > \tau^{C2}\ (v^{**})$ and thus, as stated in part 1 of the proposition, that autocracies trade less.

Part 2. Begin by observing that $\tau^{D1}\ (\theta) \leq \tau^{D2}$. This follows from the fact that $\tau^{D1}(0) = \tau^{D2}$ and the assumption that the revenue-maximizing tariff falls with θ, that is, $(\partial\tau^{D1}\ (\theta)/\partial\theta) < 0$. Moreover, $\tau^{C2}\ (v^{**}) < \tau^{D2}$, since τ^{D2} maximizes tariff revenues, and $\tau^{C2}\ (v^{**})$ is independent of θ. It follows that a $\overline{\theta}$ can be chosen such that $\tau^{D1}\ (\overline{\theta}) = \tau^{C2}\ (v^{**})$—that is, such that the official tariff is the same in democracies and autocracies. It then follows from the observation that autocracies allow red tape while democracies do not and that autocracies trade less than democracies conditional on having the same official trade policy.

The first part of the proposition shows that autocracies trade less than democracies. The source of this difference is the quality of institutions. These differences affect trade flows through two channels. First, autocracies have weak accountability institutions, as captured by $q(Q)$. This allows autocratic rulers to extract more rents than politicians in a democracy can. The implication is higher trade taxes under autocratic rule and consequently less imports (or exports). An improvement in accountability (better institutions) reduces trade taxes and encourages more trade. Second, autocracies also have weak external monitoring institutions (as captured by $z(Q)$). As a consequence, autocratic rulers have little incentive to weed out red tape and other unofficial trade obstructions introduced by the bureaucrats. In contrast, in a democracy with a free press and effective external monitoring, it is cheap to pay the efficiency wage. In other words, it is optimal for a ruler to enhance institutional quality endogenously. This reduces red tape and encourages trade flows.

One implication is that public sector wages are lower in autocracies than in democracies. In the model, this is due to two factors. First, because of poor monitoring, autocratic rulers find it too costly to pay efficiency wages. Second, autocratic rulers have a direct interest in keeping red tape because they share in

the rents. One interpretation, then, is that autocratic rulers hold public sector wages down to encourage bureaucrats to collect unofficial rents that are (partly) passed up through a chain to them. In fact, for a given quality of institutions, Q, an increase in the value of red tape for the ruler, b, increases the likelihood that the regime introduces red tape, and consequently it trades less with the rest of the world.[19] In other words, insofar as rent-seeking chains are more common in autocracies than in democracies—and empirical observations suggest that they are—the model suggests an alternative reason why autocracies trade less: autocratic rulers, unlike democratic politicians, have a direct interest in maintaining and encouraging informal mechanisms (red tape) for extracting rents from international trade. The second part of the proposition shows that precisely because of differences in the incentives for rulers to control red tape in the two types of societies, autocracies trade less than democracies for a given official trade policy. The reason is unobserved red tape.

There may, of course, be other reasons why an autocracy trades less than a democracy with the same official trade policy that are not captured by the model. For example, tariffs and export taxes may violate international trade agreements and induce autocratic rulers, who would have imposed additional trade taxes in the absence of these constraints, to seek informal ways of extracting rents from international trade. To ensure that this does not drive the empirical results, participation in international agreements is controlled for in the estimations (see section III).

Another reason why trade may be more distorted in an autocracy than in a democracy is that autocratic rulers must cater to powerful domestic elites with a particular interest in trade protection. This is likely to play an important role in autocracies where local elites have a substantial stake in import-competing sectors. However, lobbying by special interest groups for trade protection is also common in democracies (see, for example, Grossman and Helpman 1994). It is thus unclear whether this line of reasoning leads to systematic differences in trade flows according to regime.

Yet another reason relates to systematic differences in the protection of property rights in autocracies and democracies. Poor protection of property rights discourages foreign direct investment in autocracies and may have an adverse affect on trade volumes for a given trade policy.

Alternative explanations exist for why public sector wages are low in autocracies. One possibility, suggested by Besley and McLaren (1993), is that paying tax collectors or customs officials a wage below the market wage—a so-called capitulation wage—can yield more net revenue than trying to discipline officials through wage incentives. Another possibility, suggested by Olson (2000), is that autocrats, aiming to extract maximum revenues from the population, tax inframarginal units of work highly while at the same time keeping

19. Note that b reduces R^D and R^C (see lemmas 1 and 2) and thus makes it more likely that the ruler, for given Q, decides not to weed out red tape.

the tax on extra income close to zero. The result of such a system would also be very low wages for the normal amount of work done by public officials. These are valid explanations, and the focus here on the efficiency wage complements them while stressing a new reason why autocratic rulers might want to keep public sector wages low: to encourage their officials to create rents that they can share in.

II. Empirical Assessment of the Model

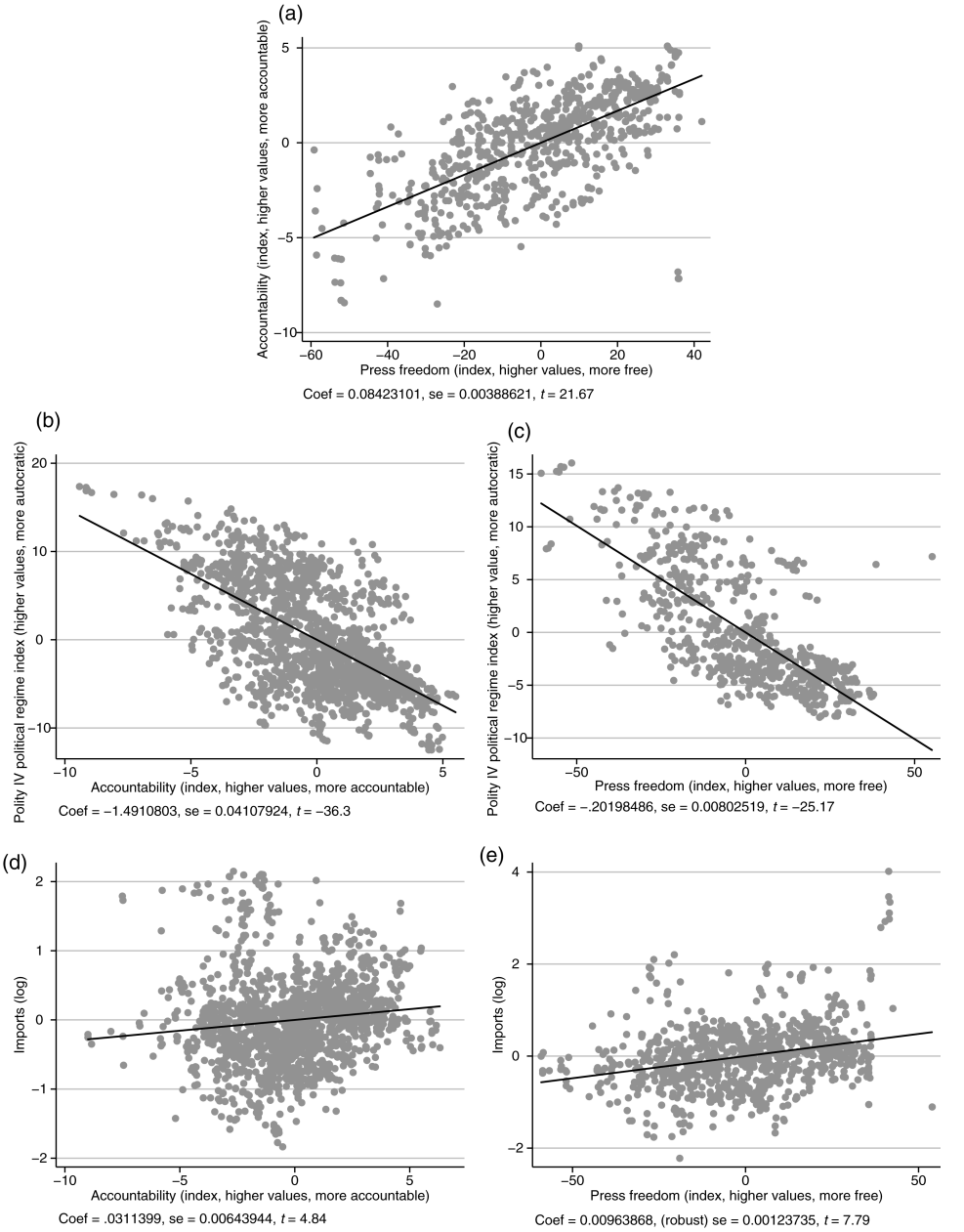
Before turning to the main econometric analysis, this section assesses the key assumptions of the model empirically and provides some suggestive but direct evidence on the relevance of the two transmission channels highlighted by the model. This is because these aspect of the model cannot be directly tested (due to lack of sufficient data)[20] in the main empirical analysis, which consequently focuses on testing whether an autocracy trades less than a democracy.

The cornerstone of the model is the parameter Q expressing institutional quality. Press freedom and political accountability go together and are features of democracies that tend to be absent in autocracies. The Freedom of the Press indicator published by Freedom House (2009) and the Democratic Accountability indicator from the International Country Risk Guide (see Knack and Keefer 1995) are used to measure these attributes empirically. Higher values of the two indicators correspond to more freedom and more accountability, respectively (figure 1). Panel a of figure 1 shows that press freedom goes hand in hand with political accountability, as expected based on Adserà, Boix, and Payne (2003).[21] All panels of figure 1 control for GDP, annual time dummy variables, and heteroskedasticity, so figure 1 depicts partial-regression leverage plots. In the econometric analysis below, data limitations force the focus onto broader measures of political regime type, including the Polity IV index. This index measures regime type on a scale from -10 (autocracy) to 10 (democracy), but to make it comparable to similar indicators, it is rescaled so that higher values correspond to more autocratic countries. To see whether the transmission channels identified by the model make sense, the partial correlation between press freedom and political accountability, on the one hand, and the Polity IV index, on the other hand, were studied. The results are shown in panels b and c of figure 1 and support the assertion that democracy (as measured by the Polity IV index) is associated with more press freedom and more political accountability. Finally, data on total real imports to a country in a given year are used to assess whether press freedom and

20. Systematic data on freedom of the press, for example, is available only from 1994 onward. As basically all political transitions took place before 1994, the fixed effects approach prevents these data from being directly incorporated.

21. Adserà, Boix, and Payne (2003) report a strong, positive relationship between a measure of free circulation of newspapers and a number of alternative measures of political accountability.

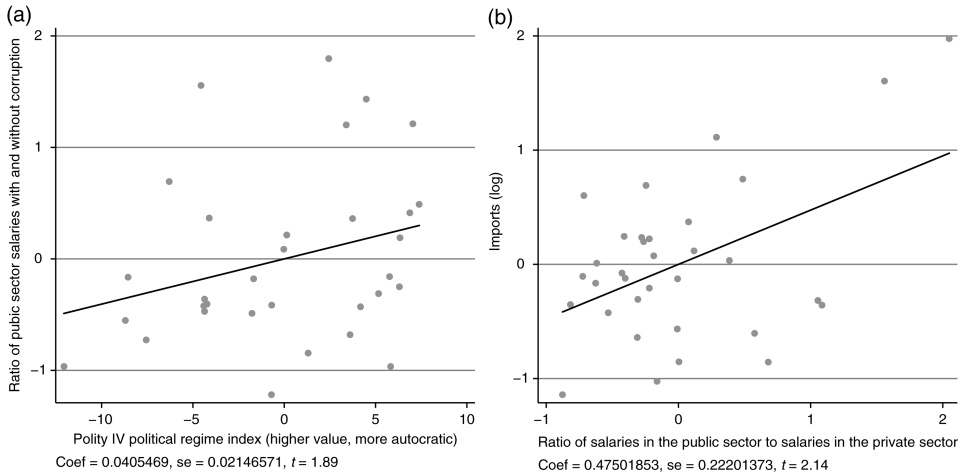
FIGURE 1. Press Freedom and Accountability Compared with Regime Type and Imports



Note: The panels depict partial-regression leverage plots (controlling for income, annual time dummy variables, and heteroskedasticity). Data are for 1984–2000 for accountability and 1994–2000 for press freedom. The coefficient of the underlying regression, its standard deviation, and *t*-statistic are reported below each graph.

Source: Authors' analysis based on data from Freedom House (2009), Knack and Keefer (1995), Gurr, Jaggers, and Moore (2003), and Feenstra (2000).

FIGURE 2. Public Salaries Compared with Regime Type and Imports



Note: The panels depict partial-regression leverage plots (controlling for income and heteroskedasticity). Data are averages for 1970–90. The coefficient of the underlying regression, its standard deviation, and *t*-statistic are reported below each graph.

Source: Authors' analysis based on data from Feestra (2000), Gurr, Jagger and Moore (2003), and Rauch and Evans (2000).

political accountability are linked directly to trade volumes. The partial-regression plots in panels d and e show that both press freedom and political accountability are associated with larger volumes of trade. This supports the relevance of the two channels highlighted by the model, albeit only through correlations.

Another key feature of the model is the efficiency wage (or lack thereof). Rulers in autocratic societies (rationally) decide not to pay the efficiency wage, which increases the probability that bureaucrats will create red tape.[22] Again, these assumptions are validated using partial-regression leverage plots. Because only cross-sectional information on public sector wages is available, only GDP and heteroskedasticity can be controlled for in figure 2. Figure 2 addresses two questions using survey data from Rauch and Evans (2000) on wages in the public sector (relative to the private sector) for 35 developing countries averaged over 1970–90. Is the fraction of total pay accounted for by bribes or

22. The observation is supported by the findings of Gorodnichenko and Sabirianova Peter (2007), who find that public sector employees (including customs officials) in Ukraine receive approximately 30 percent lower wages than those in comparable jobs in the private sector. Yet the consumption expenditures and asset holdings of the two groups are essentially identical, which indicates that bureaucrats receive sizable unofficial payments. The negative relationship between relative civil service pay and corruption is also demonstrated in van Rijckeghem and Weder (2001). Finally, Adserà, Boix, and Payne (2003) report a strong positive relationship between the level of democracy and measures of the quality of the bureaucracy in a sample of about 100 countries for 1980–95. This directly supports the notion that autocratic rulers have less incentive to build incentives for their bureaucrats.

other extralegal sources of income larger in autocracies than in democracies? And does the official public sector wage affect trade flows?

Panel a of figure 2 shows the relationship between the ratio of total to official income (higher values indicate that extralegal income is a more important source of income) and the Polity IV index. Unofficial income accounts for a larger share of total income in autocratic countries. This is consistent with the notion that public officials are more likely to supplement their (low) official wages with bribe income created through red tape in an autocracy than in a democracy. Moreover, as the model suggests, higher official wages in the public sector are positively correlated with the volume of imports, as demonstrated by the partial-regression plot in panel b of figure 2.

Taken together these empirical observations suggest that the model and its basic assumptions provide an empirically sound and consistent theory of why regime type might affect trade volumes. But the question of whether autocracies trade less, and if so, whether the relationship is causal, remains to be answered.

III. Empirical Specification

This section turns to estimating the relationship between a country's political regime and its involvement in international trade, thereby testing the two main implications of the model listed in proposition 1 and answering the question posed by the title of the article. To this end, a dyadic model of trade is used for a sample of up to 130 countries covering 1962–2000. The dependent variable is imports of country i from country e in year t rather than total trade flows between pairs of countries.[23] This choice avoids what Baldwin (2006, p. 18–19) calls the "silver-medal of gravity mistakes"—that is, the sizable upward bias that regressions with average bilateral trade flows as the dependent variable are subject to when trade is unbalanced.[24] More specifically, the baseline specification is the following dyadic panel model:

$$\begin{aligned}\ln(\text{import}_{iet}) = {} & \beta_1 \text{regime}_{it-1} + \beta_2 \text{regime}_{et-1} + \beta_3 \ln(\text{GDP}_{it}) \\ & + \beta_4 \ln(\text{GDP}_{et}) + \beta_5 \ln(\text{GDP p.c.}_{it}) + \beta_6 \ln(\text{GDP p.c.}_{et}) \\ & + \beta_7 \ln(\text{WTO}_{it}) + \beta_8 \ln(\text{WTO}_{et}) + \beta_9 \ln(\text{regional}_{iet}) \\ & + \gamma_{ie} + \delta_t + \varepsilon_{iet},\end{aligned} \tag{18}$$

where import_{iet} is imports of country i from country e in year t; regime_{it-1} and regime_{et-1} are lagged values of measures of regime type (democracy or autocracy) of the importing and exporting country (to be discussed below); GDP_{it}

23. Data on nominal import flows are taken from Feenstra (2000).

24. This follows from the fact that the log of the average is not equal to the average of the logs if the import and export flows are not identical in magnitude (Jensens's inequality). For a formal proof, refer to Baldwin (2006).

and GDP_{et} are real GDP of the importing and exporting country; GDP $p.c._{it}$ and GDP $p.c._{et}$ are GDP per capita of the importing and exporting country; WTO_{it} and WTO_{et} are dummy variables indicating whether the importer or exporter country is a member of the General Agreement on Tariffs and Trade/ World Trade Organization (WTO); and $regional_{iet}$ is a dummy variable taking the value of 1 if both the importer and the exporter are members of the same regional trade agreement.

Table 1 lists the sources, definitions, and summary statistics of all the variables used in the analysis. All regressions include fixed effects for the trading pair, γ_{ie}, as well as year fixed effects, δ_t. This is a variant of the approach adopted by Feenstra (2004), who introduced the notion of country-specific effects as "multilateral resistance terms."[25] The dyadic effects control for unobserved trading pair characteristics that are fixed over time, with the subtlety that these unobservable effects can be asymmetric, that is, $\gamma_{ie} \neq \gamma_{ei}$. Baltagi, Egger, and Pfaffermayr (2003) and Baldwin (2006) point out the importance of correcting for these trading pair and time fixed effects. Baldwin (2006, p. 15–16) calls the omission of these effects the "gold-medal of gravity mistakes, " noting that fixed effects control only for the time-invariant part of multilateral resistance. However, including time-varying fixed effects would preclude identification of the regime type effect and cannot be done.

Because of the difficulty of obtaining reliable quantitative measures of regime type, three indicators are used as proxies. Each captures different aspects of the institutional environment and has flaws and advantages. The first indicator is the regime type indicator constructed by Alvarez and others (1996) and Przeworski and others (2000). It defines democracy as a political system in which incumbents can lose elections and are forced to comply with these outcomes. More specifically, a country is classified as a democracy if the executive and the legislature are elected through contested elections in which more than one party has a chance of winning. The resulting dummy variable takes the value of 1 for autocracies and 0 for democracies. The second indicator is the Polity IV index constructed by Gurr, Jaggers, and Moore (2003).[26] The index is measured on a scale of −10 (autocracy) to 10 (democracy). As mentioned above, so that the results obtained with this indicator are

25. An alternative to the specification with the dyadic specific effects is to follow Rose (2004) and include the geographic distance between the trading partners, and dummy variables controlling for common language, common border, colonial ties, common colonizer, and landlockedness as well as fixed effects for the importers and exporters. The empirical findings remain unchanged when this is done. For further details on this type of gravity model, see Anderson and van Wincoop (2003), Rose (2004), and Brun and others (2005).

26. The Polity IV index—or more accurately the "polity2" index—summarizes different indicators of political authority patterns to measure three key aspects of a country's political system: competitiveness and openness in executive recruitment, constraints on the chief executive, and competitiveness and regulation of political participation. A weighted sum of the components is used to construct two summary variables, measuring democracy on a scale from 0 to 10 and autocracy from −10 to 0. The Polity IV index is the sum of these two subindexes.

TABLE 1. Definitions, Sources, and Summary Statistics of the Variables Used in the Analysis

Variable	Description	Source	Mean	Standard deviation
ln import	Log of nominal imports in U.S. dollars	Feenstra 2000	7.96	3.49
Polity IV[a]	Inverse of polity2 indicator: 1 = most democratic, 21 = most autocratic	Gurr, Jaggers, and Moore 2003	7.59 (*i*)	7.32 (*i*)
			7.58 (*e*)	7.30 (*e*)
Freedom House[a]	Average of "political rights" and "civil liberties" indicators: 1 = most democratic,7 = most autocratic	Freedom House 2006	3.09 (*i*)	1.94 (*i*)
			3.13 (*e*)	1.94 (*e*)
Przeworski and others[a]	Dummy variable taking the value of 1 for autocratic states	Alvarez and others 1996; Przeworski and others 2000	0.42 (*i*)	0.49 (*i*)
			0.40 (*e*)	0.49 (*e*)
ln GDP[a]	Log of GDP in constant 2000 U.S. dollars	World Bank 2006	24.18 (*i*)	2.17 (*e*)
			24.26 (*e*)	2.23 (*i*)
ln GDP p.c.[a]	Log of GDP in constant 2000 U.S. dollars divided by midyear population	World Bank 2006	7.95 (*i*)	1.59 (*i*)
			7.96 (*e*)	1.58 (*e*)
Restriction index[a]	Subindex economic restrictions of the KOF Swiss Economic Institute Index of Globalization; combines data on hidden import barriers, mean tariff rate, taxes on international trade (percent of current revenue) and capital account restrictions	Dreher 2006	6.30 (*i*)	1.97 (*i*)
			6.24 (*e*)	1.99 (*e*)
Regional trade agreement	Dummy variable for pairs that are a member of the same regional trade agreement	Rose 2004	0.01	0.11
WTO[a]	Dummy variable for members of the General Agreement on Tariffs and Trade/World Trade Organization	www.wto.org	0.77 (*i*)	0.42 (*i*)
			0.76 (*e*)	0.43 (*e*)

[a]For these variables *i* refers to importing countries and *e* refers to exporting countries.

Source: Authors' analysis based on sources described in the table.

comparable to those obtained with the two other indicators, the variable was recoded so that higher values indicate that a society is more autocratic. The third indicator is the average of the political rights and civil liberties indicators constructed by Freedom House (2006). The resulting indicator ranges from 1 to 7, with higher values indicating that a society is more autocratic.

All of the indicators have drawn criticism. Przeworski's regime type indicator uses the most clear-cut definition of the three but has the disadvantage of being a dummy variable without "shades of gray." The Polity IV index has been criticized for the way values are assigned to its various subcomponents.[27] Freedom House draws critique because its indicators are completely survey based. Furthermore, the three indicators focus on slightly different aspects of political institutions and should therefore perhaps best be viewed as complements rather than substitutes. Przeworski and others (2000) aim to capture a combination of political participation and contestability of political power. The Polity IV index basically measures political competition and ignores how widely extended the voting franchise is and other aspects of popular participation in politics. The Freedom House index focuses more on political rights and civil liberties than on de facto political competition and participation. The complementarity of the measures is another good reason to use all three in the analysis.[28]

As Milner and Kubota (2005) argue, it takes time for changes in political institutions to affect trade patterns, and the effects of democratic transitions are likely to be long lasting. For this reason, the three institutional indicators are entered either with a one-year lag or as the average of the five preceding years. This also mitigates potential endogeneity problems arising if international trade encourages the development of democratic institutions, a topic addressed in section IV.

The baseline specification allows testing of the first implication of the model—that autocratic countries trade less. The second implication is that autocratic countries trade less conditional on official trade policy. To test this, the baseline model must be extended with a proxy for trade policy. Trade restrictions can take many forms, so a multidimensional index is used. In particular, the restriction subindex from the KOF Swiss Economic Institute Index of Globalization is employed (see Dreher 2006). This restriction index combines publicly available information on nontariff import barriers, mean tariff rates, other taxes on international trade, and capital account restrictions. It ranges from 1 to 10, with higher values indicating fewer restrictions.

It is important to notice that specification (18) is designed to estimate the effect of the regime type of individual countries (as opposed to pairs of

27. Moreover, as discussed by Vreeland (2008), instances of civil violence are coded as less autocratic than peaceful autocracies because civil violence is taken as a form of political participation. Given that, for example, civil wars often lead to a regime change, this middle range of the scale is associated with regime instability (see Plümper and Neumayer 2009).

28. The correlation between the three measures is, however, high despite these differences, ranging from 0.8 to 0.9.

countries) on trade flows and to do so separately for an importing and an exporting country. This dyadic panel setup allows the results to be directly compared with those in the previous literature. However, the underlying data set is extremely large (more than 80,000 observations). For this reason, it is possible that the t-statistics associated with the estimated coefficients on the regime type indicators are spuriously large. An alternative to the dyadic panel setup that circumvents this problem and provides another avenue for testing whether autocracies trade less is a single-country panel model. In the model, the dependent variable is a country's total annual imports or exports, respectively, and the unit of analysis is a country-year observation:[29]

$$\begin{aligned} \ln(y_{jt}) = {} & \alpha_1 \text{regime}_{jt-1} + \alpha_2 \ln(\text{GDP}_{jt}) + \alpha_3 \ln(\text{GDP p.c.}_{jt}) \\ & + \alpha_4 (\text{restriction index})_{jt} + \gamma_j + \delta_t + \varepsilon_{jt} \end{aligned} \tag{19}$$

where y_{jt} is either total annual real imports or exports to or from country j at time t.[30] Country- and time-specific fixed effects are included as well as real GDP, real GDP per capita, and the restriction index. This specification is not based on the gravity model.

IV. The Main Empirical Results

Table 2 shows the results of the estimation of equation (18). To account for the particularities of trade flows, the error terms are clustered at the trading pair level.[31] All control variables have the expected signs and are highly significant except GDP per capita and WTO membership of the exporting country in some specifications. More important, all three regime type indicators yield the same result: autocracies trade significantly less. The coefficients for importing and exporting countries are of roughly similar magnitude. Moreover, most of the estimated coefficients on the regime type indicators are modestly larger in the specifications with five-year averages than in the specifications with one-year lags. This suggests that the effect of regime type on trade is persistent; a finding in line with that of Milner and Kubota (2005). It also suggests that changes in trade flows take place gradually after a regime change.

Given its dichotomous nature, Przeworski and others' (2000) regime type indicator is the easiest to interpret: a democracy that turns into an autocracy experiences an average decrease of 19.6–22.9 percent of imports and 17.3–22.4 percent of exports, all else being equal. Because the fixed effects estimator does not use information from countries that do not change their regime

29. This reduces the sample size to about 1,700 observations.

30. Because this single-country panel model is not derived from the gravity model, real-trade volumes are used. However, using nominal volumes instead leaves the results unchanged.

31. As an alternative, the standard errors were clusted on the importer and exporter levels separately. This did not affect the results.

TABLE 2. Results from the Dyadic Panel Model

Variable	Przeworski and others 2000		Polity IV		Freedom House	
$\text{Autocracy}_{i\ t-1}$	−0.196*** (0.025)	—	−0.017*** (0.002)	—	−0.052*** (0.008)	—
$\text{Autocracy}_{e\ t-1}$	−0.173*** (0.025)	—	−0.012*** (0.002)	—	−0.040*** (0.008)	—
$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	—	−0.229*** (0.033)	—	−0.021*** (0.002)	—	−0.065*** (0.012)
$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	−0.224*** (0.033)	—	−0.014*** (0.002)	—	−0.030*** (0.011)
$\log \text{GDP}_i$	1.290*** (0.086)	1.142*** (0.099)	1.303*** (0.087)	1.146*** (0.099)	1.213*** (0.105)	0.587*** (0.134)
$\log \text{GDP}_e$	1.073*** (0.088)	1.033*** (0.101)	1.088*** (0.088)	1.050*** (0.101)	1.357*** (0.106)	0.958*** (0.133)
$\log \text{GDP p.c.}_i$	0.117 (0.082)	0.272*** (0.092)	0.129 (0.083)	0.310*** (0.093)	0.164* (0.097)	0.833*** (0.121)
$\log \text{GDP p.c.}_e$	0.332*** (0.092)	0.431*** (0.107)	0.351*** (0.092)	0.454*** (0.107)	0.077 (0.11)	0.595*** (0.135)
Regional trade agreement	0.434*** (0.048)	0.399*** (0.047)	0.410*** (0.048)	0.375*** (0.046)	0.298*** (0.046)	0.255*** (0.044)
WTO_i	0.077** (0.033)	0.098*** (0.038)	0.066** (0.033)	0.087** (0.038)	0.109*** (0.033)	0.086** (0.035)
WTO_e	0.147*** (0.035)	0.104** (0.043)	0.157*** (0.036)	0.108** (0.044)	0.079** (0.039)	0.054 (0.045)
Observations	181,095	136,813	174,538	132,880	146,793	101,719
Trading pairs	10,407	8,120	9,753	7,701	10,345	7,850
R-squared (within)	0.5	0.51	0.51	0.51	0.34	0.3

**Significant at the 5 percent level; ***significant at the 1 percent level.

Note: The dependent variable is the log of imports. Numbers in parentheses are standard errors clustered at the trading pair level. $\text{Autocracy}_{i\ t-1}$ is the autocracy score lagged one year for importing countries and $\text{Autocracy}_{e\ t-1}$ is the autocracy score lagged one year for exporting countries. $\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$ is the average of the five years prior to the observation for importing countries, and $\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$ is the average of the five years prior to the observation for exporting countries. See table 1 for definitions of other variables. All regressions contain trading pair and time-specific fixed effects that are significant at the 1 percent level.

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

within the sample period—only the time variation of the political regime indicators is used to identify the effects—these figures likely underestimate the effect of autocracy on trade flows,[32] a notion supported by the the data. For example, Nigeria experienced a transition from democracy to autocracy in 1982/83 (according to Przeworski's regime type indicator). The result was a 37 percent decline in its imports and a 26 percent decline in exports.

Both the Polity IV and the Freedom House indexes are measured on an ordinal scale. On the 21 point scale of the Polity IV index, a one point move toward autocracy reduces imports 1.7–2.1 percent and exports 1.2–1.4 percent, all else being equal. This means that a hypothetical country that undergoes a transition from full democracay to complete autocracy would lose about 34 percent of imports and about 24 percent of exports. On the 7 point scale of the Freedom House index, a hypothetical country that goes through the same transition would lose about 31 percent of imports and about 24 percent of exports. To give a concrete—albeit unrealistic—example, imagine that in 2000 Switzerland suddenly had the political regime of Myanmar. The consequence would be a 28.9 percent reduction of imports and a 20.4 percent reduction of exports, according to the Polity IV index, and a 31 percent reduction of imports and a 24 percent reduction of exports, according to the Freedom House index. Although differences exist, it is striking how similar the results obtained with the three different indicators are.

For a sample of developing countries, Milner and Kubota (2005) show that democracies have lower tariff rates than autocracies do. Thus, the finding from table 2—autocracies trade less—could simply be a result of this effect. To investigate this, the restriction index, introduced above, is added to the specification in equation (18) and the estimation is re-run. Not surprising, the restriction index has a positive impact on trade flows and is highly significant for importing countries (table 3). This indicates that a country with fewer trade restrictions imports more. For exporting countries, the coefficient on the restriction index is also positive, though statistically insignificant in two specifications.

More important, the main finding from the baseline model persists even after controlling for differences in trade policy: autocracies trade less. The coefficients for Przeworski and others' (2000) regime type indicator and for the Polity IV index are somewhat lower than those reported in table 2 but still highly significant. The coefficients for the Freedom House index remain virtually unchanged. Two factors can account for the difference. First, including the restriction index reduces the sample size, which accounts for most of the difference. Second, the coefficients on the regime type indicators in the specification without the restriction index would be biased downward (more negative) if autocracies had higher trade barriers, as suggested by Milner and Kubota (2005), and trade barriers reduce trade flows.

32. Figure A-1 in the appendix shows that the regime type indicators exhibit enough time variation to allow a fixed effects approach.

TABLE 3. Results with Trade Restrictions

Variable	Przeworski and others 2000		Polity IV		Freedom House	
$\text{Autocracy}_{i\ t-1}$	−0.139*** (0.032)	—	−0.013*** (0.003)	—	−0.055*** (0.011)	—
$\text{Autocracy}_{e\ t-1}$	−0.174*** (0.031)	—	−0.012*** (0.002)	—	−0.054*** (0.01)	—
$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	—	−0.188*** (0.039)	—	−0.018*** (0.003)	—	−0.069*** (0.015)
$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	−0.207*** (0.040)	—	−0.013*** (0.003)	—	−0.047*** (0.014)
Restriction index_i	0.139*** (0.002)	0.129*** (0.020)	0.132*** (0.020)	0.123*** (0.020)	0.143*** (0.019)	0.150*** (0.020)
Restriction index_e	0.042** (0.020)	0.053*** (0.020)	0.037* (0.020)	0.051** (0.020)	0.028 (0.019)	0.006 (0.020)
log GDP_i	1.378*** (0.138)	1.222*** (0.148)	1.351*** (0.138)	1.206*** (0.147)	1.329*** (0.149)	0.803*** (0.187)
log GDP_e	1.301*** (0.136)	1.081*** (0.145)	1.313*** (0.136)	1.099*** (0.144)	1.356*** (0.145)	1.070*** (0.181)
log GDP p.c._i	−0.065 (0.135)	0.104 (0.144)	−0.005 (0.136)	0.169 (0.144)	−0.051 (0.146)	0.475*** (0.18)
log GDP p.c._e	0.1 (0.143)	0.347** (0.151)	0.128 (0.143)	0.371** (0.15)	0.046 (0.152)	0.367* (0.188)
Regional trade agreement	0.216*** (0.065)	0.223*** (0.064)	0.191*** (0.065)	0.198*** (0.063)	0.190*** (0.063)	0.167*** (0.057)
WTO_i	−0.065 (0.002)	−0.064 (0.022)	−0.065 (0.009)	−0.063 (0.015)	−0.063 (0.046)	−0.057 (0.047)
WTO_e	−0.043 (0.046)	−0.048 (0.100*)	−0.043 (0.038)	−0.048 (0.103*)	−0.043 (0.054)	−0.046 (0.033)
Observations	89,148	74,857	87,571	73,756	82,436	59,363
Trading pairs	4,974	4,255	4,779	4,157	4,972	4,184
R-squared (within)	0.48	0.52	0.48	0.53	0.41	0.38

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: The dependent variable is the log of imports. Numbers in parentheses are standard errors clustered at the trading pair level. $\text{Autocracy}_{i\ t-1}$ is the autocracy score lagged one year for importing countries, and $\text{Autocracy}_{e\ t-1}$ is the autocracy score lagged one year for exporting countries. $\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$ is the average of the five years prior to the observation for importing countries, and $\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$ is the average of the five years prior to the observation for exporting countries. See table 1 for definitions of other variables. All regressions contain trading pair and time-specific fixed effects that are significant at the 1 percent level.

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

Overall, these results show that the tariff channel, as identified by Milner and Kubota (2005), is not the only transmission mechanism at play. The model points to two alternative transmission channels (the accountability channel and the bureaucracy channel), and the findings here are consistent with the presence of both. The results are not driven by the fact that autocratic rulers would violate trade agreements if they formalized the rents from customs corruption in the form of high tariffs because trade agreements are controlled for in all the estimations. If this were the only reason why autocratic rulers distort trade more than democratic politicians do, no systematic regime differences would be found, conditional on trade agreements.[33]

The theory here suggests that the regime type of individual countries matters, whereas Morrow, Silverson, and Tabares (1998) and Mansfield, Milner, and Rosendorff (2000) argue that the congruence between the regime type of pairs of trading countries (two democracies, two autocracies, or a mixed pair) is what matters. It is thus important to test these alternatives against each other. Because including the dyadic regime type indicators and the individual country regime type indicators in the same specification leads to perfect collinearity,[34] each dyadic regime type indicator is included one at a time in specification (18). Using Przeworski's regime type indicator,[35] the following three dyadic regime type dummy variables are used: *two autocracies*, which equals 1 if the pair consists of two autocracies; *two democracies*, which equals 1 if the pair consists of two democracies; and *mixed regime*, which equals 1 if the pair consists of an autocracy and a democracy. To ensure that the findings of Mansfield, Milner, and Rosendorff (2000) can be replicated using the sample here, which covers more countries and a longer time period, a specification is also estimated without the regime type variables (not reported). In line with Morrow, Silverson, and Tabares (1998) and Mansfield, Milner, and Rosendorff (2000), pairs of democracies are found to trade more than mixed pairs (table 4). However, in contrast to Mansfield, Milner, and Rosendorff (2000), pairs of autocracies are found to trade less than pairs of democracies.

More important, when each dyadic regime type indicator is added to the baseline specification one at a time, the sign and significance of the individual country regime type indicators remain unaffected. This suggests that the regime type of an individual country matters for trade flows over and above any effect that might arise from the congruence or lack thereof with the political regime of the trading partner. So, congruence matters, but not in the way suggested by Mansfield, Milner, and Rosendorff (2000): conditional on a country's own

33. Formal trade agreements are noisy proxies of obligations toward trading partners and are thus an imperfect measure.

34. Mansfield, Milner, and Rosendorff (2000) take pairs of democracies as the reference group against which the impact of the other possible combinations of regimes is measured.

35. Prezworski's regime type indicator was also used by Mansfield, Milner, and Rosendorff (2000). Using it rather than the Polity IV or the Freedom House index avoids arbitrary decisions on how to define a particular regime type.

TABLE 4. Results of a "Horse Race" with Mansfield, Milner, and Rosendorff (2000)

Variable	Przeworski and others 2000		
Autocracy$_{i\ t-1}$	−0.213*** (0.050)	−0.096*** (0.034)	−0.154*** (0.034)
Autocracy$_{e\ t-1}$	−0.245*** (0.046)	−0.128*** (0.035)	−0.186*** (0.032)
Two democracies$_{t-1}$	−0.117** (0.051)	—	—
Two autocracies$_{t-1}$	—	−0.117** (0.051)	—
Mixed regime$_{t-1}$	—	—	0.059** (0.026)

**Significant at the 5 percent level; ***significant at the 1 percent level.

Note: The dependent variable is the log of imports. Numbers in parentheses are standard errors clustered at the trading pair level. The baseline specification is from table 3. Two democracies$_{t-1}$ is a dummy variable equal to 1 if both trading partners are democracies. Two autocracies$_{t-1}$ is a dummy variable equal to 1 if both trading partners are autocracies. Mixed regime$_{t-1}$ is a dummy variable equal to 1 if the trading pair consists of an autocracy and a democracy. All regressions contain trading pair and time-specific fixed effects that are significant at the 1 percent level.

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

political regime, pairs of autocracies and pairs of democracies trade less, while mixed pairs trade more. Of course, these results are not directly comparable to those reported by Mansfield, Milner, and Rosendorff (2000) because in this specification the coefficient on each dyadic regime type indicator can be interpreted as an interaction term rather than relative to pairs of democracies. Nevertheless, taken together, the political regime of a country is a key determinant of trade flows, while the congruence of its regime with those of its trading partners seems to be of only secondary importance.

V. ROBUSTNESS ANALYSIS

Perhaps the most important robustness check consists of estimating the single-country panel model (equation 19). To save space, the coefficients on the regime type indicators are reported only in the top two rows of table 5. As in the dyadic panel model, all three regime type indicators yield the same result: autocracies trade significantly less, conditional on trade policy. The size of the effect is, however, somewhat smaller than in the dyadic model for two of the regime type indicators (the Polity IV index and Przeworski's regime type indicator) but of comparable size for the Freedom House index. Overall, the findings based on the dyadic model are not a statistical artifact of the large number of observations. Moreover, the robust negative effect of autocracy on total imports and exports strengthens the assertion that the regime type of individual countries matters more for trade volumes than does the regime congruence between pairs of trading countries.

Numerous robustness checks were conducted to see whether the results reported in tables 2 and 3 are sensitive to changes in the specification and estimation method. The dyadic panel model with the restriction index (reported in

TABLE 5. Results from the Robustness Analysis

Technique	Variable	Przeworski and others 2000		Polity IV		Freedom House	
Single-country imports	$\text{Autocracy}_{i\ t-1}$	−0.064**	—	−0.006***	—	−0.055***	—
	$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	—	−0.106***	—	−0.011***	—	−0.066***
Single-country exports	$\text{Autocracy}_{e\ t-1}$	−0.055**	—	−0.004***	—	−0.029***	—
	$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	−0.086***	—	−0.009***	—	−0.041***
Reweighted least squares	$\text{Autocracy}_{i\ t-1}$	−0.055***	—	−0.007***	—	−0.030***	—
	$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	−0.155***	—	−0.013***	—	−0.048***	—
	$\text{Autocracy}_{e\ t-1}$	—	−0.073***	—	−0.011***	—	−0.030***
	$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	−0.128***	—	−0.010***	—	−0.043***
Least absolute value	$\text{Autocracy}_{i\ t-1}$	−0.056***	—	−0.007***	—	−0.030***	—
	$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	−0.160***	—	−0.013***	—	−0.051***	—
	$\text{Autocracy}_{e\ t-1}$	—	−0.079***	—	−0.012***	—	−0.031***
	$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	−0.128***	—	−0.010***	—	−0.042***
Autocratic dummy variables	$\text{Autocracy}_{i\ t-1}$	—	—	−0.180***	—	−0.089***	—
	$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	—	—	−0.113***	—	−0.102***	—
	$\text{Autocracy}_{e\ t-1}$	—	—	—	−0.277***	—	−0.117***
	$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	—	—	−0.220***	—	−0.050
Correcting for instability	$\text{Autocracy}_{i\ t-1}$	−0.150***	—	−0.014***	—	−0.058***	—
	$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	−0.180***	—	−0.013***	—	−0.055***	—
	$\text{Autocracy}_{e\ t-1}$	—	−0.187***	—	−0.018***	—	−0.072***
	$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	−0.205***	—	−0.013***	—	−0.048***
Poisson estimation	$\text{Autocracy}_{i\ t-1}$	−0.204***	—	−0.018***	—	−0.079***	—
	$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	−0.351***	—	−0.031***	—	−0.107***	—
	$\text{Autocracy}_{e\ t-1}$	—	−0.220***	—	−0.023***	—	−0.075***
	$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	−0.375**	—	−0.031*	—	−0.098**

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: The dependent variable is total imports or exports for single-country panel regressions. $\text{Autocracy}_{i\ t\text{-}1}$ is the autocracy score lagged one year for importing countries, and $\text{Autocracy}_{e\ t\text{-}1}$ is the autocracy score lagged one year for exporting countries. $\text{Autocracy}_{i\ t\text{-}1}$ to $_{i\ t\text{-}5}$ is the average of the five years prior to the observation for importing countries, and $\text{Autocracy}_{e\ t\text{-}1}$ to $_{e\ t\text{-}5}$ is the average of the five years prior to the observation for exporting countries. In the single-country panel regressions, the dependent variable is a country's total real import or export. The control variables are: GDP, GDP p.c., the restriction index as well as country- and time-fixed effects. In the dyadic panel regressions, the baseline specification is taken from Table 3. These regressions contain trading pair and time-specific fixed effects, all of which are significant at the 1% level with standard errors clustered at the trading pair level.

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

table 3) is used as the baseline, and several estimation techniques that reduce the risk of outliers drive the results. First, the model was reestimated using reweighted least squares, a robust regression technique that weighs observations in an iterative process.[36] Starting with ordinary least squares, estimates are obtained through weighted least squares, where observations with relatively large residuals get smaller weights. The coefficients remain highly significant, though their magnitude is somewhat reduced (see row three of table 5). When comparing the coefficients reported in tables 3 and 5, the coefficients on the political regime type indicator of importing countries are approximately halved, while the coefficients for exporting countries change only minimally. The least absolute value estimator, which minimizes the sum of the absolute deviations from the median, was also used.[37] The results are comparable to those obtained with the reweighted least squares estimator, and the regime type effect remains highly significant (see row four of table 5).

Second, the Polity IV and Freedom House indexes are often dichtomized to avoid the problem of dealing with an ordinal scale. This, of course, requires defining a threshold determining which countries to treat as democracies. For the Freedom House index a typical choice is 2.5, and for the Polity IV index, 0 (see, for example, Persson and Tabellini 2003).[38] The estimated coefficients of the two resulting regime type dummy variables remain highly significant and negative, with one exception.

Third, the results presented so far could be driven by regime instability rather than by regime type: if autocracies tend to be more unstable than democracies, the regime type indicators could simply be picking up instability. To rule out this possibility, a variable that counts the number of regime transitions to date for each country is included in the regressions. The estimated coefficients on the regime type indicators remain unaffected when political instability is controlled for in this way, so regime instability can be ruled out (see row six of table 5). Moreover, regime instability, while sometimes significant, is not a robust determinant of trade flows (not reported).

Fourth, the recent literature on estimation of dyadic gravity models stresses the importance of taking into account that many pairs of countries do not trade at all. Ignoring these (non-) flows might lead to biased estimates. Santos Silva and Tenreyro (2006) propose one way to circumvent this potential problem: using the maximum likelihood Poisson estimator instead of ordinary least squares. While the regime type effect for exporting countries loses some of its statistical significance, the results are very similar to those obtained with ordinary least squares.[39]

36. In this context, the term "robust" is used as robustness with respect to the dependent variable.

37. This is also known as mean absolute deviation or L1 norm regression.

38. This choice of threshold is not uncontroversial, however. Pevehouse (2002), for example, suggests a value of 6. Re-estimating using this stricter definition of what constitutes a democracy leaves the findings quantitatively unchanged.

39. To make the results obtained with this estimator comparable with the ordinary least squares results, the dependent variable is the volume of imports rather than the log of the volume of imports.

VI. Instrumental Variable Estimates

As argued by López-Córdova and Meisner (2008), among others, involvement in international trade may foster democratization. This could happen because trade changes relative factor rewards in such a way that the balance of power tilts in favor of social groups that benefit from democracy. If such considerations are important[40] and not effectively dealt with by lagging the regime type indicators,[41] the estimate of the effect of regime type on trade volumes suffers from a simultaneity bias: the coefficients on the regime type indicators are biased toward zero. It is also possible that, despite including fixed effects and many time-varying control variables, the estimates suffer from omitted variable bias or that measurement errors introduce significant biases.

To deal with these issues, the dyadic panel model was reestimated using instrumental variable techniques. To minimize the problem of weak instruments, four instruments for political regime type were used. The leading instrument is the number of successful assassinations of the political leader of a country. While an assassination attempt might be endogenous to regime type—for example, autocratic rulers might be more prone to such attacks—Jones and Olken (2009) contend that the outcome of the attempt is random. In particular, they show that successful assassination of an autocratic leader increases the probability of a transition to democracy. Following Jones and Olken (2009), the number of successful assassinations (per year) was used as an instrument for the regime type indicators.

A second instrument was the percentage of votes cast in line with the Group of Seven (G7) countries in the United Nations General Assembly. Dreher and Sturm (2006) show that more democratic countries are more likely to vote in line with the G7. While the G7 countries share a common set of interests, they do not always agree and vote together. So the risk that the voting pattern of a non-G7 country is driven by commercial interests with respect to one potential trading partner (such as the United States) can be avoided by using the percentage of its votes cast in line with the average vote of the G7 as the instrument. In other words, while it is possible that a non-G7 country supports the stance of a particular G7 country to further its trade interests, it is unlikely that the voting pattern of these countries with respect to the G7 as a whole is driven by trade considerations.

While these instruments have not been used in the literature before, two instruments that have been used were also included. Milner and Kubota (2005)

40. Gassebner, Lamla, and Vreeland (2009) find that trade openness is not robustly related to democratization processes. This casts some doubt on how much of a concern reverse causality is in practice.

41. All models were re-estimated using second to fifth lags of the regime type indicators separately (not reported). Doing so confirms the results reported in the text. In particular, the coefficients on all lags of the regime type indicators are highly significant and cannot be statistically distinguished from the coefficient on the one-year lag of the regime type indicator.

argue that the average age of a country's political parties and average school attainment can be used as instruments for regime type.[42] The rationale for using the average age of political parties as an instrument is that political parties tend to be younger in democracies than in autocracies because in democratic systems new parties can form as new political and economic interests emerge. An example is the wave of green parties established in Western Europe during the 1970s and 1980s. In autocratic systems, new parties typically do not form as easily. This instrument is clearly relevant. However, if trade causes democratization, the average age of parties could also be affected by trade volumes, but only going forward from the transition to democracy. By construct, the average age of political parties at a given point in time is looking backward, thereby breaking any direct correlation between contemporaneous trade volumes and the average age of the parties.

The rationale for using education as an instrument can be traced back to Lipset (1959), who argued that education is at least a necessary condition for democracy. While education through its effect on human capital certainly is related to the composition of trade through the principle of comparative advantage, it is unlikely to have a direct effect on the volume of trade.

Finding instruments for the political regime is notoriously difficult. All the instruments exhibit only limited time variability, so some endogeneity in the time domain might still prevail, and the results must be treated with caution.

Table 6 reports the second-stage regression results (with the four instruments included simultaneously), while table A1 in the appendix reports the results of the first-stage regression. The correlation between each instrument and the three regime type indicators confirms the above reasoning, as suggested in table A1. Moreover, in all specifications the first-stage *F*-statistics, indicating the relevance of the instruments, easily pass the threshold of 10 proposed by Staiger and Stock (1997).[43] The *p*-values associated with the Hansen *J*-test for overidentification show that the test always fails to reject at the 10 percent level. Perhaps the most convincing of the four instruments is the number of successful assassinations. Taking the validity of this instrument as given, the result of the *J*-test can be interpreted as a validation of the three other instruments.

For importing countries, the coefficient on the regime type indicators remains significant at the 1 percent level. In the presence of reverse causality, the instrumental variable estimates should be numerically larger than the ordinary least squares estimates. The point estimates are clearly larger in absolute size, but for the specification with a one-year lag the instrumental variable estimates are, in fact, not statistically different from the ordinary least squares results reported in table 3. This casts some doubt on whether reverse causality

42. The source of the former data is Beck and others (2001), while the source of the latter is Barro and Lee (2000).

43. The *F*-statistics range from 10.5 to 23.2. Furthermore, the Anderson canonical correlation, the Cragg-Donald statistics, and the Anderson-Rubin test suggest that the instruments are neither underidentified nor weak.

TABLE 6. Results from the Instrumental Variable Estimation

Variable	Przeworski and others 2000		Polity IV		Freedom House	
$\text{Autocracy}_{i\ t-1}$	−1.499** (0.625)	—	−0.066** (0.027)	—	−0.298** (0.122)	—
$\text{Autocracy}_{e\ t-1}$	0.374 (0.445)	—	0.027 (0.021)	—	0.129 (0.120)	—
$\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$	—	−1.575*** (0.369)	—	−0.079*** (0.02)	—	−0.355*** (0.089)
$\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$	—	0.144 (0.31)	—	0.017 (0.019)	—	0.056 (0.097)
Restriction index_i	0.203*** (0.038)	0.182*** (0.033)	0.139*** (0.027)	0.126*** (0.029)	0.193*** (0.034)	0.200*** (0.032)
Restriction index_e	−0.066** (0.033)	−0.030 (0.03)	−0.045* (0.026)	−0.017 (0.029)	−0.086** (0.036)	−0.047 (0.032)
log GDP_i	0.718** (0.292)	0.629* (0.339)	0.664** (0.294)	0.471 (0.341)	1.472*** (0.346)	1.496*** (0.401)
log GDP_e	1.018*** (0.250)	0.733** (0.323)	0.987*** (0.249)	0.692** (0.312)	0.709** (0.361)	0.533 (0.394)
log GDP p.c._i	0.807** (0.390)	0.950** (0.370)	0.826** (0.370)	1.159*** (0.375)	0.006 (0.279)	0.258 (0.354)
log GDP p.c._e	−0.177 (0.292)	0.168 (0.346)	−0.107 (0.280)	0.196 (0.342)	0.146 (0.290)	0.373 (0.353)
Regional trade agreement	0.209** (0.088)	0.125* (0.07)	0.143* (0.079)	0.061 (0.066)	0.166** (0.08)	0.057 (0.068)
WTO_i	0.032 (0.075)	0.022 (0.096)	−0.098 (0.078)	−0.190* (0.103)	−0.015 (0.07)	−0.053 (0.095)
WTO_e	−0.010 (0.092)	0.007 (0.120)	0.011 (0.087)	−0.006 (0.120)	−0.016 (0.084)	−0.022 (0.115)
Observations	32,580	22,369	32,026	22,226	33,111	22,833
Trading pairs	2,382	1,909	2,269	1,872	2,394	1,925
R-squared (within)	0.27	0.25	0.32	0.3	0.31	0.28
Hansen *J*-statistic (*p*-value)	0.84	0.98	0.39	0.17	0.37	0.17
First state *F*-statistic (*i*)	11.73	13.85	12.94	17.44	16.14	23.15
First state *F*-statistic (*e*)	12.07	14.79	10.48	16.86	15.88	20.15

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: The dependent variable is the log of imports. Numbers in parentheses are standard errors clustered at the trading pair level. $\text{Autocracy}_{i\ t-1}$ is the autocracy score lagged one year for importing countries, and $\text{Autocracy}_{e\ t-1}$ is the autocracy score lagged one year for exporting countries. $\text{Autocracy}_{i\ t-1\ \text{to}\ t-5}$ is the average of the five years prior to the observation for importing countries, and $\text{Autocracy}_{e\ t-1\ \text{to}\ t-5}$ is the average of the five years prior to the observation for exporting countries. See table 1 for definitions of other variables. All regressions contain trading pair and time-specific fixed effects that are significant at the 1 percent level. The Hansen *J*-statistic reports the *p*-value for the test of overidentification. The first stage *F*-statistic reports the first stage *F*-statistic of the excluded instruments for the importer (*i*) and exporter (*e*).

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

is a major issue (see also Gassebner, Lamla, and Vreeland 2009). By contrast, the coefficients on the regime type indicators for exporting countries become positive but are no longer significant at conventional levels. Again, this is inconsistent with reverse causality but may be driven by the fact that the ordinary least squares estimates suffer from omitted variables bias. For example, if autocracies systematically pursue an industrial policy that makes it hard for export sectors to flourish (while imports are unaffected by this), the ordinary least squares estimates will have a sizable downward bias that is then corrected by the instrumental variable estimator. Based on the instrumental variable estimates, autocracies import less than democracies, but regime type seems to matter less for exports.[44]

VII. Conclusions

The question that motivates this article is simple: does the political regime of a country systematically affect that country's involvement in international trade? The theoretical model provides two reasons why the answer is likely yes. In contrast to previous theoretical work, this article argues that the root cause of regime differences in trade flows is differences in political accountability. These differences affect trade flows directly through the impact on trade taxes (which are more prevalent in autocracies than in democracies), but they also work through a more subtle, indirect channel. Rulers of societies with weak accountability institutions have no incentive to strengthen these institutions by offering wage incentives to customs officials. The reason is complementarity between different aspects of the institutional environment. As a consequence, the theory suggests that not only do autocracies trade less but that they trade less conditional on official trade policy.

The implications of the model are tested not only within the framework of a standard dyadic (gravity) model of international trade but also within a single-country panel model where the outcome variable is total imports (or exports). These designs distinguish between the effects of the political regime of an importing and of an exporting country and differentiate the estimations from previous work on the congruence of the regime type of pairs of trading countries. Autocracies are found to trade significantly less than democracies, even after controlling for differences in trade policy. The magnitude of the effect is substantial: according to the most conservative estimates, autocracies have 5.5–22.9 percent less imports and 5.5–22.4 percent less exports, all else being equal. The results of the effect for importing countries are robust to a battery of different estimation techniques, including the use of instrumental variables. The instrumental variable estimates,

44. In principle the restriction index could also be endogenous, as trade patterns could affect trade policy. However, using second lags to instrument for the index does not suggest that this concern is important.

however, cast some doubt on the robustness of the relationship between regime type and exports.

Overall, the analysis shows that autocracies import less (and may export less) and that this effect is driven not only by differences in trade policy but also by systematic differences in political accountability. In other words, a democracy trades more with the rest of the world because democratically elected politicians are less tempted to use trade taxes to extract rents and because these politicians face the right incentives to build institutions that weed out trade-distorting red tape in the customs service. Of course, other explanations are also consistent with the results. For example, rulers must cater to powerful domestic elites who have a special interest in trade protection in many autocracies, as they often have a substantial stake in import-competing sectors. However, lobbying by special interests groups for trade protection is also common in democracies. It is, therefore, unclear if this line of reasoning leads to systematic regime differences in trade flows.

What policy conclusions can be drawn from this study? Trade integration is often seen as an engine of development. As Sachs and Warner (1995, p.2) put it, "Trade liberalization not only establishes powerful direct linkages between the economy and the world system, but also effectively forces the government to take actions on the other parts of the reform program . . . ." This study shows that autocracies are less integrated in world trade than democracies. Moreover, rather than fostering reforms, autocratic regimes are prone to red tape because of a lack of accountability. This missing trade link might be a reason—so far ignored—why many autocratic countries lag behind in development. This is why the "good governance agenda" fostered by the World Bank and many other international institutions is such an important initiative to counterbalance the negative impact that autocratic regimes have on economic integration.

Appendix

Table A-1. Instrumenting for the Regime Type Indicators: First-Stage Regression Results

	Przeworski and others 2000		Polity IV		Freedom House	
Variable	Importer	Exporter	Importer	Exporter	Importer	Exporter
$t-1$						
Assassination$_{i\ t-1}$	−0.099***	0.005	−0.652***	0.077	−0.305***	−0.001
Assassination$_{e\ t-1}$	0.011	−0.098***	0.055	−0.538***	0.011	−0.300***
Voting in line G7$_{i\ t-1}$	−0.419***	0.014	−10.612***	−0.188	−1.508***	0.003
Voting in line G7$_{e\ t-1}$	0.022	−0.429***	−0.319	−9.788***	−0.026	−1.385***
Average party age$_{i\ t-1}$	0.001***	6.0E-05	0.014***	−7.2E-05	0.003***	1.4E-04
Average party age$_{e\ t-1}$	−5.8E-05	0.001***	−0.001	0.014***	−3.7E-04	0.003***

(Continued)

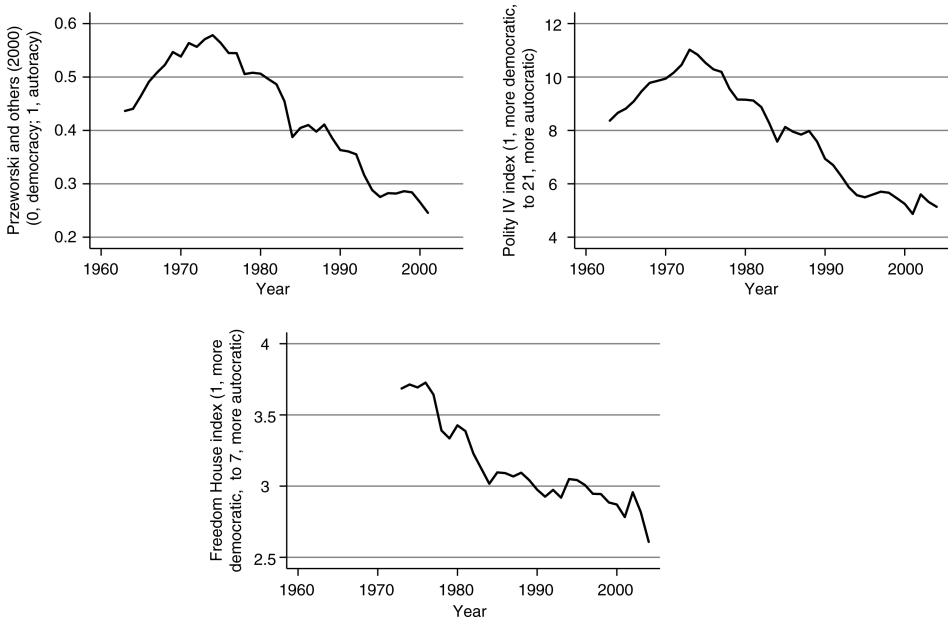
TABLE A-1. Continued

Variable	Przeworski and others 2000		Polity IV		Freedom House	
	Importer	Exporter	Importer	Exporter	Importer	Exporter
School attainment$_{i\ t-1}$	0.012	−0.004	0.092	−0.013	−0.105***	0.004
School attainment$_{e\ t-1}$	−0.001	0.023**	−0.014	0.163*	−0.005	−0.058*
t − 1 to t − 5						
Assassination$_{i\ t-1 \text{ to } t-5}$	−0.665***	−0.019	−2.994***	−0.136	−1.365***	−0.069
Assassination$_{e\ t-1 \text{ to } t-5}$	−0.018	−0.725***	−0.653	−1.890***	−0.098	−1.159***
Voting in line G7$_{i\ t-1 \text{ to } t-5}$	−1.010***	0.028	−20.257***	0.016	−2.834***	0.076
Voting in line G7$_{e\ t-1 \text{ to } t-5}$	0.048	−1.010***	0.156	−18.720***	0.107	−2.535***
Average party age$_{i\ t-1 \text{ to } t-5}$	0.001***	6.2E-05	0.022***	1.8E-04	0.006***	−4.2E-04
Average party age$_{e\ t-1 \text{ to } t-5}$	−1.5E-04	0.001***	−0.002	0.024***	−0.001	0.008***
School attainment$_{i\ t-1 \text{ to } t-5}$	0.009	−0.003	0.128	0.041	−0.175***	0.012
School attainment$_{e\ t-1 \text{ to } t-5}$	−0.002	0.025**	−0.013	0.207**	−0.006	−0.127***

Note: Results are for the first-stage of the instrumental variable regression in table 6. Only the results for the instruments are displayed. Each column represents one first-stage regression. The top part of the table shows the setup with the one period lag of the autocracy measure, while the bottom part gives the results of the approach using the average over the previous five periods.

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

FIGURE A-1. Political Regime Variables, 1962–2000



Source: Authors' analysis based on Przeworski and others (2000), Alvarez and others (1996), Gurr and others (2003), and Freedom House (2006).

References

Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2008. "Income and Democracy." *American Economic Review* 98(3):808–42.

Acemoglu, Daron, and James A. Robinson. 2001. "A Theory of Political Transitions." *American Economic Review* 91(4):938–63.

Adserà, Alícia, Carles Boix, and Mark Payne. 2003. "Are You Being Served? Political Accountability and Quality of Government." *Journal of Law, Economics, and Organization* 19(2):445–90.

Aidt, Toke S. 1997. "Cooperative Lobbying and Endogenous Trade Policy." *Public Choice* 93(3–4):455–75.

Alvarez, Mike, Jose A. Cheibub, Fernando Limongi, and Adam Przeworski. 1996. "Classifying Political Regimes." *Studies in Comparative International Development* 31(2):3–34.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93(1):170–92.

Baldwin, Richard. 2006. *'The Euro's Trade Effects.'* European Central Bank Working Paper WPS/594. Frankfurt, Germany: European Central Bank.

Baltagi, Badi H., Peter Egger, and Michael Pfaffermayr. 2003. "A Generalized Design for Bilateral Trade Flow Models." *Economics Letters* 80(3):391–97.

Barro, Robert. 1999. "Determinants of Democracy." *Journal of Political Economy* 107(S6):S158–83.

Barro, Robert, and Jong-Wha Lee. 2000. *'International Data on Educational Attainment: Updates and Implications.'* CID Working Paper 42. Cambridge, Mass: Harvard University, Center for International Development.

Beck, Thorsten, George Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh. 2001. "New Tools in Comparative Political Economy: The Database of Political Institutions." *World Bank Economic Review* 15(1):165–76.

Becker, Gary S., and George Stigler. 1974. "Law Enforcement, Malfeasance and the Compensation of Enforcers." *Journal of Legal Studies* 3(1):1–19.

Besley, Timothy, and John McLaren. 1993. "Taxes and Bribery: The Role of Wage Incentives." *Economic Journal* 103(416):119–41.

Besley, Timothy, and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand: Media Capture and Government Accountability." *American Economic Review* 96(3):720–36.

Boockmann, Bernhard, and Axel Dreher. 2003. "The Contribution of the IMF and the World Bank to Economic Freedom." *European Journal of Political Economy* 19(3):633–49.

Brennan, Geoffrey, and James M. Buchanan. 1980. *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge, U.K.: Cambridge University Press.

Brun, Jean-Francois, Céline Carrère, Patrick Guillaumont, and Jaime de Melo. 2005. "Has Distance Died? Evidence from a Panel Gravity Model." *World Bank Economic Review* 19(1):99–120.

Dai, Xinyuan. 2002. "Political Regimes and International Trade: The Democratic Difference Revisisted." *American Political Science Review* 96(1):159–65.

Daumal, Marie. 2008. "Federalism, Separatism and International Trade." *European Journal of Political Economy* 24(3):675–87.

Dreher, Axel. 2006. "Does Globalization Affect Growth? Evidence from a New Index of Globalization." *Applied Economics* 38(10): 1091–110.

Dreher, Axel, and Sarah Rupprecht. 2007. "IMF Programs and Reform—Inhibition or Encouragement?" *Economics Letters* 95(3):320–26.

Dreher, Axel, and Jan-Egbert Sturm. 2006. *Do IMF and World Bank Influence Voting in the UN General Assembly?* CESifo Working Paper 1724. Munich, Germany: CESifo.

Feenstra, Robert. 2000. *NBER-United Nations Trade Data 1962–2000*. (http://cid.econ.ucdavis.edu/data/undata/undata.html).

———. 2004. *Advanced International Trade: Theory and Evidence*. Princeton, N.J.: Princeton University Press.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1–3):5–25.

Freedom House. 2006. *Freedom in the World 2006: The Annual Survey of Political Rights and Civil Liberties*. Washington, D.C.: Freedom House.

———. 2009. *Freedom of the Press*. Washington, D.C.: Freedom House.

Gassebner, Martin, Noel Gaston, and Michael J. Lamla. Forthcoming. "The Inverse Domino Effect: Are Economic Reforms Contagious?" *International Economic Review*.

Gassebner, Martin, Michael J. Lamla, and James R. Vreeland. 2009. *Extreme Bounds of Democracy*. KOF Working Papers 224. Zurich, Switzerland: ETH Zurich, KOF Swiss Economic Institute.

Gehlbach, Scott, and Konstantin Sonin. 2008. *Government Control of the Media*. Madison: University of Wisconsin and Mascow: New Economic School. (http://ssrn.com/abstract=1315882).

Giavazzi, Francesco, and Guido Tabellini. 2005. "Economic and Political Liberalizations." *Journal of Monetary Economics* 52(7):1297–330.

Gorodnichenko, Yuriy, and Klara Sabirianova Peter. 2007. "Public Sector Pay and Corruption: Measuring Bribery from Micro Data." *Journal of Public Economics* 91(5–6):963–91.

Grossman, Gene M., and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review* 84(4):833–50.

Gurr, Ted R., Keith Jaggers, and Will Moore. 2003. *Polity Handbook IV*. Boulder, Colo.: University of Colorado Press.

Hillman, Arye L. 1982. "Declining Industries and Political-Support Protectionists Moves." *American Economic Review* 72(5):1180–90.

Jones, Benjamin F., and Olken, Benjamin A. 2009. "Hit or Miss? The Effect of Assassinations on Institutions and War." *American Economic Journal: Macroeconomics* 1(2):55–87.

Knack, Stephen, and Keefer, Philip. 1995. "Institutions and Economic Performance: Cross-country Tests Using Alternative Institutional Measures." *Economics & Politics* 7(3):207–27.

Li, Quan, and Rafael Reuveny. 2003. "Economic Globalization and Democracy: An Empirical Analysis." *British Journal of Political Science* 33(1):29–54.

Lipset, Seymour M. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53(1):69–105.

López-Córdova, J. Ernesto, and Christopher M. Meisner. 2008. "The Impact of International Trade on Democracy: A Long-run Perspective." *World Politics* 60(4):539–75.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2000. "Free to Trade: Democracies, Autocracies, and International Trade." *American Political Science Review* 94(2):305–21.

———. 2002. "Replication, Realism, and Robustness: Analyzing Political Regimes and International Trade." *American Political Science Review* 96(1):167–69.

Mayer, Wolfgang. 1984. "Endogenous Tariff Formation." *American Economic Review* 74(5):970–85.

Milner, Helen V., and Keiko Kubota. 2005. "Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries." *International Organization* 59(1):107–43.

Morrow, James D., Randolph M. Siverson, and Tressa E. Tabares. 1998. "The Political Determinants of International Trade: The Major Powers, 1907–90." *American Political Science Review* 92(3):649–61.

Nieswiadomy, Michael L., and Mark C. Strazicich. 2004. "Are Political Freedoms Converging?" *Economic Inquiry* 42(2):323–40.

Olson, Mancur. 2000. *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships*. New York: Basic Books.

Persson, Torsten, and Tabellini, Guido. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, Mass.: MIT Press.

Persson, Torsten, and Guido Tabellini. 2003. *The Economic Effects of Constitutions*. Cambridge, Mass.: MIT Press.

Pevehouse, Jon C. 2002. "With a Little Help from my Friends? Regional Organizations and the Consolidation of Democracy." *American Journal of Political Science* 46(3):611–26.

Plümper, Thomas, and Eric Neumayer. 2009. "The Level of Democracy during Interregnum and Affected Transition Periods: Recoding the polity2 Score." London: London School of Economics.

Przeworski, Adam, Mike Alvarez, Jose A. Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Regimes and Economic Well-being in the World, 1950–1990*. New York: Cambridge University Press.

Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-level Games." *International Organization* 42(3):427–60.

Rauch, James E., and Peter B. Evans. 2000. "Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries." *Journal of Public Economics* 75(1):49–71.

Rigobon, Roberto, and Dani Rodrik. 2005. "The Rule of Law, Democracy, Openness and Income: Estimating the Interrelationships." *Economics of Transition* 13(3):533–64.

Rose, Andrew K. 2004. "Do We Really Know that the WTO Increases Trade?" *American Economic Review* 94(1):98–114.

Sachs, Jeffrey D., and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1995(1):1–118.

Santos Silva, Joao M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics* 88(4):641–58.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 63(3):557–86.

Sturm, Jan-Egbert, and Jakob de Haan. 2003. "Does More Democracy Lead to Greater Economic Freedom? New Evidence for Developing Countries." *European Journal of Political Economy* 19(3):547–63.

Van Rijckeghem, Caroline, and Beatrice Weder. 2001. "Bureaucratic Corruption and the Rate of Temptation: Do Wages in the Civil Service Affect Corruption, and by How Much?" *Journal of Development Economics* 65(2):307–31.

Vreeland, James R. 2008. "The Effect of Political Regime on Civil War: Unpacking Anocracy." *Journal of Conflict Resolution* 52(3):401–25.

World Bank. 2006. *World Development Indicators 2006*. CD-ROM. World Bank, Washington, D.C.

Financial Institutions and Markets across Countries and over Time: The Updated Financial Development and Structure Database

Thorsten Beck, Aslı Demirgüç-Kunt, and Ross Levine

This article introduces the updated and expanded version of the Financial Development and Structure Database. The database includes indicators on the size, efficiency, and stability of banks, nonbank financial institutions, and equity and bond markets over 1960–2007. It also contains indicators of financial globalization.
JEL codes: G1, G2

The current financial crisis has moved the financial sector yet again to the top of the policy agenda. A large literature shows that the financial sector affects the rate of economic growth and the distribution of income. When the financial system goes awry and fails, it can devastate the lives of many people, as the world is currently experiencing.[1] As proper measurement is essential for analyzing causes and designing solutions, indicators measuring the size, activity, efficiency, and stability of the financial system are important for analysts, researchers, and policymakers alike.

This article introduces the updated and expanded version of the Financial Development and Structure Database. It provides statistics on the size, activity, efficiency, and stability of banks, nonbanks, equity markets, and bond markets across a broad spectrum of countries over time. It also contains several indicators of financial globalization, including statistics on international bond issues, international loans, off-shore deposits, and remittance flows.

Thorsten Beck (corresponding author, T.Beck@uvt.nl) is professor of economics; a fellow at CentER, the business and economics research center of Tilburg University; and Chair of the European Banking Center at Tilburg University. Aslı Demirgüç-Kunt (Ademirguckunt@worldbank.org) is a senior research manager, Finance and Private Sector Development, in the Development Economics Research Group at the World Bank. Ross Levine (ross_levine@brown.edu) is the James and Merryl Tisch Professor of Economics and the director of the William R. Rhodes Center for International Economics at Brown University.

1. For an overview of the extensive literature on the finance and growth relationship, see Levine (2005) and Beck (2009). For more recent evidence on the relationship among finance, income equity, and poverty reduction, see Beck, Demirgüç-Kunt, and Levine (2007), World Bank (2007), and Demirgüç-Kunt and Levine (2009).

THE WORLD BANK ECONOMIC REVIEW, VOL. 24, NO. 1, pp. 77–92 doi:10.1093/wber/lhp016
Advance Access Publication January 11, 2010

The database draws on a wide array of primary sources to cover different dimensions of the financial system. The database and further details on its construction are available at http://econ.worldbank.org/financialstructure.

First published in 1999, the Database of Financial Development and Structure has sparked considerable cross-country analytical work, inside and outside the World Bank.[2] The database has also sparked further efforts within the Bank to collect financial sector indicators, on both a global and a regional level, and to benchmark countries.

The revised database contains a select number of financial system indicators that are readily available for a large number of countries over 1960–2007. This necessarily excludes certain indicators that are available only for a small number of countries (such as detailed stock market liquidity or primary bond indicators) or for a few points in time (such as most indicators of how much individuals access the formal financial system). Compared with the original version of the database (as described in Beck, Demirgüç-Kunt, and Levine 2000), the revised version leaves out several indicators that were rarely used in the literature and can easily be constructed as ratios of other variables and several indicators for which it is difficult to access raw data across a large number of countries and a long period. Finally, the data are available annually and thus do not capture shorter-term trends.

This article uses the database to illustrate global financial system trends over recent decades. It shows that financial systems have continued to deepen along many dimensions, with rising values for standard indicators of financial intermediary and market development. However, progress has been uneven across income groups and regions. The deepening has been concentrated in high-income countries, with much less deepening in middle- and low-income countries.

Since the data end in 2007, they do not fully capture the recent crisis. However, indicators of banking efficiency, profitability, and stability match the trends in the boom period leading up to the recent global financial crisis. Specifically, the lower margins for traditional lines of business and the search for higher returns were possible only through high-risk taking, especially in high-income countries.

Integration into global financial markets has also increased, though the driving forces have differed for different income groups. While the increase in international lending and bond issues has been concentrated in high-income countries, low- and lower middle-income countries have benefited from higher remittance flows. Also, the ratio of off-shore deposits to domestic deposits is higher in low-income countries than in middle- and high-income countries,

2. The working paper version (Beck, Demirgüç-Kunt, and Levine 1999) is among the top 1 percent of papers cited in Research Papers in Economics (Repec), and the published version in the World Bank Economic Review is among the top 10 cited articles of the journal (Beck, Demirgüç-Kunt, and Levine 2000).

perhaps reflecting a lack of trust in domestic banking systems, though the ratio has halved over the past 12 years.

Any cross-country data collection effort is subject to biases due to different degrees of measurement quality across countries, as well as different accounting standards. However, such concerns are reduced by the fact that the raw data all come from one source, such as the International Monetary Fund's (IMF various years) International Financial Statistics, the BankScope database (Bureau van Dijk Electronic Publishing various years), and Swiss Re (various years). There are also concerns about coverage, especially for indicators based on the raw data from individual banks, as the coverage is incomplete. While this might result in certain measurement errors, internal World Bank comparisons with individual country data have mostly confirmed the reliability of the data.

Indicators of the size of financial systems are in section I; indicators of the structure, efficiency, and stability of commercial banks are in section II; indicators of the size and activity of capital markets and insurance sectors are in section III; and indicators of financial globalization are in section IV. Section V points to areas requiring additional research to provide data on important missing indicators. The working paper version of the article provides more detailed discussion of individual data series and their development across income groups and over time (Beck, Demirgüç-Kunt, and Levine 2009).

I. The Size of the Financial System

This section on the size of the financial system focuses on banks, bank-like financial institutions, equity markets, and private bond markets. (Table 1 lists the variables in the database and their coverage periods.) The indicators on financial intermediary development are based on the raw data from the International Financial Statistics from the IMF (various years), the equity market indicators on raw data from the Emerging Market Database (Standard & Poor's various years), and the bond market indicators on raw data from the Bank for International Settlements (BIS).

Liquid liabilities to GDP is a traditional indicator of financial depth. It is the ratio of currency plus demand and interest-bearing liabilities of banks and other financial intermediaries to GDP. *Liquid liabilities* is the broadest available indicator of financial intermediation; it includes all banks and bank-like and nonbank financial institutions.[3] There is wide cross-country variation in liquid liabilities to GDP, ranging from 395 percent in Luxembourg to less than

3. The International Financial Statistics of the IMF distinguish three groups of financial institutions. The first group comprises the central bank and other monetary authorities. The second group, deposit money banks, comprises all financial institutions with "liabilities in the form of deposits transferable by check or otherwise usable in making payments" (IMF 1984, p. 29). The third group, other financial institutions, comprises other bank-like institutions and nonbank financial institutions that serve as financial intermediaries, while not incurring liabilities usable as means of payment.

TABLE 1. Time Coverage of Variables

Variable	Coverage
Deposit money/central bank assets	1960–2007
Liquid liabilities/GDP	1960–2007
Central bank assets/GDP	1960–2007
Deposit money bank assets/GDP	1960–2007
Other financial institutions assets/GDP	1960–2007
Private credit by deposit money banks/GDP	1960–2007
Private credit by deposit money banks and other financial institutions/GDP	1960–2007
Bank deposits/GDP	1960–2007
Financial system deposits/GDP	1960–2007
Bank credit/bank deposits	1960–2007
Currency outside banking system/base money	1960–2007
Liquid liabilities (millions of 2000 U.S. dollars)	1960–2007
Overhead costs	1987–2007
Net interest margin	1987–2007
Bank concentration	1987–2007
Bank return on assets	1987–2007
Bank return on equity	1987–2007
Bank cost–income ratio	1987–2007
Bank *z*-score	1987–2007
Life insurance penetration	1960–2007
Nonlife insurance penetration	1987–2007
Stock market capitalization/GDP	1976–2007
Stock market total value traded/GDP	1975–2007
Stock market turnover ratio	1976–2007
Number of listed companies per 10,000 people	1975–2007
Private bond market capitalization/GDP	1990–2007
Public bond market capitalization/GDP	1990–2007
International debt issues/GDP	1988–2007
Loans from nonresident banks (net)/GDP	1993–2007
Loans from nonresident banks (amt outstanding)/GDP	1995–2007
Offshore bank deposits/domestic bank deposits	1995–2007
Remittance inflows/GDP	1970–2007

Source: See text for details.

1 percent in Sudan. Variation in the absolute size of financial systems is even greater, as illustrated by *liquid liabilities in USD*. On the one extreme, there are financial systems with trillions of U.S. dollars, such as Japan or the United States. On the other extreme, there are small and poor countries with financial systems not even as large as a single small bank in a developed country.

Currency outside banking system to base money is the share of base money that is not held as bank deposits. The level and change in currency outside the banking sector are frequently used as an estimate of the underdevelopment of the formal financial system (Schneider and Ernste 2000).

Financial system deposits to GDP is the ratio of all checking, savings, and time deposits in banks and bank-like financial institutions to economic activity and is a stock indicator of deposit resources available to the financial sector for

its lending activities. The database also contains an indicator limited to deposits of deposit monetary institutions, *bank deposits to GDP*.

While the previous indicators measure the liability side of financial intermediaries' balance sheets, indicators of the asset side capture credit allocation, one of the most important functions of financial intermediaries. *Private credit by deposit money banks and other financial institutions to GDP* is the ratio of claims on the private sector by deposit money banks and other financial institutions to GDP. It is a standard indicator in the finance and growth literature; countries with higher levels of private credit to GDP have been shown to grow faster and experience faster rates of poverty reduction (King and Levine 1993; Beck, Levine, and Loayza 2000; Beck, Demirgüç-Kunt, and Levine 2007). A somewhat narrower indicator—limited to deposit money banks—is *private credit by deposit money banks to GDP*.[4]

The size of equity markets is captured by *stock market capitalization to GDP*, or the ratio of the value of listed shares to GDP. It indicates the size of the stock market relative to the size of the economy. An indicator of the importance of private bond markets, *private bond market capitalization to GDP* is the ratio of the total amount of outstanding domestic debt securities issued by private or public domestic entities to GDP. Because of limited underlying raw data, this indicator is available for only 42 countries and only since 1990.

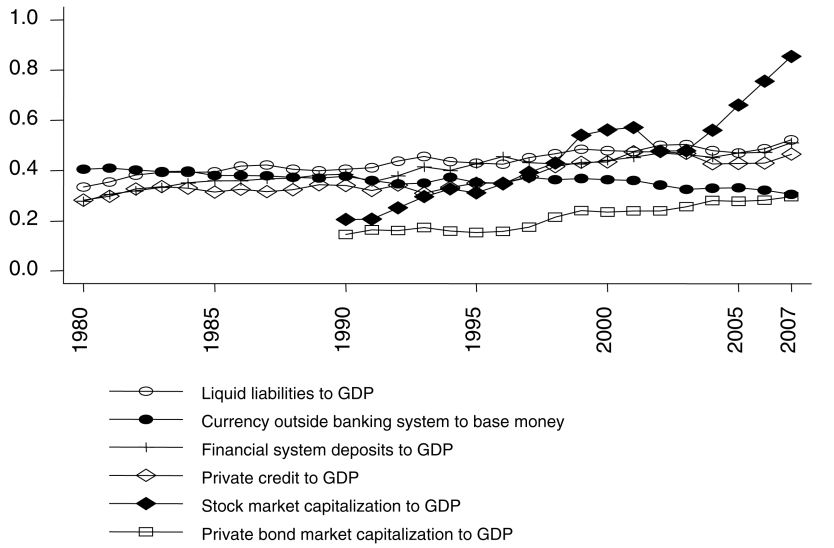
Figure 1 shows the development of these indicators between 1980 and 2007 for the sample of countries for which each indicator is available. While currency outside the banking system has decreased over time across countries, the indicators of the size of the financial system all show a positive trend line, with a rapid increase starting in 2005. Most notably, the ratio of stock market capitalization to GDP almost doubled between 2003 and 2007.

While these size indicators have been popular in the academic literature and with analysts, it is important to recognize that they are only proxies for financial sector development and that bigger is not always better, as the recent crisis has shown. Demirgüç-Kunt and Detragiache (1998, 2002) show that credit growth is a good crisis predictor, and Loayza and Ranciere (2006) show that while higher levels are associated with higher growth in the long term, there is a short-term negative relationship between credit levels and growth in GDP per capita. Critically, Beck and Levine (2002) show that the component of private credit to GDP explained by legal fundamentals is associated with economic growth, while Demirgüç-Kunt and Maksimovic (2002) show that credit beyond the level predicted by legal system efficiency and macroeconomic stability is not associated with faster firm growth.

Another important caveat is that these size measures are only proxies for the financial sector's function of allocating savings to their best uses. The size indicators do not capture directly the efficiency with which financial institutions

4. This measure does not distinguish between banks of different ownership types. Also, it does not include securitized loans, as it refers only to loans on the balance sheet of banks.

FIGURE 1. Financial System Size Indicators, Median Value over Time, 1980–2007



Source: Authors' analysis of Financial Development and Structure Database (http://econ.worldbank.org/financialstructure).

and markets undertake this role. Some indicators presented in the next sections measure the efficiency of financial institutions and markets, which in turn can proxy for allocation efficiency.

II. THE BANKING SYSTEM—SIZE, STRUCTURE, EFFICIENCY, AND STABILITY

The banking system constitutes the largest part of the financial system in most countries, especially in emerging and developing market economies. The database therefore includes an array of indicators measuring the size, structure, efficiency, and stability of banks across countries and over time. Additional indicators have been added that gauge the efficiency, profitability, and stability of banking sectors.

Based on the raw data from the International Financial Statistics, three indicators measure the claims on the whole nonfinancial real sector (including government, public enterprises, and the private sector) by three types of financial institutions: *central bank assets to GDP*, *deposit money banks assets to GDP*, and *other financial institutions assets to GDP*. Also included is a measure of the importance of commercial banks relative to the central bank, *deposit money to central bank assets*. Countries where deposit money banks have a larger role in financial intermediation than do central banks are considered to have higher levels of financial development (King and Levine 1993).

Several indicators of intermediation efficiency are included. First, bank credit to bank deposits is the ratio of claims on the private sector to deposits in deposit money banks. It gauges the extent to which banks funnel credit to the private sector. The variation in 2007 was large—between 21 percent in the Republic of Congo and 307 percent in Denmark. Obviously, deposits are not the only funding source of banks, and credits are not the only assets that banks hold, so a ratio much above one suggests that private sector lending is also funded with nondeposit sources, which could result in funding instability such as that experienced recently by many banks and countries in Central and Eastern Europe.

Second, *net interest margin* is the accounting value of a bank's net interest revenue as a share of its total earning assets, while *overhead cost* is the accounting value of a bank's overhead costs as a share of its total assets. Unlike the previous banking system indicators, these two variables are constructed from raw bank-level data from the BankScope database. Both measures are unweighted averages across all banks of a country for a given year.[5] Higher levels of net interest margins and overhead costs indicate lower levels of banking efficiency, as banks incur higher costs and there is a higher wedge between lending and deposit interest rates. Net interest margins have declined over time, with margins in high-income countries recently falling from already low levels (figure 2).[6]

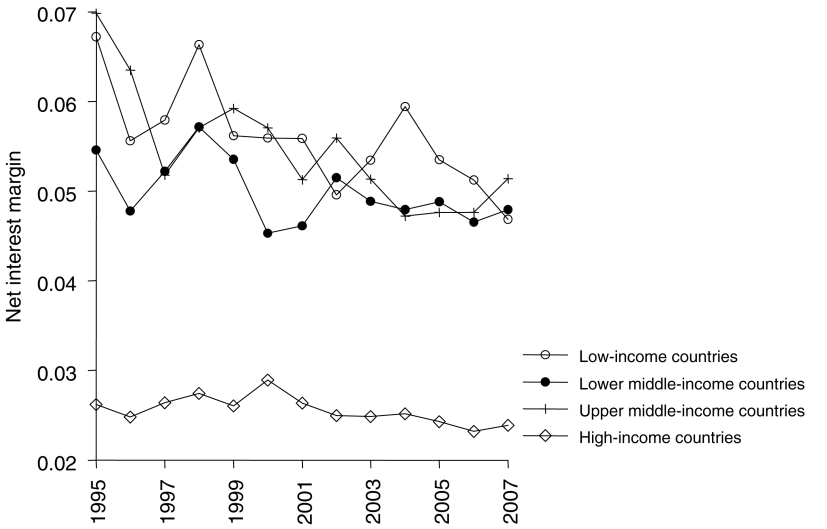
The final indicator of banking efficiency, *cost–income ratio*, measures overhead costs relative to gross revenues, with higher ratios indicating lower levels of cost efficiency. As in the case of net interest margins and overhead costs, data on cost–income ratios are based on the bank-level data. Banks in richer countries typically have lower cost–income ratios.

Concentration, an indicator of banking market structure, is the ratio of the three largest banks' assets to total banking sector assets. The indicator is based on the bank-level data from BankScope, which raises measurement concerns. Since BankScope's bank coverage is not complete, variation across countries and time might be driven by differences in coverage rather than differences in the market structure. There is no clear correlation between concentration and income levels of countries. The median country in the upper middle-income category has the lowest concentration ratio, while the median country in the low-income category has the highest ratio. Concentration is frequently used as an indicator of the lack of banking system competitiveness, though Claessens and Laeven (2004) find a very low correlation between concentration and other measures of banking competitiveness. Nevertheless, in the absence of more

5. BankScope coverage is less than 100 percent of most countries' banking sector. This poses relatively few problems for efficiency measures but causes greater difficulty for measures of market structure, as discussed later.

6. Spreads denote the difference between *ex ante* contracted loan and deposit interest rates, while margins are the interest (and noninterest) revenue actually received on loans minus the interest costs on deposits (minus noninterest charges on deposits). The main difference between spreads and margins is lost interest revenue on nonperforming loans, so spreads are normally higher than margins.

FIGURE 2. Net Interest Margin, Median Values by Income Group over Time, 1995–2007



Source: Authors' analysis of Financial Development and Structure Database (http://econ.worldbank.org/financialstructure).

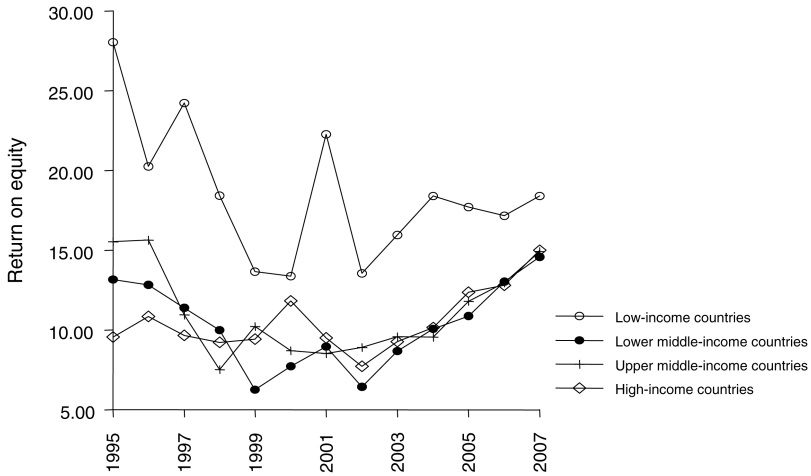
detailed banking level data, concentration ratios are still the most readily available market structure indicator across countries and over time.[7]

Two indicators of profitability, *return on assets* and *return on equity*, are computed as unweighted averages across all banks in a given year. While banks in the median country in high- and middle-income countries have a return on equity of around 15 percent, with little variation between high-, upper middle-, and lower middle-income countries, the median return on equity was more than 20 percent in 2007 in low-income countries. Return on equity shows considerable variation over time, declining from 12 percent in 1995 to 8 percent in 2002, before rising again to pass 15 percent in 2007 (figure 3). Similarly, returns on assets first declined then rose, a trend repeated across median countries of all income groups.

Finally, the z-score, an indicator of banking stability, was added to the database. The z-score is the ratio of return on assets plus the capital–asset ratio to the standard deviation of return on assets. If profits are assumed to follow a normal distribution, the z-score is the inverse of the probability of insolvency. Specifically, z indicates the number of standard deviations below the expected value of a bank's return on assets at which equity is depleted and the bank is insolvent (Roy 1952; Hannan and Henwick 1988; Boyd, Graham, and Hewitt

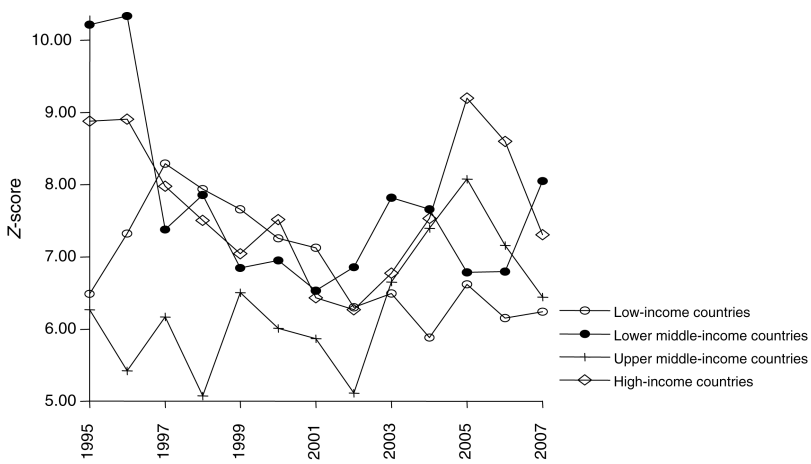
7. The original version of this database also contained indicators of foreign and government ownership of banks. However, the ownership information provided by BankScope was found to be inaccurate in many cases, so these indicators are not included. Barth, Caprio, and Levine (2008) provide ownership information on a less than annual frequency.

FIGURE 3. Bank Return on Equity, Median Values by Income Group over Time, 1995–2007



Source: Authors' analysis of Financial Development and Structure Database (http://econ.worldbank.org/financialstructure).

FIGURE 4. Bank *z*-Score, Median Values by Income Group over Time, 1995–2007



Source: Authors' analysis of Financial Development and Structure Database (http://econ.worldbank.org/financialstructure).

1993; De Nicolo 2000). Thus, a higher *z*-score indicates that the bank is more stable. As measured by the *z*-score, the bank stability varies across income groups and even more over time. The *z*-score has been declining since 1995, and following some increases in early 2000s, the declines since 2005 have been substantial for high- and upper middle-income countries (figure 4).

The banking trends documented in recent years through 2007 match those of a boom period leading up to a global financial crisis, especially in high-income countries. Among them were low and declining net interest margins (forcing banks to look for alternative income sources); rising profitability, as proxied by higher returns on assets and equity; and declining stability, evidenced by lower *z*-scores. With increasing returns on assets, the lower *z*-scores in the years leading up to 2007 can be explained either by lower capital—likely in the context of the transition toward Basel II standards in many high- and middle-income countries—or higher volatility of returns. Low- and lower middle-income countries have not shown similar banking trends, but they have also not experienced the same degree of financial deepening as high-income countries. In hindsight, these recent trends of low banking margins, a search for higher profits, and declining stability in high-income countries illustrate the circumstances surrounding the financial sector boom that led to the 2007 financial crisis.

III. Capital Markets and the Insurance Sector

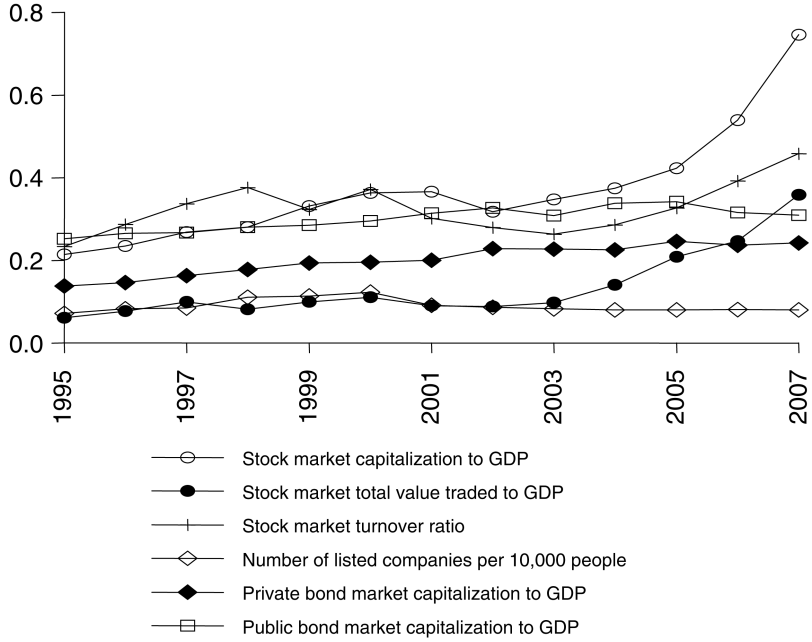
As did the original database, the new database includes several indicators of capital market development and the size of the insurance sector. The equity market indicators are based on the raw data from Standard & Poor's Emerging Markets Database; the bond market indicators are based on raw data from the Bank for International Settlements banking statistics, and the insurance data are based on the raw data from Swiss Re.

As discussed, *stock market capitalization to GDP* is the ratio of listed shares to GDP, an indicator of the size of the stock market relative to the size of the economy. *Stock market total value traded to GDP* is the ratio of total shares traded on the stock exchange to GDP, a measure of the degree of liquidity that stock markets provide to the economy. *Stock market turnover ratio* is the ratio of the value of total shares to market capitalization, a measure of the activity or liquidity of a stock market relative to its size. A small but active stock market will have a high turnover ratio, whereas a large but less liquid stock market will have a low turnover ratio. Finally, the ratio of *listed firms to population* is the share of listed companies divided by total population.

Both stock market capitalization and value traded have been increasing since 1995, while the ratio of listed firms to population does not show a clear trend (figure 5). The turnover ratio has risen slightly since 2003, but the rise has been less pronounced than the increases for the ratios of capitalization and trading to economic activity. Thus, the prices of existing stocks rather than the listing of new enterprises or greater stock market liquidity have driven stock market development in recent years. This is an important observation, as cross-country comparisons have shown that it is stock market liquidity rather than its size that matters for economic growth (Levine and Zervos 1998; Beck and Levine 2004).

Indicators of the size of the domestic bond market, *private bond market capitalization to GDP* and *public bond market capitalization to GDP*, are the

FIGURE 5. Stock, Bond, and Insurance Market Indicators, Median Values over Time, 1995–2007



Source: Authors' analysis of Financial Development and Structure Database (http://econ.worldbank.org/financialstructure).

ratio of total outstanding domestic debt securities issued by private and public domestic entities to GDP. These indicators measure the size of the market for public and private bonds relative to the real economy. While private bond market capitalization is positively correlated with country income levels, there is no clear correlation for public bond markets. There has been an upward trend in both indicators since 1995, although emphatically less than for stock markets. This is not surprising, as bonds are typically traded around their nominal value, unlike shares whose prices can be many times the original book value (see figure 5).

Indicators of the size of the insurance sector are *life insurance penetration*, the ratio of life insurance premiums to GDP, and *nonlife insurance penetration*, the ratio of nonlife insurance premiums to GDP. Both indicators measure total premium revenue in life and nonlife insurance business lines relative to economic activity. As the premium volume is the quantity of insurance coverage times its price, higher volumes can indicate either a deeper insurance market or less competition or efficiency. Both indicators increase with the income level of a country. This correlation is much stronger for life than for nonlife insurance, which is not surprising as life insurance is typically considered more income elastic than is nonlife insurance, such as motor vehicle or business

insurance policies.[8] Nonlife insurance and especially life insurance products have experienced an upward trend in recent years.

IV. Indicators of Financial Globalization

Unlike previous versions, the updated database includes several indicators of how well a country's financial system is linked to international financial markets. All these indicators are outcome variables, unlike those in much of the literature, which are *de jure* indicators of capital account or equity market liberalization. The new dataset includes only a select number of indicators of financial globalization that are not included in other datasets.[9]

International debt to GDP measures the stock of outstanding international bonds relative to a country's economic activity, while *international debt issues to GDP* measures the net flow of international bond issues relative to a country's economic activity. Both outstanding and new issues of international debt increase with countries' income level. Outstanding debt has risen annually since 1995, driven mostly by high-income countries; there have been fewer increases in middle- and low-income countries (figure 6).

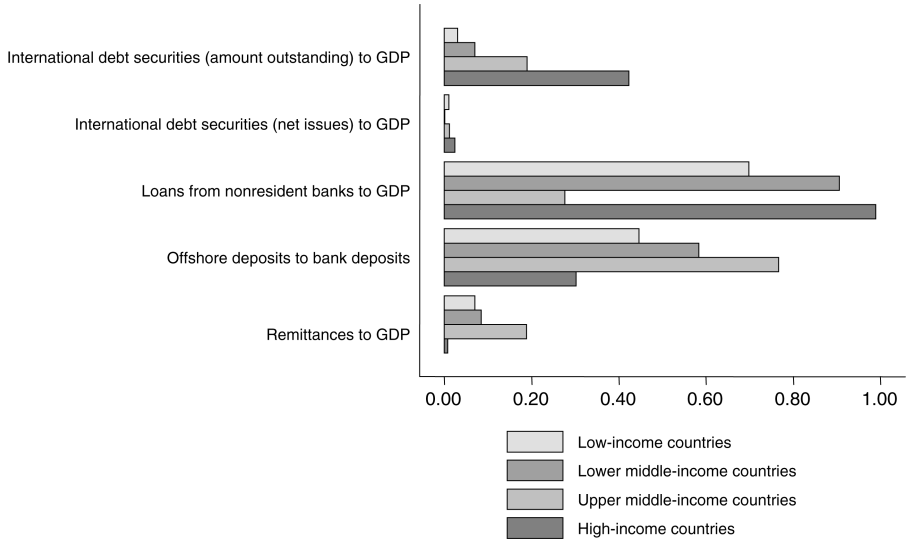
International loans from nonresident banks to GDP is the ratio of a country's loans of Bank for International Settlements reporting banks to the country's economic activity.[10] *Off-shore deposit to domestic deposits* is the ratio of deposits held by a country's nationals in off-shore banks to deposits in domestic banks. International loans increase with the country income level, while the ratio of off-shore deposits to domestic deposits is highest for low-income countries and decreases with the country income level. This can be partly explained by the lack of confidence that households and enterprises have in domestic banking systems, a phenomenon especially pronounced in Sub-Saharan Africa (Honohan and Beck 2007). International loans have been increasing in high-income countries, while remaining relatively stable in middle- and low-income countries. Off-shore deposits relative to domestic deposits have been low and stable in high- and middle-income countries, while

8. Different taxation of life insurance policies might explain variations in the size of the insurance markets even across countries at similar levels of economic development. For an in-depth study of the cross-country determinants of life insurance consumption, see Beck and Webb (2003). For an exploration of the relationship between insurance sector development and economic growth, see Arena (2008).

9. See, for example, Lane and Milesi-Ferretti (2008) on international asset and liability positions across countries; Chinn and Ito (2006) on *de jure* measures of capital account openness, and Bekaert and Harvey (2000) on equity market liberalization.

10. Bank for International Settlements reporting banks include banks residing in Australia, Austria, the Bahamas, Bahrain, Bermuda, Brazil, Cayman Islands, Chile, Denmark, Finland, Greece, Guernsey, Hong Kong (China), India, Ireland, Isle of Man, Jersey, Korea, Luxembourg, Macao (China), Malaysia, Mexico, the Netherlands Antilles, Norway, Panama, Portugal, Singapore, Spain, Taiwan (China), and Turkey.

FIGURE 6. Financial Globalization Indicators, Mean Values by Income Group, 2003–07



Source: Authors' analysis of Financial Development and Structure Database (http://econ.worldbank.org/financialstructure).

falling by half (from 4.9 to 2.4 percent) in low-income countries between 1997 and 2007.

Finally, *remittance inflows to GDP* are the flow of official remittances relative to economic activity. As migration has increased, remittance flows have become an important source of capital inflows in many developing countries, rising from less than 1 percent in 1995 to almost 2 percent in 2007. In some countries, including Moldova, Tajikistan, and Tonga, remittance inflows are more than 30 percent of GDP; in Liberia, they are 94 percent of GDP. On average, however, the lower middle-income countries have the highest ratio of remittances to GDP. The rise in remittances is driven by the doubling of remittance flows to low- and lower middle-income countries. Remittance flows to upper middle- and high-income countries have not shown a clear trend in recent years. Remittance patterns also reflect increasing migration flows from developing to developed countries. These statistics likely underestimate remittance flows, however, as they exclude informal flows (captured by the omitted category in balance of payments statistics).[11]

In summary, the trend toward globalization in financial services has been uneven across income groups. While this globalization trend has been especially pronounced in international lending and bond issues in high-income

11. According to estimates, at least a third of remittances are sent through informal channels (Freund and Spatafora 2008).

countries, low- and lower middle-income countries have benefited from increased remittance flows. The current global financial crisis could alter these trends, but it is too soon to know what the effects will be.

V. Concluding Remarks

The expanded and updated version of the Financial Structure Database will facilitate rigorous research that can enhance policy recommendations.

The current crisis has exposed the need for additional data collection, both to better understand the causes of the crisis and to design policies to mitigate the impact of financial sector fragility. For example, future research should collect information on cross-border links among banks, nonbanks, and financial markets to obtain better measures of international financial links. Understanding the asset and funding structure of the financial institutions is similarly important, as recent research has shown how balance sheet structure can affect both the profitability and the stability of banks (Caprio, Laeven, and Levine 2007; Demirgüç-Kunt and Huizinga forthcoming; Laeven and Levine forthcoming).

References

Arena, Marco. 2008. "Does Insurance Market Activity Promote Economic Growth? A Cross-Country Study for Industrialized and Developing Countries." *Journal of Risk and Insurance* 75(4):921–46.

BIS (Bank for International Settlements). Various years. "Quarterly Review on International Banking and Financial Market Development." Basel, Switzerland: Bank for International Settlements.

Bureau, van Dijk Electronic Publishing. Various years. *n.d. BankScope database*. (www.bvdep.com/en/bankscope.html).

Barth, Jim, Gerard Caprio, and Ross Levine. 2008. "Bank Regulations Are Changing: For Better or for Worse?" *Comparative Economic Studies* 50(4):537–56.

Beck, Thorsten. 2009. "The Econometrics of Finance and Growth." In Terence Mills and Kerry Patterson, eds., *Palgrave Handbook of Econometrics*. Vol. 2. Hampshire, U.K.: Palgrave Macmillan.

Beck, Thorsten, Aslı Demirgüç-Kunt, and Ross Levine. 1999. "A New Database on Financial Development and Structure." Policy Research Working Paper 2146. World Bank, Washington, D.C.

———. 2000a. "A New Database on Financial Development and Structure." *World Bank Economic Review* 14(4):597–605.

———. 2007. "Finance, Inequality, and the Poor." *Journal of Economic Growth* 12(1):27–49.

Beck, Thorsten, Aslı Demirgüç-Kunt, and Ross Levine. 2009. "Financial Institutions and Markets across Countries and over Time: Data and Analysis." Policy Research Working Paper 4943. World Bank, Washington, D.C.

Beck, Thorsten, and Ross Levine. 2002. "Industry Growth and Capital Allocation: Does Having a Market- or Bank-based System Matter?" *Journal of Financial Economics* 64(2):147–80.

———. 2004. "Stock Markets, Banks, and Growth: Panel Evidence." *Journal of Banking and Finance* 28(3):423–42.

Beck, Thorsten, Ross Levine, and Norman Loayza. 2000. "Finance and the Sources of Growth." *Journal of Financial Economics* 58(1–2):261–300.

Beck, Thorsten, and Ian Webb. 2003. "Economic, Demographic, and Institutional Determinants of Life Insurance Consumption across Countries." *World Bank Economic Review* 17(1):51–88.

Bekaert, Geert, and Campbell R. Harvey. 2000. "Foreign Speculators and Emerging Equity Markets." *Journal of Finance* 55(2):565–613.

Boyd, John H., Stanley L. Graham, and R. Shawn Hewitt. 1993. "Bank Holding Company Mergers with Nonbank Financial Firms: Effects on the Risk of Failure." *Journal of Banking and Finance* 17(1):43–63.

Caprio, Gerard, Luc Laeven, and Ross Levine. 2007. "Governance and Bank Valuation." *Journal of Financial Intermediation* 16(4):584–617.

Chinn, Menzie, and Hiro Ito. 2006. "What Matters for Financial Development? Capital Controls, Institutions, and Interactions." *Journal of Development Economics* 81(1):163–92.

Claessens, Stijn, and Luc Laeven. 2004. "What Drives Bank Competition? Some International Evidence." *Journal of Money, Credit, and Banking* 36(3):563–83.

Demirgüç-Kunt, Aslı, and Enrica Detragiache. 1998. "The Determinants of Banking Crises: Evidence from Developing and Developed Countries." *IMF Staff Papers* 45(1):81–109.

———. 2002. "Does Deposit Insurance Increase Banking System Stability? An Empirical Investigation." *Journal of Monetary Economics* 49(7):1373–406.

Demirgüç-Kunt, Aslı, and Harry Huizinga. Forthcoming. "Bank Activity and Funding Strategies: The Impact on Risk and Return". *Journal of Financial Economics*.

Demirgüç-Kunt, Aslı, and Ross Levine. 2009. "Finance and Inequality: Theory and Evidence." Annual Review of Financial Economics *1*. Palo Alto, CA: Annual Reviews.

Demirgüç-Kunt, Aslı, and Vojislav Maksimovic. 2002. "Funding Growth in Bank-based and Market-based Financial Systems: Evidence from Firm Level Data." *Journal of Financial Economics* 65(3):337–63.

De Nicolo, Gianni. 2000. *Size, Charter Value, and Risk in Banking: An International Perspective*. International Finance Discussion Papers 689. Washington, DC: Board of Governors of the Federal Reserve System.

Freund, Caroline L., and Nikola Spatafora. 2008. "Remittances: Transaction Costs, Determinants, and Informal Flows." *Journal of Development Economics* 86(2):356–66.

Hannan, Timothy H., and Gerald A. Hanweck. 1988. "Bank Insolvency Risk and the Market for Large Certificates of Deposit." *Journal of Money, Credit, and Banking* 20(2):203–11.

Honohan, Patrick, and Thorsten Beck. 2007. *Making Finance Work for Africa*. World Bank, Washington, DC.

IMF (International Monetary Fund). Various years. *International Financial Statistics database*. (www.imfstatistics.org/imf/logon.aspx).

———. 1984. "A Guide to Money and Banking Statistics in *International Financial Statistics*." Washington, DC: International Monetary Fund.

King, Robert G., and Ross Levine. 1993. "Finance and Growth: Schumpeter Might Be Right." *Quarterly Journal of Economics* 108(3):717–38.

Laeven, Luc, and Ross Levine. Forthcoming. "Bank Ownership Structure, Regulation, and Risk Taking." *Journal of Financial Economics*.

Lane, Philip, and Gian Maria Milesi-Ferretti. 2008. "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970–2004." *Journal of International Economics* 73(2):223–50.

Levine, Ross. 2005. "Finance and Growth: Theory and Evidence." In Phillippe Aghion, and Steven Durlauf, eds., *Handbook of Economic Growth*. Amsterdam: North Holland.

Levine, Ross, and Sara Zervos. 1998. "Stock Markets, Banks, and Economic Growth." *American Economic Review* 88(3):537–58.

Loayza, Norman, and Romain Ranciere. 2006. "Financial Development, Financial Fragility, and Growth." *Journal of Money, Credit, and Banking* 38(4):1051–76.

Roy, Andrew D. 1952. "Safety First and the Holding of Assets." *Econometrica* 20(3):431–49.

Schneider, Friedrich, and Dominik Ernste. 2000. "Shadow Economies: Size, Causes, and Consequences." *Journal of Economic Literature* 38(1):77–114.

Standard & Poor's. Various years. *Emerging Markets Database*. (www2.standardandpoors.com/portal/site/sp/en/us/page.product/dataservices_emd/2,9,6,0,0,0,0,0,0,0,0,0,0,0,0,0,0.html).

Swiss Re. Various issues. *Sigma*. Zurich: Swiss Re. (www.swissre.com/pws/research publications/sigma ins. research/sigma insurance research.html).

World Bank. 2007. *Finance for All: Policies and Pitfalls*. World Bank, Washington, DC.

Early Academic Performance, Grade Repetition, and School Attainment in Senegal: A Panel Data Analysis

Peter Glick and David E. Sahn

Little is known in developing country environments about how a child's cognitive skills manifested in the first years of schooling are related to later educational success, because the panel data needed to analyze this question have been lacking. This study takes advantage of a unique data set from Senegal that combines test score data for children from the second grade with information on their subsequent school progression from a follow-up survey conducted seven years later. Measures of skills from early primary school, corrected for measurement error using multiple test observations per child, are strongly positively associated with later school progression. A plausible interpretation is that parents invest more in a child's education when the returns to doing so are higher. The results point to the need for remedial policies to target lagging students early on to reduce early dropout. Grade repetition policies target poorly performing students and are pervasive in Francophone Africa. Using variation across schools in test score thresholds for promotion to identify the effects of second-grade repetition, the analysis shows that repeating students are more likely to leave school before completing primary school than students with similar ability who are not held back, pointing to the need for alternative measures to improve the skills of lagging children. JEL codes: I21, O15

Reflecting widespread recognition of the importance of human capital for individual welfare and economic growth, an enormous amount of empirical research in developing countries has examined the determinants of household investments in children's schooling.[1] However, little is known about the dynamics of schooling, in particular, the relationship of early cognitive skills to later education outcomes. In developing countries, where households are likely to be resource and credit constrained, decisions on how long to continue a child's education may depend strongly on the perceived returns to schooling

Peter Glick (corresponding author, pglick@rand.org) is a senior economist with the RAND Corporation. David E. Sahn (david.sahn@cornell.edu) is International Professor of Economics at Cornell University. The authors thank three anonymous referees and seminar participants at Cornell University and RAND for helpful comments and suggestions.

1. See Glewwe and Kremer (2006) and Glick (2008) for reviews of the literature.

THE WORLD BANK ECONOMIC REVIEW, VOL. 24, NO. 1, pp. 93–120 doi:10.1093/wber/lhp023
Advance Access Publication January 25, 2010

and, in turn, on a child's apparent ability. The effect of ability may be particularly large among poorer families in such environments, for whom resource or credit constraints are likely to be more binding. Knowledge of the nature and extent of these patterns would be highly relevant for policy. In particular, a strong connection between early achievement as manifested by test scores and subsequent school attainment would point to the need for policies directed at children who fall behind academically early in their school careers.

Knowledge of this relationship in developing country contexts is very sparse, primarily because the data needed to analyze this question have been lacking. The vast majority of studies of schooling in developing countries rely on cross-sectional data. Dynamic analyses of school attainment, however, require panel data containing both early measures of skills and later schooling outcomes. Surveys with this information are rare in developing countries, especially for Africa.

This study takes advantage of unusual panel data from Senegal that combines test score data for children from the second grade with information on their school progression from a follow-up survey conducted seven years later. With information on school and household characteristics, the analysis is able to control for confounding influences on school attainment in estimates of the effects of early academic performance on later education outcomes as well as determine whether wealth and school differences exacerbate or mitigate the impacts of early achievement gaps among children.

The study also exploits the panel to assess the impacts of early grade repetition—a pervasive policy targeting poor achievers—on later school attainment. While repetition normally must be considered as jointly determined by factors that also influence attainment, this study is able to take advantage of the fact that, conditional on observed academic ability at the end of second grade, having to repeat second grade is reasonably considered exogenous to student ability or effort.

Indicators of early cognitive ability, corrected for measurement error using multiple test observations per child, are strongly positively associated with later school progression. If interpreted causally, this result is consistent with a model of human capital investments under certain assumptions about returns to schooling in the labor market. Household wealth and parental (father's) schooling also have (conditional on early measured ability) direct positive effects on attainment, and there is some evidence that the impact of academic ability on later school outcomes differs for children in poor and well-off households, as well as between girls and boys. The large impact of early test scores on later attainment, controlling for family and school characteristics, points to the importance of school policies that direct early remedial attention to low-achieving children. Finally, conditional on academic ability, repeating a grade has a negative impact on school progression, implying that the private costs associated with stringent repetition policies exacerbate the negative effects on attainment of poor early academic outcomes.

I. Conceptual Framework

A simple model of child ability and household schooling investments helps frame the empirical analysis. The framework shows that making predictions about the relationship of ability to schooling attainment (or about interactions of household wealth and ability in schooling demand) is not straightforward without specific assumptions about the shape of the schooling returns functions. Consider a two-period model of parental investment in the education of two children of differing ability, A^i (= 1,2), as measured, for example, by tests early in school, and assume $A^1 > A^2$. For the exposition, assume that A^i represents the ability endowment of the child given exogenously to the parents (as discussed below, this cannot be assumed for the test score measures used in this study). For a pure investment model in which parents view schooling only as a means to future consumption through transfers from children's period 2 wealth (W^i), the lifetime utility of the parents is written as:

$$U = U_1\big(C_1, C_2\big[W^1(S^1, A^1), W_2(S^2, A^2)\big]\big) \tag{1}$$

where C_t is consumption in period t, and W^1 and W^2 are functions of individual child ability and the level of schooling S^i. The parents maximize equation (1) subject to the intertemporal budget constraint

$$C_1 + \frac{C_2}{1+r} = Y_1 + \frac{W^1(S^1, A^1)}{1+r} + \frac{W^2(S^2, A^2)}{1+r} - P^1S^1 - P^2S^2 \tag{2}$$

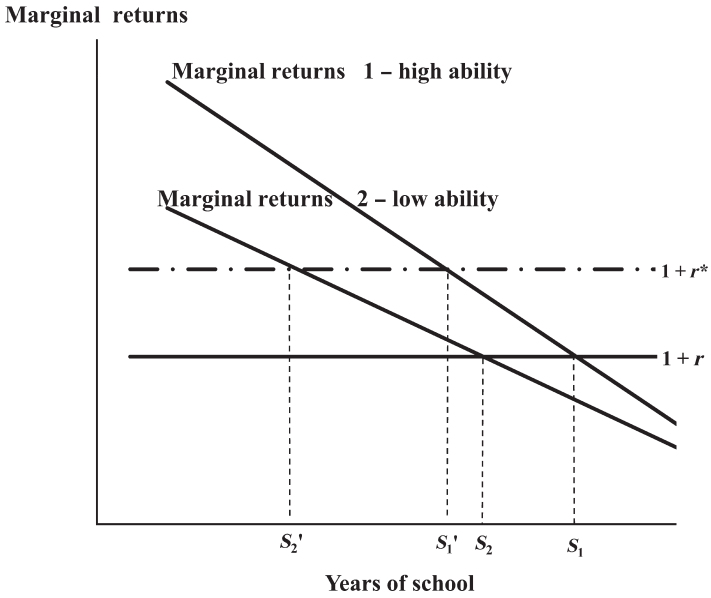
where P^i is the unit cost of schooling for child i. The first-order conditions imply (in the absence of credit constraints) that parents invest in the education of each child until the marginal return equals the market rate of interest:[2]

$$\frac{1}{P^1}\frac{\partial W^1(S^1, A^1)}{\partial S^1} = \frac{1}{P^2}\frac{\partial W^2(S^2, A^2)}{\partial S^2} = 1 + r. \tag{3}$$

If costs are the same, the difference in the optimal levels of S^1 and S^2 will depend on how ability affects the schooling return functions. A standard assumption in models of human capital investment is that returns diminish with additional schooling, which can be captured using the quadratic form $W = \alpha_1 S - \alpha_2 S^2$ ($\alpha_1 > 0$, $\alpha_2 > 0$). A simple way to allow returns to be influenced by ability is to assume that returns to the low-ability child, relative to those to the high-ability child, are reduced proportionately by some factor β ($0 < \beta < 1$), so returns for the low-ability child are $\beta(\alpha_1 S - \alpha_2 S^2)$. This yields marginal return functions $\alpha_1 - 2\alpha_2 S$ for the high-ability child and $\beta\alpha_1 - 2\beta\alpha_2 S$ for the low-ability child (figure 1). At any level of schooling, marginal returns are higher for the high-ability child, so for any interest rate, investments

2. For the exposition, direct schooling impacts on utility are ignored so this becomes a pure investment model.

FIGURE 1. Higher Marginal Returns for Higher Ability Child and Larger Income Effects for Low-Ability Child



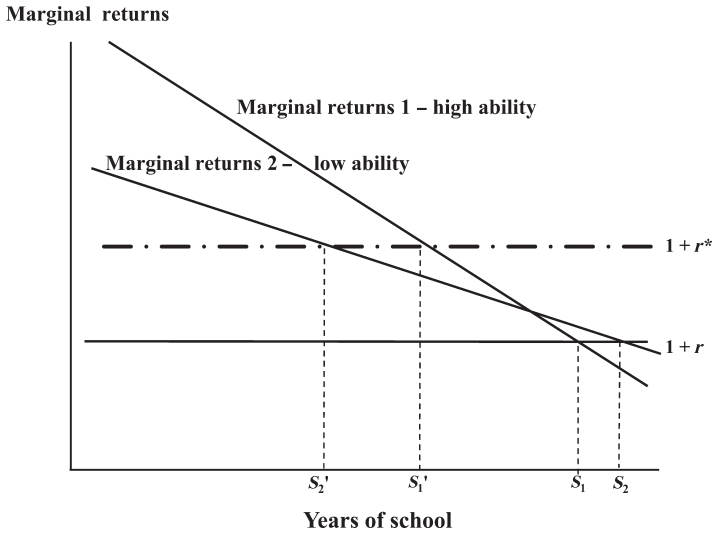
Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

in that child's schooling are greater. If households are credit constrained, a similar result would be obtained, but marginal returns would be equated to the parents' marginal rate of substitution of future for current consumption, the shadow interest rate (r^*).

While this outcome seems intuitive, it depends on the form of the returns functions, as pointed out by Behrman (1997). Alternative and no less arbitrary assumptions about how ability influences human capital production could yield a different outcome. One could flexibly represent quadratic returns to the low-ability child as $\gamma\alpha_1 S - \delta\alpha_2 S^2$, with $0 < \gamma < 1$, $0 < \delta < 1$, implying marginal returns $\gamma\alpha_1 - 2\delta\alpha_2 S$. Specific values of these parameters would produce marginal returns that are larger for the high-ability child over most of the schooling range but eventually become larger for the low-ability child, as depicted in figure 2. Thus, at a low enough interest rate (such as r in the figure), schooling investments will be greater for this child. Therefore, the relationship of total schooling and child ability is conceptually ambiguous.[3] Consumption-based considerations would add more ambiguity. For example, parental preferences for

3. The likelihood that the marginal return functions cross as in figure 2 increases if the high-ability child is also more productive in home or work activities. This raises the opportunity cost of schooling of that child relative to the low-ability child and, through the increase in cost P in equation (3), shifts down the marginal returns for the more able child.

FIGURE 2. Switch in Relative Marginal Returns for High- and Low-Ability Children, and Larger Income Effects for Low-Ability Child



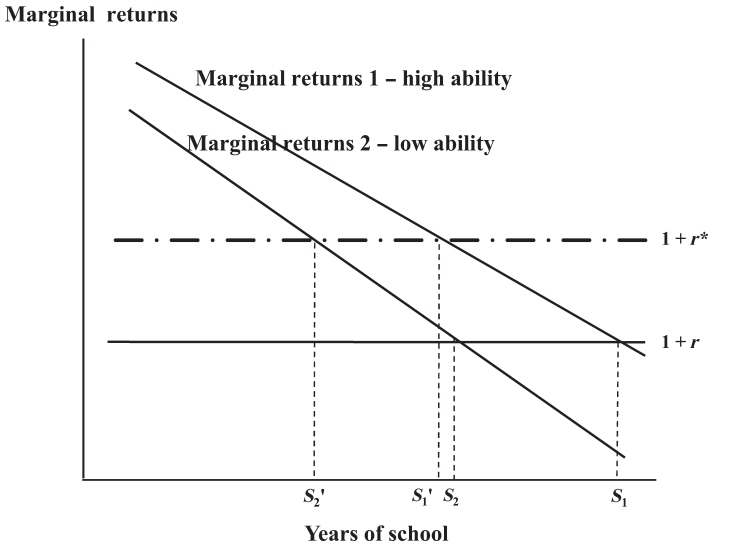
Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

equity might lead to investing more resources in low-ability children to compensate for their otherwise lower future income and welfare.

Next, consider another question of interest for the empirical analysis: how the response to ability may be affected by the level of household income—essentially, the sign on an interaction term of an ability measure and household income or assets in a model estimating schooling attainment.[4] In a pure investment framework with no borrowing constraints, there should be no interaction, since investments occur until marginal returns equal the interest rate. But credit constraints are likely, and these should impinge more on households with low current resources. For a credit-constrained (poor) household, the rate of interest is determined as indicated by the marginal rate of substitution of future for current consumption. If the constraint binds, this shadow interest rate r^* will be higher than the market rate r. This is shown in figure 1, with schooling of the two children under the binding constraint equal to $S^{1'}$ and $S^{2'}$. The effect of a sufficiently large increase in income is to relax the credit constraint so that marginal returns are equated to r rather than r^*. While in both the poor (constrained) and better-off household, the higher ability child receives more schooling, this gap gets smaller as income increases, because (given the nature of the quadratic returns functions) marginal returns are falling less steeply for the low-ability child (see figure 1). Therefore, it takes more additional investment to reach equilibrium for that child at the new lower interest rate.

4. The analysis here is formally similar to Garg and Morduch's (1998) analysis of the effects of changes in income on girl–boy gaps in health investments.

FIGURE 3. Higher Marginal Returns and Larger Income Effects for High-Ability Child



Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

The returns functions depicted in figure 2 yields this result as well. Once again, however, different assumptions about how ability affects the returns to schooling imply different outcomes. If, as in figure 2, returns to the low-ability child are represented by the quadratic $\gamma\alpha_1 S - \delta\alpha_2 S^2$ but with a sufficiently smaller absolute value of γ (or larger value of δ), returns could appear as in figure 3. In this case, the low-ability child receives less schooling both at low and high incomes, but the gap widens with income, because the marginal returns fall more steeply for this child. Again, this ambiguity occurs in a pure investment framework without resort to assumptions about parental preferences, though these would add further uncertainty to predictions. For example, if inequality aversion were a normal good, increases in income could lead to greater additions to investments in the low-ability child even if credit constraints were not operative.

Therefore, the relationship of early manifested ability and subsequent school investments, and the role of household resources in modifying the effects of ability on investments, are empirical questions. They can be addressed here because of the availability of both test score measures of early ability and information on subsequent schooling.[5]

5. However, a disadvantage of the data used for this study is that they do not have this information for multiple children in the same family, so household fixed effects cannot be used to eliminate unmeasured household factors associated with both ability and later outcomes. The heterogeneity issue is discussed in the next section.

II. Data and Empirical Approach

The 2003 Senegal Household Education and Welfare (EBMS) survey[6] was conceived as a follow-up to an earlier, school-based, study, the Program on the Analysis of Education Systems of the Conference of Francophone Ministers of Education (PASEC). The PASEC study administered tests of written math and French to a cohort of students beginning in second grade in 1995 and continuing through subsequent years of primary school (see CONFEMEN 1999; Michaelowa 2003). The 2003 survey attempted to re-interview these children, who were now of middle school age (ages 14–17). Of the original 120 PASEC clusters, 60 (28 urban and 32 rural communities) were randomly selected for the new survey. Of the 20 PASEC children per cluster/school who were tested in second grade in 1995–96, survey enumerators were able to find, on average, 15 children in rural clusters and 17 in urban ones. It should be noted that the sample, by construction, selected on children who made it into second grade, which eliminates a nontrivial share of children who do not make it that far (primarily because they never entered school).[7]

The present analysis combines the test score information and detailed school and teacher information from the PASEC survey with information from the 2003 EBMS survey, which contains detailed information on household characteristics and child schooling. Information was also collected on the schooling of children who no longer reside at home. This is important since 16 percent of the PASEC cohort was no longer living at home in 2003, and this behavior is likely to be endogenous to schooling outcomes (for example, a young person may live with relatives to attend secondary school, or leave school and home to marry or find work).[8]

In simplest form, the models estimated can be represented as follows:

$$S_i = a_0 + a_1 A_{95i} + a_2 \mathbf{X}_i + a_3 \mathbf{Q}_i + a_4 A^*_{95i} \mathbf{X}_i + a_5 A^*_{95i} \mathbf{Q}_i + e_i, \tag{4}$$

where S_i is the later schooling measure (grade attainment reported in 2003) of child i, A_{95i} is ability measured in 1995–96 (test score in second grade), $\mathbf{X}_i$ is a vector of individual and household characteristics, and $\mathbf{Q}_i$ is a vector of school inputs such as teacher background and supplies.

6. EBMS was a joint research project of Cornell University (United States), Centre de Recherche en Economie Appliquée (Senegal), and Institut National de la Recherche Agronomique (France).

7. Randomly selected households with children of similar age in the same school catchment areas were also interviewed in the EBMS survey to achieve greater representivity and a larger sample. Of this group, 24 percent did not enter second grade, almost all of whom never enrolled. Since this analysis focuses on the links of early academic success and later school outcomes, only the PASEC cohort, for which early test score information is available, is used.

8. For children not living at home, information was collected on highest completed grade and current enrollment but not on school entry and exit ages, so total years in school cannot be calculated for these children. Therefore, the highest completed grade is used as the dependent variable in the analysis. It will differ from total years of school because of grade repetition. The implications of repetition are discussed in section IV.

The coefficients must be interpreted carefully and will depend in part on how A_{95i} is represented. One option is to use actual test scores from the 1995 to 1996 school year. Using the second-grade pretest (given at the start of the school year) provides a measure of ability at the end of the child's first year of school, thus coming close to measuring ability at the time of school entry. However, in addition to measuring a child's inherent ability 'endowment,' these pretest scores also reflect school and home inputs into learning during first grade and home inputs during the preschool period that lead to better cognitive development; nutrition is likely to be particularly important (Glewwe and Miguel 2008). If, further, these unmeasured inputs also directly affect later attainment or are correlated with factors that do (such as parental preferences), this will upwardly bias the estimated effect of early achievement. Regarded as associational rather than as strictly causal, the relation of early performance to school attainment is still relevant for policy. Test score information is relatively easy to obtain and could potentially be used to direct remedial action to academically lagging children.

An alternative to test scores is the 'residual method', in which the unexplained portion in a production function regression for human capital (with the test score as the dependent variable) is taken as the individual's genetic ability endowment. In this approach an equation such as the following is estimated:

$$TS_i = b_0 + b_1\mathbf{X}_i + b_2\mathbf{Q}_{95i} + v_i. \tag{5}$$

where TS_i is the 1995 pretest score and the residual v_i is the endowment—ability purged (in contrast to the test score itself) of the effects of measured household characteristics $\mathbf{X}_i$ and grade-specific school factors $\mathbf{Q}_{95i}$such as teacher background and class variables (note that $\mathbf{Q}_{95i}$differs from Q_i in equation (4), which would incorporate all years and teachers of the child at the school). Conceptual models of parental decision-making would generally assume that parents are able to assess the child's ability as represented by v_i and make investment decisions based on it. For modeling behavior, then, this may be the preferred measure. Estimation of equation (5) requires detailed school, teacher, and household data, which the Senegal surveys have.

However, there are shortcomings with using this measure of A_{95i} in equation (4) as well. Since only measured school and household inputs are controlled, the residual will pick up the effects of unmeasured household or school factors. Moreover, these may be correlated with S_i, leading to endogenity problems in estimating the effect of "pure" ability, just as with early test scores. A second problem, relevant to either ability measure, is measurement error. It is easy to see how a single test can be a noisy measure of a student's knowledge. In particular, performance could be highly dependent on the students' disposition on the day of the test. If this source of error is random, it will, all things equal, bias toward zero the estimated impacts of early test score (or similarly, endowment as measured by the residual).

Fortunately, the PASEC data permit the use of an instrumental variables procedure to deal with this problem and correct for (nonsystematic) measurement error. For each second-grade student there are multiple—potentially four—test scores: French and math pretests, and French and math posttests (conducted at the end of the school year). Say that test score TS_1 is being used as a measure of cognitive ability and that the relationship between them takes the form $A_{95i} = \alpha_0 + \alpha_1 TS_{1i} + \mu_{1i}$, where μ_i captures the noise in the relationship. An additional test TS_2 provides another proxy for ability: $A_{95i} = \beta_0 + \beta_1 TS_{2i} + \mu_{2i}$. If measurement error is uncorrelated across these indicators ($\text{Cov}(\mu_{1i}, \mu_{2i}) = 0$), their correlation is due only to their common dependence on A_{95i}. A consistent estimate of a_1 in equation (1) can then be obtained by instrumenting for one test score using the value of the other.[9] The method is applicable as well if the residual approach is used (Pitt, Rosenzweig, and Hassan 1990); in that case, one test score residual is used as an instrument for the other.

The data provide several options: use the French second-grade pretest to instrument the math pretest score (and the reverse), or use the second-grade posttests to predict pretest scores. The first approach appears less desirable. Since the two pretests were given at the same time, measurement errors are likely to be correlated (perhaps because of the child's disposition on that day), violating the key assumption for the instrumental variable strategy. A second issue is that a nontrivial number of students are missing test scores for one of the pretests. For these reasons, the posttest average of the math and French scores (or test score regression residual) are used to predict the average of the pretest math and French scores (or residuals); averaging the two permits the inclusion of cases where only one test was taken in either or both periods. The less attractive aspect of this choice is that the posttest score instrument will incorporate the effects of school inputs in second grade (and the posttest residual will capture impacts of unmeasured inputs during this period). Finally, some 6 percent of children in the sample took the posttests but not the pretests. Estimates from the predicting equation are used to impute the pretest score for these cases.

It should be stressed that this instrumental variables strategy deals with measurement error but not with any potential endogeneity-related bias in the estimate of a_1, which would come about if unmeasured school or home inputs are correlated with both initial achievement and subsequent attainment.[10]

9. This is the multiple indicators approach (see Wooldridge 2002, for a general presentation). It is noteworthy that even with a correlation between the error terms of the indicators (systematic measurement error), Bound, Brown, and Mathiowetz (2001) show that the multiple indicator instrumental variables solution will still generally be a more consistent estimator of the true parameter of interest than will ordinary least squares estimation using the indicator with measurement error.

10. Potential instruments for early academic achievement in the data, at least for the second-grade posttest score, include second-grade teacher variables related to qualifications and pedagogical practices, provided these are truly idiosyncratic and vary across teachers/grades within a school. While some of these factors are indeed associated with the posttest outcomes, these instruments had low power and resulted in implausible or highly nonrobust second stage estimates of the effect of early test score on grade attainment.

The fact that the measured relation of later success to early achievement may be associational rather than strictly causal is not necessarily a hindrance from a policy perspective. As already noted, knowledge of this association would make it possible to target early on children who are otherwise likely to do poorly later. For a behavioral interpretation, however, heterogeneity bias is clearly a greater concern. A partial correction is to estimate school fixed effects, which purge estimates of bias from linear associations of school-level unobservables with the included covariates. It is plausible that unmeasured household preferences and child ability operate in significant part through the choice of school (and implicitly through location since school and cluster are interchangeable in the PASEC survey). The fixed effects estimates will eliminate bias from this source. They will not, obviously, account for heterogeneity among children *within* schools. Therefore, it cannot be claimed that the estimates of the effects of early skills represent purely causal effects.[11]

With respect to the functional form of equation (4), ordered probit is used to model the determinants of years of completed schooling. The ordered probit model treats grade attainment as the outcome of a series of ordered discrete choices. The model is easily extended to allow for right censoring of the dependent variable, an important concern since some 60 percent of the cohort is still attending school.[12]

Two alternate specifications are estimated, reflecting different assumptions about unobserved (school-level) heterogeneity: random effects and (as just described) fixed effects. The school random effects model directly parameterizes the within-school correlation of errors but assumes that these are not correlated with included regressors. Estimation of the random effects model involves specifying the likelihoods conditional on the school-level random effect, then integrating over all possible values of the random effect to obtain the unconditional likelihood. This is done using Hermite integration, following the approach suggested by Butler and Moffit (1982) (for further details, see Glick and Sahn 2000). The fixed effects specification, in contrast, does not require the orthagonality assumption on the school-level errors; it eliminates any linear associations of the errors with included covariates by including dummy variables for each school. If the correlation of school-level covariates

11. It is useful to compare the approach of this study with several other analyses (Glewwe and Jacoby 1994; Alderman and others 1996; Kingdon 1996) on cross-sectional data that estimate school attainment or skills using contemporaneously measured ability as captured by Raven's abstract thinking test. This is usually presented as a measure of innate (hence preschool) ability, but as Glewwe and Kremer (2006) note that interpretation is implausible because the test is likely to also reflect an individual's schooling. Therefore, these studies do not provide strictly causal estimates of the impact of innate ability (as the present study does not), nor do they permit an understanding of how later education success is related to ability measured at or near the start of school (which, in contrast, the panel data in this study do permit).

12. As almost no children report a grade higher than 9 by the 2003 survey, the top category is aggregated as grade 9 and above.

and the error term is in fact zero, however, the fixed effects model is less efficient than the random effects model.[13]

Since estimation in the fixed effects model is based only on within-school variation, the effects of school factors that are constant across children within a school are not estimated. Nevertheless, the impacts of teacher characteristics and classroom supplies can be estimated because they retain some variation across children within the same school. Specifically, the 1995–96 school surveys included detailed information on teacher background and classroom characteristics and supplies for the original second-grade PASEC class and for subsequent primary grades attended by the cohort. Basic teacher and class information was also collected on at least one class containing grade repeaters in subsequent years (for example, a second-grade class in 1996–97, when non-repeaters would be in the third grade). The ordered probit regressions include the averages for these variables for each student enrolled in classes for the school years for which complete information is available: 1995–96,1996–97, and 1997–98 (there will be fewer than three years of data for early dropouts or absent children).[14] The paths of individual children diverge, because some children repeat one or more grades while others progress smoothly to the next grade (while some drop out), so there is some within-school (across child) variation in teacher characteristics and classroom supplies measures.

Information on director background and school facilities was collected only in 1995. Because of the large number of potential covariates for classroom supplies and facilities, principal components analysis was used for data reduction. For classroom supplies, the variables used were presence of a dictionary, ruler, compass, and eraser; for school facilities, the variables were presence of a library, toilet, drinking water source, and canteen. The first principal component in each case was included in the school attainment models. Finally, as the household survey did not collect consumption or income data, factor analysis was used to construct a wealth index from information on ownership of durable goods (see Sahn and Stifel 2003 for details of the method). Table 1 presents descriptive statistics on the variables used in the analysis.

13. A potentially serious concern with fixed effects in nonlinear models is bias due to the incidental parameters problem: the inclusion of cluster dummy variables means that as total sample size increases, so does the number of parameters to be estimated (with a fixed number of observations T per cluster, increasing the sample size means increasing the number of cluster dummy variables). This renders all the estimates inconsistent asymptotically. However, Monte Carlo evidence provided by Greene (2002) indicates that for the values of T in the data set used here (about 15), the sample bias is small for an ordered probit model—slightly under 10 percent—with a small downward bias in variance as well. In view of these results, it is reasonable to use fixed effects in this study, while recognizing the likelihood of a small degree of bias.

14. In addition, for children who repeat second grade twice and hence are still in second grade in the third year of the PASEC study (1997–98), information on their teacher and class will be lacking for that year, because data in 1997–98 were collected only for fourth and third grade classes. For these cases, information on the child's second-grade class in 1996–97 was used.

TABLE 1. PASEC Cohort: Means and Standard Deviations of Selected Variables

	All		First pretest score quartile		Fourth pretest score quartile	
Variable	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Highest grade completed	6.03	1.65	5.28	1.45	7.07	1.51
Still in school in 2003	0.63	0.48	0.58	0.49	0.78	0.42
Standardized second-grade pretest score[a]	0.003	0.73	−0.895	0.20	0.971	0.33
Age (years)	15.62	0.95	15.48	0.97	15.56	0.92
Girl	0.41	0.49	0.42	0.49	0.42	0.49
Asset index[b]	−0.022	0.91	−0.293	0.81	0.204	1.00
Mother's years of schooling	1.53	3.06	1.07	2.46	2.30	3.79
Father's years of schooling	2.70	4.27	2.23	3.82	3.82	5.07
Missing mother's schooling (percent)	0.10	0.30	0.07	0.26	0.13	0.34
Missing father's schooling (percent)	0.17	0.37	0.13	0.34	0.18	0.38
Rural (percent)	0.56	0.50	0.63	0.48	0.49	0.50
School director's experience (years)	12.67	8.92	9.95	8.54	14.43	9.22
Director has baccalauréat or higher (percent)	0.54	0.49	0.58	0.50	0.50	0.50
Teachers, average years of experience	11.53	7.16	9.96	7.64	12.48	6.67
Teachers, share with a baccalauréat or higher (percent)	0.43	0.39	0.40	0.42	0.45	0.35
Teachers, share female (percent)	0.24	0.36	0.20	0.34	0.32	0.38
Classroom supplies first principal component[c]	0.01	1.38	−0.09	1.61	0.07	1.15
School facilities first principal component[c]	−0.07	1.42	−0.29	1.48	0.13	1.36
Number of observations	834		208		209	

Note: Teacher and classroom supplies variables are the child-specific averages for three years of data from the PASEC survey.

a. Average of French and math scores administered at start of second grade (1996–96 school year). Predicted using second-grade posttest.

b. Constructed employing factor analysis using information on ownership of durable goods (TV, refrigerator, bicycle, and others) and housing characteristics (including drinking water source and toilet facilities).

c. Based on principal components analysis using classroom and school attributes described in the text.

Source: Authors' analysis based on data from the PASEC surveys and the 2003 Senegal Household Education and Welfare Survey (EBMS); see text for details.

III. Early Performance and Grade Attainment: Estimation Results

This section briefly discusses the first-stage regressions used to generate the test score residuals as these results are of intrinsic interest. The full regression results are shown in appendix table A-1. The dependent variables are standardized scores on second-grade post- and pretests. Test scores are lower if the student is a girl, and higher for wealthier children. For the posttest score, there is a positive impact of mother's but not father's education. Pretest scores are positively associated with the experience and education of the school director and teachers, as well as the share of teachers that are women. The teacher variables used in the pretest model are averages over different classes for the school, since the relevant class-specific information (for most children, first grade in 1994–95) is not available. Thus, these variables capture the overall quality of teachers. The posttest regression uses actual values for teachers and classroom supplies for the child's second-grade class. The teacher variables are largely not significant in this regression, with the striking exception of a large positive interaction of student gender (being a girl) and having a female teacher, suggesting that girls respond better to female teachers or that female teachers make more of an effort to encourage girls. There is also a positive impact on the posttest score of having better equipped classrooms.

In the grade attainment-ordered probits, the estimated effects of early academic ability using either predicted test score or the predicted residual were very similar, reflecting the high correlation between these two indicators. Despite the presence of a number of highly significant covariates, the explained variation in the test score regressions is low. As a result, the unexplained portion (the regression residuals) is highly correlated with the dependent variable (test score). This carries over, of course, to the measurement error-corrected predictions—the pretest score predicted from the posttest score is highly correlated with the pretest residual predicted from the posttest residual. Given the similarity of results, only the results for the ordered probit attainment models using the predicted test score are presented (table 2).[15]

Four models are presented in table 2: base models using school-level random and fixed effects (regressions 1 and 3 in table 2) and these models with interactions added (regressions 2 and 4). There is a very strong, highly significant association of grade attainment with early academic success as measured by test score at the start of second grade. The ordered probit model estimates the parameters of the linear index function underlying the model; to get comparative static effects, it is necessary to calculate changes in probabilities using the estimates. Calculations using the random effects estimates

15. One interpretation of the low R^2s in the test score regression is that most of the variation in scores is due to variation in unmeasured ability. However, as suggested above, it is more likely a combination of ability and the influence of unmeasured household or school factors.

TABLE 2. Ordered Probit Models of Grade Attainment

Variable	School-level random effects (1)	(2)	School-level fixed effects (3)	(4)
Intercept	3.785 (3.26)***	4.012 (3.16)***	3.432 (4.43)***	3.502 (4.36)***
Second-grade pretest score[a]	0.602 (5.15)***	0.448 (2.27)**	0.689 (8.59)***	0.587 (7.01)***
Girl	−0.173 (0.92)	−0.081 (0.42)	−0.278 (1.89)*	−0.199 (1.43)
Asset index	0.360 (2.24)**	0.371 (2.15)**	0.407 (5.24)***	0.414 (5.35)***
Mother's schooling	−0.009 (0.28)	−0.012 (0.36)	−0.006 (0.36)	−0.009 (0.49)
Missing mother's schooling	−0.142 (0.36)	−0.112 (0.29)	−0.128 (0.69)	−0.071 (0.37)
Father's schooling	0.035 (1.38)	0.035 (1.26)	0.034 (2.34)**	0.033 (2.40)**
Missing father's schooling	−0.357 (1.49)	−0.357 (1.44)	−0.343 (1.86)*	−0.338 (1.76)*
Pretest score × asset index		0.120 (0.98)		0.131 (1.96)**
Pretest score × girl		0.481 (1.98)**		0.364 (2.51)**
Pretest score × father schooling		0.004 (0.10)		0.002 (0.07)
Director's experience	0.000 (0.01)	−0.004 (0.18)		
Director has baccalauréat or higher	−0.145 (0.45)	−0.206 (0.67)		
Teachers, average experience	−0.016 (0.66)	0.018 (0.68)	−0.050 (2.09)**	−0.052 (2.22)**
Teachers, share with baccalauréat or higher	0.217 (0.65)	0.182 (0.54)	0.671 (1.45)	0.624 (1.38)
Teachers, share female	−0.640 (1.44)	−0.652 (1.40)	−1.485 (4.57)***	−1.383 (4.56)***
Girl × share teachers female	0.321 (0.90)	0.179 (0.44)	0.536 (2.32)**	0.397 (1.72)*
Classroom supplies first principal component	0.041 (0.36)	0.040 (0.35)	−0.015 (0.19)	−0.022 (0.29)
School facilities first principal component	0.002 (0.01)	0.002 (0.01)	−0.097 (0.72)	0.000 (0.00)
Rural	−0.256 (0.48)	−0.314 (0.53)		
Rho	0.068 (1.41)	0.070 (1.36)		
Number of observations	834	834	834	834

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: Numbers in parentheses are aysmptotic *t*-statistics. Standard errors in the fixed effects model are adjusted for clustering at the school level. See table 1 and text for variable definitions. Models also include ethnic group dummy variables, ordered probit threshold parameters, and province dummy variables (random effects model) or school dummy variables (fixed effects model).

a. Predicted using second-grade posttest score.

Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

indicate that at the mean of the regressors, a one standard deviation improvement in the second-grade pretest score is associated with a 22 percentage point increase in the probability of completing the sixth grade (relative to a mean estimated probability of 56 percent). The result for the fixed effects model with controls for community/school is virtually the same. These are clearly large impacts. Although, as noted, it cannot be established how much of this association is causal, the robustness of the estimates to controls for school and community demonstrates that they are not simply due to an association of high achievement and attainment with better schools or other locality-specific factors.

This very large and robust association of early academic performance with later educational success has implications for policy. It argues strongly for policies to target children who display early cognitive disadvantages. A limitation of the analysis, however, is that it cannot say what policies should be implemented or at what point in the child's life they would be most effective. For example, if poor cognitive skills early in school reflect largely prior nutritional deficiencies, the best time to intervene is probably well before school age, in which case it would not be very useful to rely on early school measures. One approach to dealing with poor primary performers—grade repetition—is assessed later in the paper.

While there is little comparable evidence from other developing countries, the findings here accord with panel-based research in industrialized countries that considers how early (at school entry) disparities in children's skills are related to later achievement gaps (for example, Phillips, Crouse, and Ralph 1998; Duncan and others 2007), and how academic performance is linked to educational attainment, dropout, and labor market success (Robertson and Symons, 1990; Currie and Thomas 2001; Maani and Kalb 2007).

Conditional on early test score, household wealth is a strongly significant determinant of attainment. A one standard deviation increase in the wealth index implies an approximately 12 percentage point increase in the probability of completing the sixth grade. As with test score impacts, this effect is essentially the same with and without controls for school fixed effects. These wealth impacts may reflect a number of factors, including an income effect on the demand for education and the association of wealth with inputs such as early childhood nutrition or school supplies that improve school performance and progression. Among other household covariates, there is a positive impact of father's schooling (significant only in the fixed effects case). This is in contrast to the tendency in the test score regressions for mother's schooling and not father's to improve test score outcomes. It is possible that educated mothers spend more time helping children with schoolwork than educated fathers do (leading to larger maternal education effects on scores) but that schooling duration decisions largely reflect father's preferences, which are tied to father's

education. In the fixed effects model but not in the random effects model, girls go less far in school than do boys with similar early achievement.

There are few significant coefficients on school covariates.[16] However, it is striking that, at least in the fixed effects case, the share of classes taught by a female teacher has a negative effect on attainment that is larger for boys (recall that there is variation within schools in the share of each child's teachers who are women). A smaller negative effect for girls is understandable considering the apparent positive effect of female teachers on girl's learning, but the reason for an overall negative effect of female teachers on attainment seems puzzling at first glance. One possibility is that (male) household decision-makers believe incorrectly that female teachers are of lower quality.[17] Perhaps more probable in view of results discussed below is that female teachers are more likely to fail students (especially boys), which impedes school progression.

The ordered probit models in columns 2 and 4 in table 2 add interactions of early score with assets, gender, and paternal schooling.[18] There is a positive interaction of assets and pretest score in the fixed effects model. As noted in section I, even if the test score represented a measure of ability that was exogenous to parental behavior, theoretical considerations do not lead to firm predictions about the existence or direction of this interaction. In both the random effects and fixed effects models there is a strongly significant positive interaction of test score and being a girl, meaning that the positive effect of test score on subsequent school investments is larger for girls. Caution is necessary here, however, since a higher score for girls, even as early as the end of first grade, could reflect household preferences for girls' human capital that also lead to higher attainment. Still, the results may indicate that the returns to schooling are more sensitive to observed ability for girls than they are for boys. This could happen, for example, if parents expect that a daughter with strong academic skills will be more likely (because of higher potential pay or greater ambition) to go on to participate in the labor market or gain full-time employment. This would increase their incentive to invest further in the girl's schooling, since any earnings benefit from an additional year of education would be realized over a greater number of future years of work. Since male full-time participation is more or less universal, this effect would not occur for boys (see Glick 2008 for discussion).

16. Also included in earlier specifications was distance to the nearest lower secondary school (collected in 2003). This was also insignificant. Characteristics other than distance of lower secondary schools may affect school continuation decisions, but information on these characteristics was not available.

17. The share of female teachers is unlikely to be acting merely as a proxy for poor school quality: there is no consistent pattern of correlations of female share with observed quality indicators or teacher and director background measures.

18. In initial specifications, nonlinearities in the effect of test score itself were found to be insignificant.

TABLE 3. Predicted Probabilities of Completing Sixth Grade for First and Fourth Quartiles of Second-Grade Pretest Scores

Model and probability	Fourth pretest score quartile (1)	First pretest score quartile (2)	Eliminate wealth gaps[a] (3)	Eliminate wealth and paternal education gaps (4)
School random effects model				
Probability (grade ≥ 6)	0.787	0.326	0.393	0.414
Difference in probabilities for fourth and first quartile		0.461	0.395	0.374
School fixed effects model				
Probability (grade ≥ 6)	0.799	0.386	0.464	0.486
Difference in probabilities for fourth and first quartile		0.413	0.334	0.313

Note: Table shows predicted probabilities at test score quartile means for all variables, based on ordered probit model estimates.

a. Column 3 shows the probability for first quartile after setting wealth equal to fourth quartile means. Column 4 shows the probability for first quartile after setting both wealth and paternal education equal to fourth quartile means.

Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

To put the importance of early skills in perspective, the estimates were used to predict the attainment of children in different quartiles of the pretest score distribution (table 3). Columns 1 and 2 show the predicted probabilities of completing at least the sixth grade, evaluated at the means of the regressors for the fourth and first quartiles. The probabilities differ substantially for the two groups, though this also reflects differences in other factors such as wealth and parental education. Column 3 shows the probabilities for the bottom scoring quartile after "eliminating" wealth differences by setting the asset index for this group to the mean of children in the highest achievement quartile. As expected from the association of early test performance and wealth, if household resources were the same for the two groups, the gap in probabilities would be smaller. However, the gap remains very large, closing only from 0.46 to 0.40 in the random effects case; results are similar for the fixed effects case. Column 4 shows that also eliminating gaps in father's schooling further narrows the attainment gap, but only slightly. Controlling for differences in school factors (in the random effects models) or cluster location (fixed effects models) did not substantially reduce the large differences in expected attainment. Thus, the large gap in attainment between early low- and high-achieving children reflects largely differences in early academic ability rather than associations of ability with other factors influencing attainment. This should not be taken to mean that wealth and parental education do not contribute strongly

to attainment gaps, but rather that they appear to do so primarily through impacts early on, as captured by achievement measures in second grade.

IV. How Does Targeting Poor Performers for Grade Repetition Affect Attainment?

Interpreted as a causal relationship, the link between early measures of cognitive ability and grade attainment indicates that school investments by parents are strongly reinforcing with respect to ability: high-achieving children are kept in school longer. Since the indicators of early cognitive skills may incorporate unmeasured factors that also affect later schooling, the results should be considered suggestive of such behavior rather than conclusive evidence of it. Still, even interpreted as an association, the findings indicate clearly that students who lag behind their peers early in primary school are at substantially higher risk of early withdrawal from school. Thus, if education authorities are committed to ensuring greater equity in educational attainment, targeting poor early performers for remedial help would be a logical step.

Poor performers are currently targeted for one kind of remedial treatment: grade repetition. Repetition is particularly pervasive in Francophone Africa, where it averages 20 percent for primary students as compared with 10 percent in Anglophone Africa and 2 percent in Organisation for Economic Co-operation and Development countries (Michaelowa 2003). In the Senegalese sample, 76 percent of PASEC students report having repeated at least one primary grade, and 40 percent report two or more repetitions. While repetition is thought to help lagging students catch up with their peers, the impacts of repetition on either attainment (specifically primary completion) or learning have not been established. The costs of repetition, however, are clearly large, both for the education system (in demands on teachers and classroom resources) and for families, who have to finance additional time in school for children to reach a given grade. If labor markets rewarded grade attainment—or achievement of specific levels such as primary completion—rather than solely (less easily observed) human capital, the returns to families of additional schooling would be lower than where there is less repetition. In terms of the model discussed in section I, the marginal returns for the low-ability (repeating) child in figures 1–3 would effectively shift further downward relative to those for a high-ability (nonrepeating) child—possibly explaining in part the large negative association of low measured ability and grade attainment.

It is of interest then to see how this "targeted" policy affects school attainment in the sample. Of course, even if there are negative impacts on attainment, they may be offset by improvements in learning from repeating a grade, so information on both outcomes is needed to assess repetition policies.

Attempts to measure the impacts of repetition must deal with the fact that repetition is not exogenous to school attainment or dropout. Repeaters would

be expected to be relatively poor academic performers who may well have lower attainment than nonrepeaters, aside from any positive or negative effects of repetition itself. In a few, mostly U.S. studies, researchers have been able to handle endogeneity by taking advantage of natural policy experiments, such as changes in retention policy over time or across states, or regression discontinuity designs made possible by the presence of an official test score threshold for promotion (Jacob and Lefgren 2004; Allensworth 2005; Greene and Winters 2007). While this is not the case for this study, the panel data make it possible to control for academic achievement at the time the decision is made to repeat a student (measured by the second-grade posttest score). Conditional on a student's academic performance to that date, variation in the requirement that a student repeat the grade should reflect variation in school or teacher repetition rules or attitudes rather than parental (or child) education preferences or effort.[19] This approach is similar to that of several analyses for developing countries of the effects of repetition on student learning (Gomes-Neto and Hanushek 1994; Michaelowa 2003), though unlike those studies, the present study also estimates impacts on attainment or dropout. This is made possible by the fact that the 2003 survey, rather than being school based, interviewed members of the PASEC cohort whether they were still in school or not and gathered information on the last year and grade enrolled of those who had left school.

Because the analysis is considering the effect of a repetition occurring near the start of a child's education, it focuses on a measure of early attainment rather than ultimate attainment, as in the ordered probits. It estimates the effect of second-grade test score and second-grade repetition on the probability that the child completes grade 4, or equivalently, drops out before grade 5. This choice of dependent variable avoids censoring problems, as very few children (under 4 percent) who are still enrolled as of 2003 have not already progressed beyond this level. As indicated, this analysis uses the posttest rather than the pretest, as the posttest measures skills at the time the repetition determination is made. As before, this is corrected for nonsystematic measurement error, now using the pretest score as an instrument.

A problem with the data is attrition from the PASEC sample. Some 20 percent of the PASEC second-grade cohort of 1995–96 is not in the testing sample in the next year—that is, in either the third or second-grade classes visited by the survey at the end of the 1996–97 school year. There are two possible reasons for this: the children had already dropped out, or they were

19. It is not being assumed that a student's academic ability (perhaps relative to that of peers, as discussed below) is the only factor in a teacher's decision on whether to hold the student back; in particular, the student's social or emotional maturity may also be considered. However, this will not bias the estimate of the effect of repetition unless, controlling for academic achievement, this factor would independently lead to early dropout. There is no reason to expect a (presumably temporarily) immature child not to continue in school if this immaturity is not causing the student to lag behind academically.

still enrolled but were not in attendance on the day the PASEC team returned to the school. The 2003 data indicate that the large majority of these missing children attained third grade or higher and thus belong in the second group.

Being missing is not random with respect to academic ability: children who were absent for the 1997 follow-up scored significantly lower on the prior year tests than those present, and this is the case even for children who were still in school but merely not in attendance. Since they had lower achievement at the end of second grade, children who were absent (for either reason) would be more likely to be repeaters. Further, for children who were enrolled but absent during the 1997 visit, attendance was probably poorer on average than for nonabsent children, and factors that lead to poor attendance such as higher opportunity costs or lower motivation (hence negatively affect selection into the 1996–97 sample) might also independently result in earlier dropout. Therefore, both early (before third grade) school leaving and simple absence may lead to selection bias when estimating the impacts of repetition.

It is very difficult to come up with valid instruments for this selection.[20] Even without this, however, the data allow a test for selectivity. For children missing in 1996–97, information is unavailable on second-grade repetition, but the dependent variable of interest, dropout before grade five, is observed. If there is selection on unobservables, dropout should be associated with selection (the presence of the child in 1996–97 re-survey), conditional on test score and other measured characteristics. In probits for early dropout using the full 1995–96 sample, there is a significant negative effect of a dummy variable for being in the 1996–97 sample, controlling for posttest score and other covariates (results available from the authors). However, this effect vanishes after dropping the small number of children who reported leaving school right after second grade. This finding indicates that selection into the 1996–97 sample is related positively to the propensity for continuing in school, but this selection arises from children who drop out immediately rather than those who are simply not in attendance for the 1997 test. Of course, whatever the source, the possibility of bias must be considered. It bears noting that selection issues in modeling school outcomes in panels with high withdrawal or absenteeism are hardly unique to these study data, though the ability to test for selection is unusual.

20. The EBMS collected information on health and economic shocks to the household (parental illness, unemployment, negative and positive harvest shocks, and other events) that can predict ultimate school attainment (see Glick and Sahn 2009). However, these shocks are unconvincing as instruments in the present context: they would have to predict attendance on a given day in 1996–97 but have no independent effects on dropout in that year or the next several years. Shock-related factors that affect attendance (such as loss of income or demands on the child's time) will likely also influence the decision on whether and how long to continue schooling.

Probit regressions (presented in appendix table A-2) on the sample of children present in 1996–97 confirm that the probability of repeating second grade is negatively and very strongly affected by the posttest score (*t*-statistics greater than 5.0), whether controlling for school fixed effects or not. A 1 standard deviation reduction in the posttest score increases the repetition probability by about 11 percent—a large impact given the mean repetition of 14 percent.[21] In the fixed effects model, most of the school dummy variables are highly significant and there is substantial variation in the magnitudes of the school effects. The variation across schools in repetition probabilities conditional on test performance appears to be the case at least partly because teachers or schools assess students relative to their peers, as many have suggested (Bernard, Simon, and Vianou 2005). In a specification of the non-fixed effects model including both individual score and mean class score, the mean score has a positive and significant impact, meaning that a student with a given level of ability is more likely to be held back when ranked lower within his or her class. Repetition probabilities are also higher when the teacher is a woman. Perhaps consistent with the beneficial impact of female teachers on girls' test scores, the positive female teacher effect on repetition (controlling for test score) is significantly smaller if the student is a girl.

Among household variables, parental education has no impact on repetition controlling for score, but wealth is negatively associated with repetition. Since the wealth effect persists in the within-school model, it is evident that not all of the variation in repetition conditional on ability is due to variation across teachers or schools. It is possible that wealthier parents are able to bribe teachers to avoid repetition or, possibly, are more willing to appeal the decisions of teachers (or are more effective at it). This could pose some problems for the identification of causal repetition effects. However, wealth is an included covariate in the dropout models, and this controls for contamination from wealth-related behavior affecting both repetition and dropout (to effectively capture these effects, polynomials in wealth were added to the dropout models but were not significant). It is necessary to assume that unobserved factors not captured by wealth are not significantly affecting both the teacher repetition decision and later attainment outcomes conditional on skills.

The probit results for early dropout for this sample, including the indicator for being a repeater, are presented without school fixed effects (table 4). The fixed effects estimation is not feasible for this analysis, because the identification of repetition impacts conditional on academic performance relies on variation across classes in teacher/school repetition practices—essentially, variation in test score thresholds for passing second grade—such that children with similar scores but in different schools will have different probabilities of being

21. This is all the more noteworthy because teacher promotion decisions are not based on the results of the PASEC survey tests (which were graded by survey personnel for use by the PASEC project only) but on internal tests and other assessments of the children's skills.

TABLE 4. Determinants of Dropout before Fifth Grade and 1997 Test Score

	Dropout[a]		Test score[b]	
Variable	Coefficient	*t*-statistic	Coefficient	*t*-statistic
Intercept	−1.973	−5.961***	0.000	−0.001
Second-grade posttest score[c]	−0.237	−1.837*	0.804	14.618***
Repeated second grade	0.492	2.306**	−0.204	−1.601
Girl	0.008	0.043	0.010	0.116
Asset index	−0.240	−2.475**	0.066	1.221
Mother's schooling	−0.032	−1.036	0.011	1.001
Missing mother's schooling	0.040	0.161	0.018	0.132
Father 's schooling	−0.001	−0.051	0.014	1.691*
Missing father's schooling	0.609	3.233***	0.059	0.510
Rural	0.153	0.889	−0.045	−0.297
Director's experience	−0.001	−0.180	0.004	0.632
Director has baccalauréat or higher	−0.067	−0.466	0.190	1.322
Teachers, average experience	0.021	2.033**	−0.016	−1.833*
Teacher has baccalauréat or higher	0.338	1.837*	0.097	0.553*
Teachers, share female	0.387	1.556	0.156	0.768
Girl × share teachers female	−0.382	−0.826	−0.011	−0.067
Classroom supplies principal component	−0.023	−0.590	0.025	0.571
School facilities principal component	−0.031	−0.644	0.071	1.424
Number of observations[d]	664		556	

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: See table 1 and text for variable definitions. Model also includes ethnic group dummy variables and province dummy variables. Standard errors are adjusted for clustering at the school level.

a. Probit model estimates. Dependent variable is leaving school before fifth grade.

b. Ordinary least squares estimates. Dependent variable is the standardized score from the end of the 1996–97school year (see text for details).

c. Predicted using second-grade pretest score.

d. Sample size is lower for test score model because of missing test data.

Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

held back. Considering only differences among students within a class obviously eliminates this variation, so the repeat dummy variable would do no more than capture a possible nonlinear impact of test score on dropout.[22]

The results indicate that conditional on skills as captured by posttest score (which has the expected negative effect), repeating the second grade has a large

22. The non-fixed effects model does not control for unobserved school or community factors that affect both repetition and dropout. It is important to remember, however, that the regressions include the test score, which is an ideal control to capture factors affecting schooling outcomes; conditional on test scores and the other covariates, it seems likely that the repetition-dropout relation will not be substantially affected by unmeasured factors. It is worth recalling as well from the attainment models that the effects of test score itself are extraordinarily robust to adding school fixed effects.

and significant positive impact on the probability of early dropout. The implied increase in the probability is about 11 percent, a large impact given the rate of dropout before fifth grade of 15 percent. The other estimates for the probit model are qualitatively similar to the earlier ordered probit results, though perhaps reflecting the relatively limited variation in the dependent variable, there are fewer significant coefficients. Most notably, greater wealth reduces the probability of leaving school before grade 5.

It can be concluded, therefore, that grade repetition as a policy for poor achievers has detrimental impacts on attainment among children in Senegal. The study results, subject to the caveat about selection noted earlier, suggest that children who are made to repeat a grade are less likely to complete primary school than children with similar academic ability who are not held back. This is not surprising, given that repetition effectively raises the private costs of achieving a given grade level. It may be inferred that part of the large negative effect of poor early achievement on school attainment operates through this pathway. It is also possible that poorly educated parents find it difficult to discern by other means how much their child is learning; such parents may take a failure to be promoted as a signal that the returns to schooling are particularly low for their child. While there appear to be no similar studies in other African contexts that plausibly attempt to control for repetition endogeneity, similar negative impacts on attainment were found for Uruguay by Manacorda (2007), who used both regression discontinuity and a natural experiment approach.[23] The findings on dropout impacts from the broader U.S. literature are mixed (Grissom and Shepard 1989; Eide and Showalter 2001; Allensworth 2005), but the differences in contexts between the United States and Africa are in any case very large.

As noted, there still may be a benefit to repetition if children who repeat a grade learn more by staying behind than children with similar skills who are promoted to the next, more difficult grade. The PASEC test data make it possible to address this question in a straightforward way. Students who repeated second grade in 1996–97 were tested at the end of the school year using the same second-grade math and French posttests taken in the previous year. Students who were in the third grade took a different test that nevertheless had a substantial number of questions in common with the second-grade tests. There were 38 such anchor items on the math and French tests combined, from

23. King, Orazem, and Paterno's (2008) study using panel data from Pakistan addresses the dropout implications of repetition using a different methodology. Decomposing the promotion probability into the part determined by student merit or ability and the part due to other factors, they find that only student merit has a (positive) effect on school continuation. If the merit-based part is considered a control for student ability, the nonmerit portion essentially captures exogenous sources of variation in repetition probabilities much in the way the repetition indicator controlling for ability does in the present analysis. Thus, the finding that nonmerit factors have no effect on continuation conditional on academic merit is in contrast to the negative effect of repetition found in the this study.

which standardized 1997 scores were calculated for each child. The determinants of this score were estimated in a linear regression including the posttest second-grade score and the repeat dummy variable as well as the other controls used in the dropout probit.

The results show, not surprisingly, a very large and highly significant positive association of scores in 1997 with previous year scores (see table 4). Having repeated second grade is associated with a small (0.2 standard deviation) reduction in test score relative to not repeating, though this effect is not quite significant at the 10 percent level A negative effect of repetition is not implausible, as children left behind may feel stigmatized or get less encouragement to do well in school (note that the negative coefficient does not mean that repeating children do not increase their skills over the subsequent year, only that they gain less than those with similar initial skills who are promoted).[24]

In sum, for a given level of skill (test score) at the end of second grade, having a student repeat that grade does not lead to greater test score gains relative to being promoted and may even lead to slightly reduced improvement, though this effect is imprecisely estimated. Similar results (showing no significant benefit) reported in Michaelowa (2003) were obtained in an analysis for all the Francophone countries in the PASEC study employing a broadly similar methodology. These findings contrast with the widespread perception among teachers documented in CONFEMEN (2003) that repetition leads to greater improvement. The findings suggest that there is little academic benefit to repetition to justify the costs to schools and households.

V. Conclusion

The estimates and simulations in this study indicate that early academic performance is an important predictor of school attainment in a cohort of children in Senegal. Conditional on measured skills at the start of second grade, gender, father's education, and especially household wealth also have impacts on attainment, in expected directions. There is some evidence as well that the positive effects of early achievement are greater for wealthy children and for girls. The strong association of early achievement and later educational attainment points to the benefits of measures directed to lagging children to avoid later disadvantages in skills and labor market outcomes.[25] However, the one (quite costly) policy that is currently directed at such children, grade repetition, tends

24. A negative association might also suggest that unobservable factors such as ability or motivation are affecting both repetition and subsequent test performance. Recall, however, that as with the dropout model, the estimation conditions on test outcomes immediately prior to the repetition decision (second-grade posttest score), which should control for these factors.

25. Though as noted in section III, the appropriate measures and their timing will depend on the source of the early disadvantage.

to exacerbate the negative impacts of poor early school performance on attainment.

Since evidence from a number of African countries indicates that there is a premium to primary (and secondary) completion in terms of earnings or access to formal sector employment, the finding that having lagging learners repeat grades increases premature primary dropout is important for policy. Given this finding as well as the evidence that being held back does not improve learning outcomes relative to being promoted, it would likely be significantly more effective to devise other measures to improve the skills of lagging children rather than simply having them repeat grades. One measure might be to hire local secondary school graduates to tutor such children, along the lines of a successful intervention evaluated in India by Banerjee and others (2007).

Funding

The authors are grateful to the United States Agency for International Development, the World Bank, the United Nations Children's Fund, and Institut National de la Recherche Agronomique (France) for project funding.

Appendix

Table A-1. Determinants of Second-grade Pre- and Posttest Scores

	Pretest		Posttest	
Variable	Coefficient	*t*-statistic	Coefficient	*t*-statistic
Intercept	−1.302	−3.735***	−0.314	−1.020
Girl	−0.296	−3.180***	−0.361	−3.909***
Asset index	0.152	2.928***	0.209	2.947***
Mother's schooling	0.024	1.459	0.032	2.187**
Missing mother's schooling	0.143	1.013	0.100	0.645
Father's schooling	0.005	0.487	0.007	0.664
Missing father's schooling	0.118	0.867	0.142	1.151
Rural	0.344	2.174**	0.098	0.710
Director's experience	0.021	2.340**	0.014	1.388
Director has baccalauréat or higher	0.386	3.135***	−0.043	−0.309
Teacher experience[a]	0.053	4.642***	0.011	1.546
Teacher has baccalauréat or higher[a]	0.301	2.072**	0.237	1.485
Teacher female[a]	0.398	1.947*	−0.182	−1.045
Missing teacher information in 1995–96			1.056	3.311***
Girl × teachers share female	0.059	0.351	0.541	3.764***
Classroom supplies first principal component[a]	0.009	0.215	0.094	1.996**
School facilities first principal component	−0.005	−0.106	0.071	1.213
Adjusted R^2	0.214		0.177	

(Continued)

TABLE A-1 Continued

	Pretest		Posttest	
Variable	Coefficient	*t*-statistic	Coefficient	*t*-statistic
Number of observations	787		756	

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: Pretest and posttest scores are the standardized average scores on French and math tests administered at the beginning and end of second grade, respectively. The models also include ethnic group dummy variables and province dummy variables. Standard errors are adjusted for clustering at the school level. The varying sample sizes reflect the number of students in the PASEC cohort present when the tests were given. See table 1 and text for variable definitions.

a. For the posttest regression, the teacher and classroom supplies variables are the values for the child's second-grade class (the year ending with the posttest). For the pretest regression, they are simple averages over grades 2–4 in the school.

Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

TABLE A-2. Determinants of Second-Grade Repetition: Probit Results

			With school fixed effects	
Variable	Coefficient	*t*-statistic	Coefficient	*t*-statistic
Intercept	−0.766	−1.808*	−2.094	−4.799***
Second-grade posttest score[a]	−1.295	−6.718***	−2.186	−7.323***
Girl	0.370	2.178**	0.409	1.707*
Asset index	−0.212	−1.679*	−0.469	−2.239**
Missing mother's schooling	−0.209	−0.570	0.209	0.352
Mother's schooling	−0.005	−0.147	0.026	0.590
Father's schooling	−0.014	−0.660	−0.025	−0.701
Missing father's schooling	−0.147	−0.636	−0.206	−0.535
Rural	−0.799	−2.370**		
Director's experience	−0.004	−0.251		
Director has baccalauréat or higher	−0.316	−1.296		
Teacher has baccalauréat or higher	−0.249	−0.997		
Teacher female	1.026	2.968***		
Girl × female teacher	−1.203	−3.990***	−1.548	−3.731***
Classroom supplies principal component	−0.446	−3.352***		
School facilities principal component	−0.146	−1.179		
Number of observations	664		664	

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: For fixed effects model, school dummy variables and ethnicity dummy variables not shown. For non-fixed effects model, ethnicity and province dummy variables not shown. Standard errors are adjusted for clustering at the school level. Teacher and classroom supplies variables refer to the second-grade class.

a. Predicted using second-grade pretest score.

Source: Authors' analysis based on data from the PASEC surveys and the 2003 EBMS; see text for details.

References

Alderman, H., J.R. Behrman, D. Ross, and R. Sabot. 1996. "Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy." *Journal of Human Resources* 31(1):229–54.

Allensworth, Elaine M. 2005. "Dropout Rates after High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicago's Efforts to End Social Promotion." *Educational Evaluation and Policy Analysis* 27(4):341–64.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122(3):1235–64.

Behrman, J. 1997. "Intrahousehold Distribution and the Family." In M. Rosenzweig and O. Stark eds., *Handbook of Population and Family Economics*. Amsterdam: North-Holland.

Bernard, J.-M., O. Simon, and K. Vianou. 2005. Le redoublement: mirage de l'ecole africaine? *Programme d'analyse des systèmes éducatifs de la CONFEMEN*. : La Conférence des ministres de l'education.

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement Error in Survey Data." In J.J. Heckman, and E. Leamer eds., *Handbook of Econometrics*. Vol 5. Amsterdam: North-Holland.

Butler, J.S., and R. Moffit. 1982. "A Computationally Efficient Quadrature Procedure for a One-Factor Multinomial Probit Model." *Econometrica* 50(3):761–4.

CONFEMEN (La Conférence des ministres de l'education). 1999. "Les facteurs de l'efficacité de l'enseignement primaire: données et résultats sur cinq pays d'Afrique et de l'Océan Indien." CD-ROM. La Conférence des ministres de l'education, Dakar.

CONFEMEN PASEC (Le programme d'analyse des systèmes educatifs des etats et gouvernements de La Conférence des ministres de l'education). 2003. *Le redoublement: pratiques et conséquences; résultats de l'analyse longitudinale des données du Sénégal*. Dakar: Le programme d'analyse des systèmes educatifs des etats et gouvernements de La Conférence des ministres de l'education.

Currie, J., and D. Thomas. 2001. "Early Test Scores, School Quality, and SES: Long Run Effects on Wages and Employment Outcomes." In Solomon Polachek eds., Worker Wellbeing in a Changing Labor Market *(*Research in Labor Economics *Vol 20). Amsterdam*, New York: Elsevier JAI.

Duncan, Gregg., Amy Claessens, Aletha C. Huston, Linda S. Pagini, Mimi Engel, Holly Sexton, and Chantelle J. Dowsett *et al.* 2007. "School Readiness and Later Achievement." *Developmental Psychology* 43(6):1428–46.

Eide, Eric R., and Mark H. Showalter. 2001. "The Effect of Grade Retention on Educational and Labor Market Outcomes." *Economics of Education Review* 20(6):563–76.

Garg, A., and J. Morduch. 1998. "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana." *Journal of Population Economics* 11(4):471–93.

Glewwe, Paul., and Hanan Jacoby. 1994. "Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana." *Journal of Human Resources* 29(3):843–64.

Glewwe, P., and M. Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." In E. Hanushek and F. Welch eds., *Handbook on the Economics of Education*. Amsterdam: North-Holland.

Glewwe, P., and E. Miguel. 2008. "The Impact of Child Health and Nutrition on Education in Less Developed Countries." In T. Paul Schultz and John Strauss eds., *Handbook of Development Economics*. Amsterdam: North-Holland.

Glick, P. 2008. "What Policies Will Reduce Gender Schooling Gaps in Developing Countries: Evidence and Interpretation." *World Development* 36(9):1623–46.

Glick, P., and D.E. Sahn. 2000. "Schooling of Girls and Boys in a West African Country: The Effects of Parental Education, Income, and Household Structure." *Economics of Education Review* 19(1): 63–87.

———. 2009. "Cognitive Skills among Children in Senegal: Disentangling the Roles of Schooling and Family Background." *Economics of Education Review* 28(2):178–88.

Gomes-Neto, J., and E. Hanushek. 1994. "Causes and Consequences of Grade Repetition: Evidence from Brazil." *Economic Development and Cultural Change* 43(1):117–48.

Greene, W.H. 2002. "The Behavior of the Fixed Effects Estimator in Nonlinear Models." New York University, Department of Economics, Stern School of Business, New York.

Greene, Jay P., and Marcus A. Winters. 2007. "Revisiting Grade Retention: An Evaluation of Florida's Test Based Retention Policy." *Education Finance and Policy* 2(4):319–40.

Grissom, James B., and Lorie A. Shepard. 1989. "Repeating and Dropping Out of School." In L.A. Shepard and M.L. Smith eds., *Flunking Grades: Research and Policies on Retention*. London: Falmer Press.

Jacob, B., and L. Lefgren. 2004. "Remedial Education and Student Achievement: A Regression—Discontinuity Analysis." *Review of Economics and Statistics* 86(1):226–44.

King, E., P. Orazem, and E. Paterno. 2008. "Promotion with and without Learning: Effects on Student Dropout." Iowa State University, Department of Economics, Ames, Iowa.

Kingdon, Geeta. 1996. "The Quality and Efficiency of Public and Private Schools: A Case Study of Urban India." *Oxford Bulletin of Economics and Statistics* 58(1):55–80.

Maani, Sholeh A., and Guyonne Kalb. 2007. "Academic Performance, Childhood Economic Resources, and the Choice to Leave School at Age 16." *Economics of Education Review* 26(3):361–74.

Manacorda, Marco. 2007. *The Cost of Grade Retention*. Centre for Economic Performance Discussion Paper 878. London School of Economics and Political Science, London, U.K.

Michaelowa, K. 2003. "Determinants of Primary Education Quality: What Can We Learn from PASEC for Francophone Sub-Saharan Africa?" Association for the Development of Education in Africa, Paris.

Phillips, Meredith, James Crouse, and John Ralph. 1998. "Does the Black-White Test Score Gap Widen after Children Enter School?" In Christopher Jencks and Meredith Phillips eds., *The Black-White Test Score Gap*. Washington, DC: The Brookings Institute.

Pitt, Mark M., Mark R. Rosenzweig, and Md. Nazmul Hassan. 1990. "Productivity, Health, and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries." *American Economic Review* 80(5):1139–56.

Robertson, Donald, and James Symons. 1990. "The Occupational Choice of British Children." *The Economic Journal* 100(402):828–41.

Sahn, David E., and David Stifel. 2003. "Exploring Alternative Measures of Welfare in the Absence of Expenditure Data." *Review of Income and Wealth* 49(4):463–89.

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.

Technology Adoption and the Investment Climate: Firm-Level Evidence for Eastern Europe and Central Asia

Paulo G. Correa, Ana M. Fernandes, and Chris J. Uregian

Survey data for 7,000 firms in 28 countries in Eastern Europe and Central Asia are used to examine the correlates of technology adoption proxied by ISO certification and web use. Complementary inputs such as skilled labor, managerial capacity, research and development, finance, and good infrastructure are shown to be important correlates of technology adoption. The link between market incentives and technology adoption is more nuanced. While stronger consumer pressure is significantly associated with technology adoption, competitor pressure is not, suggesting that in developing economies where many input markets are imperfect, it is primarily firms with rents that are able to adopt new technology. Foreign-owned firms exhibit significantly better technology adoption outcomes, but privatized firms with domestic owners do not. JEL codes: F1, F2, O3

Differences in technology—defined broadly to encompass all types of knowledge relevant to the production of goods and services—are an important determinant of differences in total factor productivity across countries (Prescott 1998; Hall and Jones 1999). While new technology is generated in only a few research and development (R&D)-intensive economies, the expansion in the volume of capital goods trade indicates the broader availability of technology embodied in new machinery and equipment (Eaton and Kortum 2001). Yet, when faced with similar technological alternatives, firms in some countries choose less efficient technologies even when more efficient ones are available because barriers to new technology adoption (such as those linked to the regulatory framework) distort the relative payoffs in favor of suboptimal technologies (Parente and Prescott 1994). Accordingly, countries have widely divergent living standards not because

Paulo G. Correa (pcorrea@worldbank.org) is a senior economist in the Europe and Central Asia region of the World Bank. Ana M. Fernandes (corresponding author, afernandes@worldbank.org) is an economist in the Development Research Group of the World Bank. Chris Uregian (curegian@worldbank.org) is a consultant in the Europe and Central Asia region of the World Bank.

The authors thank three anonymous referees and the journal editor for helpful comments that substantially improved the article, as well as Brett Coleman, Itzahk Goldberg, Smita Kuriakose, and Jean-Louis Racine for their suggestions.

THE WORLD BANK ECONOMIC REVIEW, VOL. 24, NO. 1, pp. 121–147 doi:10.1093/wber/lhp021
Advance Access Publication January 12, 2010

they access different stocks of knowledge, but because of differences in constraints imposed on technology choices (Parente and Prescott 2005).

This article uses a unique data set to explore how the investment climate affects a firm's technology choices. The World Bank Enterprise Surveys for firms in the Eastern Europe and Central Asia region in 2002 and 2005 contain information on technology adoption and on aspects of the investment climate related to complementary inputs (skilled labor, finance) and the market incentives facing firms (competition, ownership) for a large number of transition economies.

Studies on technology diffusion often take a macroeconomic trade-related focus, as illustrated by Zeira (1998) and Keller (2004). Most of the applied microeconomic studies reviewed by Hall and Khan (2003) estimate diffusion curves for a few technologies in a few countries, providing limited understanding of the barriers to technology adoption (Caselli and Coleman 2001; Comin and Hobijn 2008). Microeconomic evidence on the determinants of technology adoption in developing countries is scarce because of data limitations. This article attempts to fill in this gap by studying the technology choices of firms in Europe and Central Asia after a decade of transition to a market economy.

The results show that access to appropriate complementary inputs—skilled labor, managerial capacity, R&D, finance, and to a lesser extent good infrastructure—are strongly positively associated with International Organization for Standardization (ISO) 9000 certification and web use by firms. The relationship between market incentives and technology adoption is more nuanced. While consumer pressure is related to ISO certification and web use, competitor pressure is not. Fully foreign-owned firms and joint ventures exhibit significantly better technology adoption outcomes, but privatization to domestic owners is not systematically associated with more frequent technology adoption. The results provide evidence of robust correlations but cannot be interpreted as causal relationships because of the nature of the data.

The article is organized as follows. Section I describes the technology adoption measures. Section II presents the empirical strategy and methodological issues. Section III presents the results. Section IV offers some policy implications of the findings.

I. Data and Technology Adoption Measures

The Enterprise Surveys are conducted by the World Bank and the European Bank for Reconstruction and Development in 28 Europe and Central Asia countries.[1] The samples include a cross section of 6,667 firms in 2002 and 9,655 firms in 2005, plus a panel sample of 1,446 firms surveyed in both 2002

1. The surveys were formerly known as Business Environment and Enterprise Performance Surveys and are described at www.enterprisesurveys.org.

and 2005.[2] The surveys cover manufacturing and services sectors and are representative of the universe of firms by sector and location within each country.[3] Two proxies are used for the adoption of new technology: ISO 9000 certification and use of the web (email and Internet) for business operations.

ISO 9000 is a set of internationally accepted standards and technical regulations on quality management systems in manufacturing and services firms developed by the ISO (1998).[4] ISO certification focuses on improving a firm's operating processes to enhance quality and efficiency (Benner and Veloso 2008).[5] ISO certification requires detailed review and documentation of the routines underlying delivery of the firm's products and services. The routines are subject to improvements to rationalize processes and streamline interfaces between the firm's subunits. Adopting standardized best practices throughout the firm ensures that the organizational processes are repeated, allowing for continued efficiency improvements. ISO certification is awarded based on a detailed review of a firm's processes, documentation that the processes comply with ISO quality system standards, and an audit by an accredited third party (Arora and Asundi 1999). After ISO certification is awarded, regular audits are conducted to ensure continuing compliance.

There are advantages and limitations in choosing ISO certification to proxy for new technology adoption, but on balance ISO certification captures relevant aspects of new technology adoption for developing country firms. First, ISO certification indicates adherence to consistent process standards; thus it represents adoption of advanced organizational technology by firms, a key component of technological knowledge (Lipsey and Carlaw 2004). However, ISO certification is intended to minimize variations in quality; it is not awarded for product or service quality and does not capture product design improvements by firms. Second, though ISO certification does not map into the adoption of easily identifiable technologies, it is objective and comparable across firms, sectors, and countries, which is crucial for this study of correlates of technology adoption across firms. Third, the adoption of international standards and technical regulations through ISO certification is a major channel for U.K. and U.S. firms to acquire technological information and introduce product and process technology upgrading (Blind and others 2005; Corbett, Montes-Sancho, and Kirsh 2005). For developing country firms not yet operating at the world technological frontier, this channel is likely to be even more important. Fourth, ISO certification facilitates the entry of local firms into global supply chain networks, which often brings the transfer of knowledge from technologically advanced buyers (Humphrey and Schmitz 2000).

2. Safavian and Sharma (2007) use this panel sample to study firm access to finance.

3. Correa, Fernandes, and Uregian (2008) show the country and sector composition of the samples.

4. The principle underlying ISO certification is that better defined and documented processes lead to better output (Arora and Asundi 1999).

5. An example of an operating process is a manufacturing assembly line, with discrete stages assigned to individuals and machines and a specific product resulting at the end (Benner 1999).

Web use is a proxy for a firm's use of information and communication technology in business operations. Information and communication technology, an advance that changed modes of production and operation, is considered the preeminent general purpose technology of the last two decades. It is used pervasively across sectors, has prompted further innovation (Bresnahan and Trajtenberg 1995), and can have beneficial effects on productivity growth (Indjikian and Siegel 2005).

There are also advantages and limitations in choosing web use to proxy for new technology adoption. First, web use captures the adoption of general-use information and communication technologies that allow firms to process and transmit information faster, improve management and internal organization, and achieve efficiency gains. Second, web use is an objective measure, comparable across firms, sectors, and countries, making it well suited to a study of technology adoption correlates across firms. One limitation of the web use measure is its inability to capture the adoption of important information and communication technologies applied to improve production (such as computer-aided manufacturing) or reduce coordination and transaction costs (such as local area networks). Another limitation is that it does not capture firms' intensity of usage.

Three-quarters of firms in Europe and Central Asia consider machinery and equipment acquisition as their main source of technological update, according to the 2005 survey.[6] The technological advances conveyed by information and communication technology are embodied in new capital goods by nature: for example, for a firm to use the web it needs to acquire compatible computers and telecommunication tools (Boucekkine, Del Rio, and Licandro 2003). Thus, the choice of web use as a proxy for technology adoption implicitly presumes the importance of capital-embodied technological change,[7] while that is less obvious for ISO certification.

The two proxies are the best available across firms, sectors, and countries and enable examination of two dimensions of technology adoption not exploited before in a developing country context. However, they have limitations, and future research on the determinants of technology adoption should aim to collect cross-country firm-level information on the intensity of usage, measured by total investments in high-technology capital or the percentage of

6. The other possible sources of technology updating were the hiring of key personnel/consultants with technological expertise; new license or turnkey operations from international sources; new license or turnkey operations from domestic sources; developed or adapted within the firm; transferred from the parent company; developed in cooperation with customers; developed in cooperation with suppliers; obtained from business or industry associations; or obtained from universities or public institutions. The extent to which new technology is embodied in new capital equipment, the subject of long-standing debate, is not addressed here.

7. The importance of capital-embodied technological change for output growth is shown by Hulten (1992) and Long and Summers (1993), though Hulten (1992) also shows a role for disembodied technical change.

employees using high-technology capital (such as office, computing, and accounting machinery).[8] In addition to ISO certification and web use, this study uses survey data on firm ownership, size, workforce skills, managerial capacity, research and development (R&D) expenditures, access to credit, infrastructure failures faced, competition proxies, exports, and imported inputs (see appendix table A-1 for all variables used in the analysis and their definitions).

In 2002, 13.6 percent of firms were ISO certified and 58.2 percent used the web (table 1). In 2005, 12.5 percent of firms were ISO certified and 67.4 percent used the web. For both proxies and years, technology adoption is more prevalent in the EU-8 countries (defined in appendix table A-1). There is substantial heterogeneity in technology adoption across sectors (table 2). Firms in mining, quarrying, and construction are significantly more likely to be ISO certified but no more likely than the average firm to use the web.[9] Firms in manufacturing are significantly more likely to be ISO certified and use the web than are firms in services (real estate being the exception). The ISO certification finding is consistent with global experience of greater prevalence of certification in manufacturing, where quality signals matter for export competitiveness.[10]

In ordinary least squares regressions (not reported here), ISO certification and web use are strongly correlated with firm performance in Europe and Central Asia, after controlling for sector fixed effects and GDP per capita (or country fixed effects) based on the 2002 or 2005 samples. ISO-certified firms and web users exhibit significantly higher average value added per worker and faster sales growth and pay higher wages than firms not adopting those technologies, within sectors and countries. The data do not enable establishing causality for these estimated performance premia of technology adopters, but the strong correlation validates the use of ISO certification and web use as proxies for economically relevant technology upgrading by firms.

II. Empirical Strategy and Methodological Issues

This section describes the empirical strategy, discusses the investment climate factors considered, and highlights the methodological issues associated with the empirical strategy.

Empirical Strategy

When deciding whether to adopt a new technology, firms are assumed to make a profit-maximizing cost–benefit assessment of different alternatives. A firm

8. Basant and others (2006) undertake a first effort in this direction for Brazilian and Indian firms.

9. Adoption of ISO certification in mining and quarrying is likely associated with the resource-intensive exports of some transition economies, whereas in construction it may be explained by government requirements that contractors be ISO certified (Guler, Guillen, and Macpherson 2005).

10. In particular, EU directives requiring quality system registration have made ISO certification imperative for firms in Europe and Central Asia aspiring to collaborate with EU firms through supplier relationships (Guler, Guillen, and Macpherson 2005).

TABLE 1. Summary Statistics (Percent Unless Otherwise Indicated)

	2002 sample		2005 sample		Panel sample	
Variable	Mean	Observations	Mean	Observations	Mean	Observations
ISO-certification dummy variable	13.6	6,610	12.5	9,655	14.0	2,887
Web-use dummy variable	58.2	6,667	67.4	9,655	65.4	2,892
Share of skilled labor	18.0	6,572	17.0	9,519	20.0	2,634
Manager with college education dummy variable	70.2	6,611	—	—	—	—
Manager age (years)	44.6	6,610	—	—	—	—
R&D intensity	0.0	6,667	0.0	6,984	0.4	980
Access to finance dummy variable	40.3	6,655	42.1	9,655	42.7	2,888
Number of days with power outages	11.0	6,656	9.7	9,645	15.9	2,881
Number of days with unavailable telephone lines	5.8	6,653	1.7	9,635	3.2	2,868
Dummy variable for market share less than 5 percent	65.7	6,667	—	—	—	—
Price–cost margin (percent)	18.9	5,656	22.7	8,460	20.4	2,045
Dummy variable for pressure to innovate from competitors	74.1	6,586	75.2	9,538	73.4	2,852
Dummy variable for pressure to innovate from consumers	71.7	6,560	72.2	9,466	69.2	2,838
Number of permanent workers	143	6,636	102	9,654	109	2,888
Dummy variable for privatized firm	14.4	6,667	13.7	9,655	15.9	2,892
Dummy variable for private firm (from start-up)	71.7	6,667	77.6	9,655	69.8	2,892
Dummy variable for fully foreign-owned firm	6.6	6,667	5.1	9,655	7.4	2,892
Dummy variable for joint venture	9.5	6,667	6.8	9,655	6.4	2,892
Ownership share of largest shareholder	76.7	6,180	76.8	9,563	76.3	2,788
Export share	11.4	6,636	10.1	9,642	10.6	2,886
Imported inputs share	29.8	6,667	28.0	9,655	30.9	2,892

— is not available.

Source: Authors' analysis based on 2002 and 2005 Enterprise Surveys; see text for details.

decides to adopt if the corresponding expected net benefits (benefits minus costs) are larger than those of the alternatives, including that of not adopting new technology. Let π^*_{ijc} be the net benefits for firm i in sector j and country c and *Adoption* be a dummy variable that equals 1 if firm i adopts ISO

TABLE 2. Technology Adoption across Sectors for Sample Firms in Europe and Central Asia (percent)

	ISO certification		Web use	
Sector	2002 sample	2005 sample	2002 sample	2005 sample
Region average	13.6	12.5	58.2	67.4
Mining and quarrying, energy related	36.7**	14.3	70.0	75.0
Mining and quarrying, not energy related	21.3	19.4	62.5	70.1
Manufacturing				
Food beverages and tobacco	23.6***	16.2***	54.1*	57.0***
Textiles	17.4	7.9***	58.4	63.0**
Leather	9.7	8.3	58.1	54.2*
Wood	23.8*	7.1*	50.0	60.0
Pulp and paper	11.2	13.8	81.0***	84.1***
Petroleum	8.3	16.7	91.7***	100.0***
Chemicals	26.6**	36.6***	78.8***	88.1***
Rubber and plastics	24.5*	23.3**	62.3***	88.9***
Nonmetallic minerals	17.2	20.3**	54.0	64.2
Metals	24.4***	19.0***	65.2*	77.2***
Machinery and equipment	33.8***	24.9***	79.9***	82.1***
Electrical and optical equipment	34.9***	30.4***	81.1***	88.7***
Transport equipment	42.3***	42.6***	86.8***	87.2***
Other manufacturing	20.3*	11.6	64.9	70.7
Services				
Construction	17.4***	15.9***	56.7	69.9*
Wholesale and retail trade	8.7***	8.0***	50.0***	62.3***
Hotels and restaurants	6.2***	7.1***	42.2***	53.0***
Transport, storage, and communications	10.0***	11.0	71.8***	78.5***
Real estate and business activities	10.9**	9.2***	71.4***	79.5***
Other services	5.1***	4.5***	50.1***	53.7***

*Significant at the 10 percent level; **Significant at the 5 percent level; ***Significant at the 1 percent level.

Source: Authors' analysis based on 2002 and 2005 Enterprise Surveys; see text for details.

certification or web use. Then:

$$\text{Adoption}_{ijc} = \begin{cases} 1 & \text{if } \pi^*_{ijc} > 0 \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

The unobserved π^*_{ijc} is allowed to depend linearly on two sets of investment climate factors—access to complementary inputs, Inp_{ijc}, and market incentives Inc_{ijc}—as well as firm controls, X_{ijc}, sector fixed effects, I_j, and GDP per capita, GDPpc_c:

$$\pi^*_{ijc} = \alpha\text{Inp}_{ijc} + \beta\text{Inc}_{ijc} + \gamma X_{ijc} + I_j + \text{GDPpc}_c + \varepsilon_{ijc}, \tag{2}$$

where ε_{ijc} represents unobserved firm characteristics influencing the adoption decision.

The probability of adopting new technology for firm i is given by:

$$(3)\quad \Pr(adoption_{ijc} = 1) = \Pr(\varepsilon_{ijc} > -\alpha \mathrm{Inp}_{ijc} - \beta \mathrm{Inc}_{ijc} - \gamma X_{ijc} - I_j - \mathrm{GDPpc}_c.$$

.

Complementary Inputs

Two sets of investment climate factors can influence the probability of adopting new technology—complementary inputs and market incentives (discussed below). A firm's access to complementary inputs can affect both the adjustment costs following the adoption of new technology and the benefits derived from adoption.

Both theoretical models and empirical evidence show how poor *labor skills* can delay a firm's technology adoption, whether because of an inability to operate advanced equipment—generally skill-biased—or because learning is technology-specific and retraining workers is costly (Navaretti, Soloanga, and Takacs 2001; Berdugo, Sadik, and Sussman 2003; Alesina and Zeira 2006). The lack of highly qualified *managerial capacity* can constrain a firm's adoption of new technology by reducing its acquisition of information on available technological solutions for its needs and could increase its adjustment costs. Inadequate *R&D investments* by a firm also hamper technology adoption possibilities since such investments are often made to develop its capabilities to assimilate and exploit external knowledge (Cohen and Levinthal 1989).

Although capital goods are considered sound collateral in developed countries, failure to regulate movable collateral in countries in Europe and Central Asia might constrain firms' *access to credit*, restricting technology adoption. Depending on the gap between existing and new technologies and associated adjustment costs, complementary physical investments might be needed, making access to credit a decisive input for technological update. Access to credit is even more crucial because new technologies such as information and communication technology are frequently embodied in capital goods. Moreover, productivity gains from information and communication technologies seem to depend on complementary investments and organizational changes (Bresnahan, Brynjolfsson, and Hitt 2002).

The availability and quality of physical *infrastructure*—especially electricity and telecommunications services—can be decisive in a firm's decision to adopt new technology, particularly information and communication technologies. Arnold, Mattoo, and Narcisco (2008) show that difficulty obtaining adequate infrastructure services can constrain firm performance.

The empirical specifications include all those measures as proxies for access to complementary inputs.

Market Incentives

Strong *product market competition* from domestic and foreign rivals and demand from consumers likely encourage incumbent firms to invest in new and more productive technologies—product upgrades and cost reductions—rather than to spend on rent-seeking activities (Baumol 1990; Aghion and Schankerman 2004).[11] Moreover, the entry of new competitors may foster innovation in incumbent firms as they try to beat their competitors and survive (Aghion and others 2001).

However, according to Schumpeter (1942), firms facing a lower entry threat are best placed to innovate since innovation requires the expectation of temporary monopoly rents. Strong product market competition can be detrimental to firms by reducing the rents that successful innovators can appropriate after introducing an innovation to the market. Aghion and others (2009) show that stronger competition has a negative effect on incumbent firms' innovation incentives (though only in industries far behind the world technology frontier). These arguments can be applied to new technology adoption by firms in developing countries: investments in technologies or processes that are widespread in developed countries but that are new locally can appear too risky if firms do not believe they will subsequently enjoy sufficiently high rents.

Private ownership and control establish profit maximization as the firm's objective and should in principle foster adopting new technologies with ensuing productivity gains (Brown, Earle, and Telegdy 2007). *Foreign ownership* in particular, whether full or partial, exposes firms to global best practice technology and management techniques (Djankov and Hoekman 2000).

The empirical specifications include these measures as proxies for market incentives.

Methodology

Equation (3) is estimated by maximum likelihood (probit), assuming that the residual, ε_{ijc}, is normally distributed. The probit specifications control for sector fixed effects since differences in technology, product demand, and competition across sectors can influence firms' incentives to adopt new technology (Cohen and Levin 1989). GDP per capita is included in the specifications to account for country heterogeneity in technology adoption not captured by differences in complementary inputs or market incentives.[12] Moreover, standard errors are adjusted for clustering at the country level to account for

11. Accordingly, Comin and Hobijn (2004) argue that trade openness is important to a country's speed of adoption of advanced technologies because openness introduces pressures from foreign competition, reducing incumbents' payoff from lobbying the government to deter the adoption of advanced technologies.

12. Country fixed effects cannot be used because of their collinearity with the location-specific infrastructure index.

possible correlations in technology adoption decisions across firms within a country.

The main results use probit estimation for ISO-certification and web-use regressions based on a cross section of firms in Europe and Central Asia in 2002 or 2005. The results identify systematic correlations between complementary inputs or market incentives and ISO certification or web use, but the estimated effects cannot be interpreted as causal.

The effects of some investment climate factors may reflect omitted managerial ability or other unobservable firm characteristics or suffer from reverse causality.[13] The strategy to address this problem is threefold. First, the specifications include firm controls—share of total output exported and share of total intermediate inputs imported—to minimize the risk that the estimates suffer from an omitted variables bias. The firm controls considered capture the firm's access to international knowledge through its involvement in international trade—shown in previous research to be a determinant of technology transfer and adoption (Keller 2004; Almeida and Fernandes 2008). However, the findings based on the cross sections of firms in 2002 and 2005 could still be driven partly by unobservables.

Second, to compute the infrastructure index, firm responses to the questions on electricity and telecommunications (detailed in the appendix) are averaged at the location level to correct for potential endogeneity of firm perceptions about the quality of infrastructure with respect to technology adoption decisions.

Third, the specification is estimated based on the panel sample using probit with random effects estimation. This approach shows how changes in access to complementary inputs and market incentives between 2002 and 2005 relate to ISO certification and web use by firms, controlling for unobserved firm heterogeneity. A disadvantage of this approach is that probit with random effects estimation requires assuming that the unobserved firm effects are uncorrelated with the regressors. While this assumption is verified when the unobserved firm effects capture unexpected production breakdowns suffered by firms during 2002–05, it might not be verified if those effects capture managerial risk aversion. Given the short panel dimension of the data—only two years of data per firm—and the much smaller size of the panel relative to the cross-sectional samples, other estimation methods could not be used for the panel regressions.[14] Thus, the magnitude of the panel results should be viewed with

13. For example, firms with more able managers are more likely to adopt new technology but are also more likely to hire more skilled workers. Thus, the effect of skilled labor in the technology adoption regressions may to some extent reflect omitted managerial ability.

14. Conditional fixed effects logit estimation is an alternative method for panel regressions with a binary-dependent variable (Maddala 1987), but the estimation relies on firms' changing status in the dependent variable, which would imply the use of a fraction of an already small sample, possibly resulting in biased estimates.

caution and taken only as an indication of the patterns of correlation once firm heterogeneity is imperfectly accounted for.

A final concern is that a firm's access to complementary inputs and market incentives (such as foreign ownership) could influence the ISO's decision to award certification, despite the fact that none of the variables is explicitly among the certification eligibility criteria.[15] Therefore, a positive coefficient on skilled labor, for example, would capture not only the positive effect of that input in the firm's decision to adopt ISO certification but also the influence such input may have played in the ISO's decision to award certification.

III. Results

This section presents the results of estimating several variants of equation (3).

Complementary Inputs

Firms employing a larger share of skilled labor (professionals) are significantly more likely to be ISO certified and to use the web (table 3). All else equal, an increase in a firm's skilled labor share by one standard deviation (22.1 percent) would be associated with a 1.5 percent increase in the frequency of ISO certification (regression 1) and an 8.3 percent increase in the frequency of web use (regression 4). The results for the panel sample confirm the importance of skills by showing that firms that increased their skilled labor share between 2002 and 2005 are significantly more likely to become ISO certified (regression 3) or to use the web (regression 6). These results are consistent with evidence in Guler, Guillen, and Macpherson (2002) on the importance of professionals with a technical background for ISO certification, because such professionals can more easily deal with the technical aspects of quality standards.

Managerial education is also strongly positively associated with ISO certification and web use. Firms run by managers with a college or postgraduate degree in 2002 are almost 4 percent more likely to be ISO certified and 23 percent more likely to use the web, all else held constant. While ISO-certified firms are more likely to be run by older managers, web use is more frequent in firms run by younger managers.[16] Firms with higher R&D intensity are also significantly more likely to be ISO certified and to use the web. A one standard deviation higher R&D intensity (5.6 percent) is associated with a 2.4 percent increase in the frequency of web use in 2002, all else held constant. The effects of R&D intensity on ISO certification are positive and weaker for the panel sample but strong for web use.

These findings show the importance of complementary investments in skills and R&D for technology adoption in firms in Europe and Central Asia and

15. The authors thank an anonymous referee for highlighting this possibility.

16. Managerial characteristics are not available in the 2005 survey, and so they are also excluded from the panel regressions.

TABLE 3. Correlates of Technology Adoption for ISO-Certification and Web-Use-Dependent Variables

Variable	ISO-certification dummy variable			Web-use dummy variable		
	2002 sample (1)	2005 sample (2)	Panel sample (3)	2002 sample (4)	2005 sample (5)	Panel sample (6)
Complementary inputs						
Share of skilled labor	0.065 (0.009)***	0.061 (0.000)***	0.139 (0.000)***	0.374 (0.000)***	0.319 (0.000)***	0.109 (0.002)***
Manager with college education dummy variable	0.038 (0.000)***			0.233 (0.000)***		
Manager age	0.001 (0.006)***			−0.003 (0.080)*		
R&D intensity	0.168 (0.013)**	0.547 0.000)***	0.211 (0.214)	0.436 (0.080)*	1.430 (0.019)**	0.185 (0.001)***
Access to finance dummy variable	0.046 (0.000)***	0.038 (0.000)***	0.049 (0.003)***	0.110 (0.000)***	0.140 (0.000)***	0.042 (0.000)***
Infrastructure index[a]	−0.011 (0.039)**	-0.005 (0.127)	−0.010 (0.125)	0.061 (0.007)***	0.056 (0.000)***	−0.010 (0.000)***
Market incentives						
Dummy variable for market share less than 5 percent	−0.049 (0.000)***			−0.104 (0.000)***		
Price–cost margin (percent)		0.003 (0.895)	0.030 (0.823)		−0.003 (0.969)	0.010 (0.118)
Dummy variable for pressure to innovate from competitors	−0.002 (0.870)	0.013 (0.242)	0.023 (0.338)	−0.003 (0.884)	0.039 (0.003)***	0.015 (0.606)
Dummy variable for pressure to innovate from consumers	0.019 (0.046)**	0.019 (0.055)*	0.022 (0.308)	0.046 (0.003)***	0.013 (0.336)	0.015 (0.296)

Dummy variable for medium-size firms	0.067 (0.000)***	0.082 (0.000)***	0.091 (0.001)***	0.170 (0.000)***	0.195 (0.000)***	0.073 (0.000)***
Dummy variable for large firms	0.101 (0.000)***	0.160 (0.000)***	0.214 (0.000)***	0.249 (0.000)***	0.240 (0.000)***	0.213 (0.000)***
Dummy variable for privatized firm	0.020 (0.247)	−0.028 (0.176)	−0.031 (0.338)	−0.006 (0.865)	−0.058 (0.205)	−0.023 (0.738)
Dummy variable for private firm (from start-up)	0.008 (0.603)	−0.042 (0.083)*	−0.010 (0.709)	0.097 (0.003)***	−0.006 (0.857)	−0.009 (0.193)
Dummy variable for fully foreign-owned firm	0.068 (0.000)***	0.036 (0.068)*	0.032 (0.376)	0.242 (0.000)***	0.126 (0.001)***	0.026 (0.002)***
Dummy variable for joint venture	0.047 (0.026)**	0.026 (0.082)*	0.142 (0.001)***	0.159 (0.000)***	0.104 (0.007)***	0.121 (0.071)*
Ownership share of largest shareholder	−0.002 (0.921)	−0.032 (0.068)*	−0.007 (0.942)	−0.071 (0.065)*	−0.130 (0.000)***	−0.002 (0.000)***
Controls						
Export share	0.078 (0.002)***	0.061 (0.001)***	0.095 0.075)*	0.256 (0.000)***	0.332 (0.000)***	0.318 (0.001)***
Imported inputs share	0.024 (0.087)*	0.044 (0.000)***	0.028 (0.221)	0.282 (0.000)***	0.273 (0.000)***	0.269 (0.000)***
GDP per capita	0.028 (0.004)***	0.027 (0.000)***	0.017 (0.264)	0.160 (0.000)***	0.140 (0.000)***	0.126 (0.000)***
Number of observations	5,589	7,968	1,906	5,625	7,965	1,916

*Significant at the 10 percent level; **Significant at the 5 percent level; ***Significant at the 1 percent level.

Note: Marginal effects at mean values from probit regressions are shown. Numbers in parentheses are *p*-values corresponding to robust standard errors clustered by country. All regressions include sector fixed effects. See appendix table A-1 for variable definitions.

a. Higher values indicate better infrastructure.

Source: Authors' analysis based on 2002 and 2005 Enterprise Surveys; see text for details.

support the role of R&D in helping firms develop absorptive capacities for external knowledge (Cohen and Levinthal 1989). R&D activities are also likely to have spillovers into managerial activities, as firms learn about their technological bottlenecks and possible solutions.

Access to finance is strongly associated with ISO certification and web use. In 2005 a firm with a bank loan is 4.2 percent more likely to be ISO certified and 15.8 percent more likely to use the web, all else held constant (see table 3). Firms gaining access to finance between 2002 and 2005 are significantly more likely to be ISO certified or to use the web. Access to finance can be crucial for decisions on adopting new technologies and on making the complementary investments needed to absorb and efficiently use the technologies. For web use, the importance of finance relates to the way technology is embodied in new capital goods. For ISO certification, the findings likely reflect the substantial costs related to the search for information on ISO standards, technical assistance for process improvement and the adoption of standards, and the fees to apply for the certification (Guler, Guillen, and Macpherson 2002). The findings provide evidence of an important microchannel through which finance may affect growth in firms in Europe and Central Asia, by increasing technology adoption.

Better infrastructure (indicated by a higher value for the infrastructure index) is significantly positively associated with web use (see table 3). Since the index captures the quality of the telecommunications network in the firm's location, this result reflects its importance for the efficient use of general-purpose information and communication technology by the firm. However, better infrastructure is also negatively associated with ISO certification—significantly so in 2002. The panel sample results show that in locations where infrastructure improved between 2002 and 2005, the frequency of web use increased but that of ISO certification decreased.

One possible explanation for this counterintuitive sign in the ISO-certification regressions is that the effect of infrastructure on ISO certification operates through other variables also included in the regressions. Another is that the infrastructure index does not account for the costs of remoteness and the risk of losses in transit (which could be proxied by the quality of the road infrastructure), which would be the most important infrastructure-related potential determinants of ISO certification. Nevertheless, the infrastructure index does exhibit sensible values, substantially higher for EU-8 and Southeastern European countries plus Turkey (with the exception of Albania) than for countries in the Commonwealth of Independent States, whose infrastructure is worse.

The findings on the effects of access to complementary inputs on technology adoption are not driven by any specific subgroup of countries (table 4). The importance of skills, R&D, and finance for ISO certification and web use is confirmed in all country subgroups. Better infrastructure has a positive effect on web use and a (weak) negative effect on ISO certification in all country

TABLE 4. Correlates of Technology Adoption across Country Subgroups for ISO-Certification and Web-Use-Dependent Variables

	EU-8 countries				Commonwealth of Independent States countries				Southeastern European countries and Turkey			
	ISO-certification dummy variable		Web-use dummy variable		ISO-certification dummy variable		Web-use dummy variable		ISO-certification dummy variable		Web-use dummy variable	
Variable	2002 sample (1)	2005 sample (2)	2002 sample (3)	2005 sample (4)	2002 sample (5)	2005 sample (6)	2002 sample (7)	2005 sample (8)	2002 sample (9)	2005 sample (10)	2002 sample (11)	2005 sample (12)
Complementary inputs												
Share of skilled labor	0.107 (0.007)***	0.082 (0.008)***	0.232 (0.000)***	0.147 (0.002)***	−0.025 (0.294)	0.045 (0.022)**	0.470 (0.000)***	0.345 (0.000)***	0.123 (0.000)***	0.078 (0.001)***	0.230 (0.004)***	0.390 (0.000)***
Manager with college education dummy variable	0.033 (0.000)***		0.141 (0.000)***		0.015 (0.171)		0.203 (0.000)***		0.025 (0.273)		0.272 (0.000)***	
Manager age	0.003 (0.000)***		−0.001 (0.457)		0 (0.542)		−0.006 (0.000)***		0.001 (0.105)		0.001 (0.662)	
R&D intensity	0.145 (0.303)	0.635 (0.103)	0.271 (0.000)***	3.701 (0.000)***	0.102 (0.166)	0.495 (0.029)**	0.737 (0.106)	1.226 (0.148)	0.297 (0.049)**	0.530 (0.023)**	0.031 (0.933)	1.336 (0.081)*
Access to finance dummy variable	0.057 (0.000)***	0.047 (0.001)***	0.098 (0.000)***	0.053 (0.000)***	0.037 (0.011)**	0.016 (0.154)	0.069 (0.028)**	0.161 (0.000)***	0.038 (0.000)***	0.059 (0.000)***	0.111 (0.002)***	0.146 (0.000)***
Infrastructure index[a]	−0.027 (0.579)	−0.031 (0.006)***	−0.058 (0.302)	0.022 (0.004)***	−0.013 (0.003)***	−0.003 (0.659)	0.080 (0.002)***	0.093 (0.000)***	−0.002 (0.788)	−0.001 (0.829)	0.105 (0.035)**	0.049 (0.000)***
Market incentives												
Dummy variable for market share less than 5 percent	−0.082 (0.001)***		−0.077 (0.000)***		−0.029 (0.055)*		−0.076 (0.064)*		−0.049 (0.000)***		−0.080 (0.077)*	
Price–cost margin (percent)		−0.075 (0.214)		−0.053 (0.048)**		0.045 (0.328)		0.192 (0.197)		0.003 (0.940)		−0.062 (0.360)
Dummy variable for pressure to innovate from competitors	−0.017 (0.210)	0.064 (0.002)***	0.003 (0.863)	0.022 (0.004)***	0.005 (0.640)	−0.010 (0.379)	0.003 (0.878)	0.046 (0.012)**	0.013 (0.535)	0.017 (0.358)	−0.024 (0.561)	0.019 (0.456)
Dummy variable for pressure to innovate from consumers	−0.004 (0.800)	−0.002 (0.916)	0.024 (0.172)	0.005 (0.687)	0.024 (0.087)*	0.021 (0.155)	0.040 (0.062)*	0.007 (0.712)	0.023 (0.076)*	0.029 (0.032)**	0.042 (0.333)	0.030 (0.119)

(Continued)

TABLE 4. Continued

Variable	EU-8 countries				Commonwealth of Independent States countries				Southeastern European countries and Turkey			
	ISO-certification dummy variable		Web-use dummy variable		ISO-certification dummy variable		Web-use dummy variable		ISO-certification dummy variable		Web-use dummy variable	
	2002 sample (1)	2005 sample (2)	2002 sample (3)	2005 sample (4)	2002 sample (5)	2005 sample (6)	2002 sample (7)	2005 sample (8)	2002 sample (9)	2005 sample (10)	2002 sample (11)	2005 sample (12)
Dummy variable for medium-size firms	0.125 (0.000)***	0.111 (0.003)***	0.105 (0.000)***	0.087 (0.000)***	0.051 (0.017)**	0.042 (0.000)***	0.213 (0.000)***	0.234 (0.000)***	0.042 (0.000)***	0.112 (0.000)***	0.127 (0.000)***	0.213 (0.000)***
Dummy variable for large firms	0.232 (0.000)***	0.225 (0.000)***	0.152 (0.000)***	0.079 (0.000)***	0.043 (0.136)	0.083 (0.000)***	0.310 (0.000)***	0.341 (0.000)***	0.047 (0.069)*	0.246 (0.000)***	0.227 (0.000)***	0.248 (0.000)***
Dummy variable for privatized firm	0.046 (0.094)*	−0.060 (0.027)**	−0.040 (0.206)	−0.008 (0.794)	−0.017 (0.346)	−0.013 (0.748)	0.007 (0.881)	−0.087 (0.268)	0.037 (0.450)	−0.011 (0.746)	0.016 (0.844)	0.061 (0.293)
Dummy variable for private firm (from start-up)	0.008 (0.727)	−0.125 (0.000)***	−0.004 (0.848)	−0.035 (0.001)***	0.011 (0.666)	−0.006 (0.874)	0.157 (0.000)***	0.026 (0.686)	−0.011 (0.721)	−0.027 (0.562)	0.026 (0.497)	0.014 (0.811)
Dummy variable for fully foreign-owned firms	0.046 (0.074)*	0.038 (0.351)	0.064 (0.135)	0.062 (0.063)*	0.123 (0.000)***	0.036 (0.182)	0.406 (0.000)***	0.152 (0.029)**	0.054 (0.161)	0.041 (0.395)	0.229 (0.003)***	0.089 (0.299)
Dummy variable for joint venture	0.019 (0.589)	0.047 (0.100)*	0.077 (0.078)*	0.065 (0.037)**	0.048 (0.116)	0.019 (0.355)	0.205 (0.000)***	0.076 (0.258)	0.046 (0.147)	0.017 (0.607)	0.103 (0.039)**	0.133 (0.079)*
Ownership share of largest shareholder	0.031 (0.355)	−0.066 (0.041)**	−0.069 (0.040)**	−0.104 (0.000)***	−0.046 (0.133)	−0.013 (0.536)	−0.106 (0.140)	−0.120 (0.012)**	0.017 (0.671)	−0.049 (0.082)*	−0.017 (0.811)	−0.147 (0.000)***
Controls												
Export share	0.054 (0.001)***	0.063 (0.001)***	0.145 (0.002)***	0.215 (0.000)***	0.063 (0.162)	0.050 (0.027)**	0.289 (0.000)***	0.447 (0.000)***	0.106 (0.040)**	0.064 (0.373)	0.248 (0.006)***	0.215 (0.001)***
Imported inputs share	0.009 (0.660)	0.036 (0.016)**	0.149 (0.000)***	0.083 (0.000)***	0.038 (0.028)**	0.043 (0.010)***	0.311 (0.000)***	0.341 (0.000)***	0.024 (0.467)	0.058 (0.000)***	0.217 (0.000)***	0.270 (0.000)***
GDP per capita	0.082 (0.000)***	0.030 (0.136)	0.168 (0.000)***	0.080 (0.066)*	0.046 (0.000)***	0.010 (0.130)	0.139 (0.001)***	0.153 (0.000)***	0.019 (0.317)	0.014 (0.324)	−0.092 (0.409)	0.022 (0.710)
Number of observations	1,702	2,490	1,721	2,465	2,391	3,428	2,404	3,441	1,486	2,023	1,497	2,008

*Significant at the 10 percent level; **Significant at the 5 percent level; ***Significant at the 1 percent level.

Note: Marginal effects at mean values from probit regressions are shown. Numbers in parentheses are *p*-values corresponding to robust standard errors clustered by country. All regressions include sector fixed effects. See appendix table A1 for variable definitions.

a. Higher values indicate better infrastructure.

Source: Authors' analysis based on 2002 and 2005 Enterprise Surveys; see text for details.

subgroups. Finally, estimates of complementary inputs do not seem to suffer from multicollinearity: the coefficients for the 2002, 2005, and panel samples are close to those estimated using equation (3) but excluding the market incentive proxies and firm controls (see table 3).

Market Incentives

Firms with smaller market shares in 2002 are significantly less likely to be ISO certified or to use the web (see table 3). Firms with lower price–cost margins in 2005 are also less likely to be ISO certified or to use the web, though the effects are weak.[17] The panel sample results indicate that firms reducing their price–cost margins between 2002 and 2005 are less likely to become ISO certified or to use the web—significantly so for web use (see table 3, regressions 3 and 6).

The pressure to innovate from competitors is only weakly positively correlated with technology adoption, with the exception of a significant correlation with web use in 2005. However, stronger pressure to innovate from consumers significantly increases the frequency of becoming ISO certified and using the web. The panel sample results show that firms facing increased pressure from competitors or increased pressure from consumers to innovate between 2002 and 2005 are more likely to become ISO certified and to use the web.

Because of the difficulty in measuring market competition, the robustness of the results are checked by replacing the dummy variables for the pressure to innovate from consumers and competitors with two competition measures proposed by Carlin, Schaffer, and Seabright (2004): elasticity of demand and number of domestic competitors in a firm's main product. Firms facing more elastic demand (whose customers would react to a price increase by buying the product from competitors) are less likely to be ISO certified or to use the web (table 5, regressions 1, 2, 5, and 6). There is no difference in ISO certification or web use between firms facing no domestic competitors in their main products and firms facing one to three competitors or more than four competitors (see table 5, regressions 3, 4, 7, and 8).

Larger firms are significantly more likely to adopt new technology. Within a sector in 2002, large firms (more than 250 workers) were about 25 percent more likely to use the web than were small firms (fewer than 50 workers; see table 3, regression 4). It could be that the economies of scale from which large firms benefit are associated with higher productivity and a higher return to technology adoption, allowing large firms to operate with a more efficient division of labor and creating better conditions for technological upgrade.

17. The 2002 survey also provides information on price–cost margins. The regressions including that measure instead of market share show that firms with lower price–cost margins in 2002 are less likely to be ISO certified or to use the web. Since the 2005 survey provides information only on price–cost margins, that variable is used in the panel regressions.

TABLE 5. Correlates of Technology Adoption—Alternative Competition Measures for ISO-Certification and Web-Use-Dependent Variables

	ISO-certification dummy variable				Web-use dummy variable			
Variables	2002 sample (1)	2005 sample (2)	2002 sample (3)	2005 sample (4)	2002 sample (5)	2005 sample (6)	2002 sample (7)	2005 sample (8)
Complementary inputs								
Share of skilled labor	0.068 (0.004)***	0.063 (0.000)***	0.064 (0.008)***	0.074 (0.000)***	0.372 (0.000)***	0.332 (0.000)***	0.371 (0.000)***	0.181 (0.000)***
Manager with college education dummy variable	0.037 (0.000)***		0.037 (0.000)***		0.227 (0.000)***		0.228 (0.000)***	
Manager age	0.001 (0.003)***		0.001 (0.006)***		−0.003 (0.098)*		−0.003 (0.092)*	
R&D intensity	0.159 (0.021)**	0.534 (0.000)***	0.161 (0.017)**	1.007 (0.000)***	0.426 (0.085)*	1.391 (0.020)**	0.428 (0.083)*	2.596 (0.000)***
Access to finance dummy variable	0.048 (0.000)***	0.040 (0.000)***	0.048 (0.000)***	0.039 (0.002)***	0.114 (0.000)***	0.145 (0.000)***	0.113 (0.000)***	0.091 (0.000)***
Infrastructure index[a]	−0.010 (0.066)*	−0.006 (0.146)	−0.009 (0.072)*	−0.009 (0.127)	0.062 (0.007)***	0.057 (0.000)***	0.062 (0.007)***	0.026 (0.000)***
Market incentives								
Price–cost margin (percent)		0.011 (0.701)		−0.028 (0.397)		−0.027 (0.668)		−0.054 (0.146)
Dummy variable for elasticity of demand: if prices increased by 10 percent customers would still buy from firm but slightly lower quantities	−0.008 (0.479)	−0.013 (0.193)			0.013 (0.410)	−0.002 (0.914)		

Dummy variable for elasticity of demand: if prices increased by 10 percent customers would still buy from firm but much lower quantities	−0.013 (0.286)	−0.009 (0.370)			−0.005 (0.826)	−0.073 (0.001)***		
Dummy variable for elasticity of demand: if prices increased by 10 percent customers would buy from competitors	−0.029 (0.029)**	−0.012 (0.216)			−0.026 (0.248)	−0.092 (0.000)***		
Dummy variable for firm facing 1–3 competitors in domestic market			0.028 (0.489)	0.001 (0.978)			0.124 (0.091)*	−0.007 (0.868)
Dummy variable for firm facing four or more competitors in domestic market			−0.011 (0.770)	−0.005 (0.911)			0.109 (0.174)	−0.005 (0.890)
Dummy for medium-size firms	0.065 (0.000)***	0.081 (0.000)***	0.065 (0.000)***	0.111 (0.000)***	0.173 (0.000)***	0.196 (0.000)***	0.174 (0.000)***	0.089 (0.000)***
Dummy variable for large firms	0.098 (0.000)***	0.153 (0.000)***	0.096 (0.000)***	0.194 (0.000)***	0.249 (0.000)***	0.230 (0.000)***	0.250 (0.000)***	0.101 (0.000)***
Dummy variable for privatized firm	0.027 (0.121)	−0.029 (0.156)	0.027 (0.122)	−0.036 (0.286)	−0.001 (0.984)	−0.041 (0.343)	−0.007 (0.838)	−0.018 (0.612)
Dummy variable for private firm (from start-up)	0.014 (0.396)	−0.043 (0.072)*	0.014 (0.347)	−0.054 (0.124)	0.100 (0.002)***	0.006 (0.860)	0.096 (0.005)***	0.016 (0.666)
Dummy variable for fully foreign owned	0.071 (0.000)***	0.038 (0.056)*	0.066 (0.000)***	0.047 (0.088)*	0.23 (0.000)***	0.134 (0.000)***	0.236 (0.000)***	0.076 (0.003)***

(Continued)

TABLE 5. Continued

Variables	ISO-certification dummy variable				Web-use dummy variable			
	2002 sample (1)	2005 sample (2)	2002 sample (3)	2005 sample (4)	2002 sample (5)	2005 sample (6)	2002 sample (7)	2005 sample (8)
Dummy variable for joint venture	0.048 (0.017)**	0.026 (0.084)*	0.047 (0.025)**	0.019 (0.444)	0.158 (0.000)***	0.101 (0.008)***	0.157 (0.000)***	0.036 (0.120)
Ownership share of largest shareholder	−0.005 (0.777)	−0.035 (0.033)**	−0.005 (0.793)	−0.061 (0.012)**	−0.078 (0.035)**	−0.133 (0.000)***	−0.075 (0.045)**	−0.063 (0.015)**
Dummy variable for market share less than 5 percent	−0.044 (0.001)***		−0.041 (0.002)***		−0.101 (0.000)***		−0.104 (0.000)***	
Controls								
Export share	0.072 (0.005)***	0.060 (0.002)***	0.072 (0.004)***	0.071 (0.024)**	0.264 (0.000)***	0.360 (0.000)***	0.263 (0.000)***	0.160 (0.000)***
Imported inputs share	0.025 (0.082)*	0.045 (0.000)***	0.024 (0.096)*	0.063 (0.000)***	0.283 (0.000)***	0.277 (0.000)***	0.284 (0.000)***	0.108 (0.000)***
GDP per capita	0.028 (0.002)***	0.030 (0.000)***	0.028 (0.003)***	0.045 (0.000)***	0.160 (0.000)***	0.147 (0.000)***	0.162 (0.000)***	0.076 (0.000)***
Number of observations	5,650	8,076	5,666	4,146	5,687	8,073	5,705	4,143

*Significant at the 10 percent level; **Significant at the 5 percent level; ***Significant at the 1 percent level.

Note: Marginal effects at mean values from probit regressions are shown. Numbers in parentheses are *p*-values corresponding to robust standard errors clustered by country. All regressions include sector fixed effects. See appendix table A1 for variable definitions.

a. Higher values indicate better infrastructure.

Source: Authors' analysis based on 2002 and 2005 Enterprise Surveys; see text for details.

Taken together, the findings on market shares, price–cost margins, competition, and size suggest that for firms in Europe and Central Asia, concentration is more conducive to technology adoption than is competition. This is consistent with the argument by Carlin, Schaffer, and Seabright (2004) that firms in transition economies face resource constraints that make rents important in financing technology adoption. The findings here also support Schumpeter's (1942) argument that innovation is costly and that firms facing financial market imperfections often finance technology adoption from retained earnings. These results are found despite the importance of access to external finance for technology adoption identified earlier. Finally, that market incentives for technology adoption originate more from consumers than from competitors is likely to reflect the pressure from large firms (possibly multinationals) on smaller suppliers to upgrade, accompanied by knowledge transfer and training.

For firms in Europe and Central Asia, private ownership generally has no significant effect on technology adoption, except for web use in 2002 (see table 3). Moreover, privatization does not seem to be a vehicle for technology adoption: privatized firms are no more likely than state-owned firms to be ISO certified or to use the web. One possible interpretation for this finding is that private and privatized firms have an advantage in technology adoption because of wider access to complementary inputs, but once those inputs are controlled for, the private and privatized dummy variables have no additional effect. And state-owned firms—especially large ones with access to finance—may initially lag behind private firms technologically, but they eventually catch up.

To probe the privatization result further, the privatized dummy variable is decomposed into a dummy variable for firms privatized less than three years ago and one for firms privatized more than three years ago. Firms privatized less than three years ago are significantly less likely to be ISO certified or to use the web, perhaps because newly privatized firms have other investment priorities, such as replacing old equipment (Goldberg and others 2008).

There is a strong positive association between foreign ownership—full or through a joint venture—and new technology adoption (see table 3). Foreign-owned firms are embedded in international networks, requiring frequent use of communications technology, and they compete in global markets, requiring the use of state of the art technology through internationally recognized technical standards. However, the technology adoption advantage of foreign-owned firms is strong only in the less advanced transition economies (see table 4).

The proxy for concentrated ownership is found to be generally negatively associated with technology adoption (see table 3). This finding is counterintuitive to the extent that firms with better corporate control—for which ownership concentration is a proxy—are expected to have more incentives to adopt new technology to maximize profits than are firms with more diffuse ownership and more limited exercise of control. It could be that the negative

association between technology adoption and firms with high ownership concentration is explained by how privatization programs operated in many Europe and Central Asia countries. While some shares were distributed to the general population, the government retained direct or indirect control over a large proportion of the shares.[18] More research is needed on the role of ownership control for technology adoption, exploiting panel data on firms undergoing important control changes.

Overall, the findings on ownership suggest that foreign-owned firms have significantly better technology adoption outcomes but that privatization to domestic owners brings no additional benefits. The market incentive estimates do not seem to suffer from multicollinearity since the coefficients for the 2002, 2005, and panel samples are close to those in table 3 in the variant estimate of equation (3) that excludes the complementary input proxies and firm controls.

Other Results

The coefficients on export shares and imported input shares reported in tables 3–5 show that plants engaged in these international activities are significantly more likely to be ISO certified or to use the web, mirroring the findings of Criscuolo, Haskel, and Slaughter (forthcoming) for innovation. Exporters and importers are likely to learn about new technologies through their interaction with foreign buyers and suppliers. These findings are strong for all country groups (see table 4).

The regressions reported in tables 3–5 also show a positive and significant effect of GDP per capita on ISO certification and web use. While GDP per capita accounts for an array of differences across countries, a particularly relevant interpretation here is that countries in Europe and Central Asia with higher GDP per capita have better protection of property rights, contract enforcement, and regulatory regimes, which are all potentially important correlates of technology adoption that are difficult to measure at the firm level.

IV. Conclusions

The international diffusion of technology presents an opportunity for developing countries lagging behind the world technological frontier to reduce their income gap with developed countries. But first there is a need to understand why, when facing similar technological alternatives, different firms in different countries make different technology adoption choices.

The findings presented here on correlations between investment climate factors and technology adoption for firms in 28 Europe and Central Asia countries have implications for policy reforms aimed at enabling faster and wider adoption of new technologies by private sector firms in Europe and Central Asia. Firms with better access to complementary inputs (skilled labor,

18. The authors thank an anonymous referee for suggesting this interpretation.

managerial skills R&D, finance, and to a lesser extent good infrastructure) are more likely to be ISO certified or to use the web.

The relationship between market incentives and ISO certification or web use is more nuanced. While pressures from consumers generate demand for firms to adopt new technology, pressures from competitors do not. This finding may be counterintuitive in a developed country setting but is consistent with previous studies finding that most firms in Europe and Central Asia face substantial resource constraints (particularly financial resources) so that only firms with rents are able to finance technology adoption. Accordingly, larger firms are significantly more likely to adopt new technology.

The findings also suggest that privatization is not necessarily a vehicle for technology adoption in Europe and Central Asia but that foreign ownership is strongly associated with better technology adoption outcomes, most likely through the access it provides to advanced foreign technology and management techniques.

The broad policy implication of the findings is that increasing technology adoption by firms operating under severe resource constraints requires investment climate reforms that increase the availability of complementary inputs to investments in technology adoption, while also improving market incentives.

The study identifies several robust correlates of technology adoption by examining two novel dimensions of technology in a developing country context. However, these technology adoption proxies have important limitations. A fruitful next step in data collection and research efforts would be to examine the determinants of the intensity of new technology adoption by firms.

Funding

Support from the governments of Norway, Sweden and the UK through the Multidonor Trust Fund for Trade and Development is gratefully acknowledged.

Appendix

Table A-1 Variables and their Definitions

Variable name	Definition
ISO-certification dummy variable	Dummy variable equal to 1 if the firm obtained a new quality accreditation (ISO 9000) in the three years prior to the survey
Web-use dummy variable	Dummy variable equal to 1 if the firm uses email and the internet regularly in its interactions with clients and suppliers

(Continued)

TABLE A-1 Continued

Variable name	Definition
Size categories	Small = fewer than 50 permanent workers; medium = 50–249 workers; large = 250 or more workers
Manager with college education or more dummy variable	Equal to 1 if general manager's highest level of education is a university degree or a postgraduate degree
Manager age	Age of the firm's general manager
Share of skilled labor	Percentage of the firm's current permanent full-time workers that are professionals (accountants, engineers, scientists)
R&D intensity	R&D expenditures (including wages and salaries of R&D personnel, materials, and R&D-related education and training costs) as percentage of total firm sales
Access to finance dummy variable	Equal to 1 if the firm has a bank loan or overdraft
Infrastructure index	First principal component derived from factor analysis of the negative of the average number of days with power outages or surges from the public grid in the firm's city and the negative of the average number of days with unavailable mainline telephone service in the firm's city in the year before the survey
Dummy variable for market share equal to less than 5 percent	Equal to 1 if the firm's percentage of total market sales is less than 5 percent (available only in the 2002 survey)
Price–cost margin (percent)	Percentage margin by which sales price of the firm's main product or service line in the domestic market exceeds its operating costs (materials inputs costs plus wages costs but not overhead and depreciation)
Dummy variable for pressure to innovate from competitors	Equal to 1 if the firm ranks pressure from domestic or foreign competitors as being fairly important or very important for the firm's decisions about developing new products or services and markets
Dummy variable for pressure to innovate from consumers	Equal to 1 if the firm ranks pressure from customers as being fairly important or very important for the firm's decisions about developing new products or services and markets
Elasticity of demand faced by firm	What would happen to demand if the firm were to raise the prices of its main product or services line 10 percent above the current level in the domestic market (after allowing for inflation), assuming that competitors maintained their current prices: customers would continue to buy from the firm in the same quantities as now, customers would continue to buy from the firm but at slightly lower quantities, customers would continue to buy from the firm but at much lower quantities, many customers would buy from competitors instead

(Continued)

TABLE A-1 Continued

Variable name	Definition
Number of competitors	Number of competitors for the firm's main product in the domestic market: none, one to three, or four or more
Dummy variable for privatized firm	Equal to 1 if the firm was established through privatization of a state-owned firm
Dummy variable for private firm (from start-up)	Equal to 1 if the firm was private from start-up (no state-owned predecessor)
Dummy variable for fully foreign owned	Equal to 1 if 100 percent of the firm's capital is owned by foreigners
Dummy variable for joint venture	Equal to 1 if any (but less than 100 percent) of the firm's capital is owned by foreigners
Ownership share of largest shareholder	Percentage of the firm's equity owned by the largest shareholder
Export share	Percentage of firm output exported directly or indirectly
Imported input share	Percentage of intermediate inputs used by the firm that are imported directly or indirectly
GDP per capita (log)	Values in constant 2000 dollars for 1995 (from World Bank various years)
EU-8 countries	Countries that joined the European Union in 2004: Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovak Republic, and Slovenia
CIS countries	Members of the Commonwealth of Independent States; all former Soviet republics except Estonia, Latvia, and Lithuania
Southeastern European countries and Turkey	Albania, Bosnia and Herzegovina, Bulgaria, Croatia, FYR Macedonia, Romania, Serbia and Montenegro,[a] and Turkey

a. Data are for the period before separation.

Source: Enterprise Surveys in 2002 or in 2005 unless otherwise stated.

REFERENCES

Aghion, Philippe, and Mark Schankerman. 2004. "On the Welfare Effects and Political Economy of Competition-Enhancing Policies." *Economic Journal* 114(498):800–24.

Aghion, Philippe, Christopher Harris, Peter Howitt, and John Vickers. 2001. "Competition, Imitation and Growth with Step-by-Step Innovation." *Review of Economic Studies* 68 (3):467–92.

Aghion, Philippe, Richard Blundell, Rachel Griffith, Peter Howitt, and Susanne Prantl. 2009. "The Effects of Entry on Incumbent Innovation and Productivity." *Review of Economics and Statistics* 91(1):20–32.

Alesina, Alberto, and Joseph Zeira. 2006. *Technology and Labor Regulations*. NBER Working Paper 12581. Cambridge, Mass.: National Bureau of Economic Research.

Almeida, Rita, and Ana M. Fernandes. 2008. "Openness and Technological Innovations in Developing Countries: Evidence from Firm-Level Surveys." *Journal of Development Studies* 44(5):701–27.

Arnold, Jens, Aaditya Mattoo, and Gaia Narciso. 2008. "Services Inputs and Firm Productivity in Sub-Saharan Africa: Evidence from Firm-Level Data." *Journal of African Economies* 17(4):578–99.

Arora, Arora, and Jai Asundi. 1999. *Quality Certification and the Economics of Contract Software Development: A Study of the Indian Software Service Companies*. NBER Working Paper 7260. Cambridge, Mass.: National Bureau of Economic Research.

Basant, Rakesh, Simon Commander, Rupert Harrison, and Naercio Menezes-Filho. 2006. *ICT Adoption and Productivity in Developing Countries: New Firm Level Evidence from Brazil and India*. IZA Discussion Paper 2294. Institute for the Study of Labor, Bonn, Germany.

Baumol, William. 1990. "Entrepreneurship: Productive, Unproductive, and Destructive." *Journal of Political Economy* 98(5):893–921.

Benner, Mary. 1999. "Process Management and Organizational Adaptation to Technological Change: Evidence from a Simulation Study." New York: Columbia University, Graduate School of Business.

Benner, Mary, and Francisco Veloso. 2008. "ISO 9000 Practices and Financial Performance: A Technology Coherence Perspective." *Journal of Operations Management* 26(3): 611–29.

Berdugo, Byniamin, Jacques Sadik, and Nathan Sussman. 2003. "Delays in Technology Adoption, Appropriate Human Capital, Natural Resources, and Growth." Jerusalem: Hebrew University, Department of Economics.

Blind, Knut, Andre Jungmittag, Paul Temple, Chris Spencer, Peter Swann, and Robert Witt. 2005. *The Empirical Economics of Standards*. DTI Economics Paper 12. London, U.K: U.K. Department of Trade and Industry.

Boucekkine, Raouf, Fernando Del Rio, and Omar Licandro. 2003. "Embodied Technical Change, Learning by Doing, and Productivity Slowdown." *Scandinavian Journal of Economics* 105(1): 87–97.

Bresnahan, Timothy, and Manuel Trajtenberg. 1995. "General Purpose Technologies 'Engines of Growth'?" *Journal of Econometrics* 65(1):83–108.

Bresnahan, Timothy, Erik Brynjolfsson, and Lorin Hitt. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics* 117(1):339–76.

Brown, J. D., John Earle, and Amos Telegdy. 2006. "The Productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia, and Ukraine." *Journal of Political Economy* 114(1):61–99.

Carlin, Wendy, Mark Schaffer, and Paul Seabright. 2004. "A Minimum of Rivalry: Evidence from Transition Economies on the Importance of Competition for Innovation and Growth. "Contributions article. *The B&E Journal of Economic Analysis and Policy* 3(1).

Caselli, Francesco, and Wilbur Coleman II. 2001. "Cross-Country Technology Diffusion: The Case of Computers." *American Economic Review* 91(2):328–35.

Cohen, Wesley, and Richard Levin. 1989. "Empirical Studies of Innovation and Market Structure." In *Handbook of Industrial Organization* Vol 2, ed. R. Schmalensee, and R. Willig. Amsterdam: North-Holland.

Cohen, Wesley, and Daniel Levinthal. 1989. "Innovation and Learning: The Two Faces of R&D." *Economic Journal* 99(397):569–96.

Comin, Diego, and Bart Hobijn. 2004. "Cross-Country Technology Adoption: Making the Theories Face the Facts." *Journal of Monetary Economics* 51(1):39–83.

———. 2008. *An Exploration of Technology Diffusion*. Harvard Business School Working Paper 08-093. Cambridge, Mass: Harvard Business School.

Corbett, Charles, Maria Montes-Sancho, and David Kirsch. 2005. "The Financial Impact of ISO 9000 Certification in the United States: An Empirical Analysis." *Management Science* 51(7):1046–59.

Correa, Paulo, Ana M. Fernandes, and Chris Uregian. 2008. *Technology Adoption and the Investment Climate: Firm-Level Evidence for Eastern Europe and Central Asia*. Policy Research Working Paper 4707. World Bank, Development Research Group, Trade Team, Washington, DC.

Criscuolo, Chiara, Jonathan Haskel, and Matthew Slaughter. Forthcoming. "Global Engagement and the Innovation Activities of Firms." *International Industrial Organization Journal*.

Djankov, Simeon, and Bernard Hoekman. 2000. "Foreign Investment and Productivity Growth in Czech Enterprises." *World Bank Economic Review* 14(1):49–64.

Eaton, Jonathan, and Samuel Kortum. 2001. "Trade in Capital Goods." *European Economic Review* 45(7):1195–235.

Goldberg, Itzhak, Lee Branstetter, John Goddard, and Smita Kuriakose. 2008. *Globalization and Technology Absorption in Europe and Central Asia: The Role of Trade, FDI, and Cross-border Knowledge Flows*. Working Paper 150. Central Asia Region, World Bank, Washington, DC.

Guler, Isin, Mauro Guillen, and John Macpherson. 2002. "Global Competition, Institutions, and the Diffusion of Organizational Practices: The International Spread of ISO 9000 Quality Certificates." *Administrative Science Quarterly* 47(2):207–32.

Hall, Robert, and Charles Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114(1):83–116.

Hall, Bronwyn, and Beethika Khan. 2003. *Adoption of New Technology*. NBER Working Paper 9730. Cambridge, Mass.: National Bureau of Economic Research.

Humphrey, John, and Hubert Schmitz. 2000. Governance and Upgrading: Linking Industrial Cluster and Global Value Chain Research. IDS Working Paper 120. Falmer, Brighton, U.K.: Institute of Development Studies, University of Sussex.

Indjikian, Rouben, and Donald Siegel. 2005. "The Impact of Investment in IT on Economic Performance: Implications for Developing Countries." *World Development* 33(5):681–700.

ISO (International Organization for Standardization). 1998. *Selection and Use of ISO 9000*. Geneva: International Organization for Standardization.

Keller, Wolfgang. 2004. "International Technology Diffusion." *Journal of Economic Literature* 42(3):752–82.

Lipsey, Richard, and Kenneth Carlaw. 2004. "Total Factor Productivity and the Measurement of Technological Change." *Canadian Journal of Economics* 37(4):1118–50.

Maddala, G.S. (1987). "Limited Dependent Variable Models Using Panel Data." *Journal of Human Resources* 22(3):307–38.

Navaretti, Giorgio, Isidro Soloaga, and Wendy Takacs. 2001. "Vintage Technologies and Skill Constraints: Evidence from U.S. Exports of New and Used Machines." *World Bank Economic Review* 14(1):91–109.

Parente, Stephen, and Edward Prescott. 1994. "Barriers to Technology Adoption and Development." *Journal of Political Economy* 102(2):298–321.

———. 2005. "A Unified Theory of the Evolution of International Income Levels." In P. Aghion, and S. Durlauf eds., *Handbook of Economic Growth*. Vol. 1. Amsterdam: North-Holland.

Prescott, Edward. 1998. "Lawrence R. Klein Lecture 1997: Needed: A Theory of Total Factor Productivity." *International Economic Review* 39(3):525–51.

Safavian, Mehnaz, and Siddarth Sharma. 2007. "When Do Creditor Rights Work?" *Journal of Comparative Economics* 35(3):484–508.

Schumpeter, Joseph. 1942. *Capitalism, Socialism, and Democracy*. New York: Harper.

World Bank. Various years. *World Development Indicators*. Washington, DC: World Bank.

Zeira, Joseph. 1998. "Workers, Machines, and Economic Growth." *Quarterly Journal of Economics* 113(4):1091–117.

The Effect of Refugee Inflows on Host Communities: Evidence from Tanzania

Jennifer Alix-Garcia and David Saah

Despite the large and growing number of humanitarian emergencies, there is little economic research on the impact of refugees and internally displaced people on the communities that receive them. This analysis of the impact of the refugee inflows from Burundi and Rwanda in 1993 and 1994 on host populations in western Tanzania shows large increases in the prices of nonaid food items and more modest price effects for aid-related food items. Food aid is shown to mitigate these effects, though its impact is smaller than that of the increases in the refugee population. Examination of household assets suggests positive wealth effects of refugee camps on nearby rural households and negative wealth effects on households in urban areas. JEL codes: O12, O13, F22, R23, R12

Each week seems to bring news of more humanitarian crises. In 1980, there were 5.7 million refugees and internally displaced persons worldwide; at the beginning of 2005, there were 9 million. The burden of refugees and internally displaced persons falls on the poorest countries. Almost 3 million refugees were in Sub-Saharan Africa in 2005, home to 23 percent of the world's internally displaced persons (UNHCR 2004). This article turns the spotlight on the millions uncounted in statistics: the hosts.

In addition to hosting more refugee camps than any other country, Tanzania has been the destination of two very large population flows: Burundian refugees in 1993 and Rwandan refugees in 1994. The unexpected nature and size of these population movements generate a natural experiment that allows their effects on prices and household wealth in the western Tanzanian regions hosting the refugees to be examined.

Jennifer Alix-Garcia (corresponding author) is an assistant professor in the Department of Agricultural and Applied Economics at the University of Wisconsin, Madison; her email address is alixgarcia@wisc.edu. David Saah is a principal at Spatial Informatics Group, LLC; his email address is dsaah@sig-gis.com. This article has benefited tremendously from the hard work and patience of three anonymous referees and the journal editor. The authors are very grateful for comments from participants in the development workshop in the Department of Agricultural and Resource Economics at the University of California, Berkeley. In addition, suggestions from Dick Barrett, Doug Dalenberg, Alain de Janvry, Ethan Ligon, Elisabeth Sadoulet, and Laura Schechter were indispensable. The authors also thank Ted Miguel and George Akerlof for their advice in the early stages of this project.

THE WORLD BANK ECONOMIC REVIEW, VOL. 24, NO. 1, pp. 148–170 doi:10.1093/wber/lhp014
Advance Access Publication November 12, 2009

Despite the prevalence of humanitarian crises, little research has been conducted on their impact on local economies. Williamson and Hatton's (2004) literature review reveals considerable work on the determinants of population displacement—usually civil wars (Collier and Hoeffler 1998; Hatton and Williamson 2002)—as well as on how policies in Europe and the United States have affected the direction of human flight from conflicts in developing economies. There is little mention of the effects of these crises on the refugees and internally displaced persons directly or on the communities that receive them. The exception to this trend is a recent paper by Baez (2008) that shows a substantial negative impact on health outcomes of residents living close to refugee camps in the Tanzanian region that hosted most of the refugees from the Burundian and Rwandan crises.

This article focuses on one facet of this complicated issue: the impact of refugee camps on prices in nearby markets. This interaction has received attention from development practitioners and other social scientists, with contradictory conclusions. Borton, Brusset, and Hallam (1996) and Whitaker (1999) discuss large price spikes and suggest that local populations suffer from these events. On the other hand, Landau (2002) compares a market near the refugee camps in Tanzania with one in the central part of country and finds little evidence of any impact on prices.

Two strains of literature inform the design of this study. The first is the incentive effects of food aid, and the second is the impact of immigrant flows on prices in recipient countries. Barrett (2001) thoroughly reviews the effects of food aid on local prices. The empirical results have been mixed, with much of the research focused on food for work programs rather than on free food, which is the situation in humanitarian emergencies.

Early research has shown that effectively targeted food aid, as in India, has increased consumption by the targeted population, with little or no effect on domestic food prices (Maxwell and Singer 1979; Singer, Wood, and Jennings 1987; Ruttan 1993; Insenman and Singer 1997). More recent work by Dercon and Krishnan (2004) finds food aid targeting in Ethiopia quite imperfect. Abdulai, Barrett, and Hoddinott (2005), also using data from Ethiopian households, present no evidence that households lower food production in the presence of food aid and find suggestive evidence that they increase it.

The source of the aid—foreign or domestic—is key in determining its effects on the market. Theory suggests that foreign-supplied aid is likely to depress prices, while increases in domestically produced aid could increase prices. Simulated effects of different food policy approaches in India by Binswanger and Quizon (1988) confirm this intuition. In sum, research finds that whether the supply side shock of food aid in developing economies results in local price effects depends on how the aid is targeted and where it comes from.

Food aid is only one possible impact of refugee flows. Another—the population increase—can change local prices through increased demand for goods and increased supply of inexpensive labor. Immigrant movements and their

subsequent effects on host countries are topics of considerable research, usually related to labor market outcomes (Borjas 1987; Card 1990; Cortes 2005), which are not analyzed directly here. The research has shown that immigrant inflows can have either positive or negative impacts on local populations, depending on the context. Recent work by Lach (2007) finds that the movement of refugees from the former Soviet Union to Israel in the 1990s resulted in falling prices and attributes this effect to greater price sensitivity among immigrants, who have not established the store and brand allegiance of the native population and are likely to search more intensively for lower prices.

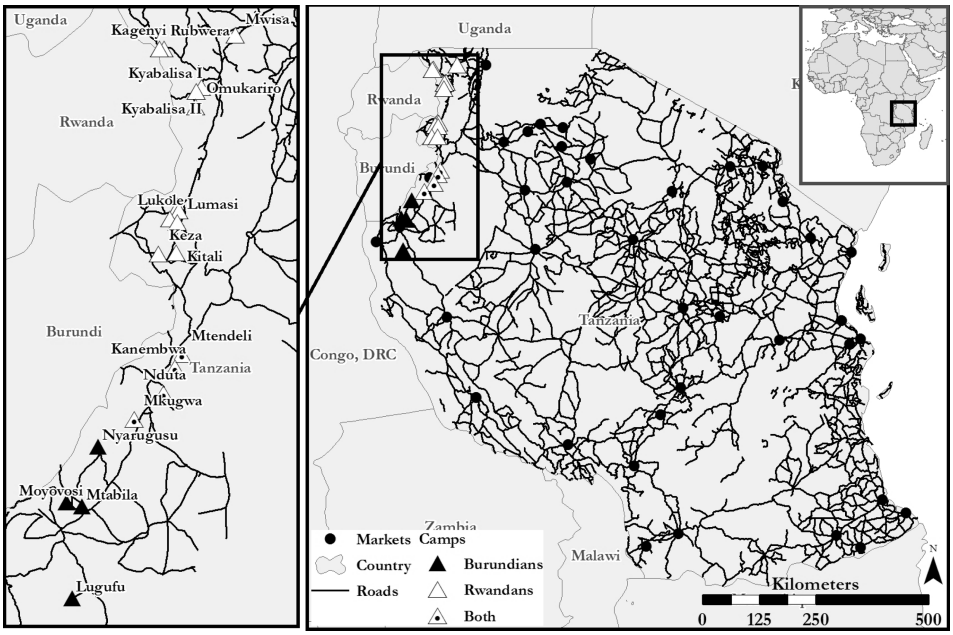
This article uses variations in refugee population and food aid over time to examine the impact of proximity to refugee camps and aid on prices of Tanzanian agricultural goods. The estimates show increases in the prices of most goods in markets closer to refugee camps as a result of the refugee inflows, though the effect is much larger for Rwandan refugees than for Burundian refugees. The differences in the effects are explained by variations in the diets of the two groups as well as by the nature and magnitude of the two crises. Food aid in the form of maize and legumes depresses the prices of these crops but does not appear to affect nonaid crops. This result is particularly strong in the short run. Suggestive evidence that rural residents living near the refugee camps may have benefited from selling home-produced agricultural products is also discussed. On the other hand, because urban households are more likely to be purchasing agricultural goods for consumption, they experience negative wealth effects.

The article is organized as follows. Section I provides background on the Tanzanian situation in 1993–94. Section II presents a framework for understanding the effects of the refugee inflow on prices. Section III describes the data. Section IV details the identification strategy and gives results from the analysis of agricultural prices. Section V discusses potential welfare effects. And section VI discusses the implications of the findings and suggests directions for future research.

I. Tanzania in 1993–94

With a GDP per capita in 2007 of about $350, Tanzania is wealthier than Burundi ($101) and Rwanda ($271) (World Bank 2009). Tanzania has a long history of accepting migrants from across Africa, and its population is known to be friendly and accepting of foreigners. Though refugee flight to Tanzania, largely from Burundi, has occurred since the 1970s, this study focuses on the largest of the recent arrivals, those in 1993 and 1994. Kagera and Kigoma, the Tanzanian regions hosting most of the refugees, have high rates of poverty, with 35–40 percent of residents living below the poverty line. In 2000, out of the 20 mainland regions, Kagera ranked 11th and Kigoma ranked 7th in poverty (Mkenda and others 2004). Both regions heavily depend on

MAP 1. Camps and Markets in Tanzania, 1993–94



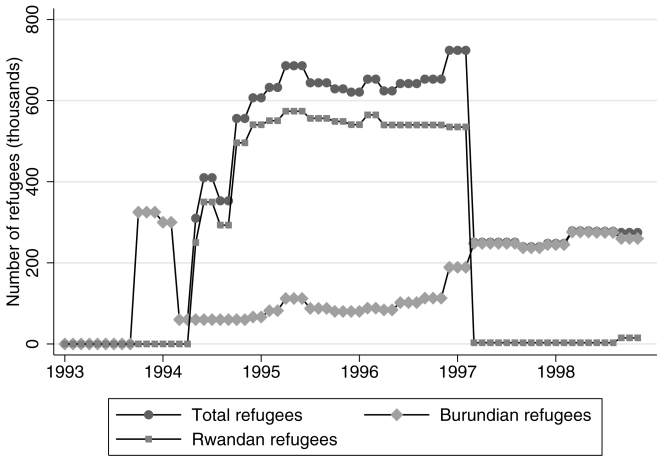
Source: UNHCR Environmental Database 1996; Reliefweb 2005; USAID 2004.

agricultural income, with about 80 percent of their regional GDP from agriculture (Tanzania Ministry of Agriculture and Food Security 2006).

The timeline of events is as follows. On October 21, 1993, the first elected president of Burundi, Melchior Ndadaye, a Hutu, was assassinated in a bloody coup led by Tutsi soldiers. Some 700,000 Hutus fled the country, many to western Tanzania. The initial influx of Burundians into Kagera and Kigoma, reported at 245,000, rose to more than 300,000 within a month (SCN 1993–98). Until 1993, refugees had largely been assimilated into Tanzanian villages. The 1993 and 1994 crises led to the construction of large refugee camps, a network of food distribution facilities, the sudden presence of multiple international agencies, and the beginning of the Tanzanian government's policy of separating the refugees from the local population (Landau 2002). Map 1 shows the road networks, location of refugee camps, and major markets in Tanzania. According to Jaspers (1994), the location of the camps was dictated by the Tanzanian government in cooperation with the International Committee of the Red Cross and the United Nations World Food Programme. Camp locations were likely chosen to facilitate the provision of food aid but were also determined by the ability of the refugees to reach them; all were within 40 kilometers of the border (Whitaker 1999).

On April 6, 1994, just as many of the Burundian refugees were preparing to return home, the presidents of Burundi and Rwanda died in an airplane crash,

FIGURE 1. Refugees in Western Tanzania, 1993–99



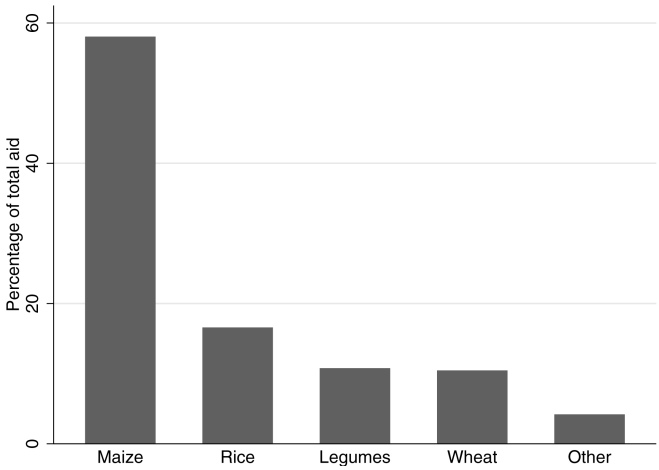
Note: Data are reported every three to four months.
Source: SCN issues 1–25.

sparking genocide in Rwanda, with 500,000–1 million people slaughtered. In a 24-hour period on April 28, nearly a quarter of a million Rwandans flooded into northwestern Tanzania's Ngara district in Kagera (UNHCR 2000). The UN Refugee Agency has called the Rwandan influx the largest and fastest movement of refugees in modern history. In 1998, the UN Office for the Coordination of Humanitarian Affairs estimated the local population of the refugee-affected regions at about 1.3 million (UNOCHA 1998). According to the United Nations Children's Fund, refugees totaled as much as 39 percent of the population in Ngara district in Kagera and Kibondo district in Kigoma (UNICEF 2000).

UN estimates of the total refugee load in western Tanzania are produced every three to four months (figure 1), based on estimates by the managers of the refugee camps used for calculating food requirements. Although data quality is uncertain, the population counts are usually revised downward, suggesting that estimates exceed the actual number of refugees.

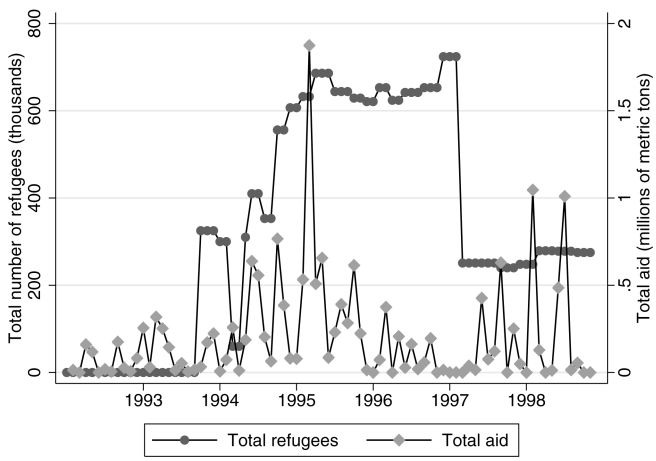
Anecdotal evidence suggests that the Rwandan refugees were relatively wealthy, especially compared with the Burundian refugees, having brought cash and other assets used to trade (Borton, Brusset, and Hallam 1996). The main source of food in the camps was maize or maize flour, which generally constituted 83 percent of the cereal distributed to refugees, with sorghum or rice making up the other 17 percent. The World Food Programme supplied 75 percent of the aid, the International Committee of the Red Cross, 22 percent. Most of the food was imported through Mombasa or Dar es Salaam, but 23,000 tons of maize and legumes given to refugees were produced in Tanzania (of the country's 270,000 tons total) from April through the end of

FIGURE 2. Food Aid to Tanzania, by Type, 1992–98



Source: SCN issues 1–25; WFP/INTERFAIS 2008.

FIGURE 3. Number of Refugees and Total Food Aid to Tanzania, 1992–98



Note: Data are reported every three to four months.
Source: SCN issues 1–25; WFP/INTERFAIS 2008.

1994. Tanzania's 1994–95 National Agricultural Census shows total production of maize in Kagera and Kigoma at 68,400 tons (Tanzania Ministry of Agriculture and Food Security 2006).

Most aid sent to Tanzania came as maize or maize products and legumes (beans, lentils, and peas), although rice and wheat deliveries were not insubstantial (figure 2). Food aid was arriving in Tanzania before the Burundian and Rwandan crises (figure 3). Its destination is unclear, but the lack of refugee camps in western Tanzania suggests that the most likely destination was food

for work programs or camps on the southern border. The aid clearly increased in response to the Rwandan crisis and stayed high even after many refugees had returned home. Other forces were likely determining the flow of aid, and potential sources of endogeneity to local prices are investigated below.

According to Whitaker (1999), refugees typically sold about 75 percent of their food rations. Jaspers (1994) found that maize was a particularly popular food for Rwandan refugees to sell in order to purchase plantains, cassava, and sweet potatoes. They "generally preferred their own staples of cassava, cooking bananas, and sweet potatoes, which were also produced by local farmers. Refugees therefore used a variety of strategies to gain access to these foods, including trading, purchasing, and stealing. With this huge increase in the market for local crops, the prices of foods such as cassava and especially cooking bananas [plantains] skyrocketed (p. 3)."

II. Theoretical Framework

This section presents a simple framework for analyzing the local price effects of population displacement and of the subsequent flows of aid. A large inflow of refugees and aid implies both supply side and demand side effects on the market. On the supply side, food aid increases the amount of aid-related goods available, which may put downward pressure on prices if the food aid is imported and upward pressure if it is provided locally. The percentage of food aid purchased in Tanzania is small relative to the total aid provided (10 percent of World Food Programme provisions), but the amount is substantial given local production capacity.

The population increase results in increased demand for all goods. These pressures can substantially change the prices of tradable goods only when trade with areas outside affected regions is limited. The model below assumes that transaction costs prohibit immediate price adjustment through the inflow of goods from other regions or countries—not unreasonable in western Tanzania, with its limited range of substitute goods and high transaction costs that may result in much more localized price effects. Kahkonen and Leathers (1999) indicate that such costs in Tanzania are due to "movement restrictions, infrastructural impediments, limited access to credit, lack of storage capacity, and contract enforcement problems" (p. page 57). They cite a 1990 World Bank study that concluded that only 24 percent of Tanzania's paved roads were in good condition, with the remaining poor or fair. Only 16 percent of maize farmers live within 5 kilometers of a market where they can sell their product, and prices of maize and cotton (the two crops considered in the study) vary considerably by city. Some 30–40 percent of maize produced in Tanzania is lost due to a lack of on-farm storage every year, and only one farmer of the 139 interviewed by Kahkonen and Leathers reported having obtained credit.

The inflow of refugees may also depress wages, which may result in falling prices where labor is an important agricultural input. Although the government

of Tanzania has restricted refugees' ability to seek employment, there is substantial anecdotal evidence that they do so nonetheless (Jacobsen 2005).

The model is formalized as follows. Suppose that households have concave utility functions dependent on their consumption of aid goods, x_a, and nonaid goods, x_n. They also have simplified budget constraints, with the sum of spending on consumption of all goods equal to income: $p_a\, x_a + p_n\, x_n = m$.[1] Refugee population income, m_r, is assumed to be different from that of the host population, m_h: $m_r \neq m_h$.

The maximization of $u(x_a, x_n)$, subject to $p_a\, x_{ak} + p_n\, x_{nk} = m_k$ with $k = r$, h, yields household demand functions of

$$x_{ik}^d(p_i, p_j, m_k) \tag{1}$$

where $i, j \in a, n$ and $j \neq i$.

The concavity of the utility function produces demand functions that are decreasing in own price and increasing in the price of the other good and in income: $\partial x_j^d/\partial p_j < 0$, $\partial x_j^d/\partial p_i \geq 0$, $\partial x_j^d/\partial m_k > 0$. With the total number of refugee households denoted by R and the total number of host households by H, market demand for goods yields

$$Hx_{ih}^d(p_i, p_j, m_h) + Rx_{ir}^d(p_i, p_j, m_r). \tag{2}$$

Both refugee and host populations may participate as laborers in the production of all goods, whose main input is labor. A concave production function will yield supply functions of the form

$$x_i^s(p_i, w; \alpha_i) \tag{3}$$

where α_i is a parameter indicating the productivity of labor, p_i is the price of good $i \in a$, n, and w is the wage. The supply function increases in own price and decreases in the wage: $\partial x_i^s/\partial p_i > 0$, $\partial x_i^s/\partial w < 0$. Aggregate supply is the sum of supply for P, individual producers. It is assumed that, in the short run, P does not depend directly on R—which is not unreasonable given the restrictions on refugee land ownership in Tanzania.

For aid-supplied goods, there is an additional component to the supply function—the aid itself, which depends on the number of refugees. Imported aid is denoted by $a_f(R)$, a function that increases in R. The quantity of aid purchased locally, $a_d(R)$, affects the market on the demand side. Equilibrium in the

1. The simple framework here assumes that household production can be separated from consumption. There is clear evidence in countries like Tanzania that such decisions are, in fact, nonseparable. However, the complications of nonseparability yield little payoff in this situation, where the predictions regarding price changes are equivalent in either case and where there is no information on household production choices that would allow causes or effects of nonseparability to be identified at the household level.

aid-related markets is then determined by

$$(4) \qquad Hx^d_{ih}(p_a, p_n, m_h) + Hx^d_{ir}(p_a, p_n, m_r) + a_d(R) = Px^s_a(p_a, w; \alpha_a) + a_f(R).$$

Rearranging and totally differentiating this expression by price and the number of refugees yields

$$(5) \qquad \frac{\partial p_a}{\partial R} = \frac{\frac{\partial a_f}{\partial R} - \frac{\partial a_d}{\partial R} - x^d_{ar}}{H\frac{\partial x^d_{ah}}{\partial p_a} + R\frac{\partial x^d_{ar}}{\partial p_a} - P\frac{\partial x^s_a}{\partial p_a}}.$$

The denominator of equation (5) is always negative, while the numerator's sign is ambiguous. Domestically produced aid and foreign-supplied aid move prices in opposite directions, and the additional demand from refugees puts upward pressure on prices in the same way that domestically produced food aid might. If the foreign aid effect ($\partial a_f/\partial R$) exceeds the other two effects, the price will decrease. The price might also increase—if, for example, the aid were not sufficient to satisfy refugee demand and refugees began to purchase these products on local markets. A positive price effect could be exacerbated by local procurement of aid. In nonaid markets, the effect of the refugees occurs without the buffer of aid and is therefore unambiguously positive.

In this framework, labor is demanded in the production of both aid-related and nonaid goods. Assuming substitutability between refugee and host labor, and a concave production function, the increase in the labor supply caused by the refugees is easily shown to depress the wage. This model ignores other factors likely to be associated with the refugees, most notably changes in the local economy as a result of the aid infrastructure itself: increases in demand, especially for luxury goods purchased by aid workers; changes in transportation costs as a result of the presence of more cars and trucks on highways delivering aid; and subsequent pressures for road improvements that might affect transactions costs. In addition, individual household responses to price changes are not explored in terms of their decision to market their home production.

However, this model generates reasonable predictions for the available data—agricultural prices in markets throughout Tanzania. None of the goods included in the analysis is likely to be strongly preferred by aid workers. There are three predictions of interest. First, the price of aid-related goods will be affected by both the increase in population caused by the refugee inflow and the availability of food aid. The net effect will be the sum of these two, and these effects move in opposite directions when food aid is foreign supplied and in the same direction when food aid is purchased locally. Second, nonaid goods are likely to experience price increases because there is no mitigating effect from foreign-supplied aid. Third, wages may decrease, although this effect is not formally tested here (section V presents suggestive evidence regarding household wealth).

TABLE 1. Percentage of Calories from Maize, Legumes, Bananas and Plantains, and Milk, by Country, 2004

Country	Maize	Legumes	Bananas and plantains	Milk
Tanzania	34.9	3.4	1.9	2.1
Burundi	11.8	22.6	13.2	0.8
Rwanda	10.8	11.9	30.4	1.8

Source: FAO 2005.

III. DATA

The data in this study come from various sources. The Famine Early Warning System set up by the U.S. Agency of International Development provides monthly prices from 44 urban markets in Tanzania, beginning in 1985 and ending in 1998 (USAID 2004; see map 1 for these markets). The data report a single price for each market every month. Until 1991, the prices of major commodities were controlled by the government, so data from before January 1992 have been discarded. Although the data contain prices for numerous crops, many of the series have large gaps. The six markets with the most serious data omissions—those without complete data during the major refugee influx (late 1993 and early 1994)—have been dropped in favor of the more complete time series for maize, legumes, bananas, plantains, and milk.[2] The first four are staple crops that are both grown and eaten in the regions of interest, though maize is preferred more by Tanzanians than by the refugee groups. Maize and legumes, unlike bananas, are also part of the standard food aid package. Milk is included because it is often supplied to refugee camps for supplemental feeding programs targeting mothers and small children.

In Kagera, the two most common agricultural systems are banana/coffee/horticulture and maize/legume. In Kigoma banana/coffee/horticulture is also common, as are cotton/maize and sweet potato/sorghum/groundnuts (Tanzania Ministry of Agriculture and Food Security 2006). These regions are not major suppliers of aid goods. In the 1994–95 growing season, Kagera and Kigoma produced only 2.4 percent of the total maize in the country, but significantly more of the legumes (15.9 percent) and bananas (28.2 percent). In Kagera, bananas and plantains make up 60 percent of agricultural production, while in Kigoma, they constitute 26 percent. Milk production in Tanzania is generally on a very small scale, and production in Kagera and Kigoma is of a traditional, low-input variety (Muriuki and Thorpe 2006).

The share of legumes, bananas, and plantains in the diet is much higher for the typical refugee than for the typical Tanzanian (table 1). However, considerable regional variation in food preferences is likely, and given the region's high production of plantains and bananas, local consumption of these products may be higher than the national average.

2. These series are complete for only 38 markets.

The dates and quantities of food aid deliveries to Tanzania were provided by the World Food Programme, which provided 75 percent of the aid to Tanzania (WFP/INTERFAIS 2008). To combine these data with the other data, total aid for each month was summed, giving monthly deliveries in hundreds of metric tons. Deliveries of maize and legumes were separated to analyze these commodities' impact on their respective market prices. The data do not contain the deliveries supplied by the International Committee of the Red Cross, the other main source of aid.

Monthly normalized difference vegetation index readings for each market were also taken from the Famine Early Warning System. The index measures vegetation vigor using satellite images and is a good proxy for agricultural productivity. The readings were extracted from geographical data with a pixel size of 8 square kilometers and were merged with price data using the reading from the pixel in which the markets are located.

Household data come from two Demographic and Health Surveys conducted in 1991–92 and 1996 (Macro International 2004). These surveys contain information on basic household characteristics, including assets and type of employment. The data have the disadvantage of not containing observations on income or expenditures, but they do cover more than 12,000 households over the two years, including more than 1,000 households in the refugee-affected regions. The two surveys were combined to make a pooled cross section that was used to analyze changes in welfare indicators across the period of interest (see section V).

IV. Impact of Refugees and Aid on Prices

The estimation of the effects of refugees and aid on prices exploits the variation in the number of refugees in Tanzania across time as well as the fact that they were present in specific parts of the country. The natural log of prices, $\log(p_{i,t})$, in market i at time t depends on the number of refugees from Burundi and Rwanda as a percentage of the population of Kagera and Kigoma, where they were located at time t. In other words, for the Burundian refugees, $B_t = \left(\frac{\text{Burundian refugees}}{\text{Kigoma population} + \text{Burundian refugees}}\right) \times 100$. A similar expression[3] is included for the Rwandan refugees (R_t).

The refugee impact is given by the interaction of these terms with a variable that is the inverse of the distance to the closest refugee camp (D_i) from market i.[4]

3. The number of refugees was divided by the population of the province to which the majority of each group went. Regressions using the refugees divided by the total population in the two provinces together were also run, with similar qualitative results and patterns of significance.

4. The inverse of distance is used rather than distance itself to reflect the isolated nature of markets in Tanzania. A quadratic form of the distance measures was also tested, with similar results, as was absolute distance. The model's R^2 was highest using the inverse of distance, the results presented here. A simpler difference-in-difference estimator, where treatment equaled 1 if the market was in Kagera or Kigoma and 0 otherwise, showed a post-treatment interaction that differed for the Burundian and Rwandan crises. The results were very consistent with those using the distance specification shown here.

D_i represents proximity to camps, and the interaction terms $R_t\ D_i$ and $B_t\ D_i$ allow the effect of the refugee inflows to vary according to the distance from the camps. The distances used to generate this variable were measured using the information in map 1 by calculating the length of the road network from the markets to the camps using geographic information system software.

Food aid, $\mathbf{F}_t$, is a vector that includes total aid, and, where the price of interest is an aid product, the amount of that product. Aid's impact is given by the interaction of $\mathbf{F}_t$ with camp proximity D_i.

It is impossible to rule out all other events that could cause spurious results, but the normalized difference vegetation index, which varies over time and space, controls for one of the main competing sources of agricultural price shocks: weather. The index measures vegetation "greenness" and thus picks up variation in both temperature and rainfall. It is included for the current period for every market along with a previous growing season average of the index to control for stocks of the crop from the previous year. These weather controls are indicated by $\mathbf{X}_{i,t}$. Market-level fixed effects $\mathbf{M}_i$ capture time-invariant market characteristics. Year-month fixed effects ψ_t are also included to control for shocks common to all markets in a given time period.

The full estimated equation is:

$$\log(p_{i,t}) = \alpha + \delta_1 B_t D_i + \delta_2 R_t D_i + \delta_3 \mathbf{F}_t D_i + \mathbf{X}_{i,t}\Gamma + \sum_{i=1}^{38} \mathbf{M}_i + \sum_{t=1}^{84} \psi_t + u_{i,t}. \tag{6}$$

For this estimation to give reasonable estimates of the effect of refugees on the local markets, the location of the refugee camps must not be affected by the markets themselves. As mentioned, the camps are likely to be have been located to facilitate the provision of food aid. But the location of the camps is random in a larger sense: the refugees entered Tanzania, rather than other border countries, as a result of directional pushes of internal conflict within their own countries, which is unlikely to have been affected by markets in Tanzania.

An additional concern with this estimation is the potential endogeneity of food aid to local food prices. Significant flows of food aid followed the refugees (see figure 3). If the influx of refugees led to increases in local food prices and donors reacted to these prices, adding aid to the regression could yield biased estimates. Conversations with World Food Programme representatives suggest that the magnitude of aid shipments is determined by the population censuses conducted in the refugee camps rather than by local prices. To test whether prices independently determine aid quantities, a simple ordinary least squares regression of total aid in a given month was run on average maize prices in Tanzania, total number of refugees in that month, and the price of maize in the United States. Despite the small number of observations ($n = 84$), the coefficients are significant for refugees (0.42, standard deviation of 0.14) and the

U.S. price of maize (−3,848, standard deviation of 1,687). The coefficient on local maize price is negative and insignificant (−14, standard deviation of 65). While an imperfect test, it clearly gives no evidence that local maize prices are important in determining the amount of aid sent to Tanzania.

Equation (6) is estimated using fixed effects ordinary least squares with robust clustered standard errors, which allows for arbitrary correlation across time within clusters and for correlation across markets. Kezdi (2004) has shown this adjustment to produce consistent standard errors in the presence of serial correlation, even in finite samples. Standard errors were also calculated using Driscoll and Kraay's (1998) method, which produced smaller estimates. The more conservative, clustered robust results are presented here.

The main results of the regression on equation (6) are as the model predicts (table 2).[5] In general, the arrival of more refugees leads to price increases that are greater the closer a market is to the refugee camps. An increase in the number of Burundian refugees affects prices of maize, maize flour, legumes, and plantains, while an increase in the number of Rwandan refugees affects all prices except for maize and maize flour. In the context of the model, these price increases are explained by the increase in demand for these products by the incoming refugees. The effect of food aid is limited to aid-related goods. Increases in the amount of maize generates an increase in both total aid, which raises price, and in the amount of maize, which lowers the price of maize. The marginal effect is 0.032–0.040, a net change of zero in the maize price. For legumes, the net effect is negative and large. This implies that changing the composition of aid can strongly affect food prices. No effect of aid on nonaid goods is observed, somewhat surprising given the anecdotal accounts of refugees trading food aid for nonaid products.

One concern with this analysis is that some of the estimated price effects could result from the longer term influence of the refugee populations on migration of Tanzanians wishing to take advantage of jobs provided by nongovernmental organizations or from interactions on the local labor market rather than from the direct impact of the refugees on demand. In addition, the price increases could also reflect income effects if a substantial number of Tanzanians experience higher incomes as a result of the refugee presence.

To examine the immediate impact of the establishment of the camps, regressions were run with data through December 1994 only, when the largest part of the Rwandan inflow had just finished (table 3). These shorter run estimates show similar impacts to the full sample results, except for maize flour, where increases in Burundian or Rwandan refugees result in price decreases. In contrast to the full sample estimates, there is no short-run effect of the Rwandan refugees on banana prices. Where the effects are significant in both the long and the short run, they are generally of similar magnitudes—except

5. A more parsimonious specification that excludes the weather controls produced nearly identical results and is available from the authors on request.

TABLE 2. Impact of Refugee Camps and Aid on Agricultural Prices, 1992–1998

Independent variable	Maize	Maize flour	Legumes	Plantains	Bananas	Milk
(Burundian refugees/ Kigoma population) × camp proximity	0.088 (0.053)*	0.128 (0.057)**	0.150 (0.034)***	0.398 (0.299)*	0.018 (0.180)	0.019 (0.073)
(Rwandan refugees/ Kigoma population) × camp proximity	−0.038 (0.065)	−0.050 (0.045)	0.081 (0.028)***	0.625 (0.182)***	0.239 (0.043)***	0.156 (0.031)***
Total aid × camp proximity	0.040 (0.005)***	0.015 (0.011)	−0.009 (0.007)	−0.037 (0.023)	−0.007 (0.006)	0.001 (0.003)
Maize aid × camp proximity	−0.032 (0.007)***	−0.013 (0.012)				
Legume aid × camp proximity			−0.078 (0.035)**			
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year/month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	2,335	2,183	2,417	1,849	2,255	2,285
R-squared	0.641	0.684	0.843	0.489	0.573	0.800

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: Numbers in parentheses are standard errors. Standard errors are robust and clustered. These are results from ordinary least squares regressions. The dependent variable is the natural log of the food price.

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

TABLE 3. Impact of Refugee Camps and Aid on Agricultural Prices, 1992–1994

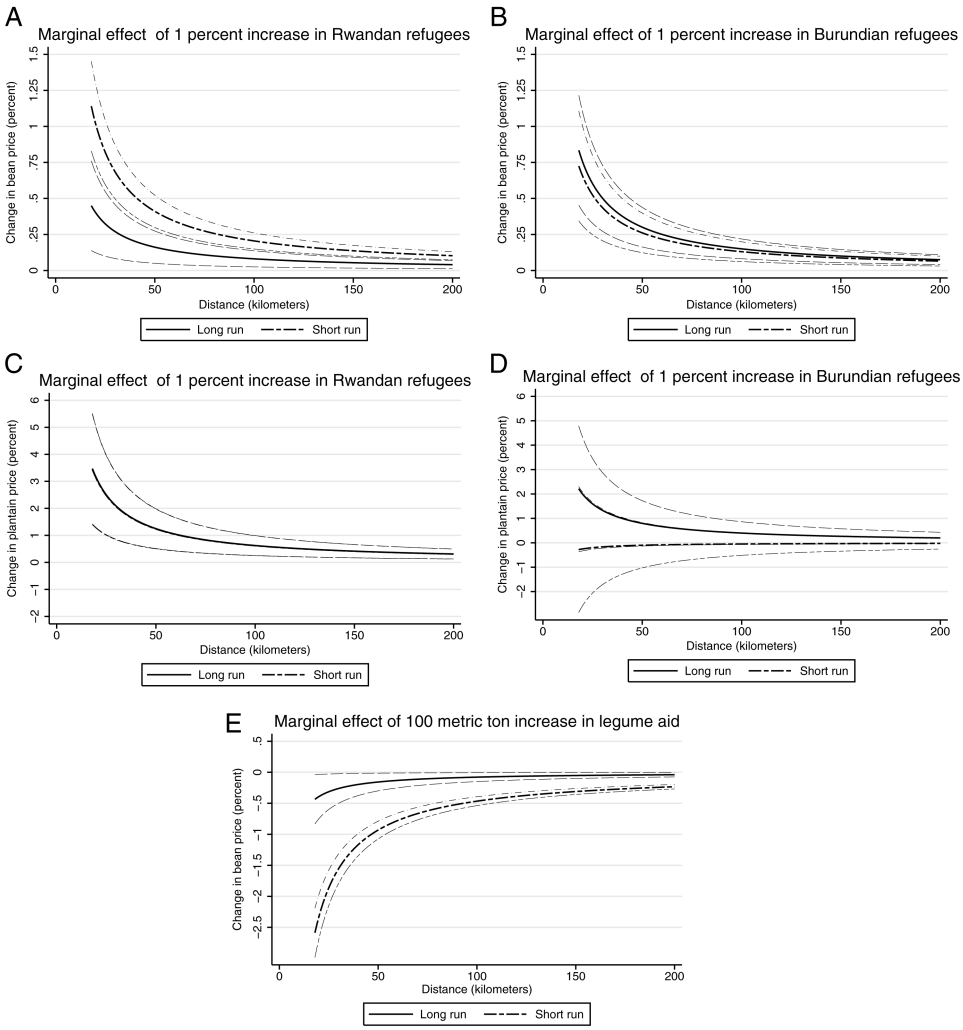
Independent variable	Maize	Maize flour	Legumes	Plantains	Bananas	Milk
(Burundian refugees/Kigoma population) × camp proximity	0.026 (0.057)	−0.183 (0.026)***	0.130 (0.049)***	−0.050 (0.066)	0.104 (0.092)	−0.043 (0.045)
(Rwandan refugees/Kigoma population) × camp proximity	−0.294 (0.161)*	−0.479 (0.103)***	0.205 (0.066)***	0.620 (0.198)***	0.076 (0.051)	0.149 (0.038)***
Total aid × camp proximity	0.089 (0.014)***	0.043 (0.027)	0.005 (0.009)	−0.044 (0.046)	0.017 (0.013)	−0.003 (0.004)
Maize aid × camp proximity	−0.087 (0.013)***	−0.044 (0.032)				
Legume aid × camp proximity			−0.466 (0.111)***			
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year/month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,060	969	1,088	953	1,020	1,004
R-squared	0.424	0.456	0.773	0.333	0.310	0.549

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: Numbers in parentheses are standard errors. Standard errors are robust and clustered. These are results from ordinary least squares regressions. The dependent variable is the natural log of the food price.

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

FIGURE 4. Impact of Refugees and Aid on Prices, by Distance from Refugee Camps



Source: Authors' calculations based on SCN issues 1–25 and WFP INTERFAIS 2008.

for legumes, for which the short-run effect of refugees is much larger than that for the full sample estimates. The estimates of the effect of specific aid—maize and legumes—are considerably larger in the truncated sample.

Because the impact of the refugee inflow is given by the interaction between two continuous variables, it is instructive to graph the marginal effect of increases in the refugee population and food aid according to distance from camps (figure 4). This highlights the change in price effects as a result of the difference in refugee inflows and aid according to distance of the market from the camps.

The impact of both refugees and aid decreases quickly with distance from camps. The largest impact of growth in the number of Rwandan refugees in Kagera is on plantains, where a 1 percentage point increase leads to a nearly 4 percent increase in price in the closest market, which quickly diminishes as distance increases. The impact of the refugees on legumes, an aid good, is smaller than it is on plaintains, a nonaid and more perishable good. Rwandan refugees have a larger impact on both types of goods than Burundian refugees do. There are three possible explanations for this. First, the Burundian refugee group was not nearly as large as the Rwandan group, and most did not stay as long. Second, the Rwandan refugees arrived with considerably more income, enhancing their ability to trade on local markets rather than relying exclusively on food aid. Third, the two groups had different food preferences (see table 1).

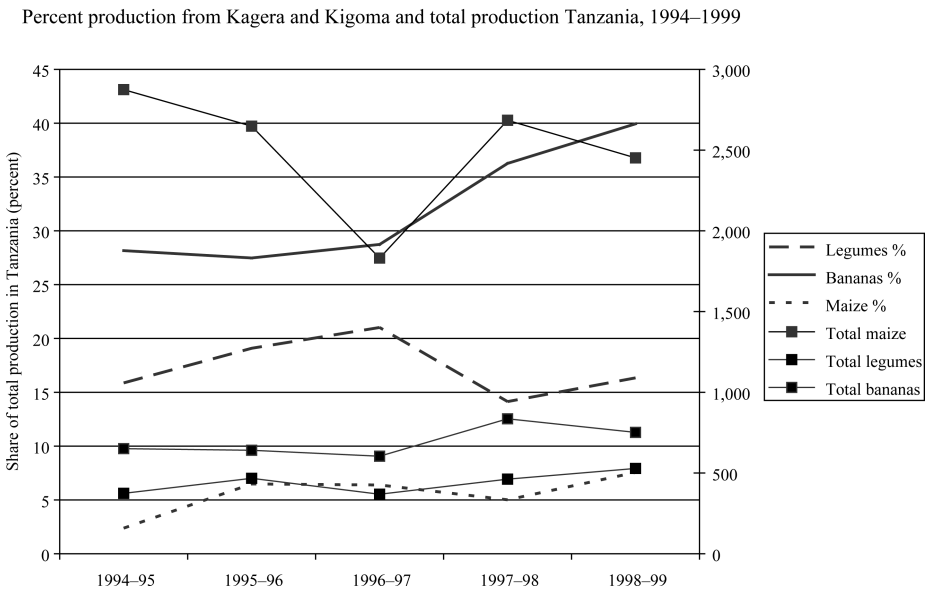
In sum, increases in the number of refugees generally pushed prices upward, with especially large effects for nonaid goods. These effects were generally similar in the short and long run, with the effect on legumes somewhat greater in the immediate aftermath of the crises. Food aid generated smaller and short-lived negative effects on prices, with no effect for the main aid product, maize. These results do not indicate how local Tanzanians may have fared as a result of these changes in their local economy. Section V discusses the potential welfare impacts of the price changes.

V. Potential Effects on Household Welfare

The increase in the refugee population in Tanzania resulted in large increases in prices in some markets—legumes, bananas, plantains, and milk—while the aid inflows put downward pressure on prices in legume markets. Assuming that the demand for staples is inelastic, price declines in these markets must result in lower revenues for producers (and increases must result in higher revenues). Therefore, the shift in demand caused by the refugee population is likely to benefit producers. The effects are the opposite for consumers in these markets—net consumers of legumes, plantains, bananas, and milk will suffer a decrease in surplus, while those purchasing maize may enjoy positive effects from the lower prices, at least temporarily. In both Kagera and Kigoma, the most common agricultural systems include both plantains and legumes, so if the loss in revenue from the food aid effect is smaller than the gain from the refugee effect, the net effect for producers is positive.

It is possible, however, that production from other regions could have been brought to the refugee-affected parts of the country to take advantage of the high prices. No data are available from before the Burundian crisis, but there are interesting trends in production in Kagera and Kigoma compared with total production in Tanzania (figure 5). In the post-refugee period, the share of Kagera and Kigoma's banana and plantain production in total production increased, while the total level of production nationwide remained relatively flat. The increase happened two years after the initial refugee inflows; bananas

FIGURE 5. Production of Maize, Legumes, and Bananas, 1994–1999



Source: Tanzania Ministry of Agriculture and Food Security 2006.

typically produce fruit 10–15 months from planting. Legume production in the western regions increased in the years just after the refugee inflows and then decreased. Part of this decrease can be explained by a decrease in productivity per hectare planted, from 1,487 Tanzania shillings in 1996–97 to 800 in 1997–98.

Although it is difficult to draw firm conclusions given the scarcity of the production data, the data provide some insight into producer decisions. First, the western regions are not the major producers of legumes in the country; the bulk of legume production comes from the more central regions, Iringa and Rukwa. But the western regions are important banana and plantain producers, more so after the refugee inflow. These simple production statistics suggest a producer reaction to the higher prices offered for these goods as a result of the crises in the region. Assessing the welfare effects on local residents requires knowing which individuals are net buyers or sellers in the markets and who receives wages or employs labor. These data are not available. However, in the 1992 Demographic and Health Survey, 77 percent of the men in rural areas of Kagera and Kigoma listed their primary occupation as farmer, whereas only 23 percent of men in urban areas did.

To investigate whether there are differential effects on net buyers and net sellers, data from the Demographic and Health Surveys as well as the fact that most rural households are agricultural producers were exploited. Using a pooled cross section constructed from the 1991 and 1996 surveys, a fixed effects estimator with effects at the regional level was applied to a regression of

the presence of wealth indicators on the effect of an interaction between proximity to the refugee camps and a dummy variable equal to 1 for observations in the 1996 survey.[6] The same proximity metric applied in section IV is applied here: the inverse of distance to the nearest refugee camp. Because the exact coordinates of each sampled cluster are not available, the estimated center point of the ward in which each sample cluster was proxied.

The results of the regressions show positive and significant effects on the presence of radios and bicycles in rural households closer to the refugee camps in the 1996 sample than in the 1991 sample (table 4). The point estimate for cement floors is positive but not significant. For urban households, proximity to the refugee camps has a negative and significant effect on the change in bike and cement floor ownership. The interpretation is that while urban households farther from camps were installing more cement floors over this period, those closer to camps were not installing any. Using a much smaller sample, the impact on households that identify farming as their primary source of income compared with nonfarm sources was analyzed.[7] The results are qualitatively similar: nonfarm households experience a negative impact of proximity to refugee camp, while farm households show increased presence of wealth indicators the closer they are to refugee camps.

These results, while consistent with the observed price changes, are merely suggestive. The Demographic and Health Survey samples are chosen to allow comparisons across regions, not necessarily on the fine scale that the distance analysis demands of the data.[8] So bias could be introduced into the estimates through the different samples in different years. In addition, household location within a ward cannot be precisely measured, and the number of households in close proximity to the refugee camps is small.[9]

These results concord with the story suggested above. Rural, farming households are likely to be net sellers of agricultural goods and thus to benefit from higher prices in key markets. The agricultural production statistics confirm that banana production, which experienced positive price shocks, increased in the post-refugee period. The increase in legume production, a small but important part of regional output, also could have benefited rural producers. Urban, nonfarming households, by contrast, lose from the higher prices, because they are

6. Controls included number of household members, number of women and children, gender and age of the household head, and the highest grade of schooling attained by the household head.

7. The smaller sample size is due to the structure of the survey; only a subsample of households was asked to respond to detailed questions regarding individual occupations.

8. An estimate that replaces the proximity variable with a dummy variable equal to 1 when a household is in Kagera or Kigoma yielded similar impact results.

9. It was impossible to match every cluster with a ward, but the final sample includes 8,687 rural households whose wards are known: 1,170 are in wards within 200 kilometers of the closest refugee camp, with 522 within 200 kilometers of the nearest camp in the 1996 sample. The closest rural ward in this sample is 2 kilometers from the nearest refugee camp, and the closest urban ward 20 kilometers.

TABLE 4. Effect of Proximity on Household Wealth Indicators

	Dependent variable		
Independent variable	Radio	Bicycle	Cement floor
Rural sample			
1996	0.091 (0.013)***	0.130 (0.015)***	−0.001 (0.007)
1996 × camp proximity	0.511 (0.195)***	0.459 (0.069)***	0.040 (0.065)
Observations	8,634	8,636	8,687
R^2	0.073	0.095	0.036
Urban sample			
1996	0.074 (0.031)**	0.071 (0.018)***	0.098 (0.035)***
1996 × camp proximity	0.482 (20.205)	−50.673 (10.208)***	−60.056 (20.633)**
Observations	2,040	2,037	2,053
R^2	0.085	0.081	0.061
Farm sample			
1996	0.095 (0.027)***	0.157 (0.023)***	0.001 (0.012)
1996 × camp proximity	−0.523 (0.481)	0.660 (0.264)**	0.092 (0.108)
Observations	1,520	1,523	1,531
R^2	0.057	0.039	0.028
Nonfarm sample			
1996	0.145 (0.038)***	0.124 (0.035)***	−0.025 (0.030)
1996 × camp proximity	−10.392 (0.424)***	−0.584 (0.296)**	0.107 (0.162)
Number of observations	605	605	608
R^2	0.081	0.075	0.051

*Significant at the 10 percent level; **significant at the 5 percent level; ***significant at the 1 percent level.

Note: Numbers in parentheses are standard errors. Standard errors are robust and clustered at the district level. These are partial results from fixed effects ordinary least squares regressions with the effect at the district level. Other included variables are proximity to camps, the number of household members, number of women and children, gender and age of the household head, and the highest grade of schooling attained by the household head.

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

net buyers of food. Urban households may also be affected by changes in the urban housing and labor markets as a result of the population influx.

To the extent that the impacts are relatively isolated—they are much larger the closer a ward is to the refugee camps—it is possible that they balance out regional disparities that leave western Tanzania poorer than much of the country. In addition, the crises may also disproportionately benefit rural residents and potentially redress long-standing rural–urban inequalities. While the price effects may positively affect some households in refugee-hosting regions, there are still many households——any that purchase items whose price has increased—vulnerable to the negative welfare effects of the humanitarian crises.

VI. CONCLUSION

Refugee situations are not likely to disappear, and understanding the impacts of refugee camps on poor host populations is imperative. This article presents

evidence that the refugee inflows into western Tanzania from 1993 to 1998 resulted in increases in the prices of agricultural goods that are consumed and produced by local populations in Tanzania. Prices in the same markets showed less impact from refugee crises from Burundi than from those from Rwanda, perhaps because of differences in the diets of these groups or because of the relatively smaller and slower nature of the first of the two crises. Food aid is shown to have a depressive effect on legume prices, but not maize prices. The magnitude of the aid effects is considerably smaller than that of the refugee effects. Household data produce suggestive evidence of increased incidence of wealth indicators in rural areas and decreases in urban areas. This is consistent with a scenario where producer households benefit from higher prices for agricultural goods and then invest that money in durable goods.

Clearly the analysis is imperfect. The price results depend on the impact of the camps on a very limited number of markets. In addition, the number of wealth indicators available for analysis was small. Despite these limitations, food aid does the job that it was intended to do: it offsets, at least partially, the impacts of increased demand created by refugee populations. Evidence is also presented that the demand side effects of refugee populations are substantial and affect markets in a way that may benefit local producers and hurt local consumers. This suggests that policy-makers should be concerned with net buyers of agricultural goods in refugee-hosting regions because they are likely to be adversely affected by the price shocks resulting from refugee demand. Although both host governments and aid agencies are often stretched to their budgetary limits, investment in mitigating negative impacts on host villages is warranted—one humanitarian crisis need not cascade into another.

This article gives insight into the effect of humanitarian emergencies on food prices. It is not able to shed light on the effects of these catastrophes on health, environmental, or labor market outcomes and has not touched on the economy internal to the camps themselves. These and other important questions—such as price volatility and coping strategies, including support from the government or from neighbors—are left for future analysis. Further research is essential for informing the policies of international agencies whose missions include supporting refugees and for the many countries that find themselves hosting refugees from other countries or large populations displaced within their own borders.

References

Abdulai, A., C. Barrett, and J. Hoddinott. 2005. "Does Food Aid Really Have Disincentive Effects? New Evidence from Sub-Saharan Africa." *World Development* 33(10):1689–1704.

Baez, J. 2008. '*Civil Wars Beyond Their Borders: The Human and Health Consequences of Hosting Refugees*.' IZA Discussion Paper 3468. Institute for the Study of Labor, Bonn, Germany.

Barrett, C. 2001. "Food Security and Food Assistance Programs." In Bruce Gardner, and Gordon C. Rausser eds., *Handbook of Agricultural Economics*. Vol. 2B. Amsterdam: Elsevier.

Binswanger, H.P., and J.B. Quizon. 1988. *Distributional Consequences of Alternative Food Policies in India*. Baltimore, MD.: Johns Hopkins University Press.

Borjas, G. 1987. "Migrants, Minorities, and Labor Market Competition." *Industrial and Labor Relations Review* 40(3):382–92.

Borton, J., E. Brusset, and A. Hallam. 1996. "The International Response to Conflict and Genocide: Lessons from the Rwanda Experience." Study 3: Humanitarian and Aid Effects. Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda, Copenhagen. (www.reliefweb.int/library/nordic/book3/pb022.html).

Card, D. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43(2):245–57.

Collier, P., and A. Hoeffler. 1998. "On the Economic Causes of Civil War." *Oxford Economic Papers* 50(4):563–73.

Cortes, P. 2005. '*The Effect of Low-skilled Immigration on US Prices: Evidence from CPI Data*.' Working Paper. Cambridge, Mass: Massachusetts Institute of Technology.

Dercon, S., and P. Krishnan. 2004. "Food Aid and Informal Insurance." Policy Research Working Paper 3217. World Bank, Washington, D.C.

Driscoll, J., and A. Kraay. 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80(4):549–60.

FAO (Food and Agriculture Organization of the United Nations). 2005. "Food and Agricultural Organization Country Profiles: Burundi, Rwanda, and Tanzania." (www.fao.org/countryprofiles/).

Hatton, T.J., and J.G. Williamson. 2002. *What Fundamentals Drive World Migration?* NBER Working Paper 9159. Cambridge, Mass.: National Bureau of Economic Research.

Insenman, P.J., and H.W. Singer. 1977. "Food Aid: Disincentive Effects and Their Policy Implications." *Economic Development and Cultural Change* 25(2):205–37.

Jacobsen, K. 2005. *The Economic Life of Refugees*. Bloomfield, Conn.: Kumarian Press.

Jaspers, S. 1994. '*The Rwandan Refugee Crisis in Tanzania: Initial Successes and Failures in Food Assistance*.' Relief and Rehabilitation Network Paper 6. Overseas Development Institute, London.

Kahkonen, S., and H. Leathers. 1999. '*Transaction Costs Analysis of Maize and Cotton Marketing in Zambia and Tanzania*.' Technical Paper 105. SD Publication Series. U.S. Agency of International Development, Bureau for Africa, Office of Sustainable Development, Washington, D.C.

Kezdi, G. 2004. "Robust Standard Error Estimation in Fixed-effects Models." *Hungarian Statistical Review* 9:95–116.

Lach, S. 2007. "Immigration and Prices." *Journal of Political Economy* 115(4):548–87.

Landau, L. 2002. "The Humanitarian Hangover." Ph.D. Dissertation. University of California, Berkeley, Department of Political Science.

Macro International. 2004. "Demographic and Health Surveys for Tanzania, 1991–92 and 1996." Macro International, Calverton, Md.

Maxwell, S.J., and H.W. Singer. 1979. "Food Aid to Developing Countries: A Survey." *World Development* 7(3):225–47.

Mkenda, A.F., E.G. Luvanda, L. Rutasitara, and A. Naho. 2004. "Poverty in Tanzania: Comparisons across Administrative Regions." Interim Report for the Government of Tanzania. Poverty and Economic Policy Network, Dakar.

Muriuki, H., and W. Thorpe. 2006. "Smallholder Dairy Production and Marketing in Eastern and Southern Africa: Regional Synthesis." International Livestock Research Institute, Nairobi.

Reliefweb. 2005. "Map: Kigoma Region Refugees, 1997". (www.reliefweb.org).

Ruttan, V. 1993. *Why Food Aid?* Baltimore, MD.: Johns Hopkins University Press.

SCN (UN Standing Committee on Nutrition). "Various issues". *Refugee Nutrition Information System*. Geneva: UN Standing Committee on Nutrition.

Singer, H.W., J. Wood, and T. Jennings. 1987. *Food Aid: The Challenge and the Opportunity*. Oxford, U.K.: Oxford University Press.

Tanzania Ministry of Agriculture and Food Security. 2006. "Basic Data: Agriculture Sector." Government of Tanzania, Dar Es Salam.

UNHCR (UN Refugee Agency). 2000. "Ch. 10: The Rwandan Genocide and its aftermath." In *The State of the World's Refugees 2000: Fifty Years of Humanitarian Action*. Geneva: UN Refugee Agency.

———. 2004. *The Global Report 2004*. Geneva: UN Refugee Agency.

UNHCR Environmental Database. 1996. "Rwandan Refugee Camps, Kagera, Tanzania". (http://geodata.grid.unep.ch/).

UNICEF (United Nations Children's Fund). 2000. "UNICEF Humanitarian Action Donor Update 13 October 2000: Tanzania." (www.unicef.org/emerg/Tanzania13Oct.pdf).

UNOCHA (United Nations Office for the Coordination of Humanitarian Affairs). 1998. "United Nations Consolidated Inter-Agency Appeal for the United Republic of Tanzania." United Nations, New York. (www.reliefweb.int/rw/rwb.nsf/AllDocsByUNID/fdbc2f914cc0a157c12566d400570773).

USAID (U.S. Agency for International Development). 2004. "United States Agency for International Development Famine Early Warning System." (www.fews.net).

WFP/INTERFAIS (United Nations World Food Programme/International Food Aid Information System). 2008. "Food Aid Deliveries to Tanzania, 1992–1998." (www.wfp.org/fais/).

Whitaker, B.E. 1999. "Changing Opportunities: Refugees and Host Communities in Western Tanzania." New Issues in Refugee Research Working Paper 11. Nairobi: UN Refugee Agency, Centre for Documentation and Research.

Williamson, J.G., and T.J. Hatton. 2004. *Refugees, Asylum Seekers, and Policy in Europe*. NBER Working Paper 10680. Cambridge, Mass.: National Bureau of Economic Research.

World Bank. 2009. *World Development Indicators 2009*. World Bank, Washington, D.C.

THE WORLD BANK
1818 H Street, NW
Washington, DC 20433, USA
World Wide Web: http://www.worldbank.org/
E-mail: wber@worldbank.org

ISBN 978-0-19-960192-3
9 780199 591923