

THE WORLD BANK ECONOMIC REVIEW

Volume 20 • 2006 • Number 3

The Primacy of Institutions Reconsidered: Direct Income Effects
of Malaria Prevalence

Kai Carstensen and Erich Gundlach

When Is External Debt Sustainable?

Aart Kraay and Vikram Nehru

Will African Agriculture Survive Climate Change?

*Pradeep Kurukulasuriya, Robert Mendelsohn, Rashid Hassan,
James Benhin, Temesgen Deressa, Mbaye Diop, Helmy Mohamed Eid,
K. Yerfi Fosu, Ghaidys Gbetibouo, Suman Jain, Ali Mahamadou,
Remeth Mano, Jane Kabubo-Mariara, Samia El-Marsafawy,
Ernest Mohia, Samiha Ouda, Mathien Ouedraogo, Isidor Sene,
David Maddison, S. Niggol Seo, and Ariel Dinar*

Microenterprise Dynamics in Developing Countries: How Similar are
They to Those in the Industrialized World? Evidence from Mexico

Pablo Fajnzylber, William Maloney, and Gabriel Montes Rojas

The "Glass of Milk" Subsidy Program and Malnutrition in Peru

David Stifel and Harold Alderman

How Endowments, Accumulations, and Choice Determine the
Geography of Agricultural Productivity in Ecuador

Donald F. Larson and Mauricio León

www.wber.oxfordjournals.org



THE WORLD BANK ECONOMIC REVIEW

EDITOR

Jaime de Melo, *University of Geneva*

EDITORIAL BOARD

Kaushik Basu, *Cornell University, USA*

Jean-Marie Baland, *University of Namur, Belgium*

Chong-En Bai, *Tsinghua University, China*

Timothy Besley, *London School of Economics, UK*

François Bourguignon, *World Bank*

Kenneth Chomitz, *World Bank*

Maureen Cropper, *University of Maryland, USA*

Jishnu Das, *World Bank*

Klaus Deininger, *World Bank*

Asli Demirgüç-Kunt, *World Bank*

Stefan Dercon, *Oxford University, UK*

Ishac Diwan, *World Bank*

Augustin Kwasi Fosu, *UN Economic Commission for Africa (ECA), Ethiopia*

Alan Harold Gelb, *World Bank*

Mark Gersovitz, *Johns Hopkins University, USA*

Paul Gertler, *World Bank*

Indermit Gill, *World Bank*

Jan W i e m Gunning, *Free University, The Netherlands*

Jeffrey Hammer, *World Bank*

Graciela Kaminsky, *George Washington University, USA*

Ravi Kanbur, *Cornell University, USA*

Peter Lanjouw, *World Bank*

Thierry Magnac, *Université de Toulouse I, France*

Juan-Pablo Nicolini, *Universidad Torcuato di Tella, Argentina*

Boris Pleskovic, *World Bank*

Martin Ravallion, *World Bank*

Ritva Reinikka, *World Bank*

Elisabeth Sadoulet, *University of California, Berkeley, USA*

Joseph Stiglitz, *Columbia University, USA*

L. Alan Winters, *World Bank*

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 20 • 2006 • Number 3

- The Primacy of Institutions Reconsidered: Direct Income Effects of Malaria Prevalence **309**
Kai Carstensen and Erich Gundlach
- When Is External Debt Sustainable? **341**
Aart Kraay and Vikram Nehru
- Will African Agriculture Survive Climate Change? **367**
Pradeep Kurukulasuriya, Robert Mendelsohn, Rashid Hassan, James Benbin, Temesgen Deressa, Mbaye Diop, Helmy Mobamed Eid, K. Yerfi Fosu, Glwady's Gbetibouo, Suman Jain, Ali Mabamadou, Renneb Mano, Jane Kabubo-Mariara, Samia El-Marsafawy, Ernest Molua, Samiha Ouda, Matbieu Ouedraogo, Isidor Sine, David Maddison, S. Niggol Seo, and Ariel Dinar
- Microenterprise Dynamics in Developing Countries: How Similar are They to Those in the Industrialized World? Evidence from Mexico **389**
Pablo Fajnzylber, William Maloney, and Gabriel Montes Rojas
- The "Glass of Milk" Subsidy Program and Malnutrition in Peru **421**
David Stifel and Harold Alderman
- How Endowments, Accumulations, and Choice Determine the Geography of Agricultural Productivity in Ecuador **449**
Donald F. Larson and Mauricio León

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

Annual Subscription Rate (Volume 20, 3 Issues, 2006): *Academic libraries*—Print edition and site-wide online access: US\$137/£87/€131, Print edition only US\$131/£83/€125, Site-wide online access only: US\$123/£78/€117; *Corporate*—Print edition and site-wide online access: US\$205/£130/€195, Print edition only: US\$196/£124/€186, Site-wide online access only: US\$185/£117/€176; *Personal*—Print edition and individual online access: US\$49/£34/€51. Please note: £ Sterling rates apply in the UK, € in Europe, and US\$ elsewhere. There may be other subscription rates available; for a complete listing, please visit www.wber.oxfordjournals.org/subscriptions. *Readers with mailing addresses in non-OECD countries and in socialist economies in transition are eligible to receive complimentary subscriptions on request by writing to the UK address below.*

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

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

CONTACT INFORMATION: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnl.cust.serv@oxfordjournals.org. Tel: +44(0)1865 353907. Fax: +44(0)1865 353485. *In the Americas, please contact:* Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA. E-mail: jnlorders@oxfordjournals.org. Tel: (800) 852-7323 (toll-free in USA/Canada) or (919) 677-0977. Fax: (919) 677-1714. *In Japan, please contact:* Journals Customer Service Department, Oxford University Press, 1-1-17-SF, Mukogaoka, Bunkyo-1a, Tokyo, 113-0023, Japan. E-mail: okudaoup@po.ijinet.or.jp. Tel: (03) 3813 1461. Fax: (03) 3818 1522.

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

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

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

ADVERTISING: Inquiries about advertising should be sent to Helen Pearson, Oxford Journals Advertising, PO Box 347, Abingdon OX14 1GJ, UK. E-mail: helen@oxfordads.com. Tel: +44 (0)1235 201904. Fax: +44 (0)8704 296864.

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

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

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

COPYRIGHT © The International Bank for Reconstruction and Development/THE WORLD BANK 2006 *All* rights reserved; no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission of the publisher or a license permitting restricted copying issued in the UK by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1P 9HE, or in the USA by the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

The Primacy of Institutions Reconsidered: Direct Income Effects of Malaria Prevalence

Kai Carstensen and Erich Gundlach

Some recent empirical studies deny any direct effect of geography on development and conclude that institutions dominate all other potential determinants of development. An alternative view emphasizes that geographic factors such as disease ecology, as proxied by the prevalence of malaria, may have a large negative effect on income, independent of the quality of a country's institutions. For instance, pandemic malaria may create a large economic burden beyond medical costs and forgone earnings by affecting household behavior and such macroeconomic variables as international investment and trade. After controlling for institutional quality, malaria prevalence is found to cause quantitatively important negative effects on income. The robustness of this finding is checked by employing alternative instrumental variables, tests of overidentification restrictions, and tests of the validity of the point estimates and standard errors in the presence of weak instruments. The baseline findings appear to be robust to using alternative specifications, instrumentations, and samples. The reported estimates suggest that good institutions may be necessary but not sufficient for generating a persistent process of successful economic development.

Economists, historians, and other social scientists have explained the large differences in the standard of living between the world's richest and poorest nations in many different ways. One strand of the literature has emphasized the preeminent role of physical geography in explaining cross-country differences in the level of development. Some recent empirical studies deny any direct effect of geography on development and conclude that institutions dominate all other potential determinants of development (Hall and Jones **1999**; Acemoglu, Johnson, and Robinson **2001**; Easterly and Levine **2003**; Rodrik, Subramanian, and Trebbi

Kai Carstensen and Erich Gundlach are research fellows at the Kiel Institute for the World Economy; their email addresses are kai.carstensen@ifw-kiel.de and erich.gundlach@ifw-kiel.de. The authors thank two anonymous referees and the editor, seminar participants at Aarhus University, Goethe University in Frankfurt, and Hamburg University, participants at the Ninth Convention of the East Asian Economic Association in Hong Kong, and Michael Funke, Charles I. Jones, Rolf J. Langhammer, Dani Rodrik, and Jeffrey Sachs for helpful comments on an earlier draft, as well as William A. Masters for providing the data on frost frequency, Gordon McCord for explaining the malaria data, Stephen Donald for sharing a sample program on instrument selection, and Jean-Louis Arcand for suggesting econometric strategies.

THE WORLD BANK ECONOMIC REVIEW, VOL. 20, NO. 3, pp. 309–339

doi:10.1093/wber/1h1001

Advance Access publication June 8, 2006

© The Author 2006. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development | THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.

2004). Engerman and Sokoloff (1997) and Acemoglu, Johnson, and Robinson (2002) examine how geographic endowments in the Americas may have shaped factor abundance (people per unit of land) and how unequal factor abundance may have shaped persistent institutions imposed by the colonizing powers that enabled the entrenchment of a small group of elites. Because the Industrial Revolution required the broad participation of the population in entrepreneurship and innovation, economies were favored that started with a more equal factor abundance (due to geographic endowments) and hence with institutions that resulted in a less unequal distribution of income and wealth. Overall, this new literature emphasizes that geographic endowments affect the level of development only through their impact on factor abundance, political economy, and institutions and not more directly.

While development economists can easily agree on the relevance of good institutions for successful development and on the indirect role of physical endowments in shaping different institutional outcomes and different paths of development, there is no agreement on the direct role of geography for development in the recent empirical literature. Partly by highlighting the arguments of the older literature and partly by presenting new empirical evidence, Jeffrey Sachs and his coauthors in particular have argued in a series of papers that measures of geography such as disease ecology may directly affect the level of economic development in addition to the undisputed effects of the institutional framework of a country (Bloom and Sachs 1998; Gallup, Sachs, and Mellinser 1999; Gallup and Sachs 2001; Sachs 2001; McArthur and Sachs 2001; Sachs and Malaney 2002; Sachs 2003). The main disagreement in the current debate concerns the robustness of the empirical evidence presented by Sachs and his coauthors, which has been directly rejected by Acemoglu, Johnson, and Robinson (2001) and Rodrik, Subramanian, and Trebbi (2004) and is in conflict with the studies that favor the primacy of institutions.

A parsimonious baseline specification is used here to reconsider the general econometric limitations of alternative empirical strategies that have been applied to derive clear-cut conclusions with regard to the deep determinants of development. Recent empirical studies have not treated geographic variables such as disease ecology in the same way as measures of institutions, thus probably reducing the chances of the geography hypothesis to prevail. The contribution of this article is to see whether a measure of disease ecology such as malaria prevalence, which is likely to be an endogenous geography variable, directly explains the level of development independent of a measure of institutions, which is also likely to be an endogenous variable. This appears to be an unsettled question in the empirical literature.

Mainly to keep the empirical analysis tractable in the presence of a limited number of candidates for instrumental variables, explanatory variables other than institutions and disease ecology are ignored, not least

because measures such as the quality of economic policies (Easterly and Levine 2003) or the level of trade integration (Rodrik, Subramanian, and Trebbi 2004) have not been found to exert a direct effect on the level of development independent from the effect of institutions. Instead, the emphasis is on the fundamental problems of statistical inference implied by instrumental variable estimation and on tracing the basic reason for the different empirical results on the direct role of disease ecology. After controlling for institutional quality, malaria prevalence is found to have quantitatively important direct negative effects on income. This finding appears to be robust to alternative specifications, instrumentations, and samples.

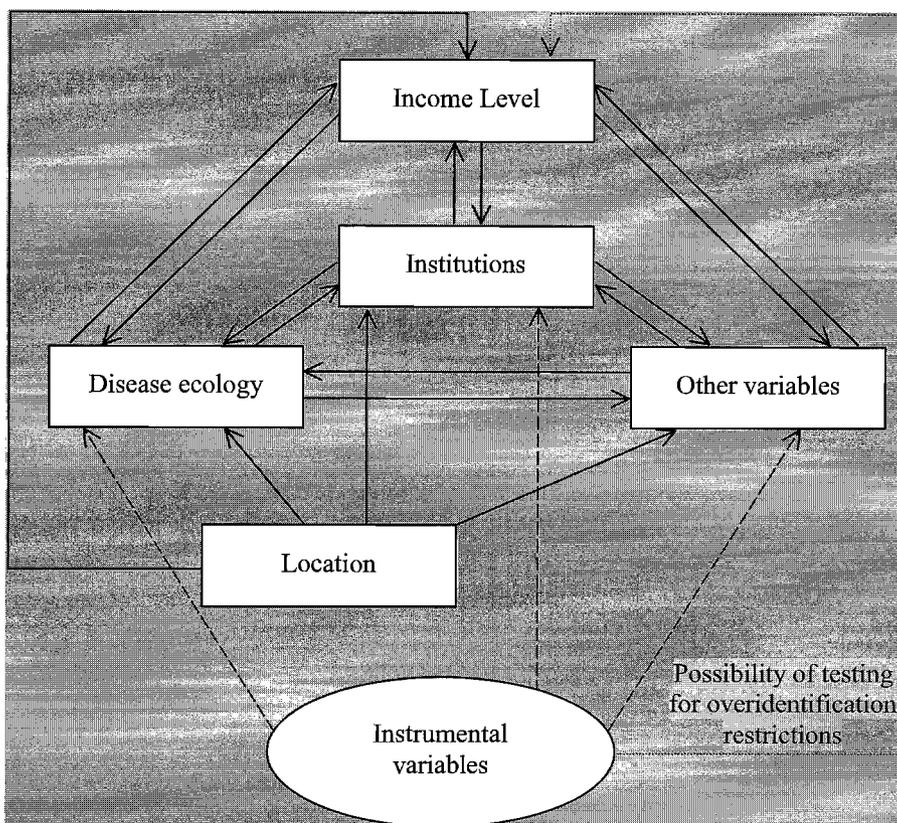
Finding a robust direct effect of malaria prevalence, which is held to be a proxy for the adverse disease ecology of a country, should matter for devising appropriate development policies. For instance, if there is no empirical evidence that malaria prevalence directly affects the level of development in impoverished countries, foreign aid may be targeted mainly to improve policies and institutions. But the finding that there are such direct effects on income means that foreign aid should also be spent on solving biophysical and technological problems that are specific to public health in tropical countries. Especially in Sub-Saharan Africa, but probably also in parts of Asia and Latin America, poor countries may need something in addition to good institutions to generate a persistent process of successful economic development.

I. THE MANY POSSIBLE LINKS BETWEEN GEOGRAPHY, INSTITUTIONS, AND DEVELOPMENT

The various approaches that have been used to identify the many possible links between geographic endowments, institutions, and level of development can be represented with the help of figure 1. The solid arrows indicate the potential directions of causality between the variables considered, the dashed arrows indicate the relation between the instrumental variables and the endogenous explanatory variables, and the dotted arrow indicates the possibility of testing for overidentification restrictions, if there are more instrumental variables than endogenous variables.

Instrumental variables are needed to identify the direct development effects of institutions and other endogenous explanatory variables. An instrumental variable can be used to identify that part of the variation in the endogenous explanatory variables that is exogenous to the variation in the dependent variable, which here is the level of income per capita. The solid arrows indicate that the instrumental variable method can identify the true causal effect of the endogenous variables on the level of income. If valid instrumental variables are available, unbiased estimates of the causal effects on income of institutions, disease ecology, and other variables may be obtained without having to identify

FIGURE 1. Alternative Links between Geography, Institutions, and Development



Source: Authors' summary.

the potential reverse causality from the level of income to the explanatory variables. In addition, the instrumental variable method can be used in principle to identify the web of relations between the endogenous explanatory variables and the indirect effects of geography on the level of income.

Indirect effects of geography are especially highlighted by Engerman and Sokoloff (1997). In terms of figure 1, they argue that geographic endowments may determine factor abundance (the arrow from "location" to "other variables"), factor abundance may entrench persistent institutions, and institutions may influence the level of development. In line with their hypothesis, Rodrik, Subramanian, and Trebbi (2004) consider whether a measure of location may affect the level of development directly or indirectly through measures of institutions and trade integration. They find only weak statistical evidence for a direct effect of a measure of geography, whereas a measure of the quality of institutions appears as the only important explanatory variable. Their conclusion is

confirmed by Easterly and Levine (2003), who do not use a measure of geography in their specification but instead employ an overidentification test to see whether their set of geography-based instrumental variables can be considered valid. They find no evidence that any of their instrumental variables should be included in their specification, which may also be interpreted as supporting the hypothesis of the primacy of institutions.¹

With all possibilities taken into account, certain variables may affect the level of income directly, through their effects on other variables, through one channel, or not at all. Empirical estimates of the various potential causal effects depend crucially on the quality of the available instrumental variables, which have to be correlated with the endogenous explanatory variables but uncorrelated with the error term in the structural equation. This condition will be satisfied if the instrumental variables affect the dependent variable only indirectly through the endogenous explanatory variables that are included in the structural equation and if the instrumental variables are not affected by any feedback from the endogenous variables. Thus, almost by definition, valid instrumental variables are difficult to come by in the cross-country empirics of development, because most economic variables are affected by the level of income. Measures of geography play a special role in this context because most of them, such as distance from the equator or temperature, can undoubtedly be considered as exogenous to the level of development and to the endogenous explanatory variables.

The problem is that once a measure of geography has a direct effect on the level of income, it no longer qualifies as a valid instrumental variable. For instance, local temperature may affect the level of income either directly through the location–income link or through its effects on other variables (figure 1). Where there is a direct location–income link, the measure of geography cannot be used as an instrumental variable. And if the remaining available instrumental variables are also mainly measures of geography such as distance from the equator, the independent variation across the exogenous variables could probably turn out to be too small to allow for an empirical identification of all causal effects of interest.

Acemoglu, Johnson, and Robinson (2001) estimate a specification where institutions and malaria prevalence determine the level of income. Since the measure of malaria prevalence is not instrumented but is nevertheless statistically insignificant, they conclude that there is no causal disease–income link. In support of their empirical results, Acemoglu, Johnson, and Robinson dismiss a priori the possibility that tropical diseases such as malaria could have a large effect on the level of development because people living in areas where such diseases are endemic may have developed immunities against the diseases. According to this view, malaria is unlikely to have strong income effects because it is a debilitating rather than a fatal disease, with the risk of severe illness and

1. See Hall and Jones (1999) for the same line of reasoning based on overidentification tests.

death limited mainly to people without any immunity such as children below the age of 5 years and adults who grew up elsewhere, like European settlers.²

However, this argument ignores other mechanisms through which malaria may affect the level of income. For instance, one form of immunity against malaria comes at a cost for the adult population. The sickle cell trait provides protection against malaria without serious health complications when inherited from only one parent, but the same allele inherited from both parents leads to sickle cell anemia.³ Sickle cell anemia generates severe episodes of pain and increasing infections, outcomes that are at least comparable to the direct negative health effects of malaria experienced by people without any immunity. These considerations suggest that due to natural selection, areas with a high prevalence of malaria are likely to be areas with a high prevalence of sickle cell anemia. Some estimates claim that up to 40 percent of the population in tropical Africa may carry the sickle cell trait.⁴ Thus, malaria may additionally cause poor health and absenteeism of the workforce through natural selection in favor of a high prevalence of the sickle cell trait.

This is not to deny that traditional studies of the economic costs of malaria find relatively small gross domestic product (GDP) losses on the basis of the total number of cases and the fixed costs of prevention and treatment. But there are at least two other mechanisms through which the pandemic nature of malaria can impose large economic costs, namely by affecting household behavior and macroeconomic variables such as foreign direct investment, trade, and tourism (Sachs and Malaney 2002). In response to the disease, households may increase fertility and thus the dependency ratio, which will reduce GDP per capita. A high fertility rate is also likely to reduce investments in human capital per child, which may reduce possibilities for long-run development. In addition, malaria infections can reduce the cognitive development and learning ability of children, which may further depress the long-run average skill level and thus the level of development. Malaria may also decrease household investment in physical

2. The *Anopheles* mosquito is the vector that transmits malaria from human to human. The mosquito must first bite an infected person who is sick with malaria. Then, the mosquito must survive several days while the malaria parasite develops in its body. Finally, the infected mosquito must bite another human to complete the circle of infection. About 40 species of the *Anopheles* mosquito are significantly involved in the transmission of malaria. These species differ substantially with respect to their feeding behavior on humans and their longevity. Thus, all other things constant, the potential for malaria transmission will be high in densely populated regions where the locally dominant *Anopheles* has developed, through biological evolution, a specific human-biting behavior and a relatively high daily survival rate and has found excellent breeding conditions. Most tropical regions combine these favorable conditions for malaria prevalence. In moderate climatic zones far away from the equator, the breeding conditions are less favorable, and the locally dominant *Anopheles* mosquitoes are less specialized on human biting and usually less robust. However, differences in the potential for malaria transmission do exist not only across but also within climatic zones, as is shown by a new measure called the stability of malaria transmission (Kiszewski and others 2004).

3. For information on sickle cell anemia, see <http://www.scinfo.org/> (March 2006).

4. See http://www.pbs.org/wgbh/evolution/library/01/2/1_012_02.html (March 2006).

capital compared with a situation with less-frequent episodes of illness, lower fertility, and lower dependency ratios.

At the macroeconomic level, malaria appears to suppress the economic linkages between malarious and nonmalarious regions of the world. Foreign investors may avoid malarious regions if the disease burden raises the costs of attracting the needed labor force, and with less foreign investment inflows, there will be fewer possibilities of exploiting comparative advantages by specialization and international trade. For instance, trade in services such as tourism cannot prosper under conditions of high rates of malaria transmission, which will reduce the possibility to import growth-enhancing investment goods. In addition, malaria might be closely related to other diseases, either as a direct causal factor or by rendering individuals more susceptible to other diseases, which would further increase its cumulative economic costs. Malaria prevalence may thus have much larger economic costs than will be visible from calculating only the direct medical costs and forgone earnings, especially if international trade in goods and services and international investment are critical factors for development in an era of closer global economic integration.

To put the geography hypothesis on a more equal footing with the primacy of institutions hypothesis, both the prevalence of malaria and the quality of institutions are used as endogenous explanatory variables, as suggested by Sachs (2003). The main interest is the relative size of the causal effects of the institutions–income link and the disease–income link (figure 1). This approach differs from the approaches by Hall and Jones (1999), Acemoglu, Johnson, and Robinson (2001), Easterly and Levine (2003), and Rodrik, Subramanian, and Trebbi (2004), which do not always instrument their measure of endogenous disease ecology, use an exogenous measure of geography, or rely on an overidentification test only. Because of previous results reported in the literature and a general shortage of plausible instrumental variables, all possible effects that might result from explanatory variables other than institutions and disease ecology are ignored, as are possible links between institutions and disease ecology. Different from Sachs (2003), additional instrumental variables are employed, recently developed econometric tests are used to check the validity of the point estimates and standard errors in the presence of weak instrumental variables, and a test for overidentification restrictions is applied to avoid a potentially unjustified exclusion of exogenous measures of geography from the baseline specification.

II. SPECIFICATION AND DATA

In line with previous empirical studies, the following cross-country regression equation is used to estimate the relative effects of institutional quality (INSTITUTIONS) and malaria prevalence (MALARIA) on economic development, which here is measured by the logarithm of GDP per capita ($\ln\text{GDPC}$):

$$(1) \quad \ln GDPC_i = \beta_1 + \beta_2 \cdot INSTITUTIONS_i + \beta_3 \cdot MALARIA_i + \varepsilon_i$$

where ε_i is an error term with zero mean and common variance, and β_2 and β_3 are the coefficients of interest. The research question is whether an estimate of β_3 , as represented by the disease–income link in figure 1, is statistically different from zero, negative, and quantitatively important. To better understand where the different results in the literature may come from, the baseline specification [(equation (1))] is re-estimated by paying particular attention to the choice of the variables, the instruments, and the country sample.

The Choice of the Variables

Indicator variables are needed to measure the effects of institutions and geography on the level of economic development. Such indicator variables are necessarily incomplete and erroneous because the three concepts are multidimensional and difficult to measure. Therefore, different indicator variables are used. The dependent variable is either the $\ln gdp$ in 1995, which is used by Acemoglu, Johnson, and Robinson (2001), Easterly and Levine (2003), and Rodrik, Subramanian, and Trebbi (2004), or the log of GDP per working age person in 1990 ($\ln gdpw$), which appears to be more closely related to the applied growth literature and is used by Hall and Jones (1999).⁵

Institutional quality is measured by one of the following three variables: an average index of the quality of governance in 1996 (rule) from Kaufmann, Kraay, and Mastruzzi (2004), the index of government antidiversion policies in 1986–95 ($gadp$) used by Hall and Jones (1999), and the index of protection against expropriation in 1985–95 (*exprop*) used by Acemoglu, Johnson, and Robinson (2001). To measure disease ecology, two measures of malaria prevalence are employed: the proportion of a country's population at risk of malaria falciparum transmission in 1994 (*malfal*) used by Acemoglu, Johnson, and Robinson (2001) or a new index of malaria risk (*malrisk*) suggested by Sachs (2003). The new index is based on the prevalence of nonfatal species of the malaria pathogen (*Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale*), where a relatively higher proportion of malaria *vivax* is reported for the Americas, Europe, and much of Asia than for sub-Saharan Africa. For international comparisons, the new index may provide a more accurate measure of the share of the population that is at risk of malaria infection than the measure used by Acemoglu, Johnson, and Robinson (2001).

The Choice of the Instrumental Variables

Two premises, both suggested by Acemoglu, Johnson, and Robinson (2001), are used to find an instrumental variable for institutional quality. First, studying the impact of institutions on the level of development has to focus on a sample of

5. See the appendix for detailed descriptions of the data and sources.

former colonies, because only this sample provides the necessary exogenous variation in measures of institutions that can be exploited to estimate a causal effect. Second, the potential endogeneity of any measure of institutional quality should be controlled for by a measure that is correlated with the current variation in the institutional frameworks without being influenced by current economic conditions,⁶ and it should only affect the current level of development through its effect on institutions but not directly. In this context, mortality among European settlers in the early nineteenth century appears to be the most plausible instrumental variable that has been suggested to date.

Differences in mortality among early settlers across colonies, which were well known in Europe at the time, may explain the differences in institutional frameworks that were created by the colonizing powers. For instance, regions with low mortality were favored for settlement, and colonies of settlers may have implemented for themselves a set of institutions that resembled the institutions of their home countries by establishing property rights, the rule of law, and checks against government power. In regions where large-scale settlement was not feasible for Europeans because of an unfavorable disease ecology and high rates of mortality, the colonial powers may have imposed a different set of institutions that did not protect private property and did not provide protection against expropriation but instead focused mainly on the extraction of natural resources.⁷ Since early settler mortality is certainly independent of current economic conditions and since early institutional frameworks have proved to be fairly persistent over time (Acemoglu, Johnson, and Robinson 2001), settler mortality across former colonies can be used as an instrumental variable that helps to identify the exogenous cross-country variation in current institutional frameworks.

To control for the endogeneity of malaria prevalence, a new measure of malaria ecology (*maleco*) is considered that was developed by Kiszewski and others (2004) and first used for cross-country regressions by Sachs (2003). Since this measure of malaria ecology is built only on the climatic factors and biological properties of each regionally dominant malaria vector, Kiszewski and others (2004) claim that *maleco* is exogenous to public health interventions and economic conditions and thus can be considered as a valid instrumental variable in regressions of economic development on malaria risk.

The index of malaria ecology measures the contribution of regionally dominant vector mosquitos to the potential transmission intensity of malaria. Thus, it includes regions where malaria is not currently transmitted but where it had

6. Technically, this means that the control measure should be uncorrelated with the error term of the income equation [equation (1)].

7. The hypothesis advanced by Acemoglu, Johnson, and Robinson (2001) that geographic and climatic conditions were decisive for the adoption of institutions that favored either settlement or resource extraction is in conflict with some historical facts for the colonization of the Americas. An alternative hypothesis favored by Engerman and Sokoloff (1997) emphasizes initial factor abundance as determinants of institutions. See Hoff (2003) for a survey of the issues.

been transmitted in the past or might be in the future.⁸ Since the region-specific dominant malaria vector reflects only the forces of biological evolution, it can be considered independent of current economic conditions. That is, terms likely to be affected by economic conditions or public health interventions (mosquito abundance, for example) do not enter the calculation of the index. The index reveals that, because of different vector properties, a given malaria intervention is likely to have a smaller impact in the tropics than in more temperate climatic zones, where the vector is less robust and does not specialize in human biting and where the parasite has less fatal infectious consequences.

However, Rodrik, Subramanian, and Trebbi (2004) doubt that *maleco* is actually exogenous to current economic conditions. They object that Sachs (2003) does not detail the construction of the index and point out that Kiszewski and others (2004) do not discuss exogeneity at all.⁹ While this critique is technically correct, doubts regarding the exogeneity of *maleco* may not be justified, as discussed in the previous paragraph. Nevertheless, three sets of further instrumental variables that relate to the climatic environment, the influence of Western European languages, and the openness of a country are considered in addition to the two baseline instruments *lnmort* and *maleco*.

Temperature, rainfall, and latitude are additional measures of the climatic environment that can be related to preconditions for the prevalence of malaria. Since a key part of the life cycle of the parasite depends on a high ambient temperature, malaria is intrinsically a disease of warm environments. Malaria also depends on adequate conditions of mosquito breeding, mainly pools of clean water from rainfall. Hence, the prevalence of frost (frost), measured as the proportion of a country's land receiving 5 or more frost days in winter, or the degree of humidity (humid), measured as the highest temperature during the month when average afternoon humidity is at its highest, may be considered as appropriate instrumental variables that are exogenous to economic conditions. In addition, distance from the equator as measured by the absolute latitude of a

8. Abstracting from all detail, the construction of the index proceeds in two basic steps. First, the regionally dominant Anopheles mosquito is identified across countries in which malaria is or has been endemic. The criteria for the identification of the dominant vector are its longevity and its human-biting habit. Second, the index of malaria ecology is calculated as $(a_i^2 p_i^E) / (-\ln p_i)$, where i is the identity of the dominant malaria vector, a is the proportion of vector i biting people $[0,1]$, p is the daily survival rate of vector i $[0,1]$, and E is the length of the extrinsic incubation period in days, which depends mainly on average temperature and differs between *Plasmodium falciparum* and *Plasmodium vivax*. Hence, the index value for a specific country is measured as a function of climatic factors that determine the required habitat of the dominant vector and of biological properties of the region-specific dominant vectors.

9. Information on the construction of the malaria transmission index (malaria ecology) is available online at <http://www.earth.columbia.edu/about/director/malaria/index.html> (March 2006). A previous version of the text describing the construction of the index may have contributed to the impression that *maleco* is not purged of endogeneity, because it stated that a measure of mosquito abundance is included in the calculation. However, observed mosquito abundance enters the index of malaria ecology only as a screen for precipitation data, where the independently identified dominant malaria vector is assumed to be absent from the specific site under consideration if precipitation falls below a certain level per month.

country (latitude) may also be used as a proxy for the climatic environment. What has to be taken into account, however, is that these three measures of climatic conditions may be instrumental variables not only for measures of disease ecology but also for measures of institutional quality. This is because settler mortality and thus the design of early institutions were influenced by the prevailing climatic conditions of the colonies. In this context, Acemoglu, Johnson, and Robinson (2001) point out that their work shows why absolute distance from the equator might matter as an instrument for a measure of institutions, as used in Hall and Jones (1999).

Other plausible instrumental variables than measures of geography are more difficult to come by. One possibility is to consider the share of the population that speaks English ($engfrac$) or another Western European language ($eurfrac$) as the first language. As suggested by Hall and Jones (1999), these variables may reflect the different degree of Western European influence on the sample countries and thus may help to identify the exogenous variation in measures of institutions. Since Acemoglu, Johnson, and Robinson (2001) generally question the exogeneity of these variables, some formal tests of exogeneity are provided when using them in the checks of the robustness of the baseline results.

Furthermore, instrumental variables that are used for checking robustness relate to a country's trade openness. More open countries may have better institutions because openness may encourage less arbitrary government behavior, especially toward property rights. Thus, exogenous measures of openness could be used as instrumental variables for measures of institutions. Two measures of openness are employed: the proportion of land area that is within 100 kilometers of the coast ($coast$), which is taken from McArthur and Sachs (2001), and the (log) predicted trade share of a country ($trade$), which is constructed by Frankel and Romer (1999) from a gravity model that uses mainly geographical variables to explain actual bilateral trade flows.

Choice of the Sample

The sample of countries is limited to former colonies for which data on early settler mortality are available. Acemoglu, Johnson, and Robinson (2001, table 7, p. 1392) estimate equation 1 for a sample of 62 countries. This sample, however, includes 14 countries that are known to provide unreliable statistics (rated as D countries in Summers and Heston 1991), two countries that are very small (less than 1 million inhabitants in 1990), and one country that depends mainly on oil production. These countries are removed from the sample. Thus, baseline results are reported for a smaller but probably more reliable sample of 45 former colonies that are not statistical terra incognita, small, or dependent on oil production. By contrast, previous studies that took issue with the Acemoglu, Johnson, and Robinson result on the primacy of institutions (McArthur and Sachs 2001; Easterly and Levine 2003; Sachs 2003; Rodrik, Subramanian, and Trebbi 2004) increased the Acemoglu, Johnson, and Robinson sample size but

disregarded data quality. As a robustness check of the baseline findings on sample size, a larger sample of countries with additional observations on settler mortality (Acemoglu, Johnson, and Robinson 2000) is also included.

III. BASELINE ESTIMATION RESULTS

To begin, equation 1 is estimated by two-step least squares (2SLS) using $\ln gdp_c$ as the dependent variable, $rule$ and $malfal$ as explanatory variables, and $\ln mort$ and $maleco$ as instrumental variables. This baseline specification is close to specifications in the literature. In particular, (a) $\ln gdp_c$ is used as the dependent variable by Acemoglu, Johnson, and Robinson (2001), Easterly and Levine (2003), and Rodrik, Subramanian, and Trebbi (2004); (b) $rule$ is used as a measure of institutional quality by Easterly and Levine (2003) and Rodrik, Subramanian, and Trebbi (2004); and (c) $malfal$ is used as a measure of malaria prevalence by Acemoglu, Johnson, and Robinson (2001). The results are presented in column 1 of table 1.

The point estimates have the expected signs and are quantitatively important. The point estimate of β_2 reflects the change in log output per capita associated with an one-unit increase in the index of governance quality. Thus, $\beta_2 = 0.89$ implies that a difference of 0.1 in the governance index is associated with a 8.9 percent cross-country difference in output per capita. To show the potential magnitude of the estimated effect of the measure of institutions on economic performance, two countries in the sample that represent about the 70th and the 30th percentile of the governance index are compared, South Africa with an index value of 0.21 and Ecuador with a value of -0.40 . This difference is predicted to result in a 0.54 log-point difference [$(0.21 + 0.40)$ times 0.89] between the log per capita GDPs of the two countries. That is, the per capita GDPs of South Africa and Ecuador are predicted to differ by a factor of about 1.7 due to institutional differences, whereas their sample per capita GDPs differ by a factor of about 2.7.

The point estimate of β_3 reflects the change in log output per capita associated with an one-unit increase in malaria prevalence. Thus, $\beta_3 = -1.04$ implies that the per capita GDPs of Paraguay and Pakistan, which represent roughly the 40th and the 70th percentile of the highly stratified distribution of the malaria index (with percent values of 0.001 for Paraguay and 0.49 for Pakistan), should differ by a factor of about 1.7 due to the differences in the proportion of the population that lives with the risk of malaria infection, whereas the sample per capita GDPs of these two countries differ by a factor of about 2.6.

The point estimates are statistically significant, with estimated standard errors of 0.18 for β_2 and 0.30 for β_3 , which imply t-statistics of 5.04 and -3.46 . These values indicate statistical significance at the 5 percent level when using t-tests based on conventional asymptotic theory. The reported conventional confidence intervals contain the unknown true parameters with a

TABLE 1. Baseline Estimation Results

| Explanatory variables | 1 | | 2 | | 3 | | 4 | |
|---|-------|--------|-------|---------|-------|--------|--------|--------|
| | rule | malfal | rule | malrisk | gadp | malfal | exprop | malfal |
| Estimated coefficients | | | | | | | | |
| Two-step least square | 0.89 | -1.04 | 0.78 | -1.31 | 3.31 | -1.48 | 0.55 | -1.03 |
| Standard error | 0.18 | 0.30 | 0.22 | 0.42 | 0.60 | 0.22 | 0.12 | 0.34 |
| Fuller ^a | 0.89 | -1.05 | 0.78 | -1.29 | 3.31 | -1.48 | 0.53 | -1.07 |
| Standard error | 0.17 | 0.30 | 0.22 | 0.40 | 0.59 | 0.21 | 0.12 | 0.33 |
| Bounds of 95 percent confidence intervals | | | | | | | | |
| Conventional | | | | | | | | |
| Upper | 0.53 | -1.65 | 0.33 | -2.16 | 2.09 | -1.91 | 0.30 | -1.72 |
| Lower | 1.25 | -0.43 | 1.23 | -0.47 | 4.52 | -1.04 | 0.79 | -0.34 |
| Conditional likelihood ratio | | | | | | | | |
| Upper | 0.44 | -1.83 | 0.05 | -2.85 | 1.53 | -2.07 | 0.29 | -1.75 |
| Lower | 1.43 | -0.18 | 1.39 | -0.23 | 4.99 | -0.87 | 1.29 | 0.75 |
| Number of observations | 45 | | 45 | | 45 | | 45 | |
| Number of instruments | 2 | | 2 | | 2 | | 2 | |
| First-stage statistics | | | | | | | | |
| F-statistic | 28.12 | 42.31 | 28.12 | 20.38 | 23.37 | 41.78 | 17.49 | 42.31 |
| p-value | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Partial R | 0.57 | 0.67 | 0.57 | 0.49 | 0.53 | 0.67 | 0.45 | 0.67 |
| Shea partial R | 0.37 | 0.43 | 0.33 | 0.28 | 0.37 | 0.46 | 0.23 | 0.34 |
| Weak-instrument test ^b | | | | | | | | |
| Cragg-Donald | 11.45 | | 7.91 | | 10.33 | | 5.74 | |
| Critical value | 7.03 | | 7.03 | | 7.03 | | 7.03 | |

Note: All specifications are estimated with the instrumental variables *lnmort* and *maleco* for the small sample of 45 countries.

^aDenotes the Fuller estimator with correction parameter $c = 1$ proposed by Hausman, Stock, and Yogo (2005).

^bThe Cragg-Donald statistic (Cragg and Donald 1993) is used by Stock and Yogo (2002) for weak-instrument tests. If the Cragg-Donald statistic exceeds the critical value, then a standard significance test with nominal size of 5 percent has a maximal size of 10 percent.

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

confidence level of 95 percent. Thus, the true coefficient of the governance index is estimated to be between 0.53 and 1.25, and the true impact of malaria prevalence is estimated to be between -1.65 and -0.43. However, these results depend on instrument relevance, as emphasized by the recent literature on weak instruments (Staiger and Stock 1997; Moreira 2003). If the instrumental variables are only weakly correlated with the endogenous explanatory variables, conventional asymptotic theory no longer holds, and statements about statistical significance and inference may lead to the wrong conclusions. The relevance of the instrumental variables has to be checked to allow for statements about statistical significance.

The first-stage regressions supply valuable information about the relevance of the instrumental variables. Highly significant F-statistics of 28.1 and 42.3 for the

first-stage regressions of rule and of malrisk are reported (see table 1). In addition, both the usual partial R^2 and the Shea (1997) partial R^2 are far above zero. These test results point to strong instruments, but even large first-stage F-statistics can be misleading. For example, the two instruments *Inmort* and *maleco* may not carry sufficient independent information, which could make it difficult to identify distinct effects of rule and malrisk. To this end, a statistic proposed by Cragg and Donald (1993) is computed, which represents the relevance of the weakest instrument. Using weak-instrument asymptotic theory, Stock and Yogo (2002) show that a conventional significance test on β with a nominal size of 5 percent has an actual size of 10 percent or more, and is thus severely distorted, if the Cragg–Donald statistic is below 7. Since the Cragg–Donald statistic equals 11.45, the results of the baseline specification are not affected by weak-instrument problems.

As a further robustness check for specifications with potentially weak instruments, a modified limited information maximum likelihood estimator (Fuller 1977) is applied. This Fuller estimator with modification parameter = 1 is more robust to the presence of weak instruments than 2SLS (Hahn, Hausman, and Kuersteiner 2004; Hausman, Stock, and Yogo 2005). The Fuller point estimates are almost identical to the 2SLS estimates, with estimates of 0.89 for β_2 and -1.05 for β_3 . In addition, 95 percent confidence intervals are computed based on inverted conditional likelihood ratio (CLR) tests that take any weak-instrument problem into account (Moreira 2003).¹⁰ The CLR intervals turn out to be only slightly larger than the confidence intervals based on conventional asymptotic theory reported above, ranging from 0.44 to 1.43 for β_2 and from -1.83 to -0.18 for β_3 . In particular, when the CLR confidence intervals are turned into significance tests, both β_2 and β_3 are individually statistically significant, because the confidence intervals do not include zero.

Before proceeding with further robustness checks, it is worth summarizing the results obtained with the baseline specification. The point estimates for the effects of institutional quality and malaria prevalence have the expected signs and are economically important. They do not appear to suffer from a weak-instrument problem and so appear statistically significant as well. Identifying a direct effect of a measure of disease ecology on the level of development conflicts with the evidence presented by Hall and Jones (1999); Acemoglu, Johnson, and Robinson (2001); Easterly and Levine (2003); and Rodrik, Subramanian, and Trebbi (2004) and confirms the evidence presented by Sachs (2003).

10. Since there are two endogenous explanatory variables, the approach of Moreira (2003) delivers a bivariate confidence region from which two univariate confidence intervals are calculated by the projection method put forward by Dufour (1997). A Matlab program that accomplishes this task is available on request.

IV. ROBUSTNESS

The results of the baseline specification are subjected to a number of robustness checks, beginning with the effects on the baseline results of alternative measures of the dependent variable and of the endogenous explanatory variables. Other robustness checks assess the inclusion of alternative and additional instrumental variables, the validity of the baseline instruments *Inmort* and *maleco*, and the impact on the results when a larger sample of countries is used.

Effects of Alternative Variables

Estimation results for specifications with alternative explanatory variables are presented in columns 2–4 of table 1. In column 2, institutional quality is still measured by the governance index (*rule*), but malaria prevalence is now measured by the risk of infection with the nonfatal malaria pathogen (*malrisk*). The main difference from the baseline specification is the smaller weight for institutional quality and the larger weight for malaria prevalence. This difference may be due simply to estimation uncertainty, which has increased compared with the baseline specification, as indicated by the larger confidence intervals. Moreover, the weak-instrument problem is of slightly greater relevance than before, as indicated by a smaller Cragg–Donald statistic, which nevertheless still exceeds the critical value of 7. Despite somewhat weaker test statistics, all general conclusions drawn from the baseline specification are confirmed by the specification with *malrisk* as well. Therefore, all further-reported specifications use *malfal* as the measure of malaria prevalence.¹¹

In column 3, the governance index is replaced by the index of government antidiversion policies (*gadp*) as a measure of institutional quality, while malaria prevalence is measured by the risk of infection with malaria *falciparum* (*malfal*). The point estimate for β_2 is considerably larger than in the baseline specification, but this is due mainly to the smaller variance of *gadp* compared with *rule*. The point estimate for β_3 is also absolutely larger than in the baseline specification, but the difference is not substantial if estimation uncertainty is taken into account. The economic significance of these estimates can be shown again for the country pairs discussed above. With a point estimate of 3.31 for β_2 , the empirical model predicts that the per capita GDPs of South Africa and Ecuador differ by a factor of 1.7 due to differences in institutional quality, whereas their sample per capita GDPs differ by a factor of about 2.7. With a point estimate of –1.48 for β_3 , the model predicts that the per capita GDPs of Paraguay and Pakistan differ by a factor of 2.1 due to differences in malaria prevalence, whereas their sample per capita GDPs differ by a factor of about 2.6. The statistical significance of these estimates can be inferred from both the conventional and the CLR confidence intervals, which do not include zero. In addition,

11. Detailed results based on specifications with *malrisk* are available on request.

there is no weak-instrument problem, as indicated by a large Cragg–Donald statistic.

In column 4, the governance index is replaced by the risk of expropriation (*exprop*), while the risk of infection with malaria falciparum (*malfal*) remains the measure of malaria prevalence. This is the specification analyzed by Acemoglu, Johnson, and Robinson (2001, table 7), who obtain an insignificant effect of malaria prevalence with their sample of countries (without instrumenting *malfal*). For the re-estimated equation, the point estimate of β_2 is smaller than in the baseline specification, but this may be explained by the estimation uncertainty and the larger variance of *exprop* compared with rule. The point estimate of β_3 is virtually unchanged. At first sight, both estimates appear statistically significant, as indicated by low standard errors (2SLS and Fuller). However, the Cragg–Donald statistic of 5.74 indicates a weak-instrument problem. The 2SLS estimator may be biased and the conventional confidence intervals may be inadequate. While the point estimates and standard errors remain virtually unchanged when the robust Fuller estimator is used, the CLR confidence intervals of Moreira (2003) indicate that the estimate of the coefficient on *exprop* is statistically significant but the estimate of the coefficient on *malfal* is not.

However, this result does not necessarily imply that malaria prevalence does not have an effect on per capita income. Rather, it indicates that it may not be possible to identify an independent effect with sufficient precision, given the relatively small sample size. This view can be supported by two observations. First, a correlation coefficient of 0.6 shows that the instrumental variables *lnmort* and *maleco* are strongly correlated, which does not leave much information in one instrument that is independent of the other. While this information appears to be sufficient for the previous specifications, it turns out to be insufficient for the current specification, as indicated by lower Shea partial R^2 s than before.

Second, the power of the significance test for β_3 (which is derived from its estimated confidence interval) might be low, probably due to the weak-instrument problem. While not much is known about the power of significance tests in the presence of weak instruments, power is certainly lower than in the conventional strong-instrument case, due to reduced estimation precision. More specifically, the power of a significance test for β_3 using conventional asymptotic theory should give an upper bound for the power of a significance test under weak-instrument asymptotic theory. Fortunately, power for the former test can be easily calculated following the approach by Andrews (1989). An interval with power below 0.5 is calculated to see in which region of true parameter values $\beta_3 \neq 0$ the test can be expected not to reject the wrong null hypothesis $\beta_3 = 0$. The interval turns out to be $[-0.67, 0.671]$. This implies that true parameter values of β_3 between -0.67 and 0.67 have a better chance to be undetected than to be detected. An interval with power below 0.95 turns out to be $[-1.23, 1.231]$. This implies that only true parameter values of $|\beta_3| > 1.23$ are likely (with

probability above 95 percent) to be found statistically significant. The 2SLS point estimate of β_3 is only -1.03 .

Since the correct power of a significance test based on weak-instrument asymptotic theory is likely to be overstated in this exercise, the lower interval limits of -0.67 for power regions below 50 percent and of -1.23 for power regions below 95 percent are only upper bounds for the correct but unknown interval limits. That is, the power of the significance test for β_3 appears to be quite low, given that parameter values for β_3 , such as the reported point estimate, are economically important but statistically difficult to distinguish from zero. Thus, finding the coefficient estimate of β_3 to be statistically insignificant is probably due to the low power of the significance test rather than to the unimportance of malaria prevalence for economic development.¹² This conclusion is also corroborated by the statistically significant point estimates of β_3 in columns 1 and 3 of table 1.

Effects of Additional Instrumental Variables

To exclude the possibility that the reported results are driven by the choice of instrumental variables, the analysis is replicated using additional sets of instrumental variables related to the climatic environment (frost, humid, and latitude), the Western European influence (*eurfrac* and *engfrac*), and openness (coast and trade). The analysis is restricted to the baseline specification, where institutional quality is measured by the governance index (rule) and malaria prevalence is measured by the risk of infection with malaria falciparum (*malfal*). The results are presented in table 2.

Once the additional climate instruments (frost, humid, and latitude) are included (table 2, column 1), the effect of institutional quality is found to be stronger than in the baseline specification (table 1, column 1), whereas the effect of malaria prevalence is weaker. Both differences are small and can be explained by estimation uncertainty. The Hansen test does not reject the three overidentifying restrictions that arise from the fact that the two endogenous regressors are now estimated with five instruments. Taken at face value, this test result would imply that the exogeneity restrictions on the instruments appear valid and that there are no direct effects on the level of economic development from the additional instruments.

The result of the Hansen test should be viewed with care if there are signs of a weak-instrument problem, because then the usual inference based on the χ^2 distribution of the test statistic would no longer hold. There is conflicting evidence on the presence of weak instruments. The Cragg–Donald statistic, which is much smaller than the critical value, does indicate a weak-instrument problem. But the weak-instrument test of Hahn and Hausman (2002), which can be applied only for overidentified equations, does not reject the null hypothesis

12. The reason for the low power is probably that the small sample is not informative enough to identify an independent effect of malaria prevalence in this specification.

TABLE 2. The Impact of Additional Instrumental Variables

| Explanatory variables | Instruments | | | | | | | |
|--|--------------------|---------|-------------------|---------|---------------------|---------|-----------------|---------|
| | 1 | | 2 | | 3 | | 4 | |
| | Baseline + climate | | Baseline + Europe | | Baseline + openness | | All instruments | |
| | rule | mal/fal | rule | mal/fal | rule | mal/fal | rule | mal/fal |
| Estimated coefficients | | | | | | | | |
| Two-step least square | 0.97 | -0.90 | 0.88 | -1.14 | 0.81 | -1.09 | 0.84 | -1.08 |
| Standard error | 0.16 | 0.29 | 0.18 | 0.27 | 0.16 | 0.29 | 0.13 | 0.25 |
| Fuller | 0.98 | -0.89 | 0.88 | -1.14 | 0.81 | -1.09 | 0.84 | -1.08 |
| Standard error | 0.16 | 0.29 | 0.18 | 0.27 | 0.16 | 0.29 | 0.14 | 0.26 |
| Bounds of 95 percent confidence interval | | | | | | | | |
| Conventional | | | | | | | | |
| Upper | 0.64 | -1.48 | 0.55 | -1.69 | 0.44 | -1.67 | 0.57 | -1.58 |
| Lower | 1.30 | -0.32 | 1.18 | -0.59 | 1.13 | -0.51 | 1.10 | -0.58 |
| Conditional likelihood ratio | | | | | | | | |
| Upper | 0.60 | -1.60 | 0.46 | -1.88 | 0.34 | -1.95 | 0.48 | -1.79 |
| Lower | 1.54 | 0.02 | 1.31 | -0.39 | 1.27 | -0.25 | 1.22 | -0.35 |
| Number of observations | 44 | | 44 | | 44 | | 44 | |
| Number of instruments | 5 | | 4 | | 4 | | 9 | |
| Hansen test (OIR) | 2.27 | | 1.05 | | 2.18 | | 6.03 | |
| p-value | 0.52 | | 0.59 | | 0.10 | | 0.53 | |
| First-stage statistics | | | | | | | | |
| F-statistic | 13.74 | 16.74 | 20.53 | 30.33 | 19.96 | 22.93 | 19.58 | 16.05 |
| p-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Partial R ² | 0.64 | 0.69 | 0.68 | 0.78 | 0.67 | 0.70 | 0.84 | 0.81 |
| Shea partial R ² | 0.48 | 0.51 | 0.49 | 0.5* | 0.46 | 0.43 | 0.67 | 0.64 |
| Weak-instrument test ^a | | | | | | | | |
| Cragg-Donald | 6.29 | | 3.49 | | 7.21 | | 5.97 | |
| Critical value | 19.45 | | 18.87 | | 16.87 | | 27.51 | |
| Hahn-Hausman test | 0.21 | -0.22 | 0.30 | -0.31 | -0.12 | 0.12 | 0.21 | -0.21 |
| p-value | 0.83 | 0.83 | 0.76 | 0.76 | 0.91 | 0.91 | 0.83 | 0.83 |

(Continued)

TABLE 2. Continued

| | Instruments | | | | | | | |
|--|--------------------|---------------|-------------------|---------------|---------------------|---------------|-----------------|---------------|
| | 1 | | 2 | | 3 | | 4 | |
| | Baseline + climate | | Baseline + Europe | | Baseline + openness | | All instruments | |
| Explanatory variables | <i>rule</i> | <i>malfal</i> | <i>rule</i> | <i>malfal</i> | <i>rule</i> | <i>malfal</i> | <i>rule</i> | <i>malfal</i> |
| Instrument selection criteria ^b | | | | | | | | |
| Donald–Newey <i>MSE</i> | 0.42 | | 0.37 | | 0.47 | | 0.10 | |
| BIC | -8.99 | | -6.50 | | -3.68 | | -18.84 | |
| HQIC | -5.67 | | -4.28 | | -1.46 | | -11.08 | |

Note: The sets of instruments are baseline (*Inmort* and *maleco*), climate (*frost*, *humid*, and *latitude*), Europe (*eurfrac* and *engfrac*), and openness (*coast* and *trade*).

^aThe Cragg–Donald statistic (Cragg and Donald 1993) is used by Stock and Yogo (2002) for weak-instrument tests. If the Cragg–Donald statistic exceeds the critical value, then a standard significance test with nominal size of 5 percent has a maximal size of 10 percent. The weak-instrument test by Hahn and Hausman (2002) is based on the normalized difference between bias-adjusted two-step least squares estimators (b2SLS) for an equation and its reverse equation, where the left-side variable and the endogenous right-side variable are interchanged.

^bThe Donald–Newey instrument selection criterion (Donald and Newey 2001) is the expected average mean-squared error of the 2SLS estimator. BIC and HQIC are the Bayesian and Hannan–Quinn information criteria (Andrews 1999) for the choice of instruments.

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

of strong instruments. On balance, it can be concluded that the results of the baseline specification are not rejected by adding additional climate instruments, because significant effects of institutional quality and malaria prevalence are obtained if the robust CLR confidence intervals are used for inference.¹³

Adding the Western European instruments (table 2, column 2) and the openness instruments (table 2, column 3) to the baseline instruments also does not change the results by much. Both institutional quality and malaria prevalence exert highly significant effects, even if a potential weak-instrument problem is taken into account, as indicated by the Cragg–Donald test. Again, the overidentifying restrictions are not rejected, and the conclusions of the baseline model are confirmed, as indicated by the CLR confidence intervals. This result is not altered when all instruments are included together (table 2, column 4). The parameters remain highly significant, and the Hansen test still does not reject the overidentifying restrictions. Moreover, a Hansen difference test cannot reject the additional overidentifying restrictions of column 4 in table 2 over those of the columns 1, 2, and 3, leading to test statistics of 3.81 (p-value 0.43), 5.03 (p-value 0.41), and 3.7 (p-value 0.59). Thus, including additional instrumental variables based on climate, Western European influence, and openness does not change the conclusions relative to those obtained from the baseline specification.

As a further robustness check to see which set of instrumental variables is favored, three formal instrument selection criteria are applied—the expected average mean squared error criterion of Donald and Newey (2001), the Bayesian information criterion, and the Hannan–Quinn information criterion of Andrews (1999). All test results suggest choosing the full instrument set.¹⁴ However, it must be taken into account that these criteria are not designed for specifications with weak instruments, where they may lead to the wrong conclusions. As a general tendency in the estimation results, the weak-instrument problem appears to increase with the number of instruments, at least if the Cragg–Donald statistic is taken as the benchmark. This may reflect an overfitting problem when using many instruments, notably in table 2, column 4, where nine instrumental variables are employed for a sample of 45 countries. Nevertheless, the point estimates are quite robust to the number of instruments included. This indicates that the results of the baseline specification are not driven by the choice of instruments and that the baseline specification should be preferred to minimize overfitting and the weak-instrument problems.

13. The 95 percent CLR confidence interval includes zero but is a borderline case. For example, a 93 percent CLR confidence interval does not include zero.

14. The values of Bayesian and Hannan–Quinn information criteria equal zero for all just-identified specifications and are therefore reported only for overidentified specifications.

Validity of *lnnort* and *maleco*

So far, the article has not questioned whether *lnnort* and *maleco* are valid instrumental variables for institutional quality and malaria prevalence. Both variables have been criticized as flawed instrumental variables in the literature. For instance, Albouy (2004) argues that Acemoglu, Johnson, and Robinson (2001) measure settler mortality (*lnmort*) imprecisely. He constructs a "high revision" variable (*lnmort2*) that he claims exhibits improved geographic relevance, statistical precision, and cross-country comparability. When Albouy (2004) re-estimates the effect of institutional quality on economic development with the revised instrument (*lnmort2*), he obtains a severe weak-instrument problem, which results in a failure to measure any statistically significant effect of institutional quality on income.

When *lnmort* is replaced by *lnmort2*, the point estimates remain almost unchanged (table 3, column 1). However, the Cragg–Donald statistic becomes quite small, indicating the presence of weak instruments. Therefore, the CLR confidence intervals are used for inference. They turn out to be very large and in the case of *maleco* to include zero. Somewhat surprisingly, the point estimate of the coefficient on institutional quality is still statistically significant, even though the first-stage statistics indicate a severe drop in explanatory power for the first-stage regression of *rule* on the instruments relative to the baseline estimate."

To improve the first-stage regression results, the set of instrumental variables is augmented with the measures of Western European influence (*eurfrac* and *engfrac*). Their inclusion increases the explanatory power in the first-stage regressions considerably (table 3, column 2). However, there is still a weak-instrument problem according to the Cragg–Donald statistic. But the CLR confidence intervals indicate statistically significant effects of both institutional quality and malaria prevalence in this specification, and the point estimates remain almost unchanged compared with the baseline model.

Thus, the revised settler mortality variable constructed by Albouy (2004) exhibits a smaller degree of instrument relevance than the original variable used by Acemoglu, Johnson, and Robinson (2001). While replacing the original with the revised variable leads to almost unchanged point estimates, the statistical significance of *maleco* becomes questionable. Nevertheless, it is still possible to come up with statistically significant and economically meaningful estimates of the effects on income of institutions and malaria prevalence without using the original mortality variable if the measures of Western European influence are added to the list of instrumental variables. Thus, the baseline results do not depend on using the original mortality variable as an instrument.

Rodrik, Subramanian, and Trebbi (2004) question the exogeneity of the second baseline instrumental variable, *maleco*. Therefore, a specification is

15. In the first-stage regression of *rule*, the revised variable *lnmort2* is highly significant with a t-value of -3.66 but much less so than in the baseline model.

TABLE 3. The Validity of *lnmort* and *Maleco*

| Explanatory variables | Instruments | | | | | | | |
|---|---------------------------------|---------------|--|---------------|--------------------------------|---------------|---|---------------|
| | 1 | | 2 | | 3 | | 4 | |
| | <i>lnmort2 + maleco</i> rule | <i>maifal</i> | <i>lnmort2 + maleco + Europe^a</i> rule | <i>maifal</i> | <i>lnmort + Europe</i> rule | <i>maifal</i> | <i>lnmort2 + Europe^a</i> rule | <i>maifal</i> |
| Estimated coefficients | | | | | | | | |
| Two-step least square | 1.12 | -0.85 | 0.89 | -1.15 | 0.75 | -1.35 | 0.77 | -1.39 |
| Standard error | 0.32 | 0.41 | 0.19 | 0.28 | 0.20 | 0.36 | 0.21 | 0.35 |
| Fuller | 1.07 | -0.90 | 0.89 | -1.15 | 0.75 | -1.33 | 0.77 | -1.37 |
| Standard error | 0.29 | 0.38 | 0.19 | 0.28 | 0.20 | 0.35 | 0.21 | 0.33 |
| Bounds of 95 percent confidence intervals | | | | | | | | |
| Conventional | | | | | | | | |
| Upper | 0.48 | -1.69 | 0.51 | -1.72 | 0.33 | -2.08 | 0.34 | -2.09 |
| Lower | 1.77 | -0.02 | 1.27 | -0.58 | 1.16 | -0.62 | 1.20 | -0.69 |
| Conditional likelihood ratio | | | | | | | | |
| Upper | 0.34 | -1.86 | 0.36 | -1.96 | -0.04 | -2.72 | 0.01 | -2.61 |
| Lower | 9.93 | 8.80 | 1.57 | -0.29 | 1.34 | -0.36 | 1.42 | -0.46 |
| Number of observations | 45 | | 44 | | 44 | | 44 | |
| Number of instruments | 2 | | 4 | | 3 | | 3 | |
| Hansen test (OIR) | | | 1.683 | | 0.149 | | 0.000 | |
| p-value | | | 0.431 | | 0.700 | | 1.000 | |
| First-stage statistics | | | | | | | | |
| F-statistic | 10.181 | 41.89 | 9.12 | 26.16 | 23.92 | 31.21 | 12.41 | 23.45 |
| p-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Partial R ² | 0.33 | 0.67 | 0.48 | 0.73 | 0.64 | 0.70 | 0.48 | 0.64 |
| Shea partial R ² | 0.13 | 0.27 | 0.35 | 0.52 | 0.29 | 0.32 | 0.27 | 0.36 |

(Continued)

TABLE 3. Continued

| | Instruments | | | | | | | |
|--|------------------|---------------|--|---------------|-----------------|---------------|-------------------------------|---------------|
| | 1 | | 2 | | 3 | | 4 | |
| | lnmort2 + maleco | | lnmort2 + maleco + Europe ^a | | lnmort + Europe | | lnmort2 + Europe ^a | |
| Explanatory variables | rule | <i>malfal</i> | rule | <i>malfal</i> | rule | <i>malfal</i> | rule | <i>malfal</i> |
| Weak-instrument test ^b | | | | | | | | |
| Cragg–Donald | 3.27 | | 5.18 | | 4.65 | | 4.76 | |
| Critical value | 7.03 | | 16.87 | | 13.43 | | 13.43 | |
| Hahn–Hausman test | | | 0.10 | −0.10 | 0.28 | −0.28 | 0.33 | 0.10 |
| p-value | | | 0.92 | 0.92 | 0.78 | 0.78 | 0.74 | 0.92 |
| Instrument selection criteria ^c | | | | | | | | |
| Donald–Newey MSE | 7.95 | | 0.65 | | 1.91 | | 1.67 | |
| BIC | | | −5.72 | | −3.57 | | −3.78 | |
| HQIC | | | −3.50 | | −2.46 | | −2.67 | |

^aRefers to the instruments *engfrac* and *eurfrac*.

^bThe Cragg–Donald statistic (Cragg and Donald 1993) is used by Stock and Yogo (2002) for weak-instrument tests. If the Cragg–Donald statistic exceeds the critical value, then a standard significance test with nominal size of 5 percent has a maximal size of 10 percent. The weak-instrument test by Hahn and Hausman (2002) is based on the normalized difference between bias-adjusted two-step least squares estimators (B2SLS) for an equation and its reverse equation, where the left-side variable and the endogenous right-side variable are interchanged.

^cThe Donald–Newey instrument selection criterion (Donald and Newey 2001) is the expected average mean-squared error of the 2SLS estimator. BIC and HQIC are the Bayesian and Hannan–Quinn information criteria (Andrews 1999) for the choice of instruments.

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

estimated that excludes maleco and uses only (original) settler mortality and the measures of Western European influence as instrumental variables (table 3, column 3). The results indicate that this specification creates a weak-instrument problem according to the Cragg–Donald statistic. But the point estimates remain statistically significant¹⁶ and quantitatively similar, compared with the baseline specification. This suggests that finding a significant influence of malaria prevalence on economic development does not hinge on the use of maleco as an instrument. Moreover, when maleco is included as an additional instrumental variable, its validity is not rejected by a Hansen difference test,¹⁷ which provides statistical evidence against the endogeneity of maleco as presumed by Rodrik, Subramanian, and Trebbi (2004). Basically, the same findings emerge if the revised settler mortality variable (*lnmort2*) is used (table 3, column 4).¹⁸

Sample Size

As a final robustness check, results for a larger sample of countries are reported in appendix (table A.1). First, the baseline specification is re-estimated (column 1). Compared with the baseline sample of 45 countries (table 1, column 1), the point estimates and interval estimates are by and large the same. Thus, independent of the sample of countries, the estimated coefficients on both institutional quality and malaria prevalence appear economically important and statistically significant. Also as before, the Cragg–Donald statistic does not signal a weak-instrument problem for this specification. In contrast to the previous estimates, there is a reduced fit of the first-stage regressions that may reflect the presumed weak data quality for some of the countries in the larger sample, especially for the governance index (rule).

Similar conclusions can be drawn from the Acemoglu, Johnson, and Robinson (2001) specification in column 2 of table A.1, where institutional quality is measured by expropriation risk (*exprop*). Compared with the estimates for the smaller sample of countries (table 1, column 4), the point estimates change slightly. The most important result, the statistical insignificance of the coefficient on malaria prevalence, also shows up in the large sample. Again, this can be traced to a severe weak-instrument problem, which affects mainly the first-stage equation for *exprop*.

Using additional instrumental variables also does not change any previous insights for the larger sample of countries. In column 3 of table A.1, results are presented for the baseline specification augmented by all available instruments. The point and interval estimates change only slightly compared with the previous estimates, both

16. The 95 percent CLR confidence interval for institutional quality includes zero but is a borderline case. For example, a 93 percent CLR confidence interval does not include zero.

17. Detailed results are available on request.

18. A potential problem with the specification reported in table 3, column 4, is signaled by the Hansen statistic that is essentially zero. This could indicate that the instruments are highly correlated with each other or that there are too many instruments. However, the correlation of the instruments is far below one, and each instrument is significant in at least one first-stage regression. In fact, *lnmort2* and *engfrac* are highly significant in the regression for rule, and *lnmort2* and *eurfrac* are highly significant in the regression for *malfal*.

institutional quality and malaria prevalence remain statistically significant, and the overidentifying restrictions are not rejected. Finally, using the revised settler mortality variable (*lnmort2*) and excluding malaria ecology (*maleco*) from the set of instrumental variables again leads to almost identical results in the large sample (table A.1, column 4) compared with the smaller sample (table 3, column 4).

The conclusion from the robustness checks is that the results of the baseline specification are not sensitive to changes in the explanatory variables, the set of instruments, or the sample of countries.¹⁹ Therefore, the hypothesis derived from the baseline specification is maintained. Both institutions and malaria prevalence appear economically important determinants of the level of development.

V. CONCLUSION

The reported empirical results suggest that the Sachs hypothesis of direct income effects of malaria prevalence cannot be dismissed as easily as claimed in recent studies by Acemoglu, Johnson, and Robinson (2001) and Rodrik, Subramanian, and Trebbi (2004). Different from Acemoglu, Johnson, and Robinson (2001), statistically significant effects on income of malaria prevalence are estimated once its potential endogeneity is controlled for, and there appears to be no empirical evidence that the instrumental variable used by Sachs (2003) is invalid, as presumed by Rodrik, Subramanian, and Trebbi (2004). For given effects of institutional quality, the estimated direct negative income effects of malaria prevalence are quantitatively important. This result appears to be robust to using alternative measures of institutions and malaria prevalence, alternative and additional instrumental variables, and an alternative sample of countries.

Taken at face value, the results imply that institutions do not dominate all other potential determinants of development. An emphasis on good governance, even if it can be successfully implemented in poor countries, will probably not suffice to achieve improved economic performance. As argued by Sachs and his coauthors in various papers, subsidized research on tropical diseases and direct assistance from foreign donors for interventions against diseases may be needed to advance the development of poor countries, which otherwise may not escape the restrictions imposed on them by adverse geographic endowments. All this is certainly not to deny that good institutions would make such interventions possible in the first place or at least would make them more productive, but the findings of this article point out that good institutions alone are not necessarily a sufficient recipe for successful economic development.

19. Regarding variations in the sample size, the inclusion of sub-Saharan African countries is key to the presented identification strategy and to the validity of the instrumental variables, as it is in Acemoglu, Johnson, and Robinson (2001). The reason is that the variance across the sample stems mainly from the difference between the sub-Saharan African countries and the other countries. The reported direct effects of malaria on income may thus explain why some previous cross-country studies reported a negative coefficient of the dummy variable for sub-Saharan Africa (Collier and Gunning 1999).

APPENDIX. Definitions and Sources of Variables

| Variable | Definition | Source |
|-----------------|---|---|
| <i>coast</i> | Proportion of land area within 100 kilometers of the sea coast | Gallup, Sachs, Mellinger (1999), here taken from McArthur and Sachs (2001) |
| <i>engfrac</i> | Proportion of the population speaking English | Hall and Jones (1999) |
| <i>eurfrac</i> | Proportion of the population speaking one of the major languages of Western Europe: English, French, German, Portugese, or Spanish | Hall and Jones (1999) |
| <i>exprop</i> | Index of protection against expropriation in 1985–95; limited to 64 countries but includes the Bahamas and Vietnam, which are not included in <i>socinf</i> ; measured on a scale of 1 to 10 | Acemoglu, Johnson, and Robinson (2001), p. 1398 |
| <i>frost</i> | Proportion of a country's land receiving 5 or more frost days in that country's winter, defined as December through February in the Northern hemisphere and June through August in the Southern hemisphere; measured on a scale of 0 to 1 | Masters and McMillan (2001) |
| <i>gadp</i> | Index of government antidiversion policies; calculated as an unweighted average of five variables: law and order, bureaucratic quality, corruption, risk of expropriation, and government repudiation of contracts; measured on a scale of 0 to 1 | Hall and Jones (1999) |
| <i>humid</i> | Highest temperature during the month when average afternoon humidity is at its highest; measured in degrees Celsius | Parker (1997) |
| <i>latitude</i> | Distance from the equator as measured by the absolute value of country-specific latitude in degrees | Hall and Jones (1999) |
| <i>lngdpc</i> | Real GDP per capita, adjusted for purchasing power parity (PPP), 1995; measured in international dollars | World Bank, Development Indicators CD-ROM (2002) |
| <i>lnmort</i> | Settler mortality rates in colonies in the early nineteenth century, fourth mortality estimate (72 countries, excluding France and the United Kingdom); measured as death rate among 1,000 settlers, where each dead settler is replaced with a new settler | Acemoglu, Johnson, Robinson (2001, p. 1398), and Acemoglu, Johnson, Robinson (2000) |
| <i>lnmort2</i> | Revised estimate of <i>lnmort</i> | Albouy (2004) |

(Continued)

APPENDIX. Continued

| Variable | Definition | Source |
|---------------|---|--|
| maleco | Combines climatic factors and biological properties of the regionally dominant malaria vector into an index of the stability of malaria transmission. which is called malaria ecology; the index of malaria ecology is measured on a highly disaggregated subnational level and then averaged for the entire country and weighted by population; the index ranges from 0 to 31.5 (Burkina Faso); for details see text; dataset as of October 27, 2003 | Kiszewski and others (2004), here taken from http://www.earth.columbia.edu/about/director/malaria/index.html#datasets |
| <i>malfal</i> | Proportion of a country's population at risk of falciparum malaria transmission in 1994; measured on a scale of 0 to 1. Revised dataset as of October 27, 2003 | Sachs (2003), here taken from http://www.earth.columbia.edu/about/director/malaria/index.html#datasets |
| malrisk | Proportion of each country's population that lives with the risk of malaria transmission, involving three largely nonfatal species of the malaria pathogen (<i>Plasmodium vivax</i> , <i>Plasmodium malariae</i> , and <i>Plasmodium ovale</i>); measured on a scale of 0 to 1; dataset as of October 27, 2003 | Sachs (2003), here taken from http://www.earth.columbia.edu/about/director/malaria/index.html#datasets |
| rule | Average governance indicator based on six aggregated survey measures: voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption | Kaufmann Kraay, and Mastruzzi (2004) |
| trade | Natural log of the Frankel–Romer predicted trade share, based on measures of population size and geography | Hall and Jones (1999) |

TABLE A-1. Results for a Larger Sample of Countries

| Explanatory variables | Instruments | | | | | | | |
|---|---------------|-----------------|---------------|-----------------|-----------------|----------------|-----------------|--------------|
| | 1 | | 2 | | 3 | | 4 | |
| | <i>lnmort</i> | <i>+ maleco</i> | <i>lnmort</i> | <i>+ maleco</i> | All instruments | <i>lnmort2</i> | <i>+ Europe</i> | |
| | <i>rule</i> | <i>malfa</i> | <i>exprop</i> | <i>malfa</i> | <i>rule</i> | <i>malfa</i> | <i>rule</i> | <i>malfa</i> |
| Estimated coefficients | | | | | | | | |
| Two-step least square | 0.92 | -0.95 | 0.61 | -0.81 | 0.78 | -1.09 | 0.77 | 4.2 6 |
| Standard error | 0.24 | 0.31 | 0.19 | 0.40 | 0.15 | 0.24 | 0.23 | 0.3 2 |
| Fuller | 0.90 | -0.96 | 0.57 | -0.88 | 0.78 | -1.09 | 0.77 | 4.2 5 |
| Standard error | 0.24 | 0.30 | 0.17 | 0.36 | 0.15 | 0.24 | 0.23 | 0.3 1 |
| Bounds of 95 percent confidence intervals | | | | | | | | |
| Conventional | | | | | | | | |
| Upper | 0.43 | -1.57 | 0.23 | -1.61 | 0.8 | -1.56 | 0.30 | 4.9 0 |
| Lower | 1.40 | -0.33 | 0.99 | -0.01 | 1.8 | -0.62 | 1.24 | 0.6 2 |
| Conditional likelihood ratio | | | | | | | | |
| Upper | 0.40 | -1.62 | 0.31 | -1.49 | 0.3 | -1.75 | 0.07 | 2.2 2 |
| Lower | 1.64 | -0.11 | 2.25 | 2.25 | 1.2 | -0.45 | 1.52 | 0.3 5 |
| Number of observations | 64 | | 62 | | 61 | | | |
| Number of instruments | 2 | | 2 | | 3 | | | |
| Hansen test (OIR) | | | | | 9 | | 0.17 | |
| p-value | | | | | 0.66 | | 0.68 | |
| First-stage statistics | | | | | | | | |
| F-statistic | 19.51 | 58.65 | 11.17 | 56.76 | 15.53 | 23.33 | 12.44 | 25.93 |
| p-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Partial R ² | 0.39 | 0.66 | 0.28 | 0.66 | 0.73 | 0.80 | 0.40 | 0.58 |

(Continued)

TABLE A-1. Continued

| | Instruments | | | | | | | |
|-----------------------------------|-----------------|---------------|-----------------|---------------|-----------------|---------------|-------------------------------|---------------|
| | 1 | | 2 | | 3 | | 4 | |
| | Inmort + maleco | | Inmort + maleco | | All instruments | | Inmort2 + Europe ^a | |
| Explanatory variables | rule | <i>malfal</i> | <i>exprop</i> | <i>malfal</i> | rule | <i>malfal</i> | rule | <i>malfal</i> |
| Shea partial R ² | 0.21 | 0.35 | 0.10 | 0.25 | 0.56 | 0.61 | 0.24 | 0.35 |
| Weak-instrument test ^b | | | | | | | | |
| Cragg–Donald | 7.97 | | 3.36 | | 6.68 | | 5.99 | |
| Critical value | 7.03 | | 7.03 | | 27.51 | | 13.43 | |
| Hahn–Hausman test | | | | | 0.37 | -0.38 | 0.29 | -0.30 |
| p-value | | | | | 0.71 | 0.71 | 0.77 | 0.77 |

^aRefers to the instruments engfrac and eurfrac.

^bThe Cragg–Donald statistic (Cragg and Donald 1993) is used by Stock and Yogo (2002) for weak-instrument tests. If the Cragg–Donald statistic exceeds the critical value, then a standard significance test with nominal size of 5 percent has a maximal size of 10 percent. The weak-instrument test by Hahn and Hausman (2002) is based on the normalized difference between bias-adjusted two-step least squares estimators (B2SLS) for an equation and its reverse equation, where the left-side variable and the endogenous right-side variable are interchanged.

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

REFERENCES

- Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2000. "The Colonial Origins of Comparative Development: An Empirical Investigation." Working Paper 00-22. Massachusetts Institute of Technology, Department of Economics, Cambridge, Mass.
- . 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5):1369–401.
- . 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117(4):1231–94.
- Albouy, David. 2004. "The Colonial Origins of Comparative Development: A Reexamination Based on Improved Settler Mortality Data." Department of Economics, University of California, Berkeley.
- Andrews, Donald W. K. 1989. "Power in Econometric Applications." *Econometrica* 57(5):1059–90.
- . 1999. "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation." *Econometrica* 67(3):543–64.
- Bloom, David E., and Jeffrey D. Sachs. 1998. "Geography, Demography, and Economic Growth in Africa." *Brookings Papers on Economic Activity* 2:207–95.
- Collier, Paul, and Jan Willem Gunning. 1999. "Explaining African Economic Performance." *Journal of Economic Literature* 37(1):64–111.
- Cragg, John G., and Stephen G. Donald. 1993. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory* 9(2):222–40.
- Donald, Stephen G., and Whitney K. Newey. 2001. "Choosing the Number of Instruments." *Econometrica* 69(5):1161–91.
- Dufour, Jean-Marie. 1997. "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models." *Econometrica* 65(6):1365–87.
- Easterly, William, and Ross Levine. 2003. "Tropics, Germs, and Crops: How Endowments Influence Economic Development." *Journal of Monetary Economics* 50(1):3–39.
- Engerman, Stanley L., and Kenneth I. Sokoloff. 1997. "Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States." In Stephen Haber, ed., *How Latin America Fell Behind*. Stanford, Calif.: Stanford University Press.
- Frankel, Jeffrey A., and David Romer. 1999. "Does Trade Cause Growth?" *American Economic Review* 89(3):379–99.
- Fuller, Wayne A. 1977. "Some Properties of a Modification of the Limited Information Estimator." *Econometrica* 45(4):939–54.
- Gallup, John Luke, Jeffrey D. Sachs, and Andrew Mellinger. 1999. "Geography and Economic Development." *International Regional Science Review* 22(2):179–232.
- Gallup, John Luke, and Jeffrey D. Sachs. 2001. "The Economic Burden of Malaria." *The American Journal of Tropical Medicine & Hygiene* 64(1–2, Suppl.):85–96.
- Hahn, Jinyong, and Jerry Hausman. 2002. "A New Specification Test for the Validity of Instrumental Variables." *Econometrica* 70(1):163–89.
- Hahn, Jinyong, Jerry Hausman, and Guido Kuersteiner. 2004. "Estimation with Weak Instruments: Accuracy of Higher-Order Bias and $m\text{SE}$ Approximations." *Econometrics Journal* 7(1):272–306.
- Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114(1):83–116.
- Hausman, Jerry, James H. Stock, and Motohiro Yogo. 2005. "Asymptotic Properties of the Hahn-Hausman Test for Weak Instruments." *Economics Letters* 89(3):333–342.
- Hoff, Karla. 2003. "Paths of Institutional Development: A View from Economic History." *The World Bank Research Observer* 18(2):205–26.

- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2004. "Governance Matters III: Governance Indicators for 1996–2002." *World Bank Economic Review* 18(2):253–87.
- Kiszewski, Anthony, Andrew Mellinger, Pia Malaney, Andrew Spielman, Sonia Ehrlich Sachs, and Jeffrey D. Sachs. 2004. "A Global Index of the Stability of Malaria Transmission Based on the Intrinsic Properties of Anopheline Mosquito Vectors." *American Journal of Tropical Medicine and Hygiene* 70(5):486–98.
- Masters, William A., and Margaret S. McMillan. 2001. "Climate and Scale in Economic Growth." *Journal of Economic Growth* 6(3):167–86.
- McArthur, John W., and Jeffrey D. Sachs. 2001. *Institutions and Geography: A Comment on Acemoglu, Johnson, and Robinson (2000)*. NBER Working Paper 8114. National Bureau of Economic Research, Cambridge, Mass.
- Moreira, Marcelo J. 2003. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica* 71(4):1027–48.
- Parker, Philip M. 1997. *National Cultures of the World. A Statistical Reference*. Westport, Conn.: Greenwood Press.
- Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9(2):131–65.
- Sachs, Jeffrey D. 2001. *Tropical Underdevelopment*. NBER Working Paper 8119. National Bureau of Economic Research, Cambridge, Mass.
- . 2003. *Institutions Don't Rule: Direct Effects of Geography on Per Capita Income*. NBER Working Paper 9490. National Bureau of Economic Research, Cambridge, Mass.
- Sachs, Jeffrey D., and Pia Malaney. 2002. "The Economic and Social Burden of Malaria." *Nature* 415(6872):680–85.
- Shea, John. 1997. "Instrument Relevance in Multivariate Linear Models: A Simple Measure." *Review of Economics and Statistics* 79(2):348–52.
- Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables with Weak Instruments." *Econometrica* 65(3):557–86.
- Stock, James H., and Motohiro Yogo. 2002. *Testing for Weak Instruments in Linear IV Regressions*. NBER Technical Working Paper 284. National Bureau of Economic Research, Cambridge, Mass.
- Summers, Robert, and Alan Heston. 1991. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988." *Quarterly Journal of Economics* 106(2):327–68.
- World Bank. 2002. *World Development Indicators 2002*. CD-ROM. Washington, D.C.

When Is External Debt Sustainable?

Aart Kraay and Vikram Nehru

The article empirically examines the determinants of debt distress, defined as periods in which countries resort to any of three forms of exceptional finance: significant arrears on external debt, Paris Club rescheduling, and nonconcessional International Monetary Fund lending. Probit regressions show that three factors explain a substantial fraction of the cross-country and time-series variation in the incidence of debt distress: the debt burden, the quality of policies and institutions, and shocks. The relative importance of these factors varies with the level of development. These results are robust to a variety of alternative specifications, and the core specifications have substantial out-of-sample predictive power. The quantitative implications of these results are examined for the lending strategies of official creditors.

This article empirically analyzes the probability of debt distress in developing countries and examines the implications of the results for the lending policies of official creditors. It defines debt distress episodes as periods in which countries resort to any of three forms of exceptional finance: substantial arrears on their external debt, debt relief from the Paris Club of creditors, and nonconcessional balance of payments support from the International Monetary Fund (IMF). Three factors—the debt burden, the quality of institutions and policies, and shocks that affect real GDP growth—are found to be highly significant predictors of debt distress, and their relative importance differs between low- and middle-income countries.

Three features of this article distinguish it from much of the large empirical literature on debt sustainability. First, one of the main interests is in understanding the determinants of debt distress among low-income countries, which have been at the center of recent debt relief efforts such as the 1996 Heavily Indebted Poor Countries (HIPC) Initiative and the 2006 Multilateral Debt Relief Initiative. In contrast, much of the existing empirical literature focuses on debt crises in middle-income countries that borrow primarily from private creditors. As shown below, the features of distress episodes and their determinants are quite different in the two groups of countries.

Aart Kraay is lead economist in the Development Research Group at the World Bank; his email address is akraay@worldbank.org. Vikram Nehru is director of the Economic Policy and Debt Department at the World Bank; his email address is vnehr@worldbank.org. The authors thank Nancy Birdsall, Christina Daseking, Gershon Feder, Alan Gelb, Indermit Gill, Rex Ghosh, Nicholas Hope, and Sona Varma for helpful comments; Carmen Reinhart for kindly sharing historical data on default episodes; and Sunyoung Lee for superb research assistance.

Second, this analysis finds that nonfinancial variables, especially the quality of policies and institutions, are key determinants of debt distress in low-income countries. The idea that policies and institutions matter for debt sustainability is not novel. But it has received relatively little attention in the empirical literature. A notable exception is Reinhart, Rogoff, and Savastano (2003), who document the importance of a country's history of nonrepayment and macroeconomic instability in driving market perceptions of the likelihood of default. The analysis here complements theirs by showing that not only does the history of nonrepayment and weak policy matter for the likelihood of debt distress but so do contemporaneous policies and institutions. Moreover, this article finds that the contemporaneous effect of improvements in policies and institutions on the probability of debt distress is quantitatively large and roughly of the same order of magnitude as reductions in debt burdens. It also finds that the role of policies and institutions is much more important in low-income countries than in middle-income countries.

Third, the article emphasizes the implications of the findings for the lending strategies of multilateral concessional creditors such as the World Bank and the IMF. In these organizations notions of debt sustainability have until recently focused almost exclusively on simple projections of debt burden indicators and their comparison with fairly arbitrary benchmarks. For example, debt relief under the HIPC Initiative was calibrated to ensure that countries emerge from the process with a present value of debt to exports of 150 percent, irrespective of other country characteristics. The results here indicate that a common single debt sustainability threshold is not appropriate because it fails to recognize the role of institutions and policies that matter for the likelihood of debt distress. In particular, the results permit summarizing striking tradeoffs between debt indicators, policies, and shocks for a given probability of debt distress. For example, the benchmark results suggest that countries at the 75th percentile of the measure of policies and institutions can have a present value of debt to exports that is two to three times higher than countries at the 25th percentile of this indicator, without increasing the probability of debt distress. These tradeoffs suggest that a country's targeted level of "sustainable" debt should vary substantially with the quality of its policies and institutions.

This work is premised on the view that avoiding debt distress is desirable. There are several reasons for this. Resolving debt distress imposes direct costs in terms of the time that debtors and creditors must spend coordinating and renegotiating claims. Excessive debt can also undercut support for policy reforms by political and civil society groups in debtor countries if they perceive that benefits from reforms will be directed to high debt service rather than to needed public services for the poor. The pressure to meet external debt service payments may also tempt debtor country governments to seek short-term solutions at the expense of fundamental, longer-term reforms. Creditors, as well, may be tempted to allocate resources according to resource needs rather than policy performance.¹ Finally,

1. For example, Birdsall, Claessens, and Diwan (2003) argue that the correlation between aid and policy performance is weak in highly indebted countries in sub-Saharan Africa.

nonrepayment of loans to multilateral lenders can have perverse distributional effects among borrowing countries. Absent new resources from donors, the failure to repay concessional loans reduces the ability of multilateral creditors to provide new loans to other developing countries. Moreover, to the extent that new lending is intended for countries with sound policies and institutions, whereas countries with poor policies and institutions are more likely to fail to service their past debts, the result can be a transfer of resources from countries with good policies to countries with bad policies.²

This article is obviously not the first to empirically investigate the determinants of debt servicing difficulties. The debt crisis of the early 1980s prompted a surge of empirical work. An early contribution is by McFadden and others (1985), who construct an indicator of debt servicing difficulties based on arrears, rescheduling, and IMF support, much like the one used here, for 93 countries over the period 1971–1982. They find that the debt burden, the level of per capita income, real GDP growth, and liquidity measures such as nongold reserves are significant predictors of debt distress, whereas real exchange rate changes are not. They also investigate the importance of state dependence and country effects and conclude that both matter, whereas the updated sample used here does not find comparable evidence of state dependence. Other studies in the early literature include that of Cline (1984), who focuses primarily on financial variables such as the determinants of debt servicing difficulties, and of Berg and Sachs (1988), who emphasize "deep" structural factors such as income inequality (which they argue proxies for political pressures for excessive borrowing) and a lack of trade openness as determinants of debt servicing difficulties among middle-income countries. In addition, Lloyd-Ellis, McKenzie, and Thomas (1990) model both the probability of debt reschedulings and their magnitude, again emphasizing financial variables. Interestingly, none of these studies focuses on direct measures of the quality of policies and institutions, as this one does.³

2. The amounts at stake are nontrivial. Consider the World Bank–administered International Development Association (IDA), which provides substantial resources to the world's poorest countries. As of 2003, IDA's portfolio consisted of highly concessional loans with a face value of roughly \$110 billion. During fiscal 2003 it disbursed \$7 billion in new loans, of which only \$1.4 billion was financed by repayments on existing loans and with most of the balance coming from infusions from rich countries. Given the long grace periods in IDA lending, this flow of repayments is anticipated to increase sharply in the future, averaging \$2.3 billion a year over 2003–2008, \$3.3 billion a year over the next five years, and \$4.2 billion in the five years after that (World Bank 2003). Holding constant future donor contributions to IDA, it is clear that any disruption in this flow of future repayment resulting from episodes of debt distress will have significant implications for IDA's ability to provide new lending to the poorest countries.

3. Another strand of this early literature tried to find a discontinuity in the relationship between debt burden indicators (usually the external debt to export ratio) and the incidence of default or market-based indicators of risk (such as the premium over benchmark interest rates on debt securities traded in the secondary market), for example, Underwood (1991) and Cohen (1996). These studies found that above a threshold range of about 200–250 percent of the present value of the debt to export ratio, the likelihood of debt default climbed rapidly. This range then became the benchmark adopted by the original HIPC Initiative in 1996 but was subsequently lowered in 1999 under the Enhanced HIPC framework.

Several more recent studies are also related to this work. Aylward and Thorne (1998) empirically investigate countries' repayment performance to the IMF, emphasizing the importance of countries' repayment histories and IMF-specific financial variables in predicting the likelihood of arrears. McKenzie (2004) studies the determinants of default on World Bank loans. Detragiache and Spilimbergo (2001) study the importance of liquidity factors such as short-term debt, debt service, and the level of international reserves in predicting debt crises.

Reinhart, Rogoff, and Savastano (2003) study the historical determinants of "debt intolerance," a term used to describe the extreme duress that many emerging market economies experience at debt levels that seem quite manageable by industrial country standards. Their most relevant finding to this work is that the *Institutional Investor* magazine's sovereign risk ratings can be explained by a very small number of variables measuring a country's repayment history, its external debt burden, and its history of macroeconomic stability. However, there are three key differences between their study and this article. First, their dependent variable, the *Institutional Investor* rating, measures perceptions of the probability of debt distress, whereas this article attempts to explain the incidence of actual episodes of debt distress.⁴ Second, their sample consists mostly of middle- and high-income countries, in contrast with the focus here on low-income countries. Third, as shown in more detail below, this article finds that contemporaneous policy matters for the incidence of debt distress, whereas a history of bad policy and nonrepayment, a key determinant in their study, matters less.

Finally, the analysis by Manasse, Roubini, and Schimmelpfennig (2003) is most closely related to that in this article. They consider a country to be in debt crisis if it is classified as being in default by Standard & Poor's or if it has access to nonconcessional IMF financing in excess of 100 percent of quota. They use logit and binary recursive tree analysis to identify macroeconomic variables reflecting solvency and liquidity factors that predict a debt crisis episode one year in advance. Once again, the key difference with the analysis in this article is that they restrict their analysis to a sample of emerging market developing countries for which such data are available (especially the Standard & Poor's data), whereas a special focus of this article is the factors affecting debt distress in low-income countries. Several of their key results, however, are broadly consistent with those found here. They find that debt burden indicators and GDP growth, as well as a somewhat different set of measures of policies and institutions, significantly influence the likelihood of debt crises.

The following section describes in detail the methodology for identifying debt distress episodes. Section II documents the relative importance of debt burdens, a measure of policies and institutions, and shocks in driving debt distress. It shows

4. As documented in Reinhart, Rogoff, and Savastano (2003), country risk ratings are only imperfect predictors of actual default episodes.

that these three variables have substantial out-of-sample forecasting power for debt distress events and that the results survive several robustness checks. Section III presents some policy implications of the results.

I. EMPIRICAL FRAMEWORK

The sample consists of all 132 low- and middle-income countries that report debt data in the World Bank's *Global Development Finance* publication and covers all years during 1970–2002 for which the necessary data are available. The *Appendix* describes the data sources.

Identifying Debt Distress Episodes

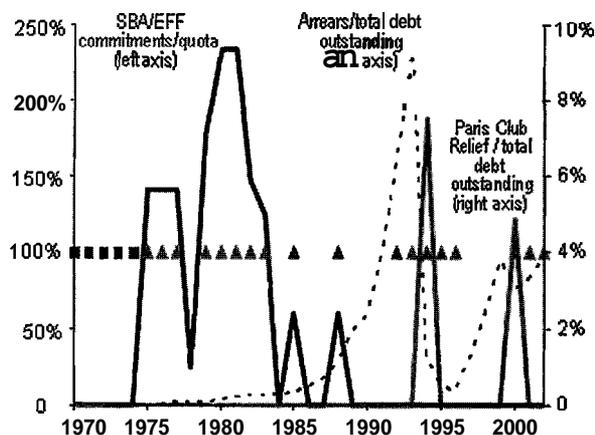
Episodes of debt distress are defined as periods in which any one or more of the following three conditions hold: the sum of interest and principal arrears is large relative to the stock of debt outstanding, a country receives debt relief in the form of rescheduling or debt reduction from the Paris Club of bilateral creditors, or the country receives substantial balance of payments support from the IMF under its nonconcessional Standby Arrangements or Extended Fund Facilities. The first condition is the basic measure of debt distress: the failure to service external obligations resulting in an accumulation of arrears. But countries that are unable to service their external debt need not fall into arrears; they can also seek debt relief from the Paris Club or balance of payments support from the IMF.⁵ This is why the arrears criterion for debt distress is complemented by the other two criteria. As a control group, nondistress episodes, or "normal times," are defined as nonoverlapping periods of five consecutive years in which none of the three indicators of debt distress is observed.⁶

To implement the rule for identifying debt distress episodes, thresholds for "large" values of arrears and "substantial" levels of IMF support have to be identified. The threshold for arrears is 5 percent of total debt outstanding. For IMF programs only those for which commitments are greater than 50 percent of the country's IMF quota are considered. Although any threshold for defining debt distress episodes is somewhat arbitrary, these values are quite high relative to the experience of the typical developing country. The threshold for arrears, for example, is roughly 10 times greater than the median value of arrears as a fraction of debt outstanding (0.4 percent) when all country-years since 1970 are pooled. Similarly, the median value of IMF commitments relative to quota is zero for pooled country-year observations, reflecting the fact that less than half the country-years in the sample correspond to an IMF program including access to nonconcessional IMF facilities. When such programs are in place, the median

5. This article does not define debt reductions under the HIPC Initiative as a separate indicator of debt distress because all debt relief under the Initiative requires parallel debt reduction by the Paris Club.

6. These episodes begin in the first year for which it is possible to find five consecutive years with no distress.

FIGURE 1. Identifying Debt-Distress Events: Example of Kenya



Source: Authors' analysis of data described in Appendix.

commitment is 52 percent of quota. This means that the threshold identifies only the top half (in terms of commitments relative to quota) of all Standby Arrangement and Extended Fund Facility programs.⁷ Finally, for Paris Club agreements the year of the agreement and the two subsequent years are identified as distress episodes because most Paris Club agreements provide relief for debt service payments falling due during a fairly short period, typically three years.

Figure 1 illustrates how normal and debt distress episodes are identified for the case of Kenya. It shows commitments under Standby Arrangements or Extended Fund Facilities (SBA/EFF) (solid black line), arrears (dashed line), and Paris Club relief (gray line). During the 1970s and 1980s Kenya received balance of payments support in excess of 50 percent of its quota for a total of 10 years, whereas during the 1990s it had four years in which arrears were more than 5 percent of debt outstanding. Finally, it also received substantial Paris Club relief in 1994 and again in 2000. This means that in total, between 1970 and 2000, Kenya experienced 17 years of debt distress (indicated by triangles). In contrast, it managed only one five-year period of normal times, beginning in 1970, in which there were no arrears, debt relief, or IMF support (indicated by squares).

In Kenya and in many other countries, episodes of debt distress are often quite short and are also often immediately preceded by other distress episodes. To be sure that episodes of prolonged debt distress rather than sporadic fluctuations in the distress indicator are being identified, all short distress episodes of less than

7. Access to the Poverty Reduction and Growth Facility of the IMF is not included as a debt distress indicator, because in many cases financing from this facility is no longer to meet temporary payments imbalances but has become a source of long-term development finance. See IMF (2002).

TABLE 1. Distress Episodes

| | | | | | |
|--------------------|-----------|------------------|-----------|---------------------|-----------|
| Albania | 1992–2002 | Ecuador | 1983–1996 | Malawi | 1979–1985 |
| Algeria | 1994–1997 | Ecuador | 2000–2002 | Niger | 1983–1990 |
| Argentina | 1983–1995 | Egypt, Arab Rep. | 1984–1995 | Nigeria | 1986–2002 |
| Bangladesh | 1979–1981 | Ethiopia | 1991–2002 | Nicaragua | 1983–2002 |
| Benin | 1983–1998 | Ghana | 1996–1998 | Pakistan | 1980–1983 |
| Brazil | 1983–1985 | Guinea-Bissau | 1981–2002 | Pakistan | 1995–2002 |
| Brazil | 1998–2002 | Guyana | 1978–2002 | Paraguay | 1986–1994 |
| Bulgaria | 1991–2000 | Honduras | 1979–2001 | Rwanda | 1994–2002 |
| Burkina Faso | 1987–1998 | Haiti | 1978–1980 | Senegal | 1980–2002 |
| Burundi | 1998–2002 | Indonesia | 1997–2002 | El Salvador | 1990–1992 |
| Cameroon | 1987–2002 | India | 1981–1983 | Somalia | 1981–2002 |
| Chile | 1983–1989 | Jordan | 1989–2002 | Seychelles | 1990–2002 |
| Côte d'Ivoire | 1981–1996 | Kenya | 1992–1996 | Thailand | 1997–1999 |
| Congo, Rep. | 1985–2002 | Kenya | 2000–2002 | Trinidad and Tobago | 1988–1992 |
| Colombia | 1999–2001 | Cambodia | 1989–2002 | Tunisia | 1986–1991 |
| Comoros | 1987–2002 | Liberia | 1980–2002 | Turkey | 1978–1984 |
| Cape Verde | 1988–2002 | Morocco | 1980–1994 | Turkey | 1999–2002 |
| Costa Rica | 1980–1995 | Madagascar | 1980–2002 | Uruguay | 1983–1986 |
| Dominican Republic | 1983–1999 | Mexico | 1983–1992 | Vietnam | 1988–2002 |
| | | | | Zimbabwe | 2000–2002 |

Source: Authors' analysis of data described in *Appendix*.

three-year duration are eliminated. So are all distress episodes preceded by periods of distress in any of the three previous years to ensure that distinct episodes of distress are being identified, as opposed to continuations of previous episodes. This procedure identifies a total of 100 episodes of debt distress and 309 episodes of normal times over 1970–2002.⁸ For Kenya this results in two distress episodes, 1992–1996 and 2000–2002. The regression analysis that follows works with a subset of 58 distress episodes and 142 normal time episodes over 1978–2002 for which data on the core explanatory variables are available. The key constraint here is the preferred measure of policy used in the analysis, the World Bank's Country Policy and Institutional Assessment (CPIA) ratings, which began in 1978.

These 58 distress episodes are listed in table 1, and the means of key variables in distress and normal time periods are listed in table 2. The list contains many familiar episodes, including those in many Latin-American countries during the

8. The criteria for defining episodes imply that not all country-year observations will belong to either a distress episode or a nondistress episode. Of the 3,553 country-year observations for which indicators of distress are available, at least one indicator of distress is observed in 1,540 of them, with the remainder corresponding to nondistress years. After short distress episodes, distress episodes preceded by other episodes, and nondistress episodes shorter than five years are discarded, 2,630 country-years remain, of which 1,085 correspond to distress episodes. The regression sample is smaller still because of limits on the availability of explanatory variables and covers 1,339 country-years, of which 629 are classified as distress.

TABLE 2. Means of Key Variables in Normal Times and Distress Events

| | All Observations | | Low-Income Countries | | Middle-Income Countries | |
|--|------------------|----------|----------------------|----------|-------------------------|----------|
| | Normal | Distress | Normal | Distress | Normal | Distress |
| Average length of episode (years) | 5.0 | 10.8 | 5.0 | 12.6 | 5.0 | 8.7 |
| Average during episode of | | | | | | |
| Arrears/debt | 0.006 | 0.094 | 0.005 | 0.133 | 0.006 | 0.046 |
| Paris Club relief/debt | 0.000 | 0.017 | 0.000 | 0.018 | 0.000 | 0.016 |
| IMF lending ^a /quota | 0.031 | 0.988 | 0.036 | 0.523 | 0.027 | 1.559 |
| Net transfers/GDP | 0.014 | 0.005 | 0.031 | 0.021 | 0.005 | -0.014 |
| Value before episode of | | | | | | |
| Present value of debt/exports | 0.818 | 1.724 | 1.089 | 1.956 | 0.666 | 1.440 |
| Country Policy and Institutional Assessment (CPIA) | 3.789 | 3.084 | 3.489 | 2.849 | 3.957 | 3.374 |
| Growth | 0.047 | 0.012 | 0.043 | 0.028 | 0.050 | -0.007 |

^aInternational Monetary Fund Standby Arrangement or Extended Fund Facility.

Source: Authors' analysis of data described in Appendix.

debt crisis of the 1980s and Thailand and Indonesia during the more recent East Asian financial crisis. There are also many lengthy episodes of debt distress in sub-Saharan Africa.⁹ A striking feature of the debt distress episodes is their length. In the regression sample the mean length of a distress episode is 10.8 years. The longest distress episode is that in the Central African Republic, which has been continuously in debt distress during the whole sample period, primarily because of high arrears. There are also very sharp differences in the values of the debt distress indicators between distress episodes and normal times. In distress episodes average arrears are 9.4 percent of debt outstanding, whereas in normal times average arrears are 0.5 percent. During distress episodes Standby Arrangement or Extended Fund Facility support averages 98 percent of quota, whereas during normal times it is 3 percent. Although by construction Paris Club relief is zero in normal times, it averages 1.7 percent of debt outstanding during distress episodes.

There are also interesting differences between low- and middle-income countries. Distress episodes tend to be longer in low-income countries (12.6 years) than in middle-income countries (8.7 years) and associated with higher levels of arrears (13.3 and 4.6 percent of debt outstanding). Net transfers on debt fall during distress episodes, but proportionately much less in low-income countries, where they decline from 3.1 to 2.1 percent of GDP on average, than in middle-income

9. One anomalous observation is Vietnam, which is identified as being in continuous debt distress since the late 1980s. This reflects continuous high levels of arrears relative to nonbilateral, non-Paris Club creditors, much of it ruble-denominated. In the vast majority of episodes of debt distress based on arrears, the arrears are primarily to multilateral and bilateral Paris Club creditors.

countries, where they decline from 0.5 to -1.4 percent of GDP. This highlights a key feature of distress episodes in low-income countries—despite experiencing severe debt servicing difficulties, these countries on average continue to benefit from positive, and only somewhat reduced, net transfers on debt.

Modeling the Probability of Debt Distress

The following probit specification is used to model the probability of debt distress:"

$$(1) \quad P[y_{ct} = 1] = \Phi(\beta'X_{ct})$$

where y is an indicator value taking a value of one for debt distress episodes and zero for normal time episodes, each beginning in country c at time t ; $\Phi(\cdot)$ denotes the normal distribution function; X_{ct} denotes a vector of determinants of debt distress; and β is a vector of parameters to be estimated. The sample consists of unbalanced and irregularly spaced observations of distress and normal times. The core specification considers a very parsimonious set of potential determinants of debt distress. As a first step to alleviating concerns about potential endogeneity biases, each variable is measured in the year prior to the beginning of the episode.¹¹

The core specifications consider three explanatory variables. The first is the present value of debt (the present value of future debt service obligations), expressed as a share of current exports.¹² This is a useful summary of the overall debt burden of a country and reflects cross-country differences in the concessionality of debt. The second is the World Bank's CPIA ratings, which is used as the preferred measure of the policy environment. Available annually since 1978, the CPIA ratings reflect the perceptions of World Bank country economists. The third variable is real GDP growth, included as a crude way of capturing the various shocks, both exogenous and endogenous, that countries experience. There are substantial differences in the means of these variables before distress episodes and normal time episodes (see table 2). The present value of debt as a

10. Because the interest here is primarily the incidence of distress episodes rather than their precise timing, this simple probit specification is adequate. Collins (2003) shows how the timing of currency crises can be modeled explicitly as the first-passage time of a latent variable to a threshold, of which the simple probit specification here is a special case. Manasse, Roubini, and Schimmelpfening (2003) suggest that binary recursive tree analysis better captures the nonlinearities in the relationship between debt crises and their determinants, in a sample of middle-income countries. We have not yet investigated whether similar nonlinearities are important in our sample.

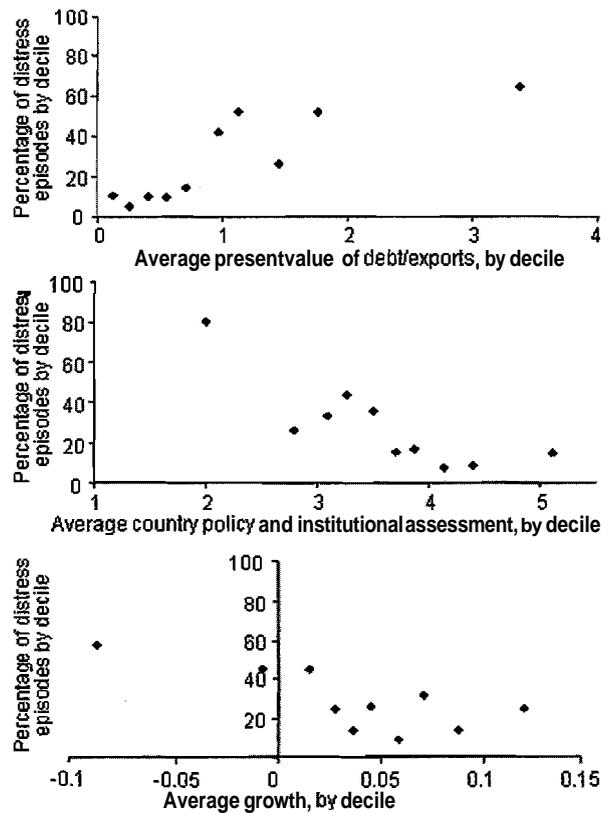
11. For example, one might expect debt burdens to increase and policy performance to deteriorate during distress episodes. This would create a spurious correlation between these variables measured during the episode and the value of the outcome variable.

12. In the working paper version of this article (Kraay and Nehru 2004), several other debt burden indicators are also considered, including total debt service as a share of exports, the face value of debt relative to exports, debt service relative to current government revenues, and debt service relative to nongold reserves. Results for these measures were qualitatively similar, with the flow debt service measures providing slightly greater predictive power for debt distress.

share of exports is more than twice as high before distress episodes (1.7) than before normal time episodes (0.8), policy is substantially worse (CPIA score of 3.1 and 3.8), and growth is considerably lower (1.2 and 4.7 percent).

Figure 2 illustrates the strong bivariate relationships between the core explanatory variables and the distress indicator. In each panel the sample of observations is divided by deciles of the explanatory variable of interest. The mean value of the explanatory variable is then computed by deciles and plotted against the mean of the distress indicator variable by decile. Thus, for example, in the first panel the mean value of the present value of debt to exports in the top decile of this variable is 3.4, and 65 percent of the observations in this decile correspond to distress. In contrast, in the bottom decile the mean value of debt relative to exports is 13 percent, and only 11 percent of the observations in this decile correspond to distress. A key feature of the data is the strong relationship between debt distress and policy performance. In the lowest decile of policy

FIGURE 2. Correlates of Debt Distress



Source: Authors' analysis of data described in Appendix.

performance, fully 80 percent of observations correspond to distress, whereas in the top three deciles of policy performance the likelihood of debt distress is about 10 percent.

II. EMPIRICAL RESULTS

This section confirms the importance of debt burdens, policies and institutions, and shocks in driving debt distress and shows the substantial out-of-sample forecasting power for debt distress events of these three variables. It also finds the results to be robust to several checks.

Results from Cove Specification

Table 3 reports the core specifications. The observations for all countries show that debt burdens, policies, and shocks as proxied by real per capita GDP growth are all highly significant predictors of debt distress. Countries with high debt burdens, low CPIA scores, and low growth in a given year are significantly more likely to experience a debt distress episode beginning in the next year. The magnitude of the effects of debt and policy is economically significant as well. Moving from the 25th percentile of indebtedness to the 75th percentile raises the probability of distress from 15 to 35 percent (holding constant the other

TABLE 3. Basic Results

| | (1) | (2) | (3) |
|---|-------------------|-----------------------------------|--------------------------------------|
| | All | Low-Income Countries ^a | Middle-Income Countries ^a |
| Present value debt/exports | 0.644 (0.152)*** | 0.143 (0.074)* | 0.262 (0.060)*** |
| Country Policy and Institutional Assessment (CPIA) | -0.557 (0.142)*** | -0.311 (0.091)*** | -0.020 (0.051) |
| Real GDP growth | -4.620 (2.085)** | -0.930 (1.199) | -2.080 (0.749)*** |
| Constant | 0.821 (0.512) | 1.911 (0.789)** | -1.375 (0.925) |
| Number of observations | 200 | 83 | 117 |
| Out-of-sample predictive power (fraction of events correctly predicted) | | | |
| All episodes | 0.71 | 0.75 | 0.78 |
| Distress episodes | 0.74 | 0.56 | 0.70 |
| Normal time episodes | 0.70 | 0.83 | 0.80 |

^aSignificant at the 10 percent level.

***Significant at the 5 percent level.

**Significant at the 1 percent level.

Note: Numbers in parentheses are standard errors.

"Marginal effects rather than slope coefficients are reported for first three variables to facilitate a comparison of the magnitude of estimated effects between these two columns.

Source: Authors' analysis of data described in *Appendix*.

variables at the median). Similarly, moving from the 25th percentile of policy to the 75th percentile lowers the probability of distress from 27 to 12 percent. The effect of growth, although significant, is not as large. Raising growth from the first to the third quartile lowers the probability of distress from 24 to 17 percent.

When the core specification is reestimated separately for low- and middle-income countries, higher debt burdens lead to significantly higher likelihood of debt distress in both groups of countries (see table 3). The magnitude of this effect is different in the two groups, however. To facilitate comparisons between these two groups, estimated marginal effects (the derivative of the probability of distress with respect to the variable of interest) are reported rather than the slope coefficients, β . This marginal effect is nearly twice as large for middle-income countries as for low-income countries. In contrast, the marginal effect of policy is much larger among low-income countries than among middle-income countries, and in the second group it is not significantly different from zero. The effect of shocks, as proxied by real GDP growth, is much larger for middle-income countries than for low-income countries and is insignificant for low-income countries. The intercepts from the probit regressions is much larger for low-income countries than for middle-income countries, suggesting that there are factors other than debt, policy, and growth that contribute to a higher rate of debt distress in low-income countries. All of these differences between the two groups are statistically significant at the 5 percent level, with the exception of the effect of growth, where the difference falls just short of significance at the 10 percent level.

As the ultimate interest is in predicting debt distress episodes based on a parsimonious set of variables, it is useful to also examine the out-of-sample predictive power of each of these first three specifications. To do this, each regression is reestimated using data through 1989. The estimated coefficients are then used together with the observed right-side variables to predict the outcome of each of the observations for the 1990s. In particular, a debt distress episode is predicted to occur if the predicted probability conditional on the observed data included in each regression is greater than the unconditional probability of distress in the pre-1990 sample. This unconditional probability is 0.38 in the full sample, 0.45 for the low-income country subsample, and 0.30 for the middle-income country subsamples. The predictive power of the forecasts is summarized by reporting the fractions of all observations after 1990 that are correctly predicted, as well as the success rate for distress episodes and normal time episodes separately.

The overall success rates are quite respectable, at 75 percent among low-income countries and 78 percent among middle-income countries. To put these success rates in perspective, note that using only the historical unconditional rate of debt distress to predict future debt distress would yield a success rate of 50 percent among low-income countries and 58 percent among middle-income

countries.¹³ The additional information in the three right-side variables therefore increases the success rate relative to this naive forecast by 20 to 25 percentage points. Note also that the success rate for predicting normal time episodes is even higher.

Overall, these results suggest that a quite parsimonious empirical model can do a reasonable job of accounting for patterns of debt distress over the past 25 years. Moreover, the out-of-sample forecasting power of the model is quite good. Before turning to the policy implications of this finding, this basic specification is first subjected to several robustness checks.

Robustness of Core Specification: Does the Type of Debt Matter?

To investigate the extent to which debt distress is affected by the type of external debt, external debt is distinguished along three dimensions. First is a variable measuring the share of external debt that is public and publicly guaranteed. Second is a variable measuring the share of external debt that is owed to official creditors, consisting of bilateral loans by governments as well as loans from multilateral organizations. Lastly, there is the concessionality of external debt, measured as one minus the ratio of the present value of debt to the nominal value of debt. Each of these variables is added in turn to the core specification for all countries and for low-income countries alone.¹⁴

All three characteristics of debt are significantly associated with the risk of debt distress (table 4). In particular, the greater the share of debt that is public or publicly guaranteed and the greater the share of debt owed to official creditors, the lower is the risk of debt distress. Also, the lower risk of debt distress, the greater the concessionality of debt. This last result may not be surprising because more concessional debt generally has lower immediate debt service obligations than less concessional debt. To the extent that debt distress is triggered by difficulties in meeting immediate debt service obligations, more concessional debt will be less likely to lead to debt distress. The finding that countries that owe more to official creditors are less likely to experience debt distress is more interesting. One interpretation is that official creditors are more likely to engage in "defensive lending," providing new loans to ensure that old loans are repaid. Another interpretation is that loans from official creditors tend to be more concessional and for reasons just given are therefore easier to service. One crude way to disentangle these two hypotheses is to put both the share owed to official creditors and the concessionality rate in the same regression. When

13. Suppose that the fraction of distress episodes observed in the past is p and that distress is randomly predicted for a fraction p of future events and no distress for the remaining fraction $1-p$. Then the success rate of such a forecast based only on the unconditional historical rate of distress would be $p^2 + (1-p)^2$. Because the historical rate of distress during the period before 1989 is $p = 0.5$ for low-income countries and $p = 0.3$ for middle-income countries, this yields the success rates given in the text.

14. Ideally, the partial effects of these three characteristics of debt would be estimated. Unfortunately, all three are quite strongly correlated at about 0.6 across observations. Given the small sample, multicollinearity problems prevent pinning down the partial effects precisely.

TABLE 4. Does the Type of Debt Matter?

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|------------------|------------------|------------------|-------------------------|-------------------------|-------------------------|
| | All | All | All | Low-Income Countries | Low-Income Countries | Low-Income Countries |
| Present value debt/exports | 1.062 (0.219)*** | 0.620 (0.167)** | 0.796 (0.186)*** | 0.917 (0.379)*** | 0.408 (0.225)* | 0.590 (0.272)*** |
| Country Policy and Institutional Assessment (CPIA) | -0.534 (0.166)** | -0.621 (0.164)** | -0.572 (0.193)** | -1.254 (0.427)** | -0.909 (0.281)** | -0.961 (0.363)*** |
| Real GDP growth | -5.166 (2.655)** | -3.590 (2.520) | -4.667 (2.588)* | -2.276 (4.474) | -3.557 (3.769) | -2.758 (3.859) |
| public or publicly guaranteed share of debt | -3.330 (0.774)** | | | -6.151 (2.015)** | | |
| Share of debt owed to official creditors | | -1.280 (0.466)** | | | -1.642 (0.963)** | |
| Concessional (1- present value/nominal debt) | | | -2.125 (0.833)** | | | -3.129 (1.279)*** |
| Constant | 2.627 (0.793)*** | 1.833 (0.741)*** | 1.064 (0.780) | 7.452 (2.383)** | 3.479 (1.290)*** | 3.080 (1.414)** |
| Number of observations | 167 | 167 | 154 | 64 | 64 | 62 |

*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.

Source: Authors' analysis of data described in *Appendix*.

this is done, concessionality is found to be significant, whereas the official creditor share is not (results not reported). This is suggestive—but hardly conclusive—evidence against the "defensive lending" hypothesis.

Policy Endogeneity and the Role of Shocks

A potential concern with the results is that the CPIA measure of policy could be endogenous, in one of two ways. One possibility is that the CPIA is simply proxying for the indicators of debt distress. For example, it could be that World Bank country economists assign poor CPIA scores to countries that are running arrears or are negotiating a Paris Club agreement. The first defense against this possibility is the use of lags of the CPIA—CPIA in the year before the distress or normal time episode begins. Nevertheless, it could be that lagging the CPIA just one year is not sufficient, if for example, the CPIA scores are based on information that a Paris Club deal is likely to happen soon. When the CPIA is lagged by three years to eliminate the possibility of a mechanical correlation between distress and CPIA scores capturing actual or imminent distress, this further lagged measure of policy remains a highly significant predictor of debt distress (columns 1 and 4 of table 5).¹⁵ As the lag lengthens, however, the CPIA score unsurprisingly becomes less significant (results not reported). However, the significance of the one-year and three-year lagged CPIA scores in predicting debt distress is unlikely to reflect primarily reverse causation from future distress outcomes to current CPIA scores because that would require remarkable foresight on the part of World Bank staff who produce CPIA scores.

The second potential endogeneity problem is that the CPIA is simply proxying for other omitted country characteristics that also matter for debt distress. These might be macroeconomic instability in the country or deep institutional characteristics such as protection of property rights. Another possibility comes from the findings of Reinhart, Rogoff, and Savastano (2003) that a country's history of default and bad policy is a robust predictor of investors' perceptions of the likelihood of sovereign default. Countries with weak property rights, high macroeconomic instability, or a history of default might be more likely both to experience debt distress and to receive worse CPIA scores. To the extent that such factors are time invariant, the usual strategy is to difference them away and focus on the within-country variation in debt distress, debt burdens, policies, and growth. As discussed in more detail below, this simple differencing strategy is not an option in the nonlinear probit specification used here. The next section uses a dynamic panel probit estimator that allows for unobserved country-specific sources of heterogeneity that help address this problem.

For now, direct controls are introduced for some of these country characteristics. Property rights protection is measured using the "rule of law" indicator

15. The CPIA variable itself is quite persistent over time. By sheer coincidence the correlation between the first and second lag of the CPIA in the sample is one, so the results using the second lag are not reported. The correlation between the first and third lag of the CPIA in the sample is 0.90.

TABLE 5. Role of Policies and Shocks

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
|--|------------|------------|------------|----------------------|----------------------|----------------------|------------|------------|------------|----------------------|----------------------|
| | All | All | All | Low-Income Countries | Low-Income Countries | Low-Income Countries | All | All | All | Low-Income Countries | Low-Income Countries |
| Present value | 0.644 | 0.655 | 0.654 | 0.373 | 0.369 | 0.368 | 0.594 | 0.644 | 0.656 | 0.419 | 0.367 |
| debt/exports | (0.151)*** | (0.156)*** | (0.153)*** | (0.187)** | (0.186)** | (0.188)* | (0.156)* | (0.148)*** | (0.158)*** | (0.186)** | (0.203)* |
| Country Policy and Institutional Assessment (CPIA) | | -4.994 | -0.533 | | -2.086 | -0.811 | -0.573 | -0.526 | -0.597 | -0.698 | -0.801 |
| | | (2.131)** | (0.142)*** | | (3.307) | (0.234)*** | (0.145)*** | (0.146)*** | (0.161)*** | (0.248)*** | (0.264)*** |
| Real GDP growth | -5.381 | -0.556 | -3.755 | -2.416 | -0.789 | -2.506 | -4.740 | | | | |
| | (2.235)** | (0.155)*** | (2.140)" | (3.110) | (0.241)*** | (3.148) | (2.128)** | | | | |
| Three-year CPIA | -0.384 | | | -0.808 | | | | | | | |
| | (0.189)*** | | | (0.234)*** | | | | | | | |
| Rule of law | | 0.014 | | | -0.097 | | | | | | |
| | | (0.176) | | | (0.332) | | | | | | |
| High inflation | | | 0.997 | | | -0.222 | | | | | |
| | | | (0.667) | | | (1.206) | | | | | |
| Default history | | | | | | | 2.101 | | | | |
| | | | | | | | (0.826)** | | | | |
| Growth in real exchange rate | | | | | | | | -0.361 | | -1.979 | |
| | | | | | | | | (0.909) | | (1.497) | |
| Growth in terms of trade | | | | | | | | | -0.321 | | -0.090 |
| | | | | | | | | | (1.438) | | (1.565) |
| Constant | 0.194 | 0.833 | 0.613 | 1.912 | 1.770 | 1.947 | 0.809 | 0.535 | 0.833 | 1.376 | 1.878 |
| | (0.531) | (0.566) | (0.528) | (0.789)** | (0.923)* | (0.811)** | (0.520) | (0.543) | (0.615) | (0.829)* | (0.899)** |
| Number of observations | 190 | 200 | 199 | 200 | 83 | 83 | 200 | 194 | 175 | 81 | 69 |

*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.

Source: Authors' analysis of data described in Appendix.

constructed by Kaufmann, Kraay, and Mastruzzi (2005). Macroeconomic instability is measured as the proportion of years in the sample period when inflation was greater than 40 percent a year. These control variables are added to the core specification for all countries and for low-income countries in columns 2, 3, 5, and 6 of table 5. Including these variables reduces the magnitude of the effect of policy somewhat compared with the results reported in table 3, but the direct effect of the CPIA remains highly significant.

To directly investigate a country's history of default on its external obligations as a predictor of debt distress, the database of default episodes compiled by Reinhart, Rogoff, and Savastano (2003) is used to identify the fraction of years between independence (or 1824, whichever is later) and 1980 in which a country was in default on its external borrowing. Although this default history variable is significant for the full sample of observations, the CPIA remains highly significant as well (column 7 of table 5).¹⁶ From all of this, as well as the results described below, there can be reasonable confidence that endogeneity of the CPIA in the sense defined here is not driving the findings.

The analysis also attempts to isolate particular shocks that might trigger debt distress. The real GDP growth variable is replaced by measures of real exchange rate movements (columns 8 and 10 of table 5) and fluctuations in the terms of trade (columns 9 and 11). The rate of change in the real exchange rate relative to the U.S. dollar is constructed using changes in the nominal exchange rate and GDP deflators. Positive values of this variable correspond to real depreciations, which would be expected to raise the risk of debt distress by making dollar-denominated debt service obligations more expensive in domestic terms. The income effect of terms of trade change is measured as the current local currency share of exports in GDP times the growth rate of the local currency export deflator minus the share of imports in GDP times the growth rate of the import deflator. Adverse terms of trade shocks lower export earnings and income and might also trigger debt servicing difficulties. Despite the prior plausibility of these two shocks, virtually no evidence is found that they are significant predictors of debt distress. For terms of trade shocks this may not be too surprising, as Raddatz (2005) documents that these shocks account for only a small share of the variation in output in low-income countries. The results for real exchange rate movements also echo the negative findings of McFadden and others (1985) mentioned earlier. The results continue to show, however, that debt burden and policy are highly significant.

Robustness of Core Specification: Dynamic Panel Probit Estimates

The robustness checks conclude with a dynamic panel probit specification used to investigate the extent to which unobserved country characteristics, as well as

16. The impact of the default history variable cannot be estimated separately in the low-income sample because this variable is by coincidence equal to zero for all of the normal time episodes among low-income countries, creating a singularity in the probit regression.

countries' history of distress, matter for the current likelihood of distress. The following dynamic probit specification with unobserved country-specific effects is estimated:

$$(2) \quad P[y_{ct} = 1] = \Phi(\beta'X_{ct} + \rho \cdot y_{c,t-1} + \mu_c)$$

where $y_{c,t-1}$ denotes the value of the distress indicator in the episode immediately prior to the one occurring at time t in country c ; ρ is a parameter capturing the persistence of distress, and μ_c is an unobserved country-specific time-invariant effect capturing unobserved country characteristics that influence the probability of debt distress.

This empirical model generalizes the one used so far in two important respects. First, it allows for serial dependence in the likelihood of debt distress by permitting the past value of the outcome variable (distress or not) to affect the probability that the current outcome will be distress. This captures in a very straightforward way the possibility that once a country has experienced debt distress it is more likely to do so in the future. Second, this model allows for unobserved country effects that influence the probability of distress in all periods for a given country. Importantly, there is no need to assume that the unobserved country effects are independent of the observed right-side variables. Thus there is no need to be concerned that the significance of the findings is being driven by omitted time-invariant country characteristics, such as property rights protection or a history of macroeconomic instability, that might affect the probability of debt distress and also be correlated with the included right-side variables.

The presence of unobserved country-specific effects complicates the estimation of the model. As noted, they cannot be eliminated by a differencing transformation common in linear panel data models. Moreover, a lagged dependent variable presents the familiar initial conditions problem: loosely, it cannot be ignored that by construction the lagged dependent variable is correlated with the unobserved country effect. This model is estimated by applying the initial conditions correction suggested by Wooldridge (2005). He proposes modeling the individual effect as a linear function of the initial observation on the dependent variable for each country, as well as time averages of all of the right-side variables. He also shows that this specification can be simply estimated using standard random-effect probit software, as long as the list of explanatory variables is augmented with the initial value of the dependent variable and time averages of all of the right-side variables for each country.

As before, debt indicators, policy, and growth remain significant predictors of the probability of debt distress in the full sample, and in the low-income country sample only debt and policies matter (table 6). The point estimates of the coefficients on debt and on policy are also quite close to those shown in table 3. No evidence is found that debt distress in the previous period significantly raises the probability of distress in the next period, after debt burdens, policy, and growth are controlled for. Taken together, these results suggest that unobserved

TABLE 6. Dynamic Probit Results

| | All | Low-Income Countries |
|--|-------------------|----------------------|
| Lagged dependent variable | -0.641 (0.451) | -0.311 (0.736) |
| Present value debt/exports | 0.628 (0.256)** | 0.557 (0.334)* |
| Country Policy and Institutional Assessment (CPIA) | -0.522 (0.260)** | -1.486 (0.520)*** |
| Real GDP growth | -8.237 (3.407)*** | -6.122 (5.791) |
| Initial dependent variable | 0.031 (0.502) | 0.381 (0.808) |
| Average present value debdexports | 0.152 (0.363) | -0.323 (0.409) |
| Average CPIA | -0.153 (0.326) | 0.700 (0.560) |
| Average real GDP growth | 5.839 (4.729) | 2.701 (7.612) |
| Constant | 1.100 (0.859) | 2.094 (1.408) |
| Number of observations | 191 | 78 |
| Number of countries | 87 | 39 |

+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.

Source: Authors' analysis of data described in *Appendix*.

time-invariant country characteristics are not responsible for the main results and that the observed persistence of debt distress is due mostly to country effects and the persistence of debt burdens, policies, and shocks rather than to a recent history of distress itself."

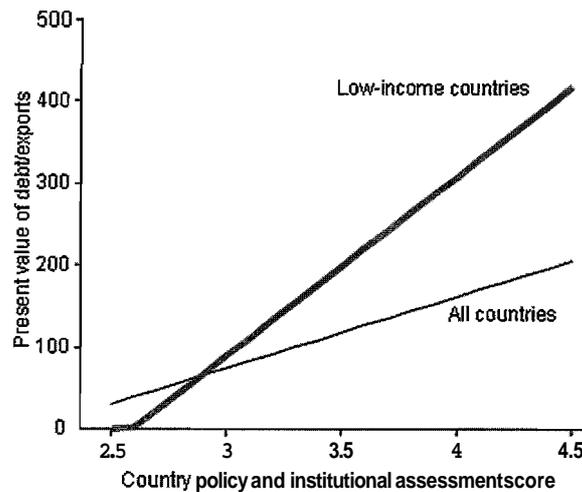
III. POLICY IMPLICATIONS

The analysis has shown that the likelihood of debt distress in low-income countries depends not only on debt burdens but also on the quality of a country's policies and institutions. This finding has important implications for the lending strategies of official creditors such as the World Bank. The basic point is that assessments of the appropriateness of a country's debt burden should reflect the quality of the country's policies and institutions. The empirical results indicate a significant tradeoff between debt burdens and policy: countries with better policies and institutions can carry substantially higher debt burdens than countries with worse policies and institutions without increasing the risk of debt distress.

Figure 3 highlights this tradeoff. Consider a hypothetical country with a growth rate of 3.6 percent (the mean of the sample). For the indicated value of the CPIA on the x-axis, the level of the present value of debt relative to exports is

17. This last result contrasts with the finding of McFadden and others (1985), who do find evidence for state dependence in episodes of debt-servicing difficulties.

FIGURE 3. Policies and Debt Distress



Source: Authors' analysis of data described in Appendix.

computed that would be consistent with a predicted probability of debt distress equal to 39 percent, which corresponds to the unconditional mean in the sample of low-income countries (truncating negative values at zero).¹⁸ The same relationship between policy and debt is reported based on the estimates pooling data for all countries.

The tradeoffs between debt and policy are considerable. For the estimates based on low-income countries, a country with average growth and poor policy (corresponding to a CPIA score of 3, roughly in the first quartile of the sample) would be able to tolerate a present value of debt to exports of about 100 percent. In contrast, a country with good policy (corresponding to a CPIA score of 4, in the fourth quartile of the sample) would be able to tolerate a debt level nearly three times higher with the same distress probability. For the estimates based on all countries, the tradeoffs are flatter. The implied debt level for a poor policy country with a CPIA of 3 would be 75 percent, whereas for a fairly good policy country with a CPIA of 4 it would be 160 percent.

Of course, for lower debt distress probabilities these lines would shift down, corresponding to lower levels of debt for any level of policy, and for higher debt-distress probabilities the line would shift up, corresponding to higher levels of debt for any level of policy. In addition, the precise magnitudes of the effects of differences in debt and policy on these implied debt levels depend on all of the estimated coefficients in the regressions on which these estimates are based, and

18. These implied debt levels are obtained by solving $p = \Phi(\beta_0 + \beta_1 \times \text{Debt} + \beta_2 \times \text{Policy} + \beta_3 \times \text{Growth})$ for debt, where p is the desired probability of debt distress.

these are subject to margins of error and vary across specifications. Thus, these numbers provide only a rough order of magnitude of the effects of policies on the level of debt consistent with a given distress probability.

The second policy implication is that the risk of debt distress should be taken into account when deciding the terms of resource transfers to low-income countries. The point here is simple. In recent years large increases in flows of development finance have been advocated to help countries meet the Millennium Development Goals. If these flows are provided in the form of concessional loans, as they have been in the past, many low-income countries are likely to see very sharp increases in their debt burdens. This could easily undo the reductions in debt burdens due to past debt relief efforts, which could thus have little lasting impact on the risk of debt distress.

Consider this simple hypothetical calculation for the 28 countries that have received debt relief under the HIPC Initiative. Between 1990 and 2003 these countries as a group received \$58 billion in disbursements of mostly concessional loans from official creditors. Given the calls for much greater aid to these countries, it is not inconceivable that these countries would receive the same amount of disbursements over the next five years. Now assume that the rate of concessionality of this new lending is the same as it is on the stock of debt outstanding as of 2003 and suppose further that it is distributed across countries in the same proportions as past official lending. This permits calculating a hypothetical present value of debt five years into the future, which can be thought of as corresponding to a rapid scaling-up in aid in the form of development lending to these countries with no change in the terms of these loans. Under this scenario the ratio of the present value of debt to exports would rise from a median of 157 percent to a median of 299 percent for these 28 countries. Based on the estimates in column 2 of table 3 and assuming no change in policies or growth performance, the estimated risk of debt distress would rise from a median of 33 percent (based on end-2003 data) to 52 percent. If exports in these countries are assumed to grow at their historical rate over the next five years, the increase in the ratio of the present value of debt to exports would be smaller, but still substantial, rising to 248 percent of exports for the median country.

This simple example illustrates how a large scaling-up in loan-based aid to low-income countries, without significant changes in the terms of these loans, is likely to result in sharp increases in external debt burdens and in the risk of debt distress. To avoid this, grants must make up a greater share of development assistance. For countries with a given quality of policies the share of grants will need to be significantly higher where debt distress probabilities are high and lower where distress probabilities are low. This implication is also consistent with the results in columns 3 and 6 of table 4, which show that the greater the concessionality of debt, the lower is the risk of debt distress.

Grants should not, however, supplant loans one for one in nominal terms in countries where the risk of debt distress is high, for two reasons. First, replacing

loans with grants equal to the face value would represent a vastly larger resource transfer than is currently envisioned by donors, and obtaining the necessary financing would be difficult. Second, such a scheme would implicitly "reward" countries implementing weak policies with greater overall resource transfers, undermining efforts to target aid to countries with good policies.

One scheme for calibrating the share of grants without exacerbating these targeting problems would be a three-step process that begins by converting the total amount of new lending into its grant equivalent from the donors' perspective (subtracting the present value of future debt-service obligations from the face value of the new lending). Next, this grant equivalent could be allocated across countries following an aid allocation rule that recognizes the importance of "needs" (the prevalence of poverty) and aid effectiveness (a function of the quality of policies and institutions of the recipient country, as is currently done by the International Development Association, the soft-loan window of the World Bank). Finally, for countries below a 'specified distress probability (where the capacity for servicing debt in the future is considered relatively good), this grant equivalent could be grossed-up into a much larger amount of concessional lending with the same grant equivalent.

Such a scheme has many advantages. It avoids the large and likely unsustainable increases in debt burdens that would follow from large-scale across-the-board new lending to low-income countries. This scheme not only ensures that resources are targeted to countries with high poverty and good policies but also provides an additional reward for good policy. Countries would prefer to be able to gross-up as much of their grant-equivalent allocation as possible into lending, and improvements in policy can create additional headroom for new borrowing by lowering the probability of debt distress. Finally, this scheme would not require any new commitments by donors to finance new grants, beyond the new transfers in grant equivalent terms implicit in donor commitments to lending at existing rates of concessionality. This is because donors would be committing to the same transfer to a country whether they provide only the grant element or they convert the grant element into a loan with the same grant equivalent. If anything, the resource transfer from the perspective of the donor might be even smaller, to the extent that calibrating the fraction of loans to the probability of debt distress results in higher actual repayment rates in the future.¹⁹

In summary, this article has shown that the risk of debt distress depends significantly on a small set of factors: debt burdens, policies and institutions,

19. In 2005 the IMF and World Bank adopted a joint Debt Sustainability Framework for low-income countries that endorsed a greater role for grants to reduce the risk of debt distress. It spelled out a set of policy-dependent debt sustainability thresholds that are based on the empirical analysis in the working paper version of this article (Kraay and Nehru 2004). IDA has chosen to implement a modified version of the proposal advanced here and in the working paper version of this article. The key difference, however, is that the IDA proposal converts the full amount of proposed lending to countries at risk of debt distress into grants, less a small discount. This results in greater resource transfers to countries at risk of debt distress and so reduces the policy selectivity of IDA resource transfers.

and shocks. This finding is robust to several checks, and the empirical model does a reasonable job of predicting future debt distress. Although at some level these results should not be too surprising, they do have important implications for how to finance resource transfers to low-income countries. The results indicate that the probability of debt distress is already high in many low-income countries and is likely to increase sharply if the large-scale development finance required to meet the Millennium Development Goals is provided in the form of concessional lending at historical levels of concessionality. The article also proposed a simple scheme of financing resource transfers to low-income countries in a way that controls the probability of debt distress, provides good incentives to borrowers, and does not involve additional donor commitments to finance large-scale new grants.

APPENDIX

Data Sources

The debt-distress indicator requires data on arrears, Paris Club deals, and IMF programs. Data on arrears are taken from the World Bank's *Global Development Finance*. The arrears data consist of arrears to official and private creditors and are expressed as a share of total debt outstanding. Paris Club deals come from the Paris Club Web site (<http://www.clubdeparis.org>). For IMF programs data on commitments under Standby Arrangement and Extended Fund Facility programs and data on the size of each country's quota, used to normalize commitments, are from the IMF's *International Financial Statistics*.

The core regressions include the present value of debt as a share of exports. Data on the numerator of this measure come from Dikhanov (2003). He applies currency-, maturity-, and time-specific market interest rates to the flow of debt-service obligations on a loan-by-loan basis, using data from the World Bank's Debtor Reporting System database to arrive at a historical series of the present value of public and publicly guaranteed debt for all developing countries since 1970. The denominator is exports in current U.S. dollars taken from the World Bank's World Development Indicators Database. The same data source is used to construct the growth rate of GDP in constant local currency units for the growth variable. The CPIA variable is a confidential policy assessment produced by World Bank country economists. Details on its structure and disclosure of recent data are available at <http://www.worldbank.org>. Data on the share of debt owed to official creditors and on the share of public and publicly guaranteed debt in total debt (at face value) are from *Global Development Finance*.

The robustness checks use the rule of law measure constructed by Kaufmann, Kraay, and Mastruzzi (2005) and countries' default history as reported by Reinhart, Rogoff, and Savastano (2003). A dummy variable for years of high inflation is constructed using consumer price index (CPI) inflation data obtained from the World Development Indicators Database, supplemented by data on the

growth rate of the GDP inflator when CPI inflation is not available. The real exchange rate index is the bilateral real exchange rate relative to the U.S. dollar, using the price index of GDP in the home country and the United States. Data on these variables also come from the World Development Indicators Database <http://www.worldbank.org>.

REFERENCES

- Aylward, Lynn, and Rupert Thorne. 1998. "An Econometric Analysis of Countries' Repayment Performance to the International Monetary Fund." IMF Working Paper 98132. International Monetary Fund, Washington, D.C.
- Berg, Andrew, and Jeffrey Sachs. 1988. "The Debt Crisis: Structural Explanations of Country Performance." *Journal of Development Economics* 29(3):271–306.
- Birdsall, Nancy, C. Claessens, and I. Diwan. 2003. "Policy Selectivity Forgone: Debt and Donor Behavior in Africa." *World Bank Economic Review* 17(3):409–35.
- Cline, William. 1984. *International Debt: Systemic Risk and Policy Response*. Washington, D.C.: Institute of International Economics.
- Cohen, Daniel. 1996. "The Sustainability of African Debt." Policy Research Working Paper 1691. World Bank, Washington, D.C.
- Collins, Susan. 2003. "Probabilities, Probits, and the Timing of Currency Crises." Georgetown University, Department of Economics, Washington, D.C.
- Detragiache, Enrica, and Antonio Spilimbergo. 2001. "Crises and Liquidity: Evidence and Interpretation." IMF Working Paper 0112. International Monetary Fund, Washington, D.C.
- Dikhanov, Yuri. 2003. "Reconstruction of Historical Present Value of Debt for Developing Countries, 1980–2001: Methodology and Calculations." World Bank, Development Economics Data Group, Washington, D.C.
- IMF (International Monetary Fund). Various years. *International Financial Statistics*. Washington, D.C.
- . 2002. "Evaluation of the Prolonged Use of Fund Resources." Independent Evaluation Office, International Monetary Fund, Washington D.C.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2005. "Governance Matters IV: Updated Governance Indicators for 1996–2004." World Bank Policy Research Working Paper 3630. World Bank, Washington, D.C.
- Kraay, Aart, and Vikram Nehru. 2004. "When is External Debt Sustainable?" Policy Research Working Paper 3200. World Bank, Washington, D.C.
- Lloyd-Ellis, H., G. W. McKenzie, and S. H. Thomas. 1990. "Predicting the Quantity of LDC Debt Rescheduling." *Economics Letters* 32(1):67–73.
- Manasse, Paolo, Nouriel Roubini, and Axel Schimmelpfening. 2003. "Predicting Sovereign Debt Crises." IMF Working Paper 031221. International Monetary Fund, Washington, D.C.
- McFadden, Daniel, Richard Eckaus, Gershon Feder, Vassilis Hajivassiliou, and Stephen O'Connell. 1985. "Is There Life After Debt? An Econometric Analysis of the Creditworthiness of Developing Countries." In Gordon Smith and John Cuddington, eds., *International Debt and the Developing Countries*. Washington, D.C.: World Bank.
- McKenzie, David. 2004. "An Econometric Analysis of IBRD Creditworthiness." *International Economic Journal* 18(4):427–48.
- Raddatz, Claudio. 2005. "Are External Shocks Responsible for the Instability of Output in Low-Income Countries?" Policy Research Working Paper 3680. World Bank, Washington, D.C.
- Reinhart, Carmen, Kenneth Rogoff, and Miguel Savastano. 2003. "Debt Intolerance." *Brookings Papers on Economic Activity* 1:1–74.
- Underwood, John. 1991. "The Sustainability of International Debt." World Bank, International Finance Division, Washington, D.C.

- Wooldridge, Jeffrey. **2005**. "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics* 20(1):39–54.
- World Bank. **2003**. "Special Purpose Financial Statements and Internal Control Reports of the International Development Association." In *The World Bank Annual Report 2003*. Vol. **2**: Financial Statements and Appendixes. Washington, D.C.
- . Various years. *Global Development Finance*. Washington, D.C.
- . Various years. *World Development Indicators Database*. Washington, D.C.

Will African Agriculture Survive Climate Change?

Pradeep Kurukulasuriya, Robert Mendelsohn, Rashid Hassan, James Benhin, Temesgen Deressa, Mbaye Diop, Helmy Mohamed Eid, K. Yerfi Fosu, Glwadys Gbetibouo, Suman Jain, Ali Mahamadou, Renneth Mano, Jane Kabubo-Mariara, Samia El-Marsafawy, Ernest Molua, Samiha Ouda, Mathieu Ouedraogo, Isidor Sène, David Maddison, S. Niggol Seo, and Ariel Dinar

Measurement of the likely magnitude of the economic impact of climate change on African agriculture has been a challenge. Using data from a survey of more than 9,000 farmers across 11 African countries, a cross-sectional approach estimates how farm net revenues are affected by climate change compared with current mean temperature. Revenues fall with warming for dryland crops (temperature elasticity of -1.9) and livestock (-5.4), whereas revenues rise for irrigated crops (elasticity of 0.5), which are located in relatively cool parts of Africa and are buffered by irrigation from the effects of warming. At first, warming has little net aggregate effect as the gains for irrigated crops offset the losses for dryland crops and livestock. Warming, however, will likely reduce dryland farm income immediately. The final effects will also depend on changes in precipitation, because revenues from all farm types increase with precipitation. Because irrigated farms are less sensitive to climate, where water is available, irrigation is a practical adaptation to climate change in Africa.

Pradeep Kurukulasuriya is a Ph.D. student in environmental economics at Yale University; his email address is pradeep.kurukulasuriya@yale.edu. Robert Mendelsohn is a professor in environmental economics at Yale University; his email address is robert.mendelsohn@yale.edu. Rashid Hassan is director and professor at the Centre for Environmental Economics and Policy in Africa (CEEPA), at the University of Pretoria; his email address is rashid.hassan@up.ac.za. James Benhin is a research fellow at CEEPA; his email address is james.benhin@up.ac.za. Temesgen Deressa is a Ph.D. student at CEEPA and a researcher at the Ethiopia Development Research Institute and the Environmental Economics Policy Forum for Ethiopia, Addis Ababa; his email address is tderessa@yahoo.com. Mbaye Diop is a lecturer at the Institut Senegalais de Recherches Agricoles, Campus universitaire de l'ESP, Dakar; his email address is mbaye.diop@isra.sn. The late Helmy Mohammed Eid was a professor at the Soil, Water, and Environment Research Institute in Cairo. K. Yerfi Fosu is a senior lecturer at the University of Ghana, Legon; his email address is yfosu@ug.edu.gh. Glwadys Gebtibouo is a Ph.D. student in environmental economics at CEEPA, University of Pretoria; her email address is ggbtibouo@postino.up.ac.za. Suman Jain is a senior lecturer in Mathematics and Statistics at the University of Zambia, Lusaka; her email address is sjain@natsci.unza.zm. Ali Mahamadou is a lecturer in Agricultural Economics at the University of Abdou Moumouni, Niamey, Niger; his email address is cresa@intnet.ne. Reneth Mano is a senior lecturer in Economics at the University of Zimbabwe, Harare; his email address is rtmano@mweb.co.zw. Jane Kabubo-Mariara is a senior lecturer in Economics at the University of Nairobi, Kenya; her email address is jmariara@uonbi.ac.ke. Samia El-Marsafawy is a researcher at the Soil, Water, and Environment Research Institute, Cairo; her email address is samiaelmarsafawy797@hotmail.com. Ernest Molua is a lecturer in Economics at the University of Buea, Cameroon; his email address is

The increasing concern about climate change has led to a rapidly growing body of research on the impacts of climate on the economy. Quantitative estimates of climate impacts have improved dramatically over the last decade (Pearce and others 1996; McCarthy and others 2001; Tol 2002; Mendelsohn and Williams 2004). Sub-Saharan Africa is predicted to be particularly hard hit by global warming because it already experiences high temperatures and low (and highly variable) precipitation, the economies are highly dependent on agriculture, and adoption of modern technology is low.

Despite the estimated magnitude of the potential impacts on Africa, there have been relatively few economic studies (Kurukulasuriya and Rosenthal 2003). Most of the quantitative projections are interpolations from empirical studies done elsewhere (Tol 2002; Mendelsohn and Williams 2004). A limited number of agronomic studies on Africa have confirmed that warming would have large effects on selected crops (Rosenzweig and Parry 1994), but these studies reflect only a small share of Africa's crops, they fail to capture how farmers might respond to warming, and they do not quantify overall economic impacts. The economic impact on the livestock sector in Africa has gone largely unstudied (Kurukulasuriya and Rosenthal 2003).

This study uses farm-level data collected across diverse climate zones in 11 African countries to explore how the current climate already affects African farmers, specifically net farm revenues. Total net farm revenue is defined as the sum of incomes from three main farm activities: dryland crops, irrigated crops, and livestock. Irrigated crops rely on at least some irrigated water (from surface flows or ground water). Dryland crops rely only on rainfall that falls on the farm. Livestock in Africa largely depend on grazing on natural lands or pasture.

emolua@gmx.net. Samiha Ouda is a researcher at the Soil, Water, and Environment Research Institute, Cairo; her email address is samihaouda@yahoo.com. Mathieu Ouedraogo is a researcher in agricultural economics at the Institut de l'Environnement et de Recherches Agricoles, Burkina Faso; his email address is oued_mathieu@yahoo.fr. Isidor Stne is a Ph.D. student in agricultural economics at the Institut Senegalais de Recherches Agricoles, Campus universitaire de IESP, Senegal; his email address is isisene@ucad.sn. David Maddison is a senior lecturer in Environmental Economics in the Department of Economics at University College London; his email address is d.maddison@ucl.ac.uk. S. Niggol Seo is a Ph.D. student in environmental economics at Yale University; his email address is niggol.seo@yale.edu. Ariel Dinar is lead economist in the Agriculture and Rural Development Department at the World Bank and the task team leader of the project leading to this article; his email address is adinar@worldbank.org. This article is based on a cooperative research effort among researchers from 11 African countries: Burkina Faso, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Niger, Senegal, South Africa, Zambia, and Zimbabwe. The research leading to this article was funded by grants from the Global Environmental Facility, the World Bank Trust Fund for Environmentally and Socially Sustainable Development, and the Swiss and Finish Trust Funds. The authors thank the U.S. National Oceanic and Atmospheric Administration for climate data, the Food and Agriculture Organization of the United Nations for soil data, and the International Water Management Institute and Allysa McCluskey and Kenneth Strzepak of the University of Colorado for hydrological data. They thank the Centre for Environmental Economics and Policy in Africa at the University of Pretoria, South Africa, for its sponsorship, leadership, and coordination of this project. They also thank the International Centre for Advanced Mediterranean Agronomic Studies in Zaragoza for helping to coordinate a meeting in Spain of all participants, and they thank the editor for insightful comments and suggestions and three anonymous referees for constructive comments.

This information is used to estimate the impacts of changing temperature and precipitation on the net revenues of African farmers using the Ricardian method (Mendelsohn, Nordhaus, and Shaw 1994). Net revenues are regressed on climate, soils, and other control variables. Separate regressions are estimated for the three main farm activities to shed light on the climate response of each of these components of farm income. The amount of land that was planted could be accurately measured for the crop regressions used to estimate net revenue per hectare. Since most African farmers rely on common land for livestock grazing, it was not possible to determine how much land was used. The livestock regressions are consequently based on revenue per farm. Although these analyses are therefore different, total farm income is still the sum of the incomes from these three sources.

I. METHODOLOGY

This study uses a Ricardian approach to measure the determinants of farm net revenues, including climate, through an econometric analysis of cross-sectional data (Mendelsohn, Nordhaus, and Shaw 1994). The approach has been applied to study the relationship between net revenues from crops and climate, including other key variables in selected countries in low latitudes (Kumar and Parikh 2001; Mendelsohn, Sanghi, and Dinar 2001; Molua 2002; Deressa, Hassan, and Poonyth 2005; Gbetibouo and Hassan 2005; Seo, Mendelsohn, and Munasinghe 2005; Kurukulasuriya and Ajwad 2006), but this study is the first application of the method to many countries across a continent (Africa). It is also the first Ricardian study to examine net revenues from livestock.

David Ricardo (1815) was the first to observe that land rents reflects the net revenue value of farmland. Farmland net revenue (R) reflects the net productivity and costs of individual crops and livestock:

$$(1) \quad R = \sum P_i Q_i(X, F, H, Z, G) - \sum P_x X$$

where P_i is the market price of crop i , Q_i is output of crop i , X is a vector of purchased inputs (other than land), F is a vector of climate variables, H is water flow, Z is a vector of soil variables, G is a vector of economic variables, and P_x is a vector of input prices. The farmer is assumed to choose inputs (X) to maximize net revenues given the characteristics of the farm and market prices.

Each farmer is assumed to choose inputs and outputs to maximize their net revenue subject to the climate and soils of each farm, in addition to other key economic variables. The observed net revenue function is therefore the locus of maximum profits given the set of exogenous climate, soil, and economic conditions. The Ricardian model is a reduced form hedonic price model of that locus of profits. Net revenue is defined as gross revenue minus the cost of transport, storage, hired labor (valued at the market wage rate), light farm tools (files, axes, machetes), heavy machinery (tractors, plows, threshers), fertilizer, pesticides,

and postharvest losses. Household labor costs are not included because the shadow wage rate that workers apply to their own time cannot be measured. This is a common problem in the development literature (Bardhan and Udry 1999). The effects of different soil types are controlled for and tested in this analysis. Water flow is also included because it is particularly important for irrigation (Mendelsohn and Dinar 2003).

In the data set used in this analysis, farmers growing crops either use irrigation or do not. Many farmers, however, combine growing crops with raising livestock. A few farmers just raise livestock. Across Africa, farmers use a combination of irrigated crops, dryland crops, and livestock. The analysis examines each of these revenue sources separately, estimating a separate Ricardian model for livestock, dryland crops, and irrigated crops. This is done, first, because each revenue source is thought to respond to climate differently and second, because information was not available about the amount of land that livestock used, and therefore revenue per hectare models for cropland cannot be used for livestock. Following Schlenker, Hanemann, and Fischer (2005), whether a farm grows dryland crops, grows irrigated crops, or raises livestock is assumed to be exogenous. Future studies will relax this assumption and predict how even the type of farm may be influenced by climate.

The standard Ricardian model relies on a quadratic formulation of climate:

$$(2) \quad R = \beta_0 + \beta_1 F + \beta_2 F^2 + \beta_3 Z + \beta_4 G + \beta_5 \log(H) + \mu$$

where μ is an error term. Both a linear and a quadratic term for climate, F (temperature and precipitation) are introduced. This quadratic functional form for climate captures the expected nonlinear shape of the relationship between net revenues and climate (Mendelsohn, Norhaus, and Shaw 1994; Mendelsohn, Sanghi, and Dinar 2001; Mendelsohn and Dinar 2003). Based on agronomic research and previously reported cross-sectional analyses, farm net revenues are expected to have a concave (hill-shaped) relationship with temperature. For each crop, there is a known average temperature that is best for crop production, but this relationship is not necessarily concave for each season. Less is known about livestock, but in general, livestock appear to be chosen to match certain climate zones (McCarthy and others 2001). So there is every reason to believe that the quadratic form of the climate variables is suitable for livestock as well. Water flow is introduced in a log form because the benefits from flow diminish as flow increases.

Past Ricardian studies have suggested that crops respond to seasonal variation in climate (Mendelsohn, Norhaus, and Shaw 1994; Mendelsohn, Sanghi, and Dinar 2001; Mendelsohn and Dinar 2003). Climate data for consecutive months are highly correlated and perform poorly in Ricardian models. In a tropical climate, even the seasons are highly correlated as there is not as much variation from one season to the next compared with temperate climate regions. Nonetheless, there was a desire to capture as much of the seasonal effects as possible. Seasons were defined as follows: winter (in the

Southern Hemisphere) is defined as May, June, and July, spring as August through October, summer as November through January, and fall as February through April. The seasons in the Northern Hemisphere are defined in the same way for the appropriate months. These seasonal definitions provide the best fit with the crop farm data, and they reflect the mid-point for key rainy seasons in Africa. All three Ricardian models use the same definition of climate.

An alternative way to measure climate is to use growing degree days (Ritchie and NeSmith 1991). Growing degree days are the sum of degrees above a specified cutoff temperature across a growing season for a particular crop. If climates are stable, growing degree days can measure land value accurately (Schlenker, Hanemann, and Fischer 2006). However, the technique was originally developed to be crop specific, and it was tied to the sowing and harvest dates of individual crops. It becomes a very vague and biased concept when applied across crops and regions with different and changing lengths of growing seasons, because, among other things, it does not capture seasonal effects or account for the impact of cold days.

The marginal impact of a single climate variable, f_j , on crop or livestock net revenue evaluated at the mean of that variable is:

$$(3) \quad E[\partial R/\partial f_j] = \beta_{1,j} + 2\beta_{2,j}E[f_j].$$

Because flow is expressed in logarithmic terms, the marginal impact of flow, H , on net revenue is

$$(4) \quad E[\partial R/\partial H] = \beta_5/H.$$

These marginal effects can be evaluated at any level of climate or flow, but the focus is on showing effects at mean climate levels for Africa. Note that the linear formulation of the model assumes that these marginal effects [equations (3) and (4)] are independent of future technological change. However, it is possible that future technological change could make crops (or other farming activity) more susceptible to temperature or precipitation changes—or less so.

The marginal change in rent is the marginal welfare effect of the change in the exogenous variable. However, with nonmarginal changes in exogenous variables, underlying prices may change. The Ricardian price schedule will overestimate welfare effects in this case because the price changes will mitigate some of the effects (Cline 1996; Adams 1999; Darwin 1999). For globally traded goods such as agricultural crops and livestock products, price changes are not likely to be a problem as local gains and losses in production are expected to offset each other for a small net change in global output (Reilly, Hohmann, and Kane 1994; Mendelsohn and Nordhaus 1999). However, a dramatic reduction in the productivity of African agriculture could affect African wage rates. For example, if productivity in a district fell substantially, local wages might fall—or if productivity rose, wages might also rise. To capture this effect, a more

complete analysis would have to model local African labor markets as well as land productivity.

The strength of the Ricardian method is that it captures the adaptation responses of farmers. The use of net revenues in the analysis reflects the benefits and costs of implicit adaptation strategies. Specifically, the analysis incorporates the substitution of different inputs and the introduction of alternative activities that each farmer has adopted in light of the existing climate. For example, the model reflects the costs of seeds, equipment, and hired labor that a farmer might pay in response to climate and the revenue that the farmer consequently earns. Farmers adapt by changing their crops, their sowing methods, their timing, and their inputs. Farms also adapt by changing their types of livestock and number of animals. All of these changes increase net revenues under the new climate conditions. Consequently, accounting for adaptation leads to much lower predicted overall damage from climate change. This is true even if all the adaptations being considered are practices currently being used in Africa.

The Ricardian method is a cross-sectional approach. It assumes that cross-sectional comparisons provide useful insights into long-term intertemporal changes. The Ricardian method does account for adaptation costs that would be associated with comparing one equilibrium state with another. However, cross-sectional analysis does not account for dynamic transition costs that might occur as farms move between two states. For example, the Ricardian model does not capture the costs of learning by doing or of decommissioning capital prematurely (Kaiser, Riha, Wilkes, Rossiter, and others 1993; Kaiser, Riha, Wilkes, and Sampath 1993; Quiggin and Horowitz 1999; Kelly, Kolstad, and Mitchell 2005). Furthermore, innovations in modern agriculture, which have been adopted in other low-latitude regions, have spread slowly in Africa, suggesting that transition costs must be examined carefully in Africa (Evenson and Gollin 2003).

The Ricardian approach has a number of other limitations. For example, the approach does not measure the effect of variables that do not vary across space. Specifically, the effect of different levels of carbon dioxide is not captured as carbon dioxide levels do not vary systematically across Africa. Controlled experiments and crop simulation modeling are required to learn about the likely positive effect of carbon fertilization. It may also be possible that some aspects of future climates do not resemble anything in the present. For example, if there is some type of extreme event in the future that does not occur in the present, the analysis will not be able to evaluate its effect. Finally, the Ricardian results can be distorted by local agricultural policies. If some countries subsidize farm inputs or regulate certain crops, they influence farmers' choices, and the empirical results will reflect these distortions. If the distortion is explicitly modeled, it can be controlled for. But if it is not carefully modeled, the climate variables may be biased. If future decision makers eliminate these subsidies or introduce new ones, the predictions from the empirical results may no longer hold.

II. DATA AND EMPIRICAL SPECIFICATIONS

The study relied on long-term average climate (normals). These long-term data for districts in Africa were gathered from two sources. Satellite data on temperature was measured by a Special Sensor Microwave Imager (SSM/I) on U.S. Department of Defence satellites (Basist and others 2001) for 1988–2003. The SSM/I detects microwaves through clouds and estimates surface temperature (Basist and others 1998; Weng and Grody 1998). The satellites conduct daily overpasses at 6 a.m. and 6 p.m. across the globe. The precipitation data come from the Africa Rainfall and Temperature Evaluation System (World Bank 2003) created by the Climate Prediction Centre of the U.S. National Oceanic and Atmospheric Administration. It is based on ground station measurements of precipitation for 1977–2000. Thus, the temperature and precipitation data cover slightly different periods. This discrepancy might be a problem for measuring variance or higher moments of the climate distribution, but it should not affect the use of the mean of the distribution.

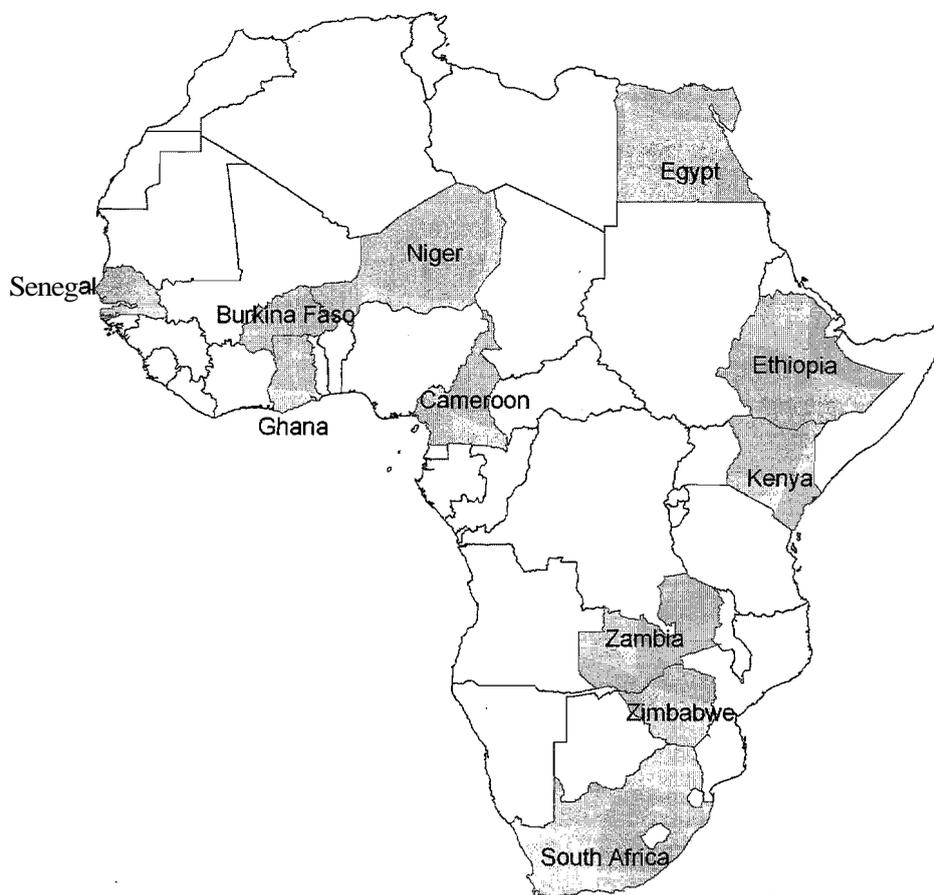
The 11 countries in this study were selected across the diverse climate zones of Africa (figure 1) and precipitation of each country in the sample. Although Africa is generally hot and dry, there is a great deal of variation across the continent (figure 2). Egypt and South Africa are much cooler than the rest of the countries in the sample. Similarly, relative to the other countries in the sample, Cameroon is very wet, followed by Kenya, Zambia, Ghana, and Ethiopia; the other countries, especially Egypt, are drier.

Within each country, districts, were selected to capture representative farms across diverse agroclimatic conditions. Between 30 and 50 districts were sampled in each country. In each district, surveys were conducted in 2002–04 of randomly selected farms (seven countries were surveyed in the 2002–03 season and four countries were added in 2003–04). Sampling was clustered in villages to reduce the cost of administering the survey. A total of 9,597 surveys were administered. The final number of surveys with usable information on crop production was 9,064. Of these, 7,238 farms had dryland crops, 1,221 had irrigated crops, and 5,062 had livestock. Many farms had both crops and livestock. The total number of farm surveys per country varied from 222 in South Africa to 1,288 in Burkina Faso.

Median net revenues per farm from dryland crops, irrigated crops, and livestock are presented in figure 3 by country. The relative importance of dryland crops and irrigated crops varies considerably. For example, Egypt is entirely dependent on irrigated crops because the climate is too dry to support crops without irrigation. In contrast, most farms in East Africa and the Sahel have very little irrigated crops. Livestock net revenue varies widely across countries, but it is particularly important in relatively dry countries.

Data on hydrology were obtained from a continental scale hydrological model of Africa (IWRMI and University of Colorado 2003). Using climate data and local topography, the model estimated the potential monthly long-term stream flow for each district. Potential water flows were used because water can be withdrawn from many places along a watershed. Water flow measures the

FIGURE 1. Map of Study Countries



amount of water coming from other districts and is an important complement to the water generated in each district from precipitation. Water flow is particularly important in Africa, where water is generally scarce. For example, the Nile delta would be completely unsuitable for farming without the water from the Nile River.

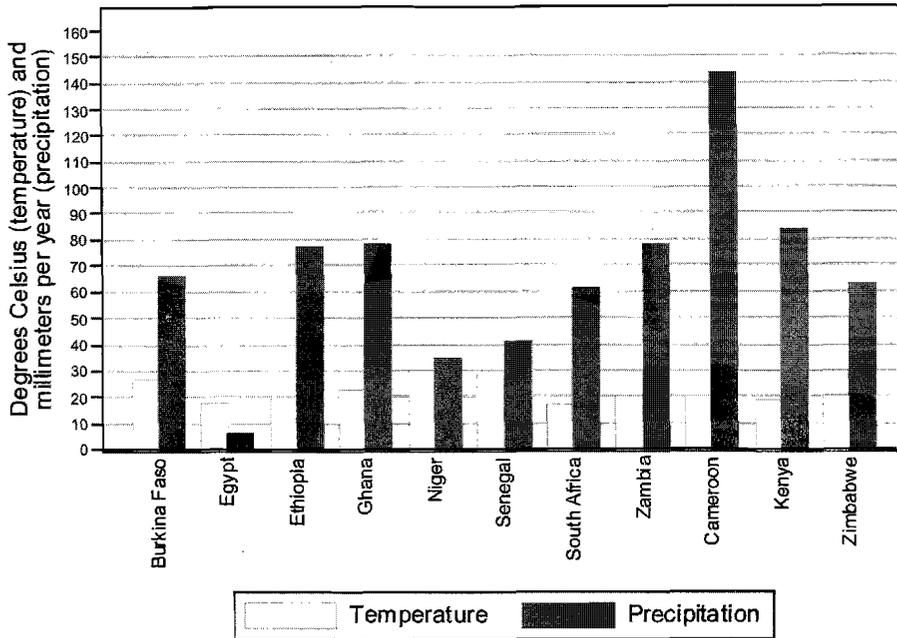
Data on the composition, coarseness, and slope of the major soils in each district were obtained from the Food and Agriculture Organization (FAO 2003).

III. RESULTS

The analysis explores three principle hypotheses. First, African net farm revenues are sensitive to climate. Second, irrigated and dryland crops have different responses to climate (Mendelsohn and Dinar 2003; Schlenker, Hanemann, and Fischer 2005). Third, crops and livestock have different climate response functions.

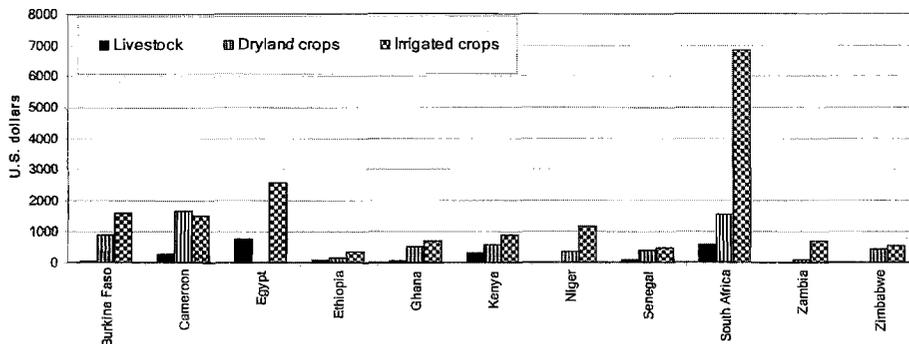
These hypotheses are tested by estimating three regressions. The net revenues per hectare for dryland crops and irrigated crops and the net revenue per farm for livestock are regressed on climate and other control variables (table 1).

FIGURE 2. Mean Annual Temperature (1988–2003 average) and Precipitation (1977–2000 average) by Country



Source: Authors' analysis based on data from Basist and others (2001) and World Bank (2003).

FIGURE 3. Median Net Revenue Per Farm from Dryland Crops, Irrigated Crops, and Livestock (2005 U.S. dollars)



Source: Surveys conducted by the authors in 2002/4.

TABLE 1. Ordinary Least Squares Regression of African Net Farm Revenues

| Variable | (1) Dryland crop (per hectare) | (2) Irrigated crop (per hectare) | (3) Livestock (per farm) |
|-----------------------------------|--------------------------------------|--|--------------------------------|
| Temperature winter | -68 (1.11) | 181 (1.06) | -8,643** (2.65) |
| Temperature winter squared | 2.5 (1.61) | -3.2 (0.62) | 191** (2.55) |
| Temperature spring | -28 (0.28) | -180 (0.85) | 6,772** (2.45) |
| Temperature spring squared | -1.0 (0.46) | 0.8 (0.15) | -136** (2.37) |
| Temperature summer | 12.5 (1.77) | 1,277*** (4.06) | -2,904** (2.24) |
| Temperature summer squared | -1.4 (1.02) | -20.2*** (3.67) | 58** (2.40) |
| Temperature fall | -58 (0.85) | -1,517** (3.32) | 4,679** (2.97) |
| Temperature fall squared | 0.4 (0.25) | 28.7** (3.38) | -95.9** (3.00) |
| Precipitation winter | -4.6*** (3.37) | 11.8 (1.54) | -3.7 (0.15) |
| Precipitation winter squared | 0.03*** (4.00) | -0.05 (1.00) | -0.18 (0.95) |
| Precipitation spring | 4.7*** (3.78) | -12.2 (1.62) | 35.2 (1.76) |
| Precipitation spring squared | -0.01** (2.56) | -0.10** (2.24) | -0.32** (2.66) |
| Precipitation summer | 3.6*** (4.71) | 27.9*** (4.86) | 34.9** (2.49) |
| Precipitation summer squared | -0.01*** (3.90) | -0.10*** (4.54) | -0.17** (2.51) |
| Precipitation fall | -2.1** (2.99) | 25.5*** (4.27) | -23.5** (2.37) |
| Precipitation fall squared | 0.01*** (5.47) | 0.08*** (4.67) | 0.11** (2.69) |
| Flow | -8.2 (2.39) | 10.2** (2.57) | -134.4** (3.11) |
| Elevation | -0.13*** (5.29) | 0.08 (0.47) | 0.65** (3.20) |
| Log household size | 23.0 (1.80) | 109 (1.61) | -244 (1.31) |
| Household electricity | 95.0*** (5.00) | 467** (3.93) | 1,236*** (3.49) |
| Eutric gleysols coarse soils | -403*** (4.15) | -1,570** (3.03) | |
| Lithosols steep soils | -234*** (5.14) | -1,130*** (3.45) | |
| Orthic luvisols medium soils | -1,950*** (4.02) | | -52,000** (3.17) |
| Chromic vertisol fine soils | -661** (2.92) | -2,105** (3.01) | |
| Chromic luvisol fine soils | -227*** (5.88) | | -8,240*** (3.35) |
| Cambic arenosols soils | 1,720** (3.05) | | |
| Luvic arenosols soils | -272*** (4.04) | | |
| Calcic yermosols medium soils | -536** (2.69) | | |
| Gleyic luvisols soils | -134** (2.78) | | |
| Rhodic ferralsols steep soils | | 5,450 (1.66) | 814,000** (3.20) |
| Chromic luvisols medium soils | | -6,040** (2.58) | -54,600** (2.91) |
| Dystric nitosols soil | | 7,110*** (4.41) | |
| Eutric cambisols fine soils | | | -1,710,000*** (3.35) |
| Calcic cambisols coarse soils | | | 3,120** (2.72) |
| Vertic cambisols fine soils | | | 913 (2.25) |
| Orthic ferralsols fine soils | | | -7,330** (3.23) |
| Rhodic ferralsols fine soils | | | -5,620** (2.94) |
| Lithosols medium steep soils | | | -2,490,000** (3.26) |
| Ferric luvisols fine soils | | | 627** (2.61) |
| Gleyic luvisols medium soils | | | 868 (2.28) |
| Chromic vertisols soils | | | 895 (2.54) |
| Eutric planasols fine steep soils | | | 1,018,000*** (3.34) |
| Constant | 665 (1.05) | 2,150 (0.85) | -8,057 (1.76) |
| Number of observations | 7,246 | 1,253 | 4,759 |
| F-statistic | 35.2 | 45.5 | 4.8 |
| R-squared | 0.16 | 0.25 | 0.22 |

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: Numbers in parentheses are robust t-statistics.

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

Revenue per hectare could not be examined for livestock because most African farmers graze their animals on communal land. Property rights on communal lands are complicated, and reliable measures of the amount of land used are not available. The marginal effect of climate on each source of revenue is examined, and climate elasticities are computed for temperature and precipitation as the percentage change in net revenue for a percentage change in temperature or precipitation.

The overall regressions in table 1 are significant at the 1 percent level, and the R-squared values are 0.16 for dryland crop, 0.25 for irrigated crop, and 0.22 for livestock models. Irrigated farms have higher average net revenue per hectare than dryland farms and respond differently to the independent variables. For example, the net revenue of irrigation rises with water flow, as expected. The soil types that affected dryland crop, irrigated crop, and livestock revenues often differ. The soil types that affect both dryland and irrigated crop revenue have a larger impact on irrigated land. Elevation has a strong positive effect on livestock revenue, a strong negative impact on dryland crop revenue, and no effect on irrigated crop revenue. Animals can adapt to high altitudes, but the large diurnal cycles associated with high altitudes are harmful to crops. Farms with more people in the household earn more crop revenue but less livestock revenue. This result implies that growing crops is more labor-intensive than tending animals. Finally, farms with electricity had higher revenue across all farm types, especially irrigated crops. Electricity and the technology associated with it may be the source of this higher value. It is also possible that farms with electricity are close to major markets (cities) and that this contributes to higher values.

The most important comparison across crops and livestock concerns the climate coefficients. Many of these coefficients are not significant because the climate variables are highly correlated with each other. Unlike temperate climates, tropical climates do not vary greatly from season to season. The four-season specification was maintained, however, to keep the study comparable with studies done in other countries. The coefficients vary across the regressions, but they are hard to interpret individually. However, the second-order terms provide a sense of what shape the response functions are taking. A negative coefficient on a squared term implies a concave shape and a positive coefficient implies a convex shape. The expectation is that the second order temperature coefficients would be negative, especially if higher temperatures were catastrophic. However, the results do not support that hypothesis. Because the observed range of precipitation is well below the maximum desired amount, the second order precipitation coefficients could have any sign. Many of these second order precipitation coefficients in table 1 turn out to be positive, suggesting that net revenue rises rapidly with precipitation.

Calculating seasonal marginal effects reveals that higher temperatures in the spring and fall are harmful and higher temperatures in the summer

and winter are beneficial. These results mirror findings from the United States, except that they are delayed by one season (Mendelsohn, Nordhaus, and Shaw 1994; Mendelsohn and Nordhaus 1999). In Africa crops are planted during the summer monsoons rather than the spring. The warmer temperatures in summer reflect the benefits of a longer growing season. The warmer temperatures in the winter help the crops to ripen. The warmer temperatures in the fall are harmful because this is when temperatures peak. The warmer temperatures in the spring are harmful because they encourage pests.

The marginal impacts of a change in annual temperature and precipitation are also evaluated at the respective sample mean. This calculation of annual effects adds a constant temperature and precipitation increment to each season. Note that the sample mean climate for dryland, irrigated land, and livestock are quite different. Irrigated land is located in drier (average annual precipitation of 33 millimeters a year) and cooler areas of the sample (average annual temperature 19° Celsius), livestock in drier (66 millimeters a year) and warmer areas (22° Celsius), and dryland crops in the warmer (22° Celsius) and wetter (72 millimeters a year) areas. Many farms in the sample that have crops also have livestock.

Warmer temperatures have very different marginal effects on dryland crops and irrigated crops. Dryland crop revenue falls an average of \$27 per hectare per 1° Celsius increase in temperature, whereas irrigated crop revenue increases an average of \$30 per hectare per 1° Celsius (table 2). Temperature has a muted effect on irrigated crops, partially because irrigation buffers the crops from rainfall shortages and partially because irrigated crops are currently planted in relatively cool locations in Africa. The change in revenue per hectare is multiplied by the mean number of hectares of each type (9 hectares of dryland and 101 hectares of

TABLE 2. Marginal Climate Impacts on Net Farm Revenue Per Hectare: Ordinary Least Squares (OLS) and Country Fixed Effects (Dollars Per Hectare)

| Marginal impact | Dryland crop | Irrigated crop | Livestock |
|-----------------|-------------------|--------------------|----------------------|
| OLS | | | |
| Temperature | -27*** (-37, -16) | 30 (-20, 80) | -379 (-775, 17) |
| Precipitation | 1.6*** (0.6, 2.8) | 3.0 (-8.8, 14.8) | 19.8*** (0.29, 39.5) |
| Fixed effect | | | |
| Temperature | -10 (-21, 0.7) | 72*** (19, 125) | -293 (-696, 110) |
| Precipitation | 1.5*** (0.2, 2.8) | -0.9 (-13.6, 11.8) | -5.2 (-20.3, 9.9) |

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: Values are calculated at the mean climate of the sample using OLS coefficients from table 1 and fixed effects coefficients from table S.5 in the supplemental appendix (available at <http://wber.oxfordjournals.org>). Numbers in parentheses are 95 percent confidence intervals. Estimates for livestock are at the farm level.

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

irrigated crops per farm) to show the change in average revenue per farm: a decline of \$239 on dryland farms and an increase of \$3,005 on irrigated farms. Livestock net revenue falls by an average of \$379 per farm per 1° Celsius. Weighting each effect by the frequency of each type of farm suggests that the mean annual impact of a 1° Celsius increase in temperature is a negligible and insignificant increase in net revenue across African farms. The initial increase in revenue from irrigated land offsets the decline in revenue for dryland and livestock.

The marginal effects of precipitation on net revenues also vary across revenue sources. The marginal effect of precipitation is \$1.66 per hectare per 1 millimeter increase in precipitation per month on dryland crops and \$2.98 on irrigated crops. However, on a farm basis the marginal effect of precipitation is \$15 per millimeter per month on dryland crops and \$302 on irrigated crops (table 3). By comparison, the marginal effect of precipitation on livestock net revenue per farm is \$15 per millimeter per month. Weighting these values by the frequency of each farm type suggests that a 1 millimeter per month increase in precipitation leads to an expected aggregate increase in net revenue of \$67 per farm.

Temperature and precipitation elasticities are also calculated (table 4). The temperature elasticity is -1.9 for dryland crops and 0.5 for irrigated crops, which, as noted are buffered from higher temperatures both by their cooler locations and the moderating effect of irrigation. The temperature elasticity for livestock is -5.4, meaning that livestock is more sensitive to temperature than crops are. Although many livestock can survive in hot locations, the most profitable livestock (beef) are limited to cool regions (South Africa and the Mediterranean). Warmer temperatures would drive these profitable beef cattle out of Africa.

TABLE 3. Marginal Climate Impacts on Net Farm Revenue Per Farm: Ordinary Least Squares (OLS) and Country Fixed Effects (Dollars Per Farm)

| Marginal impact | Dryland crop | Irrigated crop | Livestock |
|-----------------|----------------------|------------------------|--------------------|
| OLS | | | |
| Temperature | -239*** (-335, -142) | 3005 (-2040, 8048) | -379 (-775, 17) |
| Precipitation | 15*** (5.1, 25) | 301.3 (-896.6, 1499.3) | 19.9** (0.3, 39.5) |
| Fixed effect | | | |
| Temperature | -93 (-192, 7) | 7262*** (1940, 12584) | -292 (-695, 110) |
| Precipitation | 13*** (2, 25) | -93 (-1374, 1187) | -5 (-20.3, 9.9) |

***Significant at the 5 percent level.

****Significant at the 1 percent level.

Note: Values are calculated at the mean climate of the sample using OLS coefficients from table 1 and fixed effects coefficients from table S.5 in the supplemental appendix (available at <http://wber.oxfordjournals.org>) for the median size farm of each type. Numbers in parentheses are 95 percent confidence intervals.

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

TABLE 4. Comparison of Climate Elasticities: Ordinary Least Squares (OLS) and Country Fixed Effects

| Elasticity | Dryland crop | Irrigated crop | Livestock |
|---------------|-----------------------------------|--------------------------------|------------------------------|
| OLS | | | |
| Temperature | -1.9 ^{****} (-2.7, -1.1) | 0.5 (-0.3, 1.2) | -5.4 (-11.1, 0.3) |
| Precipitation | 0.4 ^{****} (0.1, 0.6) | 0.1 (-0.2, 0.4) | 0.8 ^{**} (0.0, 1.7) |
| Fixed effect | | | |
| Temperature | -0.7 (-1.5, 0.1) | 1.1 ^{****} (0.3, 2.0) | -4.2 (-10.0, 1.6) |
| Precipitation | 0.3 ^{****} (0.1, 0.6) | -0.02 (-0.4, 0.3) | -0.2 (-0.9, 0.4) |

^{***}Significant at the 5 percent level.

^{****}Significant at the 1 percent level.

Note: Values are calculated at the mean climate and mean net revenue of the sample using OLS coefficients from table 1 and fixed effects coefficients from table S.5 in the supplemental appendix (available at <http://wber.oxfordjournals.org>) for each farm type. Numbers in parentheses are 95 percent confidence intervals.

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

Precipitation elasticities are much smaller than temperature elasticities. The precipitation elasticity is 0.1 for irrigated crops, 0.4 for dryland crops, and 0.8 for livestock. Although many agronomic models focus on precipitation, these empirical results suggest that crops and livestock are more sensitive to temperature than to precipitation.

Warming may affect water flow as well. Flow has a significant effect on all three sources of farm income (see table 1). Higher flow increases the net revenue per hectare of irrigated land (the elasticity of net revenue with respect to water flow is 0.2 for irrigated land). However, flow is also likely to have a large effect on the amount of land available for irrigation. Reduced flow would limit the amount of farmland that could be converted from dryland to irrigated cropland. Flow has a negative effect on dryland and livestock net revenue even though dryland crops and to a large extent livestock do not use irrigation. In areas with good flow, farmers may use their best land for irrigation, leaving relatively poor lands for livestock and dryland crops.

A country fixed effect analysis was also conducted, with a dummy variable introduced for each country. A country fixed effect model removes unmeasurable differences between countries due to omitted variables. Again, many of the individual climate coefficients are insignificant (detailed results are presented in table S.5 in the supplemental appendix, available at <http://wber.oxfordjournals.org>). This is partly because country fixed effects remove some of the intercountry climate variation that was part of the sample design. However, the country dummy variable may also be picking up hidden factors that vary by country. In the livestock regression the only significant country dummy variable is for South Africa. This could be due to the large profitable beef cattle farms in South Africa or to the climate that supports such farms. In the regression for irrigated crops, the only significant country dummy variable is for Kenya, which

has lower than average returns per hectare. It is not clear why irrigated farms would be less profitable in Kenya than in other countries. For the dryland crop revenue regression, Cameroon has above average returns per hectare, and Ethiopia, Kenya, and Zambia have below average returns. The pattern of the dryland country coefficients may reflect the benefits of ample water in Cameroon and little water in East Africa, or they may reflect a set of hidden factors.

Tables 2 and 3 present the marginal results of the country fixed effects model along with the ordinary least squares (OLS) results already discussed. The marginal effect of temperature on dryland crops is $-\$95$ per hectare, which is a much larger loss than the OLS regression predicts. The marginal effect of temperature on irrigated crops is positive but also much larger than the OLS regression predicts. Only the livestock results are not significantly different. The introduction of country fixed effects also changes the marginal effect, increasing the benefits to dryland farmers but eliminating the effects on irrigated crops and livestock. In comparing the OLS with the fixed effects results, part of the more harmful effect of higher temperatures on dryland crops, the more beneficial effect of temperature on irrigated crops, and the more beneficial effect of higher precipitation on livestock may be due to unmeasured country level variables. When the fixed effects are introduced, these effects are moderated.

Another concern in this analysis is that Egypt is a unique case because of its dependence on the Nile River. Farmers along the Nile can irrigate and produce two seasons of crops, leading to significantly higher earnings per hectare. Because Egypt is cooler and drier than most of the sample, this could bias the climate results. Dropping Egypt has no effect on the dryland analysis, because there are no dryland observations for Egypt, but a large impact on irrigation, because a great deal of the irrigated sample comes from Egypt (table S.4 in the supplemental appendix).

Of the 1,253 observations of irrigation in the original analysis, only 531 are left when the observations for Egypt are dropped. Many of the coefficients in the new regression for irrigated crops are consequently insignificant (for example, water flow, elevation, and all the temperature coefficients). The precipitation coefficients remain significant, however. And although many observations remain in the livestock regression, many of the coefficients become insignificant except for the precipitation coefficients. Thus the observations for Egypt have a strong impact on the results for livestock and irrigated crops.

To interpret how dropping the observations for Egypt has affected the climate results, we compared the marginal effects of the climate coefficients in table 5. Dropping Egypt makes the marginal impact of warmer temperatures on irrigated land harmful, but the change in impact is not statistically significant. Without Egypt, the marginal impact of precipitation on livestock increases, but the change is also not significant. However, the marginal effect of

TABLE 5. Comparison of Marginal Impacts of Climate on Net Farm Revenue Per Farm with and Without Egypt (Dollars Per Farm)

| | Irrigated crop | | Livestock | |
|----------------------------------|------------------------|--------------------------------------|------------------------------|----------------------------------|
| | Egypt included | Egypt omitted | Egypt included | Egypt omitted |
| Marginal impact of temperature | 3005 (-2040, 8048) | -3247 (-10769, 4276) | -379 (-775, 17.2) | -642 ^{""} (-1,196, -89) |
| Marginal impact of precipitation | 301.3 (-896.6, 1499.3) | -1502 ^{***} (-2459, -545.6) | 19.9 ^{""} (0.3, 39) | 7.1 (-10.4, 24.5) |

^{""}Significant at the 5 percent level.

^{***}Significant at the 1 percent level.

Note: Effects are calculated at the mean climate of the sample using table 1 coefficients including Egypt and table S.4 coefficients in the supplemental appendix (available at <http://wber.oxfordjournals.org>) excluding Egypt. The mean farm size (in hectares) is assumed. Numbers in parentheses are 95% confidence intervals.

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

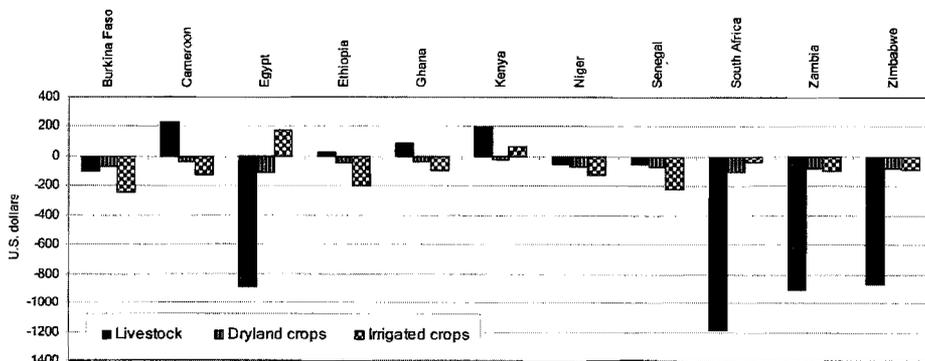
precipitation on irrigated land changes from \$301 to -\$1502 per farm, which is a significant change (the precipitation elasticity changes from +0.1 to -0.4). In Sub-Saharan Africa increased precipitation reduces the net revenues of irrigated farms. Irrigation is a better investment in drier locations. The data for Egypt, despite the country's low precipitation and high productivity, were hiding this effect.

The marginal impact of temperature and precipitation on each country is also calculated (figures 4 and 5). This calculation differs from the previous analysis in that it uses the mean temperature and mean rainfall values for each country. The analysis reveals that the impacts of climate change differ across countries. Cooler countries such as Egypt, South Africa, Zambia, and Zimbabwe are likely to suffer livestock losses from warmer temperatures because of the loss of beef cattle (figure 4). Irrigated crops in currently hot regions such as Ethiopia and West Africa will suffer with warming, whereas irrigated crops in the Nile Delta and Kenyan highlands will gain. However, some effects are fairly universal. Dryland crops in all countries throughout Africa will be damaged by any warming. Figure 5 suggests that the marginal impact of precipitation is mostly beneficial, compared with that of warming, and that livestock and irrigated farms will mostly benefit from rising precipitation and lose from declining precipitation.

IV. CONCLUSION AND POLICY IMPLICATIONS

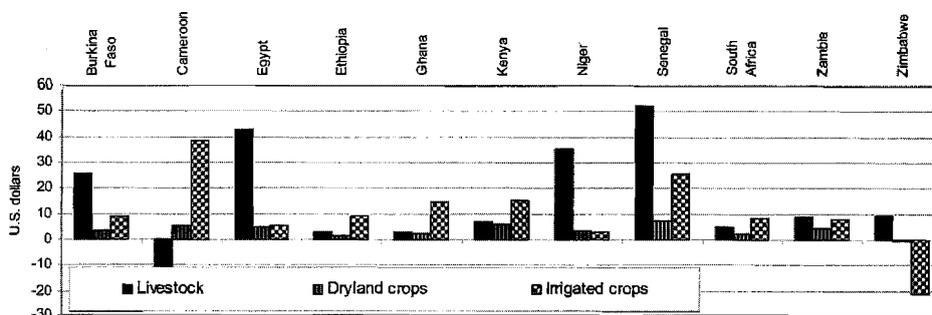
This study examined the net revenues of farmers in 11 African countries and provided quantitative confirmation of what scientists have long suspected. Although African dryland farmers have adapted to local conditions, net revenues would fall with more warming or drying. Dryland crop and livestock farmers are especially vulnerable, with temperature elasticities of -1.9 and -5.4, respectively. Irrigated cropland benefits slightly from marginal warming because irrigation

FIGURE 4.. Marginal Impact of Temperature by Country



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

FIGURE 5. Marginal Impact of Precipitation by Country



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

reduces climate impacts and because these farms are currently located in relatively cool places in Africa.

With precipitation elasticities of 0.4 for dryland crops and 0.8 for livestock across Africa, net revenues for dryland crops and livestock will increase if precipitation increases with climate change and decrease if precipitation decreases with climate change. Net revenues for irrigated land will follow in the same direction but to a much smaller extent (elasticity of 0.1). Increases in precipitation will have an unambiguously beneficial effect on African farms on average, whereas decreases in precipitation will have a harmful effect. However, country effects and within-country effects can differ.

The revenue effects for dryland crops, irrigated crops, and livestock are assessed independently. When the marginal temperature effects across all three sources of revenue are summed, increases in revenues on irrigated cropland at first offset losses for dryland crops and livestock. As temperatures continue to rise, however, the net effect on African farms becomes steadily more harmful. Total farm revenue decreases as precipitation falls but rises as precipitation increases. Climate scenarios that entail either significant warming or substantial drying will consequently be quite harmful. However, climate scenarios that entail only mild increases in temperature and more rainfall may actually be beneficial. The total impact on African agriculture will depend on the climate scenario.

The analysis reveals that net farm revenue has a quadratic relationship with both temperature and precipitation. The marginal impact of climate change consequently will depend on each farm's initial temperature and precipitation. Farms that are located in hotter and drier areas are at greater risk because they are already in a precarious state for agriculture. Dryland farming throughout Sub-Saharan Africa is vulnerable to warming. Dryland farming in the East, West, and Sahel regions of Africa are especially at risk. In contrast, irrigated crops in places that are relatively cool now, such as the Nile delta and the highlands of Kenya, enjoy marginal gains from warming. Finally, drier locations

such as Egypt, Niger, and Senegal get big livestock gains from increased precipitation relative to wetter locations in Africa.

Because Sub-Saharan African economies as a whole depend more heavily on agriculture, total GDP and per capita income is also vulnerable. In contrast, nonagricultural GDP in Northern Africa is more diversified, and so the economies of these countries are less vulnerable to climate change.

This study measures the marginal impact of climate change. It does not predict the future. Simulating likely future climate impacts is a large undertaking. First, one must examine the projections of several climate models to get a sense of the range of plausible climate scenarios. Second, one must project how agriculture is likely to change in the future, both in technology and in land use. For example, the average dryland farmer currently earns \$319 a hectare and the average irrigated land farmer earns \$1,261 a hectare. The more technologically advanced irrigated farms earn even more. The adoption of technology and capital is very important to the future of agriculture in Africa. Third, one must estimate by how much carbon fertilization is likely to increase crop productivity over time (Reilly and others 1996). These gains will reduce the magnitude of the damages in Africa, although it is not clear by how much.

Will Africa survive climate change? The results of this study suggest that Africa will be hit hard by severe climate change scenarios. Some countries are more vulnerable than others, so it is important to focus on the countries that really need help. In fact, in several scenarios, many African farmers gain whereas others lose from climate change. This study also notes that African farmers already practice some forms of climate adaptation. Policymakers may want to pay special attention to these successful adaptation practices. For example, irrigation water (including related inputs) and livestock are already used in some areas to alleviate climate hardships such as droughts and low levels of precipitation.

One adaptation that has moved very slowly in Africa is technology adoption. Africa lags behind the rest of the world in adopting irrigation, capital, and high-yield varieties (Evenson and Gollin 2003). Some technologies may help farmers adapt to drier or hotter conditions, such as the development of new soybean varieties in Brazil. However, even climate-neutral technical advances will help farmers increase productivity and counterbalance losses from climate change. Through research and outreach, governments could encourage the development and use of varieties with more tolerance for the hot and dry conditions of many of Africa's agroclimatic zones.

The quantitative results, especially the sizable differences between irrigated and dryland agriculture and livestock in Africa, suggest that promoting irrigation could help alleviate the likely effects of climate change in Africa. Where water is available, moving from dryland to irrigated agriculture would increase not only average net revenue per hectare but also the resilience of agriculture to climate change. Governments could make public investments in infrastructure

and canals for water storage and conveyance, where appropriate and where the public good nature of these investments prevent adequate private sector investment. Investment in successful irrigation in Sub-Saharan Africa ranges between \$3,600 and \$5,700 a hectare in 2000 prices (Inocencio and others 2005). This analysis suggests that the difference between dryland and irrigated agriculture runs between \$150 and \$5,000 a hectare, depending on the country. This range of investment values implies that farmers in some countries could repay irrigation investments within a very reasonable period. Policy-makers may want to consider supporting such coping interventions for climate change, where appropriate.

Finally, in addition to encouraging direct adaptations, both local and national governments and international organizations could invest in infrastructure and institutions to ensure a stable environment to enable agriculture to prosper. Such policy interventions may not only achieve the long-term goal of helping vulnerable populations adapt to climate change, but may also increase the likelihood of achieving the more immediate Millennium Development Goals, such as halving hunger, reducing poverty, and improving health.

REFERENCES

- Adams, R. M. 1999. "On the Search for the Correct Economic Assessment Method." *Climatic Change* 41(3-4): 363-70.
- Bardhan, P., and C. Udry. 1999. *Development Microeconomics*. Oxford: Oxford University Press.
- Basist, A., N. C. Grody, T. C. Peterson, and C. N. Williams. 1998. "Using the Special Sensor Microwave Imager to Monitor Land Surface Temperatures, Wetness, and Snow Cover." *Journal of Applied Meteorology* 37(9): 888-911.
- Basist, A., C. N. Williams, N. Grody, T. Ross, and S. Shen. 2001. "Using the Special Sensor Microwave Imager to Monitor Surface Wetness." *Journal of Hydrometeorology* 2(3): 297-308.
- Cline, W. R. 1996. "The Impact of Global Warming on Agriculture: Comment." *American Economic Review* 86(5): 1309-11.
- Darwin, R. 1999. "The Impact of Global Warming on Agriculture: A Ricardian Analysis: Comment." *American Economic Review* 89(4): 1049-52.
- Deressa, T., R. Hassan, and D. Poonyth. 2005. "Measuring the Economic Impact of Climate Change on South Africa's Sugarcane Growing Regions." *Agrekon* 44: 524-42.
- Evenson, R., and D. Gollin. 2003. "Assessing the Impact of the Green Revolution, 1960-2000." *Science* 300: 758-62.
- FAO (Food and Agriculture Organization). 2003. *The Digital Soil Map of the World*. Version 3.6. CD-ROM. Rome, Italy: FAO.
- Gbetibouo, G., and R. Hassan. 2005. "Measuring the Economic Impact of Climate Change on Major South African Field Crops: A Ricardian Approach." *Global and Planetary Change* 47(2-4): 143-52.
- Inocencio, A., M. Kikuchi, D. Merrey, M. Tonosaki, A. Maruyama, I. de Jong, H. Sally, and F. Penning de Vries. 2005. "Lessons from Irrigation Investment Experiences: Cost-Reducing and Performance-Enhancing Options for Sub-Saharan Africa." Report 6. International Water Management Institute, Colombo. African Water Investment Strategies. [www.iwmi.cgiar.org/africanwaterinvestment/].

- IWMI (International Water Management Institute and University of Colorado). 2003. Hydroclimatic data for the Global Environment Facility/Centre for Environmental Economics and Policy in Africa/World Bank Project "Climate, Water and Agriculture: Impacts and Adaptation of Agro-ecological Systems in Africa." Colombo and Boulder.
- Kaiser, H. M., S. J. Riha, D. S. Wilkes, D. G. Rossiter, and R. K. Sampath. 1993. "A Farm-Level Analysis of Economic and Agronomic Impacts of Gradual Warming." *American Journal of Agricultural Economics* 75(2): 387–98.
- Kaiser, H., S. Riha, D. Wilkes, and R. Sampath. 1993a. "Adaptation to Global Climate Change at the Farm Level." In H. Kaiser and T. Drennen, eds., *Agricultural Dimensions of Global Climate Change*. Delray Beach, FL: St. Lucie Press.
- Kelly, D. L., C. D. Kolstad, and G. T. Mitchell. 2005. "Adjustment Costs from Environmental Change." *Journal of Environmental Economics and Management* 50(3): 468–95.
- Kumar, K., and J. Parikh. 2001. *Indian Agriculture and Climate Sensitivity*. *Global Environmental Change* 11(2):147–52.
- Kurukulasuriya, P., and M. Ajwad. 2006. "Application of the Ricardian Technique to Estimate the Impact of Climate Change on Smallholder Farming in Sri Lanka." *Climatic Change*. doi: 10.1007/s10584-005-9021-2.
- Kurukulasuriya, P., and S. Rosenthal. 2003a. "Climate Change and Agriculture: A Review of Impacts and Adaptations." *Climate Change Series 91*. Environment Department Papers, World Bank, Washington, D.C.
- McCarthy, J., O. Canziani, N. Leary, D. Dokken, and K. White, eds. 2001. *Climate Change 2001: Impacts, Adaptation, and Vulnerability*. Cambridge: Cambridge University Press.
- Mendelsohn, R., and A. Dinar. 2003. "Climate, Water, and Agriculture." *Land Economics* 79(3): 328–41.
- Mendelsohn, R., and W. Nordhaus. 1999. "The Impact of Global Warming on Agriculture: A Ricardian Analysis: Reply." *The American Economic Review* 89(4): 1053–55.
- Mendelsohn, R., and L. Williams. 2004. "Comparing Forecasts of the Global Impacts of Climate Change." *Mitigation and Adaptation Strategies for Global Change* 9(4): 315–33.
- Mendelsohn, R., W. Nordhaus, and D. Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *American Economic Review* 84(4): 753–71.
- Mendelsohn, R., A. Sanghi, and A. Dinar. 2001. "The Effect of Development on the Climate Sensitivity of Agriculture." *Environment and Development Economics* 6(1): 85–101.
- Molua, E. 2002. "Climate Variability, Vulnerability and Effectiveness of Farm-level Adaptation Options: The Challenges and Implications for Food Security in South-Western Cameroon." *Environment and Development Economics* 7(3): 529–45.
- Pearce, D., W. R. Cline, A. N. Achanta, S. Fankhauser, R. K. Pachauri, R. S. J. Tol, and P. Vellinga. 1996. "The Social Costs of Climate Change: Greenhouse Damage and Benefits of Control." In J. Bruce, H. Lee, and E. Haites, eds., *Climate Change 1995: Economic and Social Dimensions of Climate Change*. Cambridge: Cambridge University Press.
- Quiggin, J., and J. Horowitz. 1999. "The Impact of Global Warming on Agriculture: A Ricardian Analysis: Comment." *American Economic Review* 89(4): 1044–45.
- Reilly, J., N. Hohmann, and S. Kane. 1994. "Climate Change and Agricultural Trade: Who Benefits, Who Loses?" *Global Environmental Change* 4(1): 24–36.
- Reilly, J., W. Baethgen, F. E. Chege, S. C. van de Greijn, L. Ferda, A. Iglesias, C. Kenny, D. Patterson, J. Rogasik, R. Rotter, C. Rosenzweig, W. Sombroek, and J. Westbrook. 1996. "Agriculture in a Changing Climate: Impacts and Adaptations." In R. Watson, M. Zinyowera, R. Moss, and D. Dokken, eds., *Climate Change 1995: Intergovernmental Panel on Climate Change Impacts, Adaptations, and Mitigation of Climate Change*. Cambridge: Cambridge University Press.
- Ricardo, D. 1815. *An Essay on Profits*. London: John Murray. [www.econlib.org/library].
- Ritchie, J. T., and D. NeSmith. 1991. "Temperature and Crop Development." In J. Hanks and J. Ritchie, eds., *Modeling Plant and Soil Systems* 31. Madison, Wis.: American Society of Agronomy.

- Rosenzweig, C., and M. Parry. **1994**. "Potential Impact of Climate Change on World Food Supply." *Nature* 367(6459): **133–38**.
- Schlenker, W., M. Hanemann, and A. Fischer. **2005**. "Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach." *American Economic Review* 95(1): 395–406.
- . **2006**. "The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions." *Review of Economics and Statistics* 88(1): **113–25**.
- Seo, N., R. Mendelsohn, and M. Munasinghe. **2005**. "Climate Change and Agriculture in Sri Lanka: A Ricardian Valuation." *Environment and Development Economics* 10(5): **581–96**.
- Tol, R. **2002**. "Estimates of the Damage Costs of Climate Change. Part 1: Benchmark Estimates." *Environmental and Resource Economics* 21: **47–73**.
- Weng, F., and N. Grody. **1998**. "Physical Retrieval of Land Surface Temperature Using the Special Sensor Microwave Imager." *Journal of Geophysical Research* 103(D8): 8839–48.
- World Bank. **2003**. "Africa Rainfall and Temperature Evaluation System (ARTES)." Washington, D.C.

Microenterprise Dynamics in Developing Countries: How Similar are They to Those in the Industrialized World? Evidence from Mexico

Pablo Fajnzylber, William Maloney, and Gabriel Montes Rojas

A rich panel data set from Mexico is used to study the patterns of entry, exit, and growth of microenterprises and to compare these with the findings of the mainstream theoretical and empirical work on firm dynamics. The Mexican self-employment sector is much larger than its counterpart in the United States, which is reflected in higher unconditional rates of entry into the sector. The evidence for Mexico points to the significant presence of well-performing salaried workers among the likely entrants into self-employment, as opposed to the higher incidence of poorer waged workers among the entrants into the U.S. self-employment sector. Despite these differences, however, the patterns of entry, survival, and growth with respect to age, education, and many other covariates are very similar in Mexico and the United States. These strong similarities suggest that mainstream models of worker decisions and firm behavior are useful guides for policymaking for the developing-country microenterprise sector. Furthermore, they suggest that, as a first approximation, the developing-country microenterprise should probably be viewed as they are in the advanced countries as offering potentially desirable job opportunities to low-productivity workers.

This article examines whether microenterprises in developing countries behave similarly to their industrial country counterparts or whether they represent a separate phylum altogether. In industrial countries, the last two decades have seen the emergence of a set of stylized facts about the personal and firm characteristics associated with entry into self-employment, the survival and growth of existing microenterprises, and theoretical frameworks to explain them. However, notwithstanding the increasing importance given to the

Pablo Fajnzylber is senior economist in the Private Sector Development unit for Latin America and the Caribbean at the World Bank; his email address is pfajnzylber@worldbank.org. William F. Maloney is lead economist in the Office of the Chief Economist for Latin America at the World Bank; his email address is wmaloney@worldbank.org. Gabriel Montes Rojas is a PhD student in economics at the University of Illinois, Urbana-Champaign; his email address is rmontes@uiuc.edu. This research was financed by the regional studies program, Office of the Chief Economist for Latin America and the Caribbean at the World Bank. The authors are grateful for helpful discussions with Guillermo Perry, Chris Woodruff, and members of the 2006 Latin American Flagship Team on Informality. The article has also been much improved, thanks to the comments of the journal editor and three anonymous referees.

promotion of microenterprises and small enterprises in development policy circles, there has been no systematic attempt to see how their dynamics in developing countries approximate those described in the mainstream literature.'

This represents a loss on two fronts. First, if behavioral differences are not so great, development policymakers have a wealth of analytical frameworks at their disposal for informing analytical work and policymaking on microenterprises. Second, a finding of kinship with industrial country counterparts would provide additional evidence on the debate on how to conceive of the role of the informal microenterprise in the developing world. In Organization for Economic Cooperation and Development (OECD) countries, opening a business and being one's own boss is often celebrated as a desirable alternative to salaried work, whereas in the developing world, the very large unregulated (informal) self-employed sectors are frequently seen as the disadvantaged segment of a dual labor market where workers queue for good jobs. The two views have different implications for entry and firm dynamics.

This article uses a unique data set for Mexico that permits linking detailed microenterprise survey data with rotating household panel data sets. These data offer a select number of firm and individual characteristics that allow comparison with findings in the OECD firm dynamics literature: firm characteristics such as size and time in business; individual traits of entrepreneurs such as age and education; and, more speculatively, detailed firm information such as that on capital stocks and access to credit. This permits the examination of patterns of entry and exit and determinants of performance and survival. Although comparisons with the canonical studies for industrial countries are necessarily imperfect, there is a surprising degree of similarity.

I. CONCEPTUAL CONSIDERATIONS

The mainstream theoretical and empirical literatures have proceeded in tandem and offer both stylized facts and conceptual foundations that can be used in the analysis of the Mexican data. The brief review that follows serves mostly to motivate the variables included in the analysis and to provide benchmarks

1. Only 6 of the 53 articles mentioned by Blanchflower (2004) in his self-employment literature review are on developing countries, and they focus mostly on the determinants of earnings. To our knowledge, the only previous evidence on the determinants of entry, exit, and growth of microenterprises in developing countries are the papers on Africa by Mead and Liedholm (1998), Liedholm and Mead (1999), Liedholm (2002), McPherson (1995, 1996), and Goedhuys and Sleuwaegen (2000). There have been other recent studies on firm dynamics in developing countries, but they have focused mostly on larger firms. Roberts and Tybout (1997), for instance, worked mainly with industrial surveys, which generally had limited coverage of microenterprises. Aw, Chung, and Roberts (2003) studied Korean and Taiwanese firms with more than 10 employees. Bartelsman, Haltiwanger, and Scarpetta (2004) looked at creative destruction in 24 industrial and developing countries. For most countries (including Mexico), their data cover all plants with at least one employee. However, at least in the case of Mexico, the survey is based on social security registers that, by definition, lead to the exclusion of most microenterprises found in the informal sector.

against which to compare the estimated Mexican parameters. Somewhat more tentatively, it highlights some contrasting predictions with the less-formalized dual labor market literature that permits distinguishing between two broad conceptions of the role of the sector.

The dominant view of the role of self-employment in industrial countries stresses the risk-taking, entrepreneurial nature of the sector, with the celebrated Silicon Valley high-technology start-up at its apex. In the classic framework proposed by Lucas (1978), individuals are endowed with a given—and known—level of entrepreneurial or managerial ability, which determines the returns from self-employment. Individuals with a sufficiently high level of managerial ability become entrepreneurs, whereas the rest become wageworkers. Moreover, there is evidence that, other things equal, some individuals may derive a larger utility from entrepreneurship than from wage work, thus reducing the net opportunity cost of entering self-employment. Evidence on this has been provided by Blanchflower and Oswald (1998), who reported that the self-employed report higher levels of job and life satisfaction and that half of U.S., U.K., and German workers would prefer to be self-employed, and by Hamilton (2000), who finds that nonpecuniary benefits—such as being your own boss—explain the lower conditional earnings generally found in the sector.

This view stands in contrast to the bulk of the developing-country literature that sees self-employment in the informal sector as a holding pattern for those rationed out of better jobs in the salaried sector, particularly young people.² Rather than the patterns of entry and exit and growth associated with entrepreneurship dynamics, self-employment in developing countries is often viewed as disguised unemployment. There is some parallel with the industrial country sociological literature that sees the numerous self-employed among certain ethnic minorities as recruited from among "misfits"^m—individuals who lack access to salaried employment, for instance, because of language barriers, a history of unemployment, or limited labor market experience (see Evans and Leighton 1989; Carrasco 1999 and the references they cite). However, an emerging view with roots in Hart's (1972, 1973) early work on Ghana and Kenya stresses the mounting evidence of entrepreneurial dynamism, voluntary entry, and relative job satisfaction (DeSoto 1989; Maloney 1999, 2004; Bhattacharaya 2002).

2. The dualistic view has intellectual roots perhaps best distilled in Harris and Todaro's (1970) vision of a market segmented by wage setting in the formal sector that leaves the traditional sector rationed out of modern salaried employment. The view of informal urban workers as the inferior or excluded segment became influential in the International Labour Organization and its Latin America affiliate, the Latin America Regional Employment Program (PREALC). See, for instance, Tokman (1978). Generally, this view partitions urban employment into a modern or formal sector (characterized by high productivity growth and job benefits) and a traditional or informal sector. These models typically view the informal sector as essentially stagnant and unproductive, serving as a refuge for the urban unemployed and as a receiving station for newly arriving rural migrants. However, sociological research on Mexico stresses the higher status and desirability of the sector (Balán, Browning, and Jelin 1973; Maloney 1999).

Clearly, as in industrial countries, the sector is very heterogeneous. Fields (1990) noted the presence of "upper" and "lower" tiers among the informal firms, and Cunningham and Maloney (2001), using the same data employed here, identified several distinct subsectors that differ greatly by, among other characteristics, productivity, demographics, and reason for entry. Fully cognizant, therefore, of the need for nuance in approaching the analysis, what this article seeks to establish is whether, as a first approximation, the sector more closely approximates the mainstream view or whether, in fact, it appears as a pathological outcome of regulatory distortions requiring entirely different analytical tools.

Growth and Failure

The literature commonly finds that self-employment is extraordinarily risky, and several mainstream articles on industry dynamics have attempted to provide conceptual support for the observed patterns of microenterprise mortality. Jovanovic (1982) added dynamics to Lucas's view by further assuming that managerial ability is uncertain and that individuals can only gradually learn about a firm's true cost structure by opening and operating a business. Entry into self-employment involves a fixed cost, which only those with high expected ability and profits may be willing to pay.

After entry, entrepreneurs incorporate the information from their actual profits, revise their ability estimates, and adjust the level of profit-maximizing output accordingly. Firms with consistently lower-than-expected profits tend to contract and eventually go out of business, whereas firms with unexpectedly high profits cause entrepreneurs to revise their ability estimates upward and expand. Over time, survivors obtain more precise ability estimates and approach their steady-state size. Their behavior becomes less volatile: they fail less frequently, but they also expand less rapidly. Empirical evidence favoring these basic predictions on the negative links between time in business and size, on one hand, and firm exit and growth, on the other, was obtained for the United States by Evans (1987a, 1987b) and Dunne, Roberts, and Samuelson (1988, 1989); for Germany by Wagner (1994); and for the United Kingdom by Geroski (1991).

Other dynamic models also generate these patterns, although with different analytical structures. Ericson and Pakes (1989, 1995) proposed a model of active exploration—as opposed to the passive learning assumption in Jovanovic's (1982) model—that incorporates firm-specific sources of uncertainty derived from stochastic outcomes of investments made by firms to improve their profitability. Favorable outcomes from the firms' own investments—including those that led to entry into the industry—tend to move them toward "better" states, whereas good outcomes of direct competitors move them to less profitable states.³ In this context, profits and, somewhat in contradiction to the Jovanovic

3. As in Jovanovic's (1982) model, entry, exit, and investment decisions are made to maximize the expected discounted value of future net cash flows conditional on the current information set.

model, higher capital stocks should decrease the probability of failure and be positively related to firm growth, all else being equal.

Entry

Hopenhayn's (1992) model of industry dynamics also generates broadly similar implications to Jovanovic's (1982) model and is especially notable for its analysis of the effect of changes in the cost of entry, which could be interpreted as an outside opportunity cost for some resources (for example, managerial ability) used by the firm. Higher costs of entry lead to a lower turnover rate because more *ex ante* selection occurs. This could be particularly relevant in developing-country contexts characterized by higher levels of informality and lower productivity in the salaried sector, which in Hopenhayn's model would lead to a lower entrepreneurial ability threshold for entering self-employment and thus to higher entry and exit rates.

Offsetting this effect, however, may be other constraints to entry that may be exacerbated in the developing world, and this provides another contrasting prediction that enables distinguishing between different visions of the sector. Johnson (1978) and Jovanovic (1979) postulated that young people are likely to be less risk-averse and hence should be overrepresented among those entering self-employment, a pattern that would be observationally equivalent to the standard queuing view of developing-country labor markets. Young people rationed out of formal labor would tend to enter informal self-employed work. However, the reverse pattern is found in the United States: entry increases with age. As an explanation, Evans and Jovanovic (1989) offered a variation of Lucas's (1978) model in which binding liquidity constraints may lead individuals to delay or forgo profitable business opportunities, reducing entry rates and increasing exit rates among those with low personal assets—disproportionately the young. Offering something of a contrasting prediction to that of Ericson and Pakes (1989, 1995), they argue that because credit-constrained individuals are more likely to start small businesses with a suboptimal amount of capital, returns to capital will be higher, and smaller firms (with lower capital stocks) will grow faster than firms that entered closer to their steady state.

Either model of credit constraints to voluntary entry is largely inconsistent with the dualistic view of self-employment as an easy entry-holding pattern in several ways. Both the view of the self-employed as misfits and the dualistic view of informal labor markets would predict lower entry rates into self-employment coming from salaried work than from unemployment or from out of the labor market. However, the opposite prediction could be derived from the Evans and Jovanovic's (1989) view if individuals acquire more capital and knowledge of business opportunities—and to some extent their own managerial ability—while working than while unemployed or out of the labor market. In particular, if formal schooling is relatively poor and most relative human capital is accumulated on the job, salaried employment may be a logical stepping-stone to self-employment.

Furthermore, the impact of the level of remuneration in previous jobs also potentially offers some insights that may help distinguish mainstream and dualistic models. All things equal, workers earning higher wages in salaried work would be less likely to be misfits, or unsuited to formal work, and so less prone to move into self-employment. However, in the presence of credit constraints, workers earning higher wages in the salaried sector may also be able to accumulate capital faster and so be more likely to enter self-employment. Moreover, there may be a correlation between previous productivity in the salaried sector, and thus level of remuneration, and entrepreneurial ability or at least competence in the chosen field of entrepreneurship. To the degree that this would imply a higher probability of success in self-employment, people with conditionally higher earnings in the formal sector might be expected to enter self-employment, whereas people with less skills may choose not to take the risks.

In sum, the impact of personal characteristics of existing and would-be entrepreneurs can help distinguish between the two views of the sector. In the segmented labor market scenario, unemployed individuals, young workers, those out of the labor force, and those with less schooling and lower wages should all be more likely to be self-employed, as they would be worse positioned for finding formal salaried jobs. In contrast, the mainstream literature suggests that older, better-educated, and well-paid workers with experience in the salaried sector should have a higher probability of entering, staying in, and growing in the self-employment sector as they should be more likely to have accumulated the assets required to start a business and be better positioned to find, assess, and take advantage of good business opportunities.

Finally, two additional covariates appear in the mainstream literature but yield ambiguous predictions. First, workers with more schooling might be expected to find better matches as salaried workers in larger firms that could better use specialized skills. By contrast, Rees and Shah (1986) and Cressy (1996) argued that the costs of assessing business opportunities may be lower for more educated individuals and that human capital may be a complement to managerial ability. This is suggested by Bates (1999), who showed that in the United States, the probability of survival of small businesses is positively related to the level of education of their owners.

Second, Carrasco (1999) argued that men who are married could be less willing to take risks. By contrast, Mexican sociologist González de la Rocha (1994) suggests that the possibility of combining the self-employment earnings of the household head or spouse with the salaries of other family members could reduce overall household income risk.

II. DATA AND METHODOLOGY

The analysis employs two data sets provided by the Mexican Statistical Institute (INEGI). The first is the National Urban Employment Survey [Encuesta Nacional

de Empleo Urbano (ENEU)], which follows workers across a five-quarter period in a rotating panel framework. This survey records age, gender, education, marital status, labor market status, earnings, and some general characteristics of the jobs held by economically active individuals, such as firm size. The first and fifth interviews of individuals covered in this survey between 1987 and 2001 are linked to construct a two-period panel data set that permits analysis of entry, survival, and growth in the self-employment sector. The labor market status of the surveyed individuals is established based on the characteristics of their main job during the week preceding the survey. Individuals who report that in their main job they were either "employers" (patrones) or "own-account workers" (*trabajador por su cuenta*) are defined as self-employed.⁴

The second data source is the National Survey of Microenterprises [Encuesta Nacional de Micronegocios (ENAMIN)], which in 1992, 1994, 1996, and 1998 reinterviewed a sample of the self-employed individuals covered in previous rounds of the ENEU. These surveys ask detailed questions on the characteristics of firms with up to five employees (15 in manufacturing), including information on capital stock, time in business, and access to credit from formal and informal sources, both for starting the business and at a later time. To study the effects of these characteristics on the survival and growth of microenterprises, we linked the ENAMIN sample with the panel data set constructed from the ENEU data.

To study the determinants of entry into self-employment, we first constructed a random sample of 30 percent of all men aged 15–65 years covered in the longitudinal data set constructed from the ENEU.⁵ Analysis is restricted to the behavior of self-employed men and male-headed microenterprises, which represent, respectively, 65 percent of Mexican self-employed and 70 percent of the microenterprises surveyed by ENAMIN. This sample encompasses individuals who are out of the labor force, unemployed, or working. The working group includes both salaried and contract workers, as well as individuals working without pay on cooperatives and in other unspecified jobs. The subsample of salaried individuals is also used to investigate the role of conditional wages and firm size.

In the same vein as Evans and Leighton (1989), these data are used to estimate a probit model of the determinants of entry into self-employment, looking in particular at the role of the age of potential entrepreneurs, their education levels,

4. In a nutshell, the self-employed category includes all individuals whose main job consists of working in their own businesses. This is similar to the standard definition used in the U.S. self-employment literature. Evans and Leighton (1989), for instance, define as self-employed all sole proprietors, partners, and sole owners of incorporated businesses. ENEU provides information on whether individuals have a second job but provides no income information for that job. Thus, it is not possible to calculate the fraction of total income derived from self-employment, which is why the self-employed are distinguished from other individuals based only on whether self-employment is the primary job and not on a particular cutoff for the ratio of self-employment to wage income, because that information is not available.

5. Preliminary explorations suggest some differences in the dynamics of female-headed microenterprises, but these are not examined in this article.

previous labor market positions, and conditional wages. As argued above, many of these personal characteristics have a priori ambiguous effects on the likelihood of entering and succeeding in the self-employment sector, and there is little evidence on the signs of these effects in a developing-country context. The estimated model is of the form:

$$SE_i = 1 \quad \text{if } \alpha_1 \text{Age}_i + \alpha_2 \text{Educ}_i + \alpha_3 \text{Married}_i + \alpha_4 \text{Labor}_i + \alpha_5 \text{Empsize}_i \\ + \alpha_6 \text{Wage}_i + \alpha_7 \text{Sector}_i + \alpha_8 \text{State}_i + \alpha_9 \text{Year}_i + \varepsilon_i > 0$$

$$(1) \quad SE_i = 0 \quad \text{otherwise}$$

where SE takes the value 1 if individual *i* becomes self-employed between the initial and the final ENEU interview and 0 otherwise, *Age* is a vector of categorical variables for various age ranges, *Educ* is a set of dummy variables that represent individuals' levels of schooling, *Married* is a dummy variable for married individuals, and *Labor* represents a vector of variables that describe individuals' labor market status, including whether they are out of the labor force, unemployed, or working. The vector *Labor* also describes whether inactive individuals are studying or not; whether the unemployed have been in that status for one, two, or more quarters; and whether those who are working are doing so for a salary, as contract workers, as workers without pay, or in some other modality (for example, through cooperatives). When the estimation is performed using only the sample of salaried workers, the variables *Empsize* and *Wage* are added to represent the size of the firm for which they worked at the time of the initial ENEU interview and their monthly salary in that firm. All specifications include sector dummy variables (for those who are employed), as well as state and year dummy variables, and a random error term.⁶

A maximum likelihood two-stage probit model with selection correction is also estimated for the determinants of entering self-employment conditional on an individual having left an initial labor market position. This approach is motivated by the fact that some personal characteristics may have a different effect on the probability of an individual changing labor market positions and entering self-employment conditional on having changed positions. Thus, for instance, unemployed workers could have a larger nonconditional probability of entering self-employment than their salaried counterparts, but it could be driven by the larger turnover rate among unemployed individuals and not by a larger conditional probability of entering self-employment—only 12 percent of the unemployed remain in that status after one year compared with almost 68 percent for salaried workers. Thus, this second estimation approach has the purpose of testing the robustness of the probit results with respect to the critique that they could be driven by the larger turnover of disadvantaged workers.

6. A more detailed description of the variables used in the article is in the Appendix.

Incidentally, this critique could also apply to Evans and Leighton's (1989, p. 58) probit result that individuals "who have changed jobs frequently" have a higher likelihood of entering self-employment.⁷

The first stage estimates the probability of changing labor market status between the initial and the final survey interview. The second stage, for individuals who did change labor market status, estimates the probability of entering self-employment while controlling for possible correlation between the error terms of the first- and second-stage models. The right-side variables are the same as those in equation (1), for both the first- and the second-stage equations, but added to the first-stage selection equation to identify the model is a dummy variable for individuals who have migrated to different municipalities or states; indicator variables for the second, third, and fourth quarters of each year; and dummy variables for the 99 industry groups in the Mexican classification of economic activities.⁸

Two data sets are used to explore the personal and firm characteristics associated with survival and growth of employment in the self-employment sector: the first comprises all self-employed individuals covered in the ENEU data set and the second a sample restricted to those in the linked ENEU–ENAMIN data set. While smaller, the second sample has the advantage of providing information not only on age, educational attainment, marital status, earnings, and firm size but also on capital stocks, access to credit, and time in business. When ENAMIN data are used to estimate the impact of these variables on survival and growth in self-employment, firms for which unpaid workers represent more than 50 percent of total employment are excluded.

A probit model is used to estimate the determinants of staying in self-employment, and a simple ordinary least squares estimator is used to study the correlates of microenterprise growth. The estimated model in the case of firm survival is of the following form:

$$\begin{aligned}
 \text{Survival}_i &= 1 \text{ if } \beta_1 \text{ Age}_i + \beta_2 \text{ Educ}_i + \beta_3 \text{ Married}_i + \beta_4 \text{ Earnings}_i \\
 &\quad + \beta_5 \text{ Size}_i + \beta_6 \text{ Time}_i + \beta_7 \text{ Capital}_i + \beta_8 \text{ Credit}_i + \beta_9 \text{ Sector}_i \\
 (2) \quad &\quad + \beta_{10} \text{ State}_i + \beta_{11} \text{ Year}_i + \varepsilon_i \\
 &\quad > 0 \\
 \text{Survival}_i &= 0 \text{ otherwise}
 \end{aligned}$$

where *Survival* takes the value 1 if individual *i* remains self-employed between the initial and the final ENEU interviews and 0 otherwise, *Earnings* represents the net

7. It must be noted, however, that the robustness check does not correct for the lack of information on the turnover of salaried workers within the salaried sector. Thus, as pointed out by one referee, in the sample of workers who changed labor market position, the salaried sector is excluded from the possible end-points of workers that begin in that sector, whereas it remains an important potential destination—relative to self-employment—for other types of workers, such as the unemployed.

8. Changes in labor market position are found to be more likely for individuals with a migration history and in the third and fourth quarters of the year.

monthly income from self-employment, Size is a set of dummy variables that represent the employment size of the corresponding microenterprise, Time is the number of years during which it has been in business, Capital is the value of the firm's capital stock, and Credit is a vector of variables representing whether the firm received credit from formal or informal sources, either when starting up or later. The variables Age, Educ, and Married are defined as before and so are the sector, state, and year dummy variables, which are included in all specifications.

A very similar model, with the same explanatory variables, is estimated for employment growth:

$$(3) \quad \begin{aligned} \text{Growth}_i = & \gamma_1 \text{Age}_i + \gamma_2 \text{Educ}_i + \gamma_3 \text{Married}_i + \gamma_4 \text{Earnings}_i \\ & + \gamma_5 \text{Size}_i + \gamma_6 \text{Time}_i + \gamma_7 \text{Capital}_i + \gamma_8 \text{Credit}_i \\ & + \gamma_9 \text{Sector}_i + \gamma_{10} \text{State}_i + \gamma_{11} \text{Year}_i + \varepsilon_i > 0 \end{aligned}$$

where Growth is the rate of employment growth of the firm owned by the self-employed individual i . This equation was also estimated simultaneously with a first-stage model of selection into the sample of microenterprises that remain in business over the one-year period during which firm growth is being measured. These results are not reported as they are virtually identical to those obtained by the simpler method described above.⁹

III. EMPIRICAL RESULTS

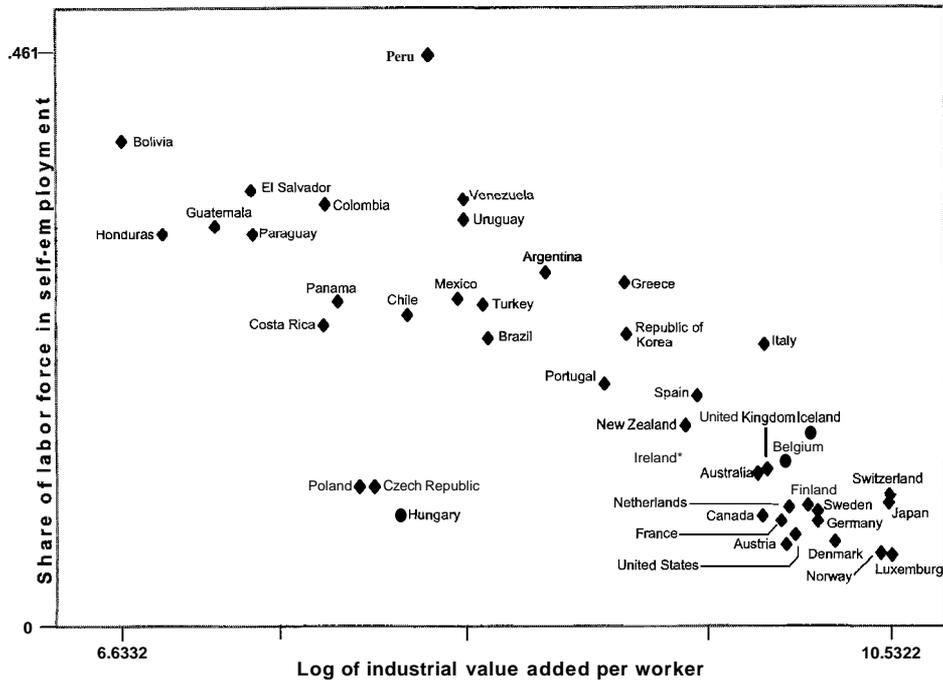
Figure 1 shows the size of the self-employed sector in Mexico relative to that in other developing and advanced countries. What is immediately apparent is that Mexico has a large share of its work force in self-employment, although it is not particularly larger than would be predicted for a country of its level of development." Although it may be that high levels of labor market distortion or poor regulation are correlated with underdevelopment and give rise to the large sector size, for several reasons, it seems more likely that it is the lower opportunity cost of entering a desirable sector, self-employment, that drives it.

First, Blau (1987) argues that this negative correlation of self-employment with development appears clearly in the United States over the century preceding the 1970s as well as in France, the Federal Republic of Germany, Italy, and Japan because of the emergence of economies of scale and specialization of labor possible only in relatively large enterprises. Consistent with this view and Maloney's (2004) estimates based on figure 1, relative total factor productivity

9. Models were also estimated that control for possible attrition biases associated with the fact that not all the individuals could be linked across the ENEU surveys. Once again, this correction leads to virtually no changes in estimation results.

10. Some care is in order in interpreting figure 1. First, it is drawn from surveys of differing degrees of coverage of rural and urban areas. Among surveys that were only urban were those in Argentina, Bolivia, Colombia, Mexico, Paraguay, and Uruguay or roughly half of the Latin American and Caribbean data set. The data from the OECD are economy-wide, but the share of the work force in rural areas is quite small.

FIGURE 1. Self-Employment and Industrial Productivity



Source: Authors' calculations using household surveys from the mid-1990s and World Bank Institute data.

emerges as statistically important, explaining the long-term secular decline as well as the rise in self-employment after the mid-1970s.

Second, although Blanchflower (2004) documents that self-employment is risky and stressful, as noted earlier, Blanchflower and Oswald (1998) also note that 63 percent of salaried workers in the United States, 48 percent in the United Kingdom, and 49 percent in Germany report that they would prefer to be self-employed, a fraction similar to the more than 60 percent of Mexican entrants to self-employment who report doing so voluntarily (see Maloney 1999). Furthermore, Blanchflower (2004) reiterates the robust finding that those who enter self-employment report higher levels of job satisfaction than employees. The fact that industrial country workers do not actually pursue their desire to become self-employed may reflect, again, the higher opportunity cost of doing so.

Finally, and consistent with the documented heterogeneity of the microenterprise sector, there is clearly room for segmenting labor market distortions or other regulatory distortions to drive sector size as well, although it is not yet clear how large a role these play. As a crude first pass, the celebrated difference in social models in France and the United States leads to no difference in the share of the work forces in self-employment in the two countries. That said,

Blau (1987) does find rising marginal income tax rates, along with rising total factor productivity in self-employment, to drive the U.S. increase in male self-employment, from 10 to 13 percent across the decade of the 1970s. Antunes and Centeno (2005) find that higher labor flexibility in the OECD countries decreases the probability of entry into self-employment, although they do not offer an impact on sector size. Maloney (2001), explaining the variance in figure 1, and Loayza, Oviedo, and Serven (2005), explaining cross-country variance in levels of informality more generally, find statistically significant impacts of labor regulation, but the level of development plays a much larger explanatory role."

Table 1 succinctly summarizes the structure and dynamics of the Mexican labor market. In addition to summarizing the breakdown of self-employment by the number of workers of the corresponding firms, Table 1 describes the composition of the Mexican labor force by employment status (e.g. self-employment, salaried work, and unemployment). The table includes a separate category for the aggregate of self-employed individuals (regardless of firm size). The lower panel shows the probabilities of transitioning across different labor market statuses. Rows represent individuals' initial labor market positions and columns their labor market statuses one year later. Each cell shows the percentage of individuals who start in a given row category and end in the corresponding column category.

If the United States is taken as a benchmark, the rates of entry and exit into self-employment in Mexico are of very similar orders of magnitude. Mexican entrepreneurs are slightly more likely to return to wage work—at a rate of 15 percent compared with 13.8 percent reported by Evans and Leighton (1989) for the United States. A slightly higher fraction of Mexican waged workers enter self-employment: 6.2 percent compared with 4 percent a year found by Evans and Leighton, and perhaps this leads to a poorer selection of entrepreneurs at the margin, as Hopenhayn (1992) suggests. Note, however, that because in the steady state, workers exiting self-employment must be replaced each year by the same number of new entrants, the higher rate of entry into self-employment from the wage sector that is found in Mexico is partly because of the relatively smaller size of the wage sector there.

Table 1 also summarizes that many microentrepreneurs originate and move to labor market positions other than salaried work. Thus, for instance, 12.4 percent of contract workers enter self-employment in a given year—twice the fraction of their salaried counterparts—and so do 11 percent of the unemployed. Moreover, although 15 percent of the self-employed move to salaried work during a given year, the same number of individuals migrates to other labor market positions—to contract work (6.2 percent) and out of the labor force (5.7 percent). As for microenterprise growth, table 1 reveals that very few microenterprises are able to expand their employment, at least over one-year periods. As an example, only

11. Antunes and Centeno (2005) also find regulation issues affecting the transition into and out of self-employment in OECD countries but do not follow through on the implications for sector size.

TABLE 1. Structure and Dynamics of the Mexican Labor Market (percent)

| | Own Account | 1-4 Workers | 5-9 Workers | 10 or More Workers | Total Entrepreneurs | Salaried ^a | Contract ^a | Salaried Plus Contract ^a | Unemployment | Other Work | Out of Labor Force | Total |
|---------------------------|----------------|-----------------------|-----------------------|----------------------------------|------------------------|-----------------------|-----------------------|---|--------------|---------------|--------------------------|-------|
| Share of population | 9.6 | 5.7 | 0.6 | 0.6 | 16.5 | 46.8 | 5.8 | 52.6 | 2.0 | 2.3 | 26.6 | 100 |
| Share of labor force | 13.1 | 7.8 | 0.8 | 0.8 | 22.4 | 63.8 | 7.9 | 71.7 | 2.7 | 3.2 | | 100 |
| Own account | 51.9 | 12.4 | 0.5 | 0.2 | 65.0 | 17.0 | 7.4 | 24.5 | 1.5 | 1.9 | 7.2 | 100 |
| 1-4 workers | 22.1 | 49.2 | 3.9 | 1.5 | 76.6 | 11.7 | 5.0 | 16.7 | 0.8 | 2.0 | 3.9 | 100 |
| 5-9 workers | 7.8 | 35.1 | 22.6 | 13.1 | 78.6 | 13.5 | 3.2 | 16.8 | 0.6 | 1.3 | 2.7 | 100 |
| 10 or more workers | 4.1 | 15.2 | 14.4 | 44.6 | 78.3 | 15.2 | 2.3 | 17.5 | 0.3 | 1.2 | 2.7 | 100 |
| Total entrepreneurs | 38.4 | 26.0 | 2.9 | 2.6 | 69.9 | 15.0 | 6.3 | 21.3 | 1.2 | 1.9 | 5.7 | 100 |
| Salaried | 4.2 | 1.7 | 0.2 | 0.2 | 6.2 | 67.7 | 3.9 | 71.6 | 1.6 | 3.7 | 17.0 | 100 |
| Contract | 9.0 | 3.1 | 0.1 | 0.1 | 12.4 | 29.8 | 35.8 | 65.7 | 1.5 | 3.4 | 17.0 | 100 |
| Salaried plus contract | 4.7 | 1.8 | 0.2 | 0.2 | 6.9 | 63.5 | 7.4 | 70.9 | 1.6 | 3.6 | 17.0 | 100 |
| Unemployment | 9.3 | 1.5 | 0.1 | 0.1 | 11.0 | 37.8 | 5.0 | 42.8 | 12.0 | 3.6 | 30.5 | 100 |
| Other work | 4.6 | 3.4 | 0.2 | 0.1 | 8.3 | 19.1 | 3.4 | 22.5 | 1.3 | 31.4 | 36.5 | 100 |
| Out of labor force | 2.6 | 0.8 | 0.0 | 0.0 | 3.5 | 9.7 | 1.3 | 11.0 | 1.7 | 2.9 | 80.9 | 100 |

^aBecause of the potential similarities between salaried and contract workers, these categories are included separately and as a combined category.

Source: Authors' analysis based on pooled Encuesta Nacional de Empleo Urbano (ENEU) **1987-2001** data.

The entries that are in bold in Table 1 correspond to the diagonal of the lower panel; the numbers are the percentage of workers that remain in the same labor market status between the initial and final period.

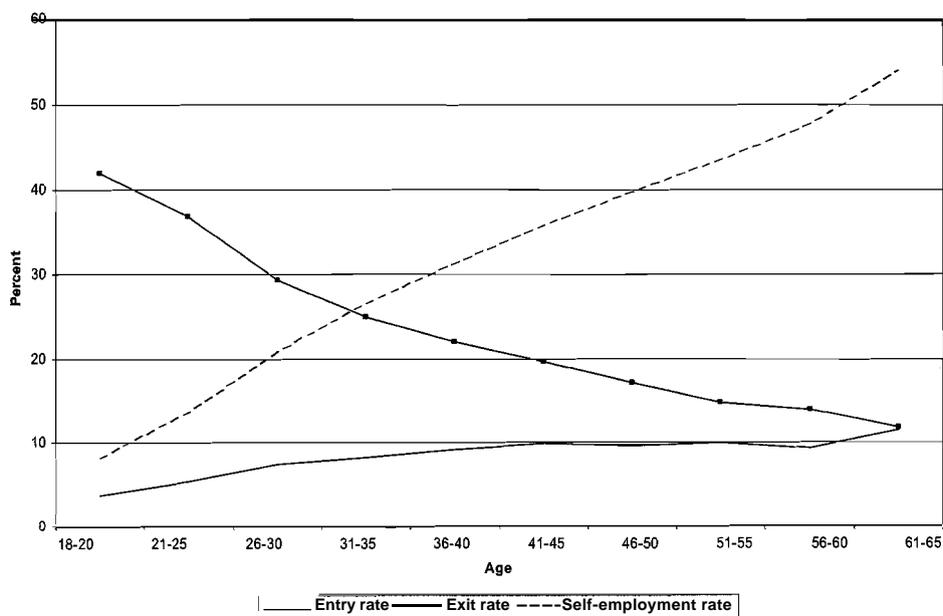
13.1 percent of microentrepreneurs with five to nine employees exhibit any expansion (to having at least 10 workers), 22.6 percent remain in the same size range, and 42.9 percent report a reduction in the number of employees. Own-account workers have similar rates of expansion, offset by a substantial decrease in employment among firms with one to four workers.

Figures 2 and 3 show respectively for Mexico and the United States, by age group, the fraction of the labor force in self-employment, the rate of entry from salaried work into self-employment, and the rate of exit from self-employment. Broadly consistent with figure 1, the self-employment rate in Mexico is at least twice as high as in the United States in every age group. The overall age patterns of self-employment rates as well as those of entry and exit are strikingly similar in the two countries. As a foreshadowing of the more detailed exercises below, the common upward and downward sloping relationships of age with rates of entry and exit are consistent with the view that older entrepreneurs get a more precise view of their underlying entrepreneurial capacity and are more likely to enter and less likely to fail than younger individuals.

Patterns of Entry

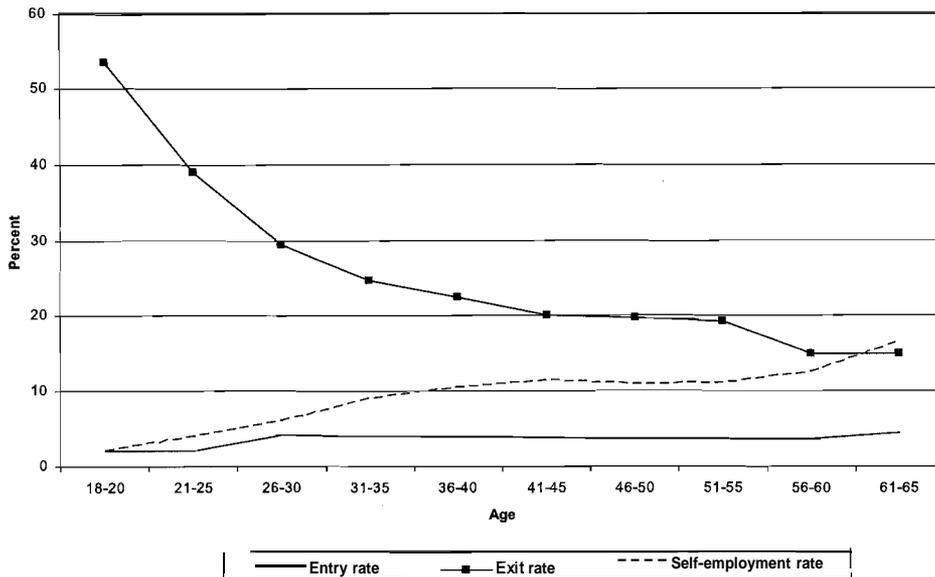
Table 2 presents the results of a more systematic examination of patterns of entry into self-employment. Columns 1–3 report probit estimates of a model of

FIGURE 2. Patterns of Entry and Exit into Self-Employment in Urban Mexico



Source: Authors' calculations using Encuesta Nacional de Empleo Urbano (ENEU) data.

FIGURE 3. Patterns of Entry and Exit into Self-Employment in the United States



Source: Evans and Leighton (1989).

the determinants of becoming self-employed over a one-year period, as measured in the final ENEU interview. It uses a sample that excludes all individuals who were self-employed in the initial ENEU interview and includes both active—although not self-employed—and out of the labor force individuals. Columns 4–6 report probit estimates of a similar entry model by using a restricted sample of individuals who were salaried workers at the time of the first ENEU interview. While the reported results are obtained without differentiating those who enter self-employment with or without employees, similar models were also estimated that define entry more restrictively as the act of becoming self-employed with at least one employee. With a few exceptions (discussed below) the results are substantially similar to those obtained with the broader definition of self-employment.

Columns 1 and 4 present standard probit estimates of the effects of personal characteristics on the probability of entering self-employment for the wider sample of all non-self-employed individuals and for the sample of salaried workers. Rather than probit coefficients, the two columns report marginal changes in the probability of entering self-employment resulting from discrete changes in the independent dummy variables and from infinitesimal changes in the only

TABLE 2. Determinants of Entry into Self-Employment

| | Entry (1) | Entry/Moving (2) | Moving (3) | Entry (4) | Entry/Moving (5) | Moving (6) |
|----------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Ages 21–35 years | 0.057*** (0.000) | 0.660*** (0.000) | -0.068*** (0.000) | 0.044*** (0.000) | 0.415*** (0.000) | -0.130*** (0.000) |
| Ages 36–50 years | 0.090*** (0.000) | 0.910*** (0.000) | -0.044** (0.012) | 0.070*** (0.000) | 0.551*** (0.000) | -0.030 (0.200) |
| Ages 51–65 years | 0.083*** (0.000) | 0.753*** (0.000) | -0.049** (0.017) | 0.076*** (0.000) | 0.446*** (0.000) | 0.142*** (0.000) |
| Schooling, 6–12 years | -0.016*** (0.000) | -0.092*** (0.000) | -0.044*** (0.000) | -0.007*** (0.001) | -0.060*** (0.009) | 0.008 (0.608) |
| Schooling, 13 plus years | -0.012*** (0.000) | 0.020 (0.455) | 0.022 (0.145) | 0.002 (0.471) | 0.066** (0.038) | 0.127*** (0.000) |
| Married | 0.010*** (0.000) | 0.413*** (0.000) | -0.228*** (0.000) | 0.014*** (0.000) | 0.223*** (0.000) | -0.189*** (0.000) |
| Out of labor force, not studying | 0.075*** (0.000) | -0.061 (0.283) | 1.139*** (0.000) | | | |
| Out of labor force, studying | 0.006" (0.099) | -0.307*** (0.000) | 0.643*** (0.000) | | | |
| Unemployed | 0.092*** (0.000) | -0.749*** (0.000) | 2.223*** (0.000) | | | |
| Unemployed for 4–6 months | 0.023** (0.023) | 0.183*** (0.039) | 0.071 (0.391) | | | |
| Unemployed for 7 plus months | 0.012 (0.170) | 0.115 (0.149) | -0.050 (0.471) | | | |
| Contract worker | 0.055*** (0.000) | -0.505*** (0.000) | 0.939*** (0.000) | | | |
| Nonpaid worker | 0.098*** (0.000) | -0.341*** (0.000) | 1.129*** (0.000) | | | |
| Other worker | 0.072*** (0.000) | -0.345*** (0.000) | 0.908*** (0.000) | | | |
| Employer size, 11–50 workers | | | | -0.033*** (0.000) | -0.394*** (0.000) | -0.281*** (0.000) |
| Employer size, 51–250 workers | | | | -0.042*** (0.000) | -0.578*** (0.000) | -0.357*** (0.000) |
| Employer size, 251 plus workers | | | | -0.081*** (0.000) | -0.721*** (0.000) | -0.517*** |
| Salaried wage (log) | | | | 0.007*** (0.000) | 0.076*** (0.000) | -0.071*** (0.000) |
| Commerce | 0.034*** (0.000) | 0.173*** (0.000) | 0.272*** (0.000) | 0.001 (0.749) | 0.035 (0.228) | 5.721*** (0.000) |
| Agriculture | 0.038*** (0.000) | 0.229*** (0.001) | 0.176 (0.179) | 0.021*** (0.002) | 0.284*** (0.000) | 4.621*** (0.000) |
| Construction | 0.126*** (0.000) | 0.538*** (0.000) | 0.396*** (0.000) | 0.075*** (0.000) | 0.563*** (0.000) | 6.036*** (0.000) |
| Services | 0.025*** (0.000) | 0.144*** (0.000) | 0.302*** (0.000) | 0.008*** (0.000) | 0.085*** (0.001) | 5.838*** (0.000) |
| Government | -0.005 (0.204) | 0.254*** (0.000) | -0.031 (0.781) | 0.020*** (0.000) | 0.229*** (0.000) | 5.632*** (0.000) |

(Continued)

TABLE 2. Continued

| | Entry (1) | Entry/Moving (2) | Moving (3) | Entry (4) | Entry/Moving (5) | Moving (6) |
|------------------------|-----------|-----------------------|------------|-----------|----------------------|------------|
| Log-likelihood | -24,257 | -67,235 | | -13,348 | -33,896 | |
| Pseudo R-squared | 0.081 | | | 0.096 | | |
| Rho statistic | | -0.198 ^{***} | | | 0.883 ^{***} | |
| Number of observations | 104,411 | 104,411 | 104,411 | 61,418 | 61,418 | 61,418 |

^{*}Significant at the 10 percent level.

^{***}Significant at the 5 percent level.

^{****}Significant at the 1 percent level.

Note: Numbers in parenthesis are p-values. Columns 1–3: 30 percent random sample of all non-self-employed individuals in their initial ENEU interview. Columns 4–6: 30 percent random sample of all salaried workers in their first Encuesta Nacional de Empleo Urbano (ENEU) interview. Columns 1–4: probit estimates; the marginal change in the probability of entering self-employment is reported (discrete change for dummy variables). Columns 2 and 3 and 5 and 6: maximum likelihood two-stage probit models with a selection correction; coefficient estimates are reported.

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

continuous explanatory variable (Wage). Columns 2 and 5 report the results of probit models estimated for restricted versions of the samples of all non-self-employed and all salaried workers, keeping only individuals who changed labor market status between the first and second ENEU interviews. These equations are estimated simultaneously with first-stage probit models of the determinants of changing labor market status between the two interviews, for which results are reported in columns 3–6. (Columns 2, 3, 5, and 6 report coefficient estimates.)

A clear age pattern emerges across specifications and samples reflecting the rising likelihood of entry until the 36–50 age bracket. Thus, keeping other personal characteristics constant and comparing with individuals aged 15–20 years (for which the rate of entry into self-employment is 2.4 percent; not shown in the table), the probability of entering self-employment is 5.7 percentage points higher for those aged 21–35 years and 9 percentage points higher for those aged 36–50 years. Although the corresponding increases in the probability of entering self-employment are slightly smaller in the sample of salaried workers (column 4), the findings are not consistent with the view of the sector as a point of entry into the labor market. They are, however, consistent with the U.S. data, Evans and Jovanovic's (1989) liquidity constraints hypothesis, and the view that older workers have a more precise measure of their underlying entrepreneurial capacity.

Educational attainment has a negative, if quantitatively small, effect on the probability of entering self-employment. Thus, while the average worker with at most some primary education has a 10.3 percent probability of entering self-employment, similar workers with some secondary education have 1.6 percentage points lower entry probability and those with some tertiary education a 1.2 percentage points lower probability (table 2, column 1). However, when restricting the sample to salaried workers (column 4) or conditioning on changing labor market status (columns 2 and 5), the negative relationship between education and entry into self-employment levels off and breaks up at higher levels of schooling, suggesting that college graduates may find the sector attractive. Moreover, when the definition of self-employment is restricted to business owners who employ at least one worker, both secondary and tertiary schooling are positively linked to entry into self-employment. This result is broadly consistent with Evans and Leighton (1989), Carrasco (1999), and Rees and Shah (1986), who found that entry rises monotonically with education.

Also consistent with Rees and Shah (1986) and Carrasco (1999) is the finding that being married is positively associated with entry, which may reflect either that the sector is not riskier than alternatives available in the labor market (salaried employment) or that being married helps diversify risk. Quantitatively, marriage has the effect of increasing by 1 percentage point the probability of entry into self-employment, which is 4.5 percent on average for nonmarried individuals. Because, as columns 3 and 6 suggest, married men are less likely to

change their initial labor market status, the effect of marital status is much stronger—about four times as large—for conditional estimates.

The importance of conditioning on changes in labor market positions becomes most clear in looking at the impact of initial employment and labor force participation status. The unconditional results suggest that individuals out of the labor force and not studying and those unemployed are more likely to enter self-employment, consistent with traditional dualistic views of the sector as disguised unemployment. Thus, while the rate of entry is 6.2 percent for the average salaried worker, the probability of entry increases for individuals who have been unemployed, for nonpaid workers, for those who are out of the labor force and not studying, for contract workers, and for other types of workers (table 2, column 1).

However, as column 3 shows, these are also labor market positions with very high rates of turnover; so, workers may well be more likely to move to all sectors at disproportionately high rates. After conditioning on changing labor market status, unemployed and nonsalaried workers (contract or nonpaid workers) are less likely to enter self-employment than those in formal salaried employment. Put differently, given that a worker will change an initial labor market position, those in formal salaried employment are more likely than most other groups to enter informal self-employment. It is worth noting, however, that among the unemployed, workers with longer unemployment spells—especially those who have been unemployed for four to six months—have higher conditional and unconditional probabilities of entering self-employment. This suggests that, even if the sector does not function predominantly as a holding pattern for misfits or those rationed out of salaried work, it does offer income opportunities for the long-term unemployed. Nonetheless, those opportunities appear to be associated with opening businesses that are owner-only, as estimates of the probability of opening microenterprises with at least one worker show nonsignificant effects for the duration of unemployment.¹²

The final three columns of table 2 present estimates for the sample of salaried workers and suggest two perhaps contradictory effects of firm size and earnings. While salaried workers employed in firms with at most 10 employees have a 12.8 percent probability of entering self-employment, that probability diminishes by between 3.3 and 4.2 percentage points for those coming from firms with 11–250 workers and by 8.1 percentage points for those previously employed in firms with more than 250 employees. This finding may be because of nonpecuniary benefits (such as social security) and job stability offered by larger firms. That the same pattern emerges conditionally although workers in larger firms are significantly less likely to change labor market position—the

12. The other results obtained with the more restrictive definition of self-employment are generally similar to those reported in this article.

negative effect of firm size is only slightly reduced in the results reported in column 5—might suggest that the results are capturing the impact of unmeasured firm- or sector-specific human capital. However, conditional on firm size and individual characteristics, workers with higher conditional wages are more likely to start a microenterprise. Whether this reflects a relaxation on the liquidity constraint or entrepreneurial ability cannot be known at this point. However, the finding that overperformers in the salaried sector are more likely to enter self-employment is at odds with predictions of both the dualistic approach to the sector and the sociological view of the sector as a preferred destination for misfits.

Sector of economic activity also has an impact. Entry into self-employment is most likely for those who hold jobs in the construction sector, and it is least likely for those employed in manufacturing. Thus, whereas a salaried worker in manufacturing has a 4.3 percent probability of entering self-employment, the probability for a salaried construction worker with similar personal characteristics is 7.5 percentage points higher (table 2, column 4). Salaried workers with jobs in agriculture and government also have—probably for different reasons—a higher probability of entering self-employment than their counterparts in manufacturing.

Who Survives?

The determinants of firm survival are estimated using a probit specification that links individual and firm characteristics to the probability of staying in self-employment over a one-year period (table 3).¹³ The sample is restricted to individuals who reported that they were self-employed in the first ENEU interview. Data from both the larger ENEU sample (column 1) and the smaller ENAMIN–ENEU-linked database (columns 2–4) are used to capture more firm-specific characteristics.

The age patterns shown in figure 2 are confirmed with the probability of survival increasing with age until the 36–50 age bracket (column 1). The average self-employed individual in the age bracket 15–20 years has a 33.3 percent probability of staying self-employed, whereas for individuals with similar personal characteristics, the probabilities of survival are 18.7 percentage points higher in the 21–35 bracket and 27.2 percentage points higher in the 33–50 bracket. A similar age pattern, although with smaller increases in survival derived from age alone, is found when the same specification is estimated with the smaller albeit richer ENAMIN–ENEU sample (columns 2–4).

13. The results are subject to the caveat that we cannot observe whether self-employed individuals are still running the same businesses as in their initial interviews. Thus, in practice, the estimate is of the determinants of an individual staying in self-employment, as opposed to the determinants of an enterprise staying in business. To the extent that failed entrepreneurs tend to switch to a different labor market position before moving to a new business venture, the above assumption should have a minor impact on the results.

TABLE 3. Determinants of Survival in Self-Employment

| | Survival (1) | Survival (2) | Survival (3) | Survival (4) |
|-----------------------------------|-------------------|-------------------|-------------------|-------------------|
| Ages 21–35 years | 0.187*** (0.000) | 0.124*** (0.003) | 0.102** (0.018) | 0.100** (0.020) |
| Ages 36–50 years | 0.272*** (0.000) | 0.190*** (0.000) | 0.148*** (0.001) | 0.143*** (0.001) |
| Ages 51–65 years | 0.238*** (0.000) | 0.173*** (0.000) | 0.130*** (0.002) | 0.127*** (0.003) |
| Schooling, 6–12 years | –0.022*** (0.000) | –0.003 (0.847) | 0.002 (0.907) | 0.002 (0.909) |
| Schooling, 13 plus years | –0.024*** (0.000) | 0.022 (0.373) | 0.031 (0.207) | 0.033 (0.182) |
| Married | 0.084*** (0.000) | 0.052** (0.014) | 0.052** (0.016) | 0.054** (0.013) |
| Self-employment earnings (log) | 0.042*** (0.000) | 0.019 (0.110) | 0.006 (0.605) | 0.005 (0.683) |
| Firm size, 2–5 workers | 0.065*** (0.000) | 0.029* (0.067) | 0.013 (0.440) | 0.012 (0.471) |
| Firm size, 6 plus workers | 0.058*** (0.000) | 0.125** (0.038) | 0.088 (0.175) | 0.085 (0.189) |
| Commerce | 0.056*** (0.000) | 0.055** (0.025) | 0.059** (0.016) | 0.055** (0.024) |
| Agriculture | –0.009 (0.391) | –0.146 (0.666) | –0.210 (0.549) | –0.208 (0.555) |
| Construction | –0.104*** (0.000) | –0.136*** (0.000) | –0.147*** (0.000) | –0.147*** (0.000) |
| Services | –0.005 (0.358) | 0.008 (0.753) | –0.011 (0.649) | –0.012 (0.631) |
| More than one job | –0.056*** (0.000) | –0.055 (0.189) | –0.071* (0.099) | –0.069 (0.110) |
| Time in business | | | 0.043*** (0.000) | 0.043*** (0.000) |
| Capital stock (log) | | | 0.007* (0.066) | 0.006 (0.147) |
| Informal start-up credit | | | | –0.001 (0.941) |
| Formal start-up credit | | | | 0.136*** (0.008) |
| Informal credit after established | | | | 0.029 (0.226) |
| Formal credit after established | | | | –0.004 (0.931) |
| Log-likelihood | –44,251 | –1,674 | –1,470 | –1,465 |
| Pseudo R-squared | 0.071 | 0.052 | 0.072 | 0.075 |
| Number of observations | 77,388 | 3,218 | 3,037 | 3,037 |

*Significant at the 10 percent level.

***Significant at the 5 percent level.

****Significant at the 1 percent level.

Note: Numbers in parenthesis are p-values. The table reports the results of a probit model on the marginal changes in the probability of staying in self-employment resulting from discrete changes in the independent dummy variables and from infinitesimal changes in the continuous explanatory variables. Column 1: sample of all self-employed individuals in their initial Encuesta Nacional de Empleo Urbano (ENEU) interview. Columns 2–4: ENEU–Encuesta Nacional de Micronegocios (ENAMIN)-linked sample.

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

Education has a negative impact on the probability of staying in self-employment (column 1). However, this effect is quantitatively small, with a reduction of about 2 percentage points in the survival probability for individuals with secondary or tertiary schooling compared with a 70 percent probability of staying for those with at most primary schooling, and it disappears in the smaller sample. These results seem counterintuitive in that more educated workers would seem better able to evaluate the probability of success in potential business opportunities and also to do better in business, as suggested by Bates (1999). In fact, when the sample is restricted to self-employed individuals who report at least one worker, the effect of educational attainment on firm survival becomes positive and significant, especially for those with some college education. As for the larger sample that includes owner-only businesses, to the degree that better-educated workers may find better opportunities in the formal sector, this may be the effect of a higher "pull" effect of better employment alternatives for more educated individuals dominating a lower "push" effect associated with smaller probabilities of business failure.

However, consistent with the relevance of this effect and the various mainstream models of firm dynamics reviewed above, higher conditional wages lead to a higher probability of survival. Thus, using the larger sample and including all self-employed individuals regardless of whether they have employees, a doubling of the earnings from self-employment is associated with a 4.2 percentage point increase in the probability of staying in that sector. The effect is quantitatively larger when the sample is restricted to businesses with at least one employee, for which a doubling of self-employment earnings is related to an 8.8 percentage point increase in the probability of maintaining that initial status.

Firm size is also positively and significantly related to survival in self-employment in both the larger and the restricted sample, which is consistent with the predictions of mainstream firm dynamics models.¹⁴ However, when time in business and capital stocks are also controlled for using the smaller ENEU-ENAMIN sample (table 3, column 3), the variables representing employment size and age of the entrepreneur cease to be significant, suggesting that they were merely proxying for the former variables, which do exhibit significant effects on self-employment survival. In particular, a doubling of the time that a microenterprise has been in business is associated with a 4.3 percentage point increase in its likelihood of survival, whereas a tripling of its capital stock is related to a 1.4 percentage point increase. Taken together, the results are quite consistent both with Jovanovic's (1982) "noisy selection" view and with the bulk of the mainstream empirical evidence that firms get a more precise estimate of

14. In the larger ENEU sample, the effect of having at least one employee is that of increasing by about 6 percent the probability of survival in self-employment. In the smaller ENEU-ENAMIN sample, the increase is smaller for those with at most five workers (2.9 percent) than for firms with at least six workers including the owner (12.5 percent). A similar increase (13 percent) is obtained for firms with at least six employees when the sample is further restricted to firms with at least one employee.

their cost structures with experience and that past measures of success are informative for the future evolution of the firm.

Survival in self-employment is 5–8 percentage points higher for married individuals, possibly because they can count on unpaid family workers, and those who report a second job are about 6 percentage points less likely to remain in self-employment. Moreover, self-employed workers in the construction sector have a probability of staying in self-employment that is 10–15 percentage points lower than for those engaged in agriculture, manufacturing, or services, whereas the self-employed workers engaged in commercial activities have a 5.5 percentage points higher survival probability than their peers in the above three sectors.

Finally, access to formal credit at start-up seems to affect whether firms survive, consistent with Evans and Jovanovic's (1989) view: firms with access to credit exhibit a 13.6 percentage points higher probability of survival than their peers with similar owner and enterprise characteristics (table 3, column 4). Once firms are established, however, neither formal nor informal channels of credit have a significant effect on survival in self-employment. However, because credit variables are positively correlated with unobservable characteristics of the firm that also affect the prospects of firm survival and growth, these estimates should be interpreted with caution. Thus, for instance, the presence of unobserved heterogeneity in managerial ability, not captured by conditional self-employment earnings and positively related to access to credit, could upwardly bias the impact of access to credit on firm survival. Moreover, the bias could also go in the opposite direction if some covariates already embody effects derived from credit access (for example, higher capital stock, employment size, and time in business). Provided that sensible instruments for credit access can be generated, future research should further explore the importance and direction of those possible estimation biases.

Growth

Employment growth is somewhat difficult to analyze because the aggregate categorical employee coding of the ENEU prohibits direct comparisons of employment dynamics with those of the United States. However, the transition matrix in table 1 suggests that, as in the United States, microenterprises create relatively more jobs but destroy roughly equivalent numbers and generate little net job creation. Individuals starting microenterprises have a higher probability of simply remaining self-employed than of increasing their firm size, whereas a firm with more than one employee has a higher probability of decreasing its size than of increasing it. Only 12 percent of owner-only firms expand to one to four employees, whereas 22.1 percent of firms of one to four employees contract to owner-only. The share of firms that go out of business decreases with firm size and ranges from 35 percent for owner-only to 22 percent for the largest category.

These dynamics are broadly consistent with a steady-state distribution dominated by very small firms. Again drawing on the United States as a benchmark,

the overall size distribution of microenterprises in Mexico is similar in terms of owner-only firms. According to U.S. Small Business Administration and Census Bureau data, 67 percent of registered U.S. firms are owner-only, and another 12 percent have at most five workers, including the owner.¹⁵ In Mexico, 62 percent of businesses operated by the self-employed have no employees, and 32 percent have at most five workers.¹⁶ Although Mexico and the United States are clearly different at the upper tail of the size distribution, at the lower tail, both the countries exhibit a remarkably large and similar share of owner-only firms.

The larger absolute number of small firms may reflect simply that, as the opportunity cost of starting a small firm falls, more workers enter, the relative distribution of ability remains similar or, in line with Hopenhayn's (1992) argument, falls, leading, in line with Lucas's (1978) insight, to a large share of very small firms. Alternatively, barriers to credit or other services may keep firms below their optimal size, a hypothesis investigated briefly later and in more detail in Fajnzylber, Maloney, and Montes Rojas (2006). The fact that most U.S. microenterprises, which exist in a relatively well-functioning economic environment, also remain very small does, perhaps, lend support to the first hypothesis.

The analysis also examines what characteristics seem correlated with a transition between size brackets over a one-year period, considering only surviving firms." The dependent variable is the imputed percentage difference between the mean value of employment in the corresponding firm size brackets. By definition, firms that remained in the same size interval have a value of zero. The results are summarized in table 4, with the results for the full ENEU sample presented in column 1 and the results for the smaller joint sample presented in the other three columns.

The relationships between firm growth and the age and marital status of the entrepreneurs are broadly similar, at least in the large sample estimates, to those obtained in the survival analysis: business owners who are married and in the age bracket 36–50 years are most likely to expand their firms, exhibiting employment growth rates that are, respectively, 11.1 and 9.4 percentage points higher than those of their peers.¹⁸ Consistent with mainstream models, entrepreneurs with conditionally higher earnings (better performance) seem to show a higher

15. Authors' calculations using U.S. Small Business Administration (2001).

16. Authors' tabulations based on the Mexican National Urban Employment Survey (ENEU). It is worth noting that while self-employment rates hover between 5 and 15 percent in most OECD countries, Latin America exhibits rates that are generally between 20 and 35 percent (figure 1). Even within the OECD, as reported by Blanchflower (2004, p. 10), "self-employment rates are generally higher in poorer countries."

17. Firm size is divided into owner-only firms, 2–5 employees, 6–10 employees, 11–15 employees, 16–50 employees, 51–100 employees, 101–250 employees, and 250 and more employees.

18. While the positive sign encountered in the marital status variable may well reflect the use of nonpaid family workers, the results are virtually unchanged when the firms with mostly nonpaid workers are kept in the sample.

TABLE 4. Determinants of Employment Growth

| | Growth (1) | Growth (2) | Growth (3) | Growth (4) |
|-----------------------------------|-------------------|-------------------|-------------------|-------------------|
| Ages 21–35 years | 0.040** (0.017) | 0.075 (0.392) | 0.072 (0.418) | 0.073 (0.412) |
| Ages 36–50 years | 0.094*** (0.000) | 0.147* (0.095) | 0.140 (0.123) | 0.136 (0.134) |
| Ages 51–65 years | 0.087*** (0.000) | 0.148* (0.098) | 0.132 (0.154) | 0.131 (0.156) |
| Schooling, 6–12 years | 0.027*** (0.000) | 0.039 (0.195) | –0.009 (0.769) | –0.011 (0.769) |
| Schooling, 13 plus years | 0.161*** (0.000) | 0.141*** (0.002) | 0.059 (0.201) | 0.050 (0.283) |
| Married | 0.111*** (0.000) | 0.115*** (0.001) | 0.102*** (0.005) | 0.100*** (0.006) |
| Self-employment earnings (log) | 0.122*** (0.000) | 0.135*** (0.000) | 0.101*** (0.000) | 0.099*** (0.000) |
| Firm size, 2–5 workers | –0.770*** (0.000) | –0.863*** (0.000) | –0.911*** (0.000) | –0.913*** (0.000) |
| Firm size, 6 plus workers | –1.026*** (0.000) | –1.352*** (0.000) | –1.421*** (0.000) | –1.417*** (0.000) |
| Commerce | –0.010 (0.319) | –0.023 (0.628) | –0.003 (0.949) | 0.006 (0.897) |
| Agriculture | 0.052** (0.013) | 1.034*** (0.000) | 1.006*** (0.000) | 1.018*** (0.000) |
| Construction | –0.102*** (0.000) | –0.226*** (0.000) | –0.145** (0.020) | –0.145** (0.020) |
| Services | –0.157*** (0.000) | –0.193*** (0.000) | –0.182*** (0.000) | –0.178*** (0.000) |
| More than one job | –0.011 (0.537) | –0.017 (0.823) | –0.017 (0.813) | –0.029 (0.700) |
| Time in business | | | –0.002 (0.888) | –0.005 (0.696) |
| Capital stock (log) | | | 0.048*** (0.000) | 0.049*** (0.000) |
| Informal start-up credit | | | | –0.086** (0.012) |
| Formal start-up credit | | | | –0.211** (0.011) |
| Informal credit after established | | | | 0.045 (0.289) |
| Formal credit after established | | | | 0.123** (0.047) |
| R-squared | 0.233 | 0.310 | 0.326 | 0.331 |
| Number of observations | 53,729 | 2,503 | 2,381 | 2,381 |

"Significant at the 10 percent level.

""Significant at the 5 percent level.

""Significant at the 1 percent level.

Note: Numbers in parenthesis are p-values. Ordinary least squares estimator. Column 1: sample of all self-employed individuals in their initial and last Encuesta Nacional de Empleo Urbano (ENEU) interview. Columns 2–4: ENEU–Encuesta Nacional de Micronegocios (ENAMIN)-linked sample.

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

propensity to grow. Moreover, the educational attainment of the owner now has a positive and significant effect on firm growth, at least when detailed firm characteristics are not controlled for.

Given that the estimates are conditional on surviving, firm size appears with the negative sign predicted by the mainstream literature: bigger firms are more likely to have achieved their optimal size. Time in business also has a negative effect, as expected from the literature on industrial countries, but its coefficient is not statistically significant. Similarly, consistent with Ericson and Pakes (1989, 1995), although less so with Evans and Jovanovic (1989), higher capital stocks are positively correlated with employment growth.

The effect of credit variables depends on the type of credit and whether it was granted at start-up or afterward (table 4, column 4). Consistent with Evans and Jovanovic (1989), the lack of access to start-up credit may imply suboptimal start-up size, with more rapid subsequent growth compared with firms that started closer to their steady-state size. This does not contradict the finding that formal sector loans granted at a later stage, when firms may be closer to their steady state and entrepreneurs better able to reveal their managerial ability, have a positive and significant impact on growth. Again, endogeneity issues urge circumspection and more careful future research.

IV. CONCLUSIONS

This article has documented patterns of entry, survival, and growth in the Mexican self-employment sector. The results suggest that the Mexican self-employment sector is much larger than its counterpart in the United States, which is reflected in higher unconditional rates of entry. Moreover, the evidence points to the significant presence of well-performing salaried workers among the likely entrants into self-employment in Mexico, as opposed to the higher incidence of poorer waged workers among the entrants into the U.S. self-employment sector. Despite these differences, however, the evidence for Mexico is very similar to that for the United States in terms of patterns of entry, survival, and growth with respect to age, education, and many other covariates. As a first approximation, microenterprises in Mexico show dynamic patterns consistent with the entrepreneurial risk-taking view usually applied to firm dynamics in the industrial world. In particular, the noisy selection and learning by doing theoretical features seem to be consistently corroborated in the sample estimates.

The patterns of entry into self-employment show that even if liquidity constraints are important for determining the future performance of new firms, self-employment does not appear to be less desirable than other labor market alternatives. Indeed, once the analysis is conditioned for the probability of changing labor market status, salaried workers are more likely to enter self-employment than individuals who are unemployed or out of the labor force.

Moreover, there is evidence that being employed as a salaried worker in a larger firm reduces the probability of entry—presumably because of benefits and job stability—but once the analysis is conditioned for firm size, higher wages increase the likelihood of becoming a microentrepreneur. This result runs counter to the view of self-employment as a holding pattern for misfits or workers rationed out of the salaried sector but supports the hypothesis that unobserved ability plays a crucial role in selecting workers in and out of self-employment.

The results on firm survival and growth also confirm the relevance of mainstream models, with size and time in business being negatively related to exit and growth. Moreover, while concerns regarding the possible endogeneity of variables related to access to credit must be kept in mind, receiving a loan after having started a business is associated with higher odds of growing, whereas receiving a loan at the time of start-up is associated with lower rates of both exit and employment growth. Presumably, this is because entrepreneurs who receive loans at the time of entry achieve an optimal size sooner.

Taken together, the results suggest first that the insights of the mainstream theoretical and empirical literature are relevant to analysis of developing-country microenterprises and thus provide helpful guides to policy. Second, they suggest that, overall, the sector corresponds closely to the dynamic models of voluntary entrepreneurship described for industrial countries.

This is not to downplay the vast heterogeneity of the sector. Firms with at least one employee correspond most neatly to the mainstream literature, whereas the dominant group of own-account businesses differs in some suggestive respects. Their entry rates fall with education, and they have lower potential for growth in response to conditionally higher earnings, suggesting that these firms may be relatively less "dynamic." But this is consistent with a combination of Lucas' (1978) and Blau's (1987) views, discussed earlier, in a developing-country context: very low formal sector productivity, particularly for low-skilled workers, implies a low opportunity cost for many workers of very poor entrepreneurial ability to enter the sector who may have no plans or capability to expand beyond one person. By contrast, the rising exit rates with education and higher rates of entry from unemployment may suggest, reasonably, that involuntary entrants to the sector are likely to be found among single-person firms. Nonetheless, the commonality of most covariate signs across subsamples of firm sizes, the fact that reported rates of voluntary entry, while below those of larger microenterprises are still close to 60 percent, and the procyclical pattern of gross labor flows into the sector overall (Bosch and Maloney 2005) suggest that the dominant character of the subsector is broadly in line with a mainstream view.¹⁹

19. In 1992, those reporting voluntary entry were 57.5 percent of own-account firms and 68.1 percent of firms with one employee or more. During the crisis, as might be expected, the rate of voluntary entry fell substantially in both the groups.

APPENDIX: VARIABLE DESCRIPTION

ENEU variables

- *Self-employment*: dummy variable that takes a value of one if the individual's primary occupation is self-employment or owner of a firm.
- *Age*: dummy variables for individual age brackets. Base category: ages 15–17 years.
- *Schooling*: dummy variables for years of schooling completed. Base category: less than six years of schooling.
- *Married*: dummy variable for married individuals.
- *Out labor*: dummy variable for individuals who are out of the labor force.
- *Unemployed*: dummy variable for unemployed individuals.
- *Unemployed for four to six months* and *for seven plus months*: dummy variables for individuals who have been unemployed for four to six months or for at least seven months.
- *Salaried wage* (log): logarithm of real monthly income from main job (salaried, contract worker, and others). The bottom and top percentiles are excluded from the sample.
- *More than one job*: dummy variable for having an income source other than the main job.
- *self-employed earnings* (log): logarithm of real monthly income from the main job (if self-employed or owner).
- *Contract worker*: dummy variable for contract workers.
- *Nonpaid worker*: dummy variable for individuals who are working without pay.
- *Other work*: dummy variable for paid workers in either a cooperative or other unspecified types of jobs (not salaried, not contract).
- *Size of firm*: dummy variable for the number of employees in the firm that corresponds to the main job.
- *Employment growth*: imputed percentage difference (in logarithm) between the mean value of the corresponding firm size brackets in a one-year period. By definition, firms that remained in the same size interval have a value of 0.

ENAMIN variables

- *Capital stock*: sum of the replacement cost of all owned or borrowed physical capital and the market price of all firm inventories.
- *Time in business*: number of years since owner began the activity or became head of the business.
- *Informal credit*: dummy variables for microenterprises that report receiving credit from clients, suppliers, friends, or family at the time of start-up or after.
- *Formal credit*: dummy variables for microenterprises that report receiving credit from a bank at the time of start-up or after.

REFERENCES

- Antunes, Antonio, and Mario Centeno. 2005. *Do Labor Market Policies Affect Employment Composition? A Look at European Countries*. Lisbon: Banco de Portugal.
- Aw, Bee Yan, Sukkyun Chung, and Mark J. Roberts. 2003. "Productivity, Output, and Failure: A Comparison of Taiwanese and Korean Manufacturers." *Economic Journal* 113(491):F485–510.
- Balán, J., H. L. Browning, and Jelin, E. 1973. *Men in a Developing Society*. Austin: Texas: University of Texas Press.
- Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2004. "Microeconomic Evidence of Creative Destruction in Industrial and Developing Countries." Policy Research Working Paper 3464. World Bank, Washington, D.C.
- Bates, Timothy. 1999. "Entrepreneur Human Capital Inputs and Small Business Longevity." *The Review of Economics and Statistics* 72(4):551–59.
- Bhattacharaya, Prabir C. 2002. "Rural-to-Urban Migration in LDCs: A Test of Two Rival Models." *Journal of International Development* 14(7):951–72.
- Blanchflower, David G. 2004. "Self-Employment: More May Not Be Better." *Swedish Economic Policy Review* 11(2):15–74.
- Blanchflower, David G., and Andrew J. Oswald. 1998. "What Makes an Entrepreneur?" *Journal of Labor Economics* 16(1):26–35.
- Blau, David M. 1987. "A Time-Series Analysis of Self-Employment in the United States." *Journal of Political Economy* 95(3):445–67.
- Bosch, Mariano, and William F. Maloney. 2005. "Labor Market Dynamics in Developing Countries: Comparative Analysis using Continuous Time Markov Processes." Policy Research Working Paper 3583. World Bank, Washington, D.C.
- Carrasco, Raquel. 1999. "Transitions to and from Self-Employment in Spain: An Empirical Analysis." *Oxford Bulletin of Economics and Statistics* 61(3):315–41.
- Cressy, Robert. 1996. "Are Business Startups Debt-Rationed?" *Economic Journal* 106(438):1253–70.
- Cunningham, Wendy V., and William F. Maloney. 2001. "Heterogeneity among Mexico's Microenterprises: An Application of Factor and Cluster Analysis." *Economic Development and Cultural Change* 50(1):131–56.
- De Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World*. New York: Harper and Row.
- Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1988. "Patterns of Firms Entry and Exit in U.S. Manufacturing Industries." *RAND Journal of Economics* 19(4):495–515.
- Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1989. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics* 104(4):671–98.
- Ericson, Richard, and Ariel Pakes. 1989. "An Alternative Theory of Firm and Industry Dynamics." Working Paper 445. Columbia University, Department of Economics, New York.
- Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62(1):53–82.
- Evans, David S. 1987a. "The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries." *Journal of Industrial Economics* 35(4):567–81.
- Evans, David S. 1987b. "Tests of Alternative Theories of Firm Growth." *Journal of Political Economy* 95(4):657–74.
- Evans, David S., and Boyan Jovanovic. 1989. "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." *Journal of Political Economy* 97(4):808–27.
- Evans, David S., and Linda S. Leighton. 1989. "Some Empirical Aspects of Entrepreneurship." *American Economic Review* 79(3):519–35.

- Fajnzylber, Pablo, William F. Maloney, and Gabriel Montes Rojas. 2006. "Releasing Constraints to Growth or Pushing on a String? The Impact of Credit, Training, Business Associations and Taxes on the Performance of Mexican Micro-Firms." Policy Research Working Paper 3807. World Bank, Washington, D.C.
- Fields, Gary S. 1990. "Labor Market Modelling and the Urban Informal Sector: Theory and Evidence." In David Turnham, Bernard Salomé, and Antoine Schwarz, eds., *The Informal Sector Revisited*. Paris: Organization for Economic Co-operation and Development.
- Geroski, Paul A. 1991. *Market Dynamics and Entry*. Oxford: Blackwell Publishing.
- Goedhuys, Micheline, and Leo Sleuwaegen. 2000. "Entrepreneurship and Growth of Entrepreneurial Firms in Cote d'Ivoire." *Journal of Development Studies* 36(3):123–46.
- González de la Rocha, Mercedes. 1994. *The Resources of Poverty: Women and Survival in a Mexican City*. Oxford: Blackwell Publishing.
- Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *Journal of Political Economy* 108(3):604–31.
- Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment, and Development: A Two-Sector Analysis." *American Economic Review* 60(1):126–42.
- Hart, Kieth. 1972. *Employment, Income and Inequality: A Strategy for Increasing Productive Employment in Kenya*. Geneva: International Labour Office.
- Hart, Keith. 1973. "Informal Income Opportunities and Urban Employment in Ghana." *Journal of Modern African Studies* 11(3):61–89.
- Hopenhayn, Hugo A. 1992. "Entry, Exit and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60(5):1127–50.
- Johnson, William R. 1978. "A Theory of Job Shopping." *Quarterly Journal of Economics* 92(2):261–78.
- Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87(5):972–90.
- . 1982. "Selection and the Evolution of Industry." *Econometrica* 50(3):649–70.
- Liedholm, Carl. 2002. "Small Firm Dynamics: Evidence from Africa and Latin America." *Small Business Economics* 8(1–3):227–42.
- Liedholm, Carl, and Donald C. Mead. 1999. *Small Enterprises and Economic Development: The Dynamics of Micro and Small Enterprises*. London: Routledge Press.
- Loayza, Norman V., Ana Maria Oviedo, and Luis Servén. 2005. "The Impact of Regulation on Growth and Informality, Cross-Country Evidence." Policy Research Working Paper 3623. World Bank, Washington, D.C.
- Lucas, Robert E. Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9(2):508–23.
- Maloney, William F. 1999. "Does Informality Imply Segmentation in Urban Labor Markets? Evidence From Sectoral Transitions in Mexico." *World Bank Economic Review* 13(2):275–302.
- . 2001. "Self-Employment and Labor Turnover in Developing Countries: Cross-Country Evidence." In S. Devarajan, F. Halsey Rogers, and L. Squire, eds., *World Bank Economists Forum*. Washington, D.C.: World Bank.
- . 2004. "Informality Revisited." *World Development* 32(7):1159–78.
- McPherson, Michael A. 1995. "The Hazards of Small Firms in Southern Africa." *Journal of Development Studies* 32(1):31–54.
- . 1996. "Growth of Micro and Small Enterprises in Southern Africa." *Journal of Development Economics* 48(2):253–77.
- Mead, D. M., and C. Liedholm. 1998. "The Dynamics of Micro and Small Enterprises in Developing Countries." *World Development* 26(1):61–74.
- Rees, Hedley, and Anup Shah. 1986. "An Empirical Analysis of Self-Employment in the U.K." *Journal of Applied Econometrics* 1(1):95–108.

- Roberts, Mark J., and James R. Tybout, eds. 1997. *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*. New York: Oxford University Press.
- Tokman, Victor E. 1978. "An Exploration into the Nature of the Informal-Formal Sector Relationships." *World Development* 6(9/10):1065-75.
- U.S. Small Business Administration. 2001. *Small Business Economic Indicators: 2000*. Washington, D.C.: Office of Advocacy.
- Wagner, Joachim. 1994. "The Post-Entry Performance of New Small Firms in Germany Manufacturing Industries." *Journal of Industrial Economics* 42(2):141-54.

The "Glass of Milk" Subsidy Program and Malnutrition in Peru

David Stifel and Harold Alderman

This study of the Vaso de Leche ("Glass of Milk") feeding program in Peru looks for evidence that this in-kind transfer program aimed at young children furthers nutritional objectives. The study links public expenditure data with household survey data to substantiate the targeting and to model the determinants of nutritional outcomes. It confirms that the social transfer program targets poor households and households with low nutritional status. Nevertheless, the study fails to find econometric evidence that the nutritional objectives are being achieved.

In designing transfer programs, governments are motivated by equity or efficiency objectives or both. Das, Do, and Ozler (2005) discuss these objectives for conditional cash transfer programs, but their analysis also applies to in-kind transfers and commodity price subsidies. While such subsidies or transfers may be politically pragmatic or administratively more feasible where markets or banks are rudimentary, the choice of in-kind or conditional transfers over direct unconditional cash transfers is generally based on the assumed presence of a market failure. For example, a food price subsidy or commodity transfer may be designed to improve the nutritional status of vulnerable groups—as well as to augment the real incomes of constituents—based on the possibility that intrahousehold allocations do not reflect the rates of return to investments in children. Food subsidies may also be motivated by the view that past underinvestments in education led to current inefficiencies in the allocation of inputs into the production of health.

No direct measure of behavior is necessary for assessing equity-driven transfers to households. The equity-improving objective can be assessed in terms of its effect on the distribution of household incomes or on poverty reduction. In contrast, as Das, Do, and Ozler (2005) point out, the evaluation of transfers differs if the main motivation is increasing efficiency rather than addressing equity. If a conditional

David Stifel is an assistant professor at Lafayette College; his email address is stifeld@lafayette.edu. Harold Alderman is lead human development economist at the World Bank; his email address is halderman@worldbank.org. The authors thank Emanuela Galasso, Stephen Younger, and three anonymous referees for extensive comments on an earlier draft. They also thank Jose Roberto Lopez-Calix and Norbert Schady for support with the data and Erik Wachtenheim (Instituto Apoyo) and José Carlos Arca for their assistance.

THE WORLD BANK ECONOMIC REVIEW, VOL. 20, NO. 3, pp. 421-448

doi:10.1093/wber/lhl002

Advance Access publication July 10, 2006

© The Author 2006. Published by Oxford University Press on behalf of the International

Bank for Reconstruction and Development/ THE WORLD BANK. All rights reserved. For permissions,

please e-mail: journals.permissions@oxfordjournals.org.

transfer is designed to increase consumption of a commodity or use of a service, one approach is to look at the net increase (after any substitution) of consumption of that good. An alternative or additional approach is to look at the outcome that the conditionality is designed to affect. Nutritional objectives of in-kind transfers are often expressed as incremental consumption of one or more goods or one or more nutrients. However, since the transferred or subsidized good can be substituted for other items in the diet, it is preferable to focus the evaluation on the impact on child growth, as, for example, in evaluations of the impact of conditional transfers on nutrition in Mexico (Behrman and Hoddinott 2005; Rivera, Sotres-Alvarez, and others 2003). These individual-specific measures are behavior-induced outcomes that are distinct from the standard welfarist measure of total consumption.

When it comes to meeting nutritional objectives through in-kind transfers, milk is often believed to be a particularly effective commodity.¹ While exclusive breast-feeding is widely advocated for children under six months old, the value of supplementation with other milk at a later age is less clear. There is some clinical evidence that milk supplementation contributes to child growth but mainly in communities where the diet is based almost entirely on root crops or when milk supplements are combined with specific interventions to shift behavior (Rivera, Hotz, and others 2003). Thus, nutritionists generally do not advocate milk as a candidate for subsidies because of its high nutrient costs and low energy density (Kennedy and Alderman 1987).

Milk subsidy programs are nonetheless prevalent, and so it is important to assess their ability to achieve their nutritional objectives. It is surprising therefore that while there are several published studies on the distributional incidence of milk or milk product subsidies, evaluations of the nutritional impact of subsidy programs are hard to find.

The literature generally takes three forms. First, there are studies of milk programs that do not include evidence on nutritional impacts. These include Tuck and Lindert's (1996) study of milk consumption in Tunisia's subsidy program (accounting for 10 percent of overall subsidies at their peak), Esanu and Lindert's (1996) analysis of Romania's milk program, and the World Bank's (2003) report on the distribution of fluid milk in the Brazilian state of Rio Grande do Sul. Similarly, while Alderman and del Ninno (1999) estimate that exempting milk from South Africa's value added tax—similar to a consumer subsidy costing more than \$150 million a year—leads to a 0.18 percent increase in protein consumption (0.03 percent for the poorest 40 percent of the population), they do not provide evidence for its effect on malnutrition rates.

Second, there are analyses of programs that include milk as one of many subsidized foods for which nutritional impacts are documented but for which the effect cannot be singled out. For example, Rush and others (1988) and Carlson and Senauer

1. For example, as reported in the December 13, 2003, issue of *The Economist*, China recently instituted a school milk program after noting the comparatively small stature of its citizens.

(2003) find that the Women, Infants, and Children Feeding Program in the United States is clearly beneficial. That program is not confined to either milk or milk substitutes, however, and these studies do not single out the role of milk subsidies. Nor is milk generally distinguished in the literature on school feeding, even though it is often included in such programs. Powell and others (1998) report the impact of one such successful program that included milk. However, the nutritional experience of such programs is generally mixed, in part because of irregular implementation (Levinger 1986).

Third, there are a handful of studies of milk programs that do present nutritional evidence—for example, studies of Mexico's Liconsa fluid milk distribution program by Gundersen and others (2000), Kennedy and Alderman (1987), and Grosh (1994). Although Grosh (1994) finds the subsidies to be distributed progressively, none of these studies shows an impact on child growth.

Simply stated, it is not known whether policies to subsidize milk are effective at achieving nutritional objectives, and without knowing this it is hard to fully understand the motivation for the in-kind transfer. This article addresses this question by studying Peru's Vaso de Leche ("Glass of Milk") program, which provides primarily milk and milk substitutes to low-income households and is motivated by nutritional objectives. The program is well suited for this analysis because its benefits are distributed progressively (Stifel and Alderman 2005), thus eliminating one common reason for a commodity distribution program to have a limited nutritional impact. This permits focusing on the nutritional outcomes that might have motivated the subsidy program. Addressing the question of nutritional impact is not straightforward, however, since randomized evaluations of full-scale interventions are often hard to implement in politically popular transfer programs. Therefore, the approach applied here links public expenditure data with household survey information to assess the program's impact on nutritional outcomes.

I. DESCRIPTION OF THE VASO DE LECHE PROGRAM

At a cost of \$97 million in 2001, Vaso de Leche is the largest social transfer in Peru and the second largest component of transfers from the central government to municipalities (Instituto Apoyo and World Bank 2002). Introduced as a pilot in Lima in 1984, the program expanded nationally during the economic crises in the late 1980s and early part of the 1990s. By 1998 the program had expanded to reach 44 percent of households with children aged from 3 to 11 through earmarked monthly transfers to municipalities (Younger 2002). By law, these municipalities are required to have an administrative committee composed of elected representatives of beneficiaries, the mayor, another local official, and a representative from the ministry of health. In addition to this administrative committee, each community has an elected Vaso de Leche mothers committee. This committee, which has a fair degree of discretionary decision-making (Instituto Apoyo and World Bank 2002), identifies the beneficiaries, the timing of deliveries, and, within limits, the commodities to be distributed.

Despite its name, the Vaso de Leche program distributes more than milk and milk substitutes. In some cases, cereals or a combination of commodities are distributed instead of or in addition to milk products. For example, 46 percent of recipient households receive one product (67 percent of them receive milk or milk substitutes), while 51 percent receive two products (88 percent of them receive milk or milk substitutes).² Nonetheless, according to calculations of this study using data from the Vaso de Leche Public Expenditure Tracking Survey, milk and milk substitutes (such as powdered milk and soymilk) account for an average of 77.5 percent of the value of total transfers. Furthermore, for households in the two poorest quintiles, milk accounts for 93.3 percent of the value of the transfers and milk substitutes for 80.4 percent.

Priority is given to households with children six years old or younger or with pregnant or lactating women. Once these first-tier beneficiaries are attended to, households with children aged from 7 to 13 and people with tuberculosis may participate. Within both categories, priority is based on need.³

There have been many excellent recent studies on the distribution of social expenditures in Peru and of Vaso de Leche in particular. For example, Younger (2002) finds a pattern of progressive distribution of Vaso de Leche benefits, with improved targeting between 1994 and 1997 as coverage increased. Using a different methodology and one of the data sources employed in the current analysis (a 1997 household survey), Ruggeri Laderchi (2001) also examines the overall distribution of food transfers and their impact on food consumption and nutrition. She finds that the transfers are slightly progressive, although the poorest 40 percent of households received only 46 percent of total transfers. She also finds that while the total share of income from food-related transfers had no impact on the height of children, the income share from participation in the Vaso de Leche program had a significant impact on standardized child height (Ruggeri Laderchi 2001, p. 36). Her specification, which treats participation as exogenous, yields a positive effect only when income is instrumented and when district fixed effects are included at the same time. The impact appears negative and, in some specifications, statistically significant in the absence of these recommended econometric procedures.

A recent Public Expenditure Tracking Survey followed the budget trail from the central government to the Vaso de Leche beneficiaries (Instituto Apoyo and World Bank 2002; World Bank and Inter-American Development Bank 2002). The study finds an appreciable variation between communities in the timing of delivery, the commodities chosen, and the administrative fees charged. Virtually all the funds released by the center were transferred to municipal Vaso de Leche

2. Powdered milk is considered a milk substitute in this context. Although the law states that the distributed products should be in prepared form, this occurs in only 39 percent of the committees outside of Lima, and only 7 percent of the recipients report consuming the products at the point of pickup (Instituto Apoyo and World Bank 2002).

3. While the laws on the Vaso de Leche indicate that malnourished individuals are to receive priority, nutritional measures (such as anthropometric indicators) are not used for targeting purposes. See, for example, Law 27470 in *El Peruano* (2001).

administrative budgets and further down to the mothers committees, with only some documented small-scale leakage in the allocations. The study finds more substantial discrepancies between the commodity allocations reported by the committees and by the household, however. The study could not account for a quarter of the product transferred. Most of the unexplained gap was in urban districts (particularly provincial capitals).⁴

II. RESEARCH STRATEGY

This section describes the approach to modeling the determinants of nutritional status⁵ and discusses the estimation strategy in the presence of endogenous program placement.

Modeling the Determinants of Nutritional Status

To determine the impact of the Vaso de Leche program on nutrition, the determinants of child nutritional status are estimated using program expenditures as an explanatory variable. The approach is to estimate the intention to treat rather than the effect of the treatment on the treated. The intention to treat can be conceptualized as the effect of the Vaso de Leche transfers being offered regardless of actual participation or dropout. In this analysis, the counterfactual of interest is the state of the world if the program had not existed, which is compared with the state of the world in the presence of the program (Heckman, Lalonde, and Smith 1999). This is distinct from the counterfactual for the effect of the treatment on the treated, which is the state of the treated if the program had not existed compared with the state of the treated in the presence of the program.

Evaluation of the intention to treat looks at the difference between outcomes among the eligible population where the treatment is available compared with the same population where it has not been made available, preferably controlling for site selection. Evaluation of the treatment on the treated looks at differences in the expected outcomes, conditional on participation. It is generally not possible to go directly from one form of evaluation to the other without additional assumptions since it is not usually possible to ascertain the participation of members of the control group had they had the same opportunities as the treatment group.

Both types of comparisons convey useful information. But, as Heckman, Lalonde, and Smith (1999) observe in their review of methodologies for

4. The study also found leakage or dilution in the sense that children did not always receive the milk that was obtained by the household. However, this is not only a difficult topic to quantify, but the welfare interpretations of this so-called leakage also differ from those of leakages in the public expenditure allocation chain. As argued in Alderman and others (1995), expecting a transferred good to be consumed entirely by one targeted individual within a household unit is not easily reconciled with any standard household model. Nor is the intrahousehold allocation as likely to be influenced by program administration as are errors of inclusion and exclusion in targeting on poverty.

5. The model is fairly standard in the literature (Strauss and Thomas 1995). This exposition draws on Sahn and Stifel (2002).

evaluation, it is often evaluation of the intention to treat that is of policy relevance (see also Rouse 1998). So, while this analysis does not measure the marginal contribution of milk consumption itself to nutritional status, the impact that is measured allows one to assess whether government expenditures on milk subsidies improve nutrition. This focus, then, differs from that of other studies on feeding programs and in-kind transfers, such as Ruggeri Laderchi (2001), which attempt to measure the effect of the treatment on the treated.

The theoretical framework for the estimation is derived from a household model in the tradition of Becker (1981). Assume that the household maximizes a quasi-concave utility function that takes as its arguments consumption of milk, x_m , all other commodities and services, x_o , leisure, l , and the health status, θ (of which a child's anthropometric measurement, h , is one dimension) of each household member. The household solves the following problem

$$(1) \quad \max_{x_m, x_o, l, \theta} u(x_m, x_o, l, \theta; A, Z)$$

where A and Z , respectively, represent household and community characteristics, some of which are not observed. Allocation choices are made conditional on the budget constraint:

$$(2) \quad p_m x_m + p_o x_o = w(T - l) + y$$

where p_m is the price of milk, p_o a vector of prices, w a vector of household members' wages, T a vector of the household members' maximum number of work hours, and y the household nonwage income.

The nutritional status of children, h , is determined by a biological health production technology:

$$(3) \quad h_i = H(I, A, Z, \mu_i)$$

where I is a vector of health inputs and μ_i represents the unobservable individual, family, and community characteristics that affect the child's nutritional outcomes. Household characteristics (such as demographics and educational levels), A , can have an impact on health by affecting household allocation decisions. Community characteristics (such as access to clean water), Z , can also have direct impacts on nutritional outcomes. Note that the input vector I includes consumption goods (such as milk) that contribute positively to household welfare both directly through x_m and x_o and indirectly through h . This represents the simultaneous choice of consumption goods and health inputs.

Given this simultaneity, the household's optimization problem can be solved to get a set of demand equations for goods and services (x), leisure (l), and health (θ). A subset of the health demand equations is the reduced-form demand equation for child nutrition, represented as follows:

$$(4) \quad h_i = \tilde{h}(A, Z, \varepsilon_i)$$

where ε_i is the child-specific random disturbance term, which is assumed to be uncorrelated with the other elements of the demand function.

The dependent variable is the standardized anthropometric height-for-age z -score (HAZ) for children under five years of age. HAZ is defined as $(h - h_r)/\sigma_r$, where h is the observed height of a child of a specified sex and age group, h_r the median height in the reference population of children of that sex and age group, and σ_r the standard deviation of height measurement for the reference population of that sex and age group. The standard reference population recommended by the World Health Organization is that of the United States National Center for Health Statistics. As several studies have indicated that less than 10 percent of the worldwide variance in height can be ascribed to genetic or racial differences (Martorell and Habicht 1986), this reference population is appropriate. Children with a HAZ score of less than -2 are usually classified as stunted.

The set of predictors consists of characteristics of the child (such as age, sex, and birth order), household demographic variables (such as household size and age and sex composition), characteristics of the parents (such as educational attainment and mother's age and height), access to public services (such as piped drinking water), and a dummy variable for living in an urban area. Given the pro-poor targeting of the Vaso de Leche program, predicted log per capita household expenditure was also included in the estimated model to control for household wealth.

Endogenous Program Placement and Explanatory Variables

The primary purpose of this exercise was not to model the overall determinants of nutrition but to see whether the Vaso de Leche program has an impact on nutrition. In terms of the model described in the previous section, this effect can be transmitted in two ways: by increasing household income by the value of the milk transfer (if this is the entire effect, the transfer is said to be inframarginal) and by directly increasing the level of milk consumption above what would have been consumed had the transfer been made in the form of cash by influencing the marginal price. Thus, the reduced-form health demand function is adapted to include the Vaso de Leche transfers as an explanatory variable to pick up the direct effect of the program on child health independent of its role as an income transfer:

$$(5) \quad h_i = h(VL, A, Z, \varepsilon_i).$$

As noted, the program is evaluated based on the intention to treat—in this case, conditional on the funding for the Vaso de Leche at the local level—and not on the household choice to take up this opportunity. Modeling the impact on self-selecting participants would require making a set of additional assumptions to determine the impact on the random eligible participant. While conditioning on the Vaso de Leche allocation to the community in lieu of participation means not having to solve the issue of endogenous household choice, the problem of potential bias from endogenous program placement remains (Rosenzweig and Wolpin 1986). Not even the sign of any potential bias can be established since an

estimated impact may be overestimated if programs are placed where the anticipated return is higher than average or be underestimated (or even negative) if programs go to favored but more developed communities.

This issue is addressed by using observations of Vaso de Leche expenditures from two different rounds of Demographic and Household Surveys (1996 and 2000). Thus, fixed effect estimations can control for the initial conditions in the communities. The general form of the models to be estimated is as follows:

$$(6) \quad h_{i,d,t} = \alpha + \phi x_{i,d,t} + \beta A_{i,d,t} + \lambda Z_{i,d,t} + \gamma VL_{d,t} + \sum_d \delta D_d + \varepsilon_{i,d,t}$$

where i is the index for individual children, d the indicator for the district in which the child resides, and t the year (1996 or 2000). VL is the district level per capita Vaso de Leche expenditure. The fixed effect version also includes, D , the set of district dummy variables. The inclusion of these dummy variables removes the influence of any time-invariant district effects, including any that might correlate with the allocation of Vaso de Leche funds.

This approach compares the differences in the changes in health status when Vaso de Leche transfers change, controlling for other community characteristics. In effect, γ is the difference in differences estimator (Moffitt 1991) of the effect of Vaso de Leche transfers on child health. Per capita expenditure on the program increased between survey years from 29.4 soles in 1996 to 37 soles in 2000, or more than 25 percent. The coefficient of variation for the change in expenditures is 0.47, indicating substantial variation in the rates by which coverage increased to identify a first difference at the community level.⁶) The expenditure data are based on total expenditures in each district and not a sample and are thus analogous to a census of expenditures.

As an additional precaution for site selection bias, instrumental variable methods were also employed with the fixed effects models to account for the possibility of any remaining unobserved factors affecting malnutrition that vary over time and are also correlated with the change in Vaso de Leche expenditures. This was approached in two ways.

First, standard two-stage least squares models were used in which the identifying instrument is the district-level Peruvian Social Fund Fondo Nacional de Compensación y Desarrollo (FONCODES) index of unmet basic needs, a composite of various measures—including access to schooling, electricity, water, sanitation, adequate housing, and measures of illiteracy—based on the 1993 census (Schady 1998). As shown in the results, the FONCODES index is correlated with district-level Vaso de

6. There were 315 districts in the study. While the relatively small number of observations per district implies that the district dummy variable will not be measured with precision, the estimates are unbiased. If the aim were to make a statement about the level of malnutrition in any given district, then the sample size in that district would be critical. However, for making a statement about the nutritional status of children at a given level of per capita program expenditures, it is the overall sample, adjusted for cluster sampling, as well as the variance of the regressor, the district means and the covariance between them, that determines this precision.

Leche expenditures, satisfying one condition for valid instruments. The other condition, which is uncorrelated with the error term, is plausible given that the index was formulated based on the 1993 census, three years before the 1996 Demographic and Health Survey (DHS). The FONCODES index may be correlated with the levels of the unobserved factors, but since the analysis also includes fixed effects estimates that are, in effect, based on the change in Vaso de Leche expenditures, the properties of the instrument in these estimates are based on the assumption that the index is uncorrelated with changes in unobserved factors. If the parameter of interest—the impact of Vaso de Leche expenditures—points to the same conclusion over the set of estimates, there can be reasonable confidence that the conclusion is robust. Although the surveys are pooled, implicitly restricting the parameters of individual and household characteristics to be constant over time, the instrumenting equations are allowed to vary between periods. This is done in two ways: by including a time dummy variable as a shifter and by allowing all of the parameters to vary over time (in which case province-level fixed effects models are estimated).

Moreover, because the basis for the FONCODES index will remain problematic, a second instrumental variable-type method is also employed using a different means of identification. In this method, proposed by Lewbel (2004) (see also Rigobon 2003), the identification of γ comes from exploiting the heteroskedasticity of the first-stage equation (Vaso de Leche expenditures). To illustrate, begin by defining the first-stage equation as

$$(7) \quad VL = \alpha_2 + \zeta X + v$$

where X can include all or a subset of the explanatory variables in the main equation and can include instruments such as the FONCODES index. If $\text{cov}(X, v^2)$ is nonzero (i.e., if the data are heteroskedastic), then γ and the other parameters in the main equation can be estimated consistently without external instruments by an ordinary linear two-stage least squares regression in which all the exogenous right-side variables and $(X - \bar{X})v^2$ are used as instruments for Vaso de Leche expenditures. The requirement that $\text{cov}(X, v^2) \neq 0$ is tested by applying a Breusch and Pagan (1979) test for heteroskedasticity to the first-stage equation.

The district-level Vaso de Leche expenditure data are merged with the DHS data for 1996 and 2000 to create a data set with 19,053 observations on child heights, which is used to estimate the model. The per capita Vaso de Leche district expenditure variable is the district average amount spent in the two years before and including the 1996 and 2000 surveys.

Thus, five variations of the model are estimated using individual child nutritional status as the dependent variable. First, ordinary least squares (OLS) model is used to estimate the basic nonfixed effects model. Second, a series of fixed effects models are estimated, starting with an OLS model. Third, this is followed by a time-varying instrumental variable model that is run with province-level (not district-level) dummy variables, since time-varying district dummies in the

instrumenting equation would perfectly predict the district-level Vaso de Leche expenditure values. The fourth is a fixed effects model using instrumental variable methods in which a time dummy variable is included in the instrumenting equation. Lastly, Lewbel's (2004) method of taking advantage of the heteroskedastic nature of the data for identification is used to verify that the results are robust. The particular advantage of this fifth model is that it does not use the FONCODES index classifications at all and thus is free of any possible problems associated with that instrument. In summary, the following models are estimated: (a) basic OLS; (b) district fixed effects, OLS; (c) province fixed effects, time-varying instrumental variable; (d) district fixed effects, time dummy variable in instrumental variable equation; and (e) district fixed effects, heteroskedasticity identification.

In all of the estimates, Huber–White standard errors are estimated to correct for homogeneity among observations in the 1,364 primary sampling units (clusters).⁷

Finally, while selective migration into high Vaso de Leche districts is theoretically a possibility, it is unlikely that the small transfer (1.8 percent of the income of the poor on average in 1997) is a major determinant of migration. To get an indication of whether this is a major concern, survey data were used to examine the probability that an individual migrated to the district of current residence from another district. While the Vaso de Leche allocation in the district of origin is not known, the marginal effect of current Vaso de Leche expenditures on the probability of migrating in the past ten years is known to be -0.0004 ($z = -0.73$). Thus, the results are unlikely to be biased if current residence is taken as exogenously determined.

III. DESCRIPTION OF DATA SOURCES

This analysis of the Vaso de Leche program benefits from a wealth of data sources available in Peru. The data come from four main sources: information on the geographic allocation of Vaso de Leche program expenditures, national household living standard surveys, national DHS, and the Public Expenditure Tracking Survey. While having multiple data sources is preferred, for evaluating nutritional impacts as illustrated here analysts need only program expenditure and household survey data with child anthropometrics.

Vaso de Leche Expenditures

The Vaso de Leche program has maintained monthly records of expenditures allocated to each administrative (department, province, and district) region in Peru since 1994. This information, along with district population sizes from the 1993 census and the 2000 pre-census, is used to determine real

7. There is slight change in standard errors if the Huber-White standard errors correction is based on districts rather than the less aggregated sample units.

annual total and per capita program expenditures in each of the recipient districts for 1994–2000. Allocations to the district program committees do not translate fully into benefits to recipients, but considering the small scale of the leakages found at the committee level by the Public Expenditure Tracking Survey, they likely represent a reasonably accurate proxy for the value of benefits available to district residents.

Living Standard Surveys

Two sources of household living standard surveys were available for this study. The first is the National Household Survey (Encuesta Nacional de Hogares, ENAHO), collected by the National Institute of Statistics and Information (INEI) in 1998, 1999, and 2000. These nationally representative surveys of more than 6,500 households (2,000 for the 2000 survey) were carried out quarterly, with each quarter's survey focusing on a different theme. This analysis concentrates on data from the second quarter module, which focuses on social services and includes information on participation in the Vaso de Leche program. Household income information is also available for each module.

The second source is the 1994 and 1997 National Living Standards Survey (Encuesta Nacional de Hogares sobre Medicion de Niveles de Vida, ENNIV), collected by Instituto Cuanto. These nationally representative surveys of more than 3,500 households collect multiple indicators of household and individual well-being (e.g., education, housing, health, economic activity, consumption, and assets). The 1994 ENNIV includes information on Vaso de Leche participation by household, and the 1997 data also include estimates of the values of the transfers made to the household. Anthropometric measurements of heights and weights of young children were also recorded.

Demographic and Health Surveys

DHS were carried out in Peru in 1996 and 2000. These nationally representative surveys of more than 28,000 households each are part of a program funded by the United States Agency for International Development and implemented by Macro International Inc., which has included more than 70 nationally representative household surveys in more than 50 countries. The surveys are conducted in single rounds with two main instruments: an individual questionnaire for women of reproductive age (15–49 years old) and a household schedule. Child anthropometric measurements are recorded in the individual module. The household schedule collects information on household members, assets, and access to public services. Since income or expenditure data are not collected, the asset data in the survey were used to predict household per capita expenditures. This was done by estimating a model of log household per capita expenditures in the 1997 ENNIV data, including the value of in-kind transfers. The explanatory variables in this model are assets in the 1997 ENNIV data that are also

available in the DHS data. This model was then used to predict the values of household expenditures in the DHS data.⁸

Public Expenditure Tracking Survey

The Public Expenditure Tracking Survey for the Vaso de Leche program was conducted by Instituto Apoyo at the end of 2001 and early in 2002, to quantify leakages and delays in public expenditure disbursements and to assess the effects of deficiencies in the system on the quality of the services provided. Thus, interviews were conducted at three levels: the municipality, the mothers committee, and the household. One hundred municipalities were sampled, and four mothers committees were randomly selected from each. Lastly, four beneficiary households were selected randomly from each mothers committee in the sample. Because there are fewer than four committees in some municipalities, only 393 committees and 1,587 beneficiary households were interviewed. The household survey includes information on household demographics, assets, and participation in the Vaso de Leche program, including the values of transfers, products received, and additional purchases made.

As with the DHS, neither income nor expenditure data were collected. Nonetheless, the share of the total program received, by wealth quintiles, was established by constructing a wealth index from the households' asset information using a factor-analysis methodology that is regularly applied to the DHS data sets (Filmer and Pritchett 2001; Sahn and Stifel 2003). Because information on households that do not participate in the Vaso de Leche program is not included in the Public Expenditure Tracking Survey data, asset weights are derived from the nationally representative 2000 DHS and applied to the tracking survey data. This permits determining how households sampled in the tracking survey rank relative to the overall national population.

The purposes for which the various data sets are used are summarized in appendix Table A.1. Vaso de Leche targeting is evaluated using the DHS, ENNIV, and ENAHO data. The Public Expenditure Tracking Survey data are used to examine the degree to which Vaso de Leche transfers are inframarginal, and the DHS⁹ and the district-level Vaso de Leche expenditure data are used in the child nutrition models.

8. The results of the first-stage regressions estimated with the ENNIV 1997 data are available on request from the authors.

9. Only the DHS data are employed for the nutrition models for two main reasons. First, two comparable data sets with anthropometric measurements are needed, so that district dummy variables as well as Vaso de Leche subsidies can be included in the models (see World Bank 1999 for a discussion of some of the comparability issues related to the ENNIV surveys). Second, although both the 1994 and 1997 ENNIV data sets have anthropometric data, the earliest year for which Vaso de Leche expenditure data are available is 1994. As explained in the text, the models use the average subsidies for the two years prior to and including the survey as explanatory variables, which would not be available for the 1994 ENNIV.

IV. RESULTS

Before examining the impact of Vaso de Leche expenditures on nutritional outcomes, this section clarifies earlier statements regarding the distribution of Vaso de Leche transfers and targeting.

Distribution

The results confirm that the Vaso de Leche program is reasonably well targeted in terms of both household incomes and child nutritional status, though there have been some leakages. For incomes, this is done by comparing the coverage rates of households by their per capita income levels¹⁰ for five household surveys (Table 1). The percentage of households with children aged six and under (tier I target group) who receive Vaso de Leche transfers declines sharply with the level of income. For example, in 1994, coverage rates declined from 38 percent ($[39.3 + 37.0112]$) of the households in the two poorest quintiles to less than 8 percent in the richest. As the coverage for all households with children increased from 28 percent in 1994 to 48 percent in 2000, coverage in the two poorest quintiles rose from 38 to 68 percent ($[68.2 + 66.9]/2$). While there was a concurrent increase for the more well-off people in the population, the poorest 40 percent of eligible households received more than three times as much as the richest 20 percent on average. These coverage rates compare favorably to the experiences in other Latin American countries (Grosh 1994) and other developing countries (Coady, Grosh, and Hoddinott 2004).

The Public Expenditure Tracking Survey data show that the mean transfer to households in the poorest national asset index quintile is 23 percent larger than to households in the richest quintile. Notably, the bulk of this comes in the form of milk products. The mean value of milk products transferred to the poorest quintile is 135 soles and 18 soles for milk substitutes and other products. Conversely, the mean values of other products received in the other quintiles are between 52 and 100 percent of the mean value of the milk products they receive (see Stifel and Alderman 2005 for more details). Therefore, milk product transfers are generally progressive in the values received by beneficiaries, while transfers of nonmilk products are not.

In nutrition-based targeting, the Vaso de Leche program also is concentrated on households with children of low nutritional status, as illustrated by coverage rates of all children under five years of age by quintile of HAZ for the three household surveys with information on both Vaso de Leche participation and anthropometric status of children (Table 2). To give a sense of program leakage to nonmalnourished children, the percentage of the children in each of the quintiles who are stunted is also shown (HAZ below -2). In 1997, for example, 64 percent of children in the least well-nourished quintile (those who are all stunted) lived in households that received Vaso de Leche food transfers, while just over 30 percent

10. Household per capita consumption is used for the 1994 and 1997 ENNV data.

TABLE 1. Vaso de Leche Coverage Rates by Quintiles of Per Capita Income (Percent)

| Quintile | ENNIV ^a 1994 | ENNIV ^a 1997 | ENAH0 1998 | ENAH0 1999 | ENAH0 2000 | Annual transfers per capita (1997 soles) ENNIV 1997 |
|-------------|----------------------------|----------------------------|---------------|---------------|---------------|---|
| 1 (poorest) | 39.3 | 60.5 | 65.5 | 59.4 | 68.2 | 26 |
| 2 | 37.0 | 52.4 | 61.5 | 50.0 | 66.9 | 30 |
| 3 | 34.3 | 44.6 | 48.2 | 39.4 | 49.4 | 19 |
| 4 | 20.1 | 30.7 | 36.0 | 29.3 | 37.3 | 22 |
| 5 (richest) | 7.8 | 15.8 | 20.2 | 15.8 | 15.2 | 7 |
| Total | 27.7 | 40.8 | 46.3 | 38.8 | 47.5 | 21 |

^aExpenditure per capita rather than income quintiles.

Note: Domain is the set of households with at least one child aged six or younger. Encuesta Nacional de Hogares sobre Medicion de Niveles de Vida (ENNIV) is a National Living Standards Survey; Encuesta Nacional de Hogares (ENAH0) is a National Household Survey.

Source: Authors' analysis based on data described in the text; see also appendix table A.1.

TABLE 2. Vaso de Leche Coverage Rates and Child Malnutrition by Quintiles of Height for Age z-scores (HAZ)(Percent)

| Quintile | ENNIV 1994 | DHS 1996 | ENNIV 1997 |
|-----------------------------------|------------|----------|------------|
| Share of children in program | | | |
| 1 | 42.8 | 41.6 | 64.0 |
| 2 | 33.8 | 33.2 | 49.2 |
| 3 | 28.4 | 26.4 | 41.8 |
| 4 | 25.4 | 21.7 | 34.8 |
| 5 | 20.0 | 20.8 | 30.5 |
| Total | 30.1 | 28.7 | 44.1 |
| Share of children who are stunted | | | |
| 1 | 100.0 | 100.0 | 100.0 |
| | 39.1 | 29.5 | 19.0 |
| | 0.0 | 0.0 | 0.0 |
| | 0.0 | 0.0 | 0.0 |
| | 0.0 | 0.0 | 0.0 |

Note: Domain is the set of children with HAZ. ENNIV (Encuesta Nacional de Hogares sobre Medicion de Niveles de Vida) is National Living Standards Survey; DHS, Demographic and Household Survey.

Source: Authors' analysis based on data described in the text; see also appendix table A.1.

in the most nourished quintile lived in households that received transfers. Nonetheless, despite the fact that the primary stated objective of the Vaso de Leche program is to reduce the levels of malnutrition in Peru, over a third of the intended beneficiaries in the most malnourished quintile were missed.

It is possible that targeting of children based on ex ante nutritional needs would have resulted in improved ex post outcomes. This could explain the low

TABLE 3. Inframarginality of Vaso de Leche Transfers by Quintile of Per Capita Expenditure (Percent of Beneficiary Households)

| | Total | 1 (Poorest) | 2 | 3 | 4 | 5 (Richest) |
|---|-------|-------------|------|------|------|-------------|
| Share that receive | | | | | | |
| Fluid milk/dairy products | 29.4 | 18.4 | 20.1 | 23.0 | 32.9 | 41.3 |
| Milk substitutes ^a | 53.3 | 74.8 | 48.8 | 43.2 | 51.5 | 48.7 |
| Milk and milk substitutes ^a | 79.5 | 89.6 | 67.3 | 63.9 | 81.7 | 85.4 |
| Other products | 58.3 | 22.3 | 58.7 | 62.6 | 62.9 | 74.9 |
| Share that purchase additional ^b | | | | | | |
| Fluid milk/dairy products | 42.5 | 20.7 | 14.6 | 30.4 | 52.1 | 52.6 |
| Milk substitutes ^a | 2.6 | 2.2 | 3.0 | 4.1 | 1.5 | 3.0 |
| Milk and milk substitutes ^a | 48.6 | 36.1 | 38.9 | 36.4 | 57.2 | 58.7 |
| Other products | 26.5 | 9.7 | 19.7 | 15.0 | 26.6 | 37.1 |

^aIncludes powdered milk.

^bShare of beneficiary households that receive the product.

Source: Authors' analysis based on data described in the text; see also appendix table A.1.

levels of coverage of malnourished children. However, if targeting based on ex ante needs is persistently effective, then as the nutritional status of participants improves over time, deterioration in the degree of targeting on malnutrition should be observed. This appears not to be the case; coverage rates for malnourished children rose from 42.5 percent in 1994 to 63.5 percent in 1997.

The Public Expenditure Tracking Survey data offer further indication of whether the quantities of milk provided to households by the program are extramarginal.¹¹ If so, they are expected to have a larger impact on milk consumption than an inframarginal program might have. While inframarginal transfers and extramarginal transfers have the same income effect, extramarginal transfers have a price effect as well.

Nearly half of recipients consume additional amounts of the products distributed to them through the Vaso de Leche program (Table 3). For example, for the 80 percent of recipients who receive milk and milk substitutes from the program, 49 percent purchase additional milk and milk substitutes. For 29 percent of households that receive milk and dairy products, the program is inframarginal for 43 percent of them with respect to these products. While only 3 percent of households that receive milk substitutes (53 percent of recipient households) purchase additional milk substitutes, most of these households also purchase milk and dairy products. For half of these households, the Vaso de Leche program

11. The transfer is extramarginal if ex post consumption (what is observed) is exactly equal to the transfer - the recipient would consume less of the product if the transfer were in the form of cash (the recipient is consuming at the kink in the budget constraint). Alternatively, if the recipient purchases additional amounts of the product, the transfer is inframarginal.

is inframarginal over the more broadly defined category of milk and milk substitutes (hence the 49 percent figure above) but not for milk substitutes alone.

Thus, although the Vaso de Leche program is found to be reasonably well targeted to the expected beneficiary groups, it is unclear *ex ante* what effect the program has had on reducing child malnutrition. The econometric analysis, discussed below, sheds some light on this issue.

Impact of Vaso de Leche Transfers on Nutritional Outcomes

This section assesses the impact of the Vaso de Leche food transfer program on nutrition by examining how the transfers affect child nutritional outcomes. This is done by estimating reduced-form models with standardized HAZ of children less than five years of age as the dependent variable. The summary statistics of the variables used in the model are shown in Table 4.

The stunting rate dropped only marginally from 26.0 percent in 1996 to 25.8 percent in 2000. This difference is not statistically significant. Because of the many confounding influences, however, the lack of progress in reducing child malnutrition is not sufficient in itself to assess the impact of the Vaso de Leche program. So models are also used.

These reduced-form models are conditioned on predicted log per capita household expenditures as a proxy for the potentially endogenous actual household expenditures.¹² In all of the models, the parameter estimates on expenditures are statistically significant at the 99 percent level of confidence (Table 5), confirming that household wealth has a positive impact on child nutritional status, a finding consistent with results on the impact of instrumented expenditure for Peru (Haddad and others 2003). For the basic models (without fixed effects), the parameter estimate is approximately 0.43; in the fixed effects models, these parameter estimates drop to between 0.16 and 0.19. Using the income response in the fixed effects models and an average income growth of 3.5 percent per capita, neutrally distributed, a counterfactual can be constructed using the 1996 DHS data. They indicate that the 26 percent malnutrition rate in that year would have declined to 25.2 percent in 2000, which is somewhat below the observed level. Moreover, the small transfer embodied in the Vaso de Leche would by itself have a negligible impact of roughly 0.003 on average z-scores.

However, as discussed in Das, Do, and Ozler (2005), in-kind or conditional transfers are expected to have a greater impact on behavior than indicated by an income transfer alone. If so, an additional direct impact from the Vaso de Leche expenditures would be expected above any impact on the level of expenditures. But the direct effect of program expenditures is negative in all of the models (Table 5), although it is not statistically significant in any. Moreover, the parameter estimates are substantively small. Thus, overall there is no evidence that expenditures on the Vaso de Leche program have a direct positive impact on the

12. These models were also estimated using an asset index constructed using factor analysis (Sahn and Stifel 2002) as a control for wealth. The results, which are qualitatively the same, are available on request from the authors.

TABLE 4. Means of Variables Used in HAZ Models, DHS 1996 and 2000

| Variable | Pooled sample | | 1996 | | 2000 | |
|---|---------------|--------------------|-------|--------------------|-------|--------------------|
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| Percent stunted | 25.9 | | 26.0 | | 25.8 | |
| HAZ ^a | -1.11 | 1.34 | -1.07 | 1.34 | -1.18 | 1.32 |
| Per capita Vaso de Leche district expenditures | 31.90 | 7.63 | 29.43 | 5.90 | 37.00 | 8.23 |
| Log per capita household expenditures (predicted) | 8.22 | 0.39 | 8.23 | 0.40 | 8.21 | 0.38 |
| Male dummy variable | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Multiple birth dummy variable | 0.01 | 0.11 | 0.01 | 0.12 | 0.01 | 0.11 |
| Birth order, second child | 0.22 | 0.42 | 0.22 | 0.42 | 0.23 | 0.42 |
| Birth order, third child | 0.16 | 0.37 | 0.16 | 0.37 | 0.17 | 0.37 |
| Birth order, fourth child | 0.10 | 0.30 | 0.11 | 0.31 | 0.10 | 0.29 |
| Birth order, fifth child | 0.07 | 0.25 | 0.07 | 0.25 | 0.07 | 0.25 |
| Birth order, sixth child and above | 0.14 | 0.35 | 0.15 | 0.35 | 0.14 | 0.34 |
| Age 0-6 months | 0.10 | 0.30 | 0.10 | 0.31 | 0.09 | 0.29 |
| Age 7-12 months | 0.10 | 0.31 | 0.11 | 0.31 | 0.10 | 0.29 |
| Age 13-18 months | 0.10 | 0.30 | 0.10 | 0.30 | 0.10 | 0.30 |
| Age 25-35 months | 0.19 | 0.39 | 0.19 | 0.39 | 0.18 | 0.39 |
| Age 36-59 months | 0.41 | 0.49 | 0.41 | 0.49 | 0.42 | 0.49 |
| Share household members age 0-5 (%) | 0.30 | 0.13 | 0.30 | 0.13 | 0.29 | 0.12 |
| Share household girls age 6-15 (%) | 0.10 | 0.12 | 0.10 | 0.12 | 0.10 | 0.12 |
| Share household boys age 6-15 (%) | 0.10 | 0.12 | 0.10 | 0.12 | 0.10 | 0.12 |

(Continued)

TABLE 4. Continued

| Variable | Pooled sample | | 1996 | | 2000 | |
|-------------------------------------|---------------|--------------------|--------|--------------------|-------|--------------------|
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| Father's education, postsecondary | 0.20 | 0.40 | 0.20 | 0.40 | 0.20 | 0.40 |
| House floor dirt dummy variable | 0.51 | 0.50 | 0.51 | 0.50 | 0.51 | 0.50 |
| Piped drinking water dummy variable | 0.55 | 0.50 | 0.54 | 0.50 | 0.58 | 0.49 |
| Flush toilet dummy variable | 0.40 | 0.49 | 0.41 | 0.49 | 0.39 | 0.49 |
| Urban dummy variable | 0.68 | 0.47 | 0.71 | 0.45 | 0.61 | 0.49 |
| Number of observations | 19,053 | | 12,045 | | 7,008 | |

^aStandardized anthropometric height for age z-score (HAZ).

Source: Authors' analysis based on data described in the text; see also appendix table A.1.

TABLE 5. Reduced-Form Models of Height for Age z-score (HAZ) (Ages 0–59), Peru Demographic and Household Surveys 1996 and 2000

| | District fixed effects | | | | | | | | | |
|---|------------------------|-------------|-------------|-------------|----------------------------------|-------------|-------------------|-------------|---------------------------------------|-------------|
| | OLS (1) | | OLS (2) | | Time varying IV (3) ^a | | Time dummy IV (4) | | Heteroskedasticity Identification (5) | |
| | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic |
| Per capita Vaso de Leche district expenditures | -0.0015 | -0.90 | -0.0022 | -0.90 | -0.0019 | -0.79 | -0.0018 | -0.75 | -0.0025 | -0.89 |
| Log per capita household expenditures (predicted) | 0.422 | 7.61*** | 0.164 | 2.90*** | 0.192 | 3.36*** | 0.164 | 2.90*** | 0.164 | 2.90*** |
| Male dummy variable | -0.083 | -4.27*** | -0.075 | -3.96*** | -0.080 | -4.23*** | -0.075 | -3.96*** | -0.075 | -3.96*** |
| Multiple birth dummy variable | -0.415 | -4.03*** | -0.459 | -4.80*** | -0.461 | -4.86*** | -0.459 | -4.80*** | -0.459 | -4.81*** |
| Birth order, second child | -0.069 | -2.06*** | -0.062 | -1.86* | -0.067 | -2.01** | -0.062 | -1.86* | -0.062 | -1.86* |
| Birth order, third child | -0.160 | -3.99*** | -0.140 | -3.53*** | -0.151 | -3.82*** | -0.140 | -3.52*** | -0.140 | -3.53*** |
| Birth order, fourth child | -0.274 | -5.53*** | -0.256 | -5.29*** | -0.267 | -5.57*** | -0.256 | -5.29*** | -0.256 | -5.29*** |
| Birth order, fifth child | -0.382 | -5.97*** | -0.333 | -5.42*** | -0.351 | -5.68*** | -0.333 | -5.41*** | -0.333 | -5.42*** |
| Birth order, sixth child and above | -0.434 | -6.69*** | -0.364 | -5.82*** | -0.378 | -6.02*** | -0.364 | -5.81*** | -0.364 | -5.82*** |
| Age 0–6 months | 1.163 | 25.12*** | 1.138 | 24.47*** | 1.143 | 24.79*** | 1.138 | 24.47*** | 1.138 | 24.47*** |
| Age 7–12 months | 0.662 | 14.02*** | 0.638 | 13.76*** | 0.644 | 13.96*** | 0.638 | 13.76*** | 0.638 | 13.75*** |
| Age 13–18 months | 0.148 | 3.24*** | 0.139 | 3.13*** | 0.146 | 3.25*** | 0.140 | 3.13*** | 0.139 | 3.13*** |
| Age 25–35 months | 0.283 | 6.90*** | 0.265 | 6.72*** | 0.270 | 6.89*** | 0.265 | 6.72*** | 0.265 | 6.72*** |
| Age 36–59 months | 0.054 | 1.51 | 0.044 | 1.29 | 0.046 | 1.34 | 0.044 | 1.29 | 0.044 | 1.29 |
| Share household members age 0–5 (%) | -0.670 | -3.04*** | -0.599 | -2.74*** | -0.600 | -2.78*** | -0.598 | -2.73*** | -0.600 | -2.74*** |
| Share household girls age 6–15 (%) | -0.3067 | -1.38 | -0.211 | -0.95 | -0.198 | -0.90 | -0.210 | -0.94 | -0.212 | -0.95 |
| Share household boys age 6–15 (%) | -0.373 | -1.62 | -0.300 | -1.33 | -0.297 | -1.33 | -0.299 | -1.33 | -0.301 | -1.33 |
| Share household women 16–25 (%) | -0.029 | -0.12 | 0.019 | 0.08 | 0.022 | 0.09 | 0.020 | 0.08 | 0.019 | 0.08 |
| Share household women 26–65 (%) | 0.107 | 0.43 | 0.225 | 0.94 | 0.231 | 0.97 | 0.226 | 0.95 | 0.224 | 0.94 |
| Share household men 16–25 (%) | 0.214 | 0.91 | 0.290 | 1.25 | 0.278 | 1.21 | 0.291 | 1.25 | 0.290 | 1.24 |

(Continued)

TABLE 5. Continued

| | District fixed effects | | | | | | | | | |
|------------------------------------|------------------------|-------------|-----------------------|-------------|----------------------------------|-------------|-------------------|-------------|---------------------------------------|-------------|
| | OLS (1) | | OLS (2) | | Time varying IV (3) ^a | | Time dummy IV (4) | | Heteroskedasticity Identification (5) | |
| | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic |
| Share householdmen 26–65 (%) | 0.348 | 1.46 | 0.424 | 1.79* | 0.428 | 1.83* | 0.424 | 1.79* | 0.424 | 1.78* |
| Number of householdmembers | -0.026 | -5.24*** | -0.023 | -4.64*** | -0.023 | -4.74*** | -0.023 | -4.63*** | -0.023 | -4.64*** |
| Head is male | -0.133 | -3.41*** | -0.095 | -2.46** | -0.097 | -2.53** | -0.095 | -2.45** | -0.096 | -2.46** |
| Head is indigenous | -0.226 | -4.68*** | -0.272 | -4.91*** | -0.254 | -4.77*** | -0.272 | -4.91*** | -0.272 | -4.91*** |
| Mother's age | 0.063 | 3.75* | 0.057 | 3.43*** | 0.058 | 3.51*** | 0.057 | 3.43*** | 0.057 | 3.43*** |
| Mother's agesquared | -0.001 | -2.67*** | -0.001 | -2.52** | -0.001 | -2.54** | -0.001 | -2.52** | -0.001 | -2.52** |
| Mother'sheight (centimeters) | 0.054 | 23.07*** | 0.051 | 22.65*** | 0.052 | 23.10*** | 0.051 | 22.65*** | 0.051 | 22.65*** |
| Mother's education, primary | 0.013 | 0.25 | -0.021 | -0.44 | -0.017 | -0.36 | -0.022 | -0.45 | -0.021 | -0.44 |
| Mother's education, secondary | 0.072 | 1.19 | 0.020 | 0.35 | 0.028 | 0.49 | 0.020 | 0.34 | 0.020 | 0.35 |
| Mother's education, postsecondary | 0.116 | 1.70* | 0.113 | 1.75* | 0.120 | 1.89* | 0.112 | 1.74* | 0.113 | 1.75* |
| Father's education, primary | -0.002 | -0.04 | 0.013 | 0.26 | 0.017 | 0.36 | 0.013 | 0.27 | 0.012 | 0.26 |
| Father's education, secondary | 0.108 | 2.38** | 0.113 | 2.52** | 0.113 | 2.53** | 0.113 | 2.53** | 0.113 | 2.52** |
| Father's education, postsecondary | 0.012 | 0.24 | 0.104 | 2.00** | 0.099 | 1.90* | 0.104 | 2.00** | 0.104 | 1.99** |
| House floor dirtdummy variable | -0.052 | -1.85* | -0.088 | -2.99*** | -0.080 | -2.82*** | -0.089 | -2.99*** | -0.088 | -2.99*** |
| Piped drinking waterdummy variable | -0.029 | -0.97 | 0.028 | 0.95 | 0.031 | 1.05 | 0.028 | 0.95 | 0.028 | 0.96 |
| Flush toilet dummy variable | 0.107 | 2.93**** | 0.111 | 3.08**** | 0.096 | 2.7**** | 0.111 | 3.08**** | 0.110 | 3.07*** |
| Urban dummyvariable | 0.211 | 5.30*** | District fixed effect | | Fixed effect | | Fixed effect | | Fixed effect | |

(Continued)

TABLE 5. Continued

| | District fixed effects | | | | | | | | | |
|--|------------------------|-------------|-----------------|-------------|----------------------|-------------|-------------------|-------------|---------------------------------------|-------------|
| | OLS (1) | | OLS (2) | | Time varying IV (3)" | | Time dummy IV (4) | | Heteroskedasticity Identification (5) | |
| | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic |
| Constant | -13.75 | -22.12"" | dummies omitted | | dummies omitted | | dummies omitted | | dummies omitted | |
| FONCODES index (t= 0) in IV equation | | | | | 0.222 | 49.21*** | | | | |
| FONCODES index (t= 1)in IV equation | | | | | 0.311 | 53.63*** | | | | |
| FONCODES index inIV equation | | | | | | | 0.150 | 2.43** | | |
| F-statistic (H_0 :instruments jointly 0) | | | | | 153.6** | | 4329.0** | | | |
| χ^2 (H_0 : OLS and IVestimates same) | | | | | 5.56* | | 6.27* | | | |
| χ^2 (Breusch–Pagan test) | | | | | | | | | | 699.90** |
| Number of observations | 19,053 | | 19,053 | | 19,053 | | 19,053 | | | 19,053 |
| R ² | 0.300 | | 0.356 | | 0.341 | | 0.355 | | | 0.355 |

OLS, ordinary least squares.

"Significant at the 90 percent level of confidence.

""Significant at the 95 percent level of confidence.

""""Significant at the 99 percent level of confidence.

"Province-level fixed effects.

Note: Instrument in IV models is district-level FONCODES index.

Source: Authors' analysis based on data described in the text; see also appendix table A.1.

nutritional outcomes of young children—the group to whom the program is directed—using either the preferred approach (controlling for the initial conditions in communities with district fixed effects) or other models.

In both of the standard instrumental variable models, the identifying instrument (FONCODES index of unmet needs) is significantly correlated with per capita district expenditures, and the instruments overall (including the time dummy variable) are jointly significant at the 1 percent level.¹³ The first-stage parameter estimates for the FONCODES index for both 1996 and 2000 in the time-varying instrumental variable model (model 3) are positive and strongly significant, with a larger effect in 2000. While this confirms that the instrument is valid in terms of its correlation with Vaso de Leche expenditures, it also implies that marginal program targeting is pro-poor (Lanjouwand Ravallion 1999). The positive parameter estimates for the instruments suggest that the incidence of inframarginal Vaso de Leche spending benefits districts with higher FONCODES indices—the poor benefit more from marginal increases in program spending that may not be distributed homogeneously across all districts.

In model 5, the chi-squared statistic for the Breusch–Pagan test of heteroskedasticity is 699.9, implying that the data in the first-stage equation are heteroskedastic. Thus, following Lewbel (2004), the condition is met for consistent estimation of the impact of Vaso de Leche expenditures on child health. Although the methodology differs from that of models 3 and 4, the parameter estimate is similar.

A few additional specifications were also tried (these are not shown here but are available from the authors). For example, while no average impact is observed in these regressions, it is possible that the impact is greater among the poor, where malnutrition rates are higher. Thus, the regressions were rerun for only the poorest 40 percent. The point estimates for the coefficient of per capita Vaso de Leche expenditures remained negative and were greater in absolute value for all of the models than the results reported in Table 5.¹⁴ Thus, there is no indication that the impact on the poor was masked by aggregation. Similarly, the results are unchanged when children under six months of age are excluded, to rule out the possibility that children who were being breast-feed would respond less to an in-kind subsidy¹⁵ and when the

13. Durbin-Wu-Hausman chi-square tests that the OLS and instrumental variable estimates are the same are all rejected at the 5 percent level (table 5).

14. While probit analysis does not use all of the information available in the data since the variation in the continuous dependent variable is ignored when it is converted to a binary variable, the model was also tested with a probit analysis as the threshold to keep the focus on the malnourished. The coefficient of Vaso de Leche expenditures remained nonsignificant.

15. In addition, the models were reestimated using other measures of nutritional status. The results were similar to the estimates presented here. More specifically, two measures were used: weight-for-height z-scores (WHZ), a measure of short-term nutritional status, and weight-for-age z-scores (WAZ), a composite of weight for height and height for age. A further test of robustness involved splitting the sample into urban and rural samples and estimating separate models. Since program leakage is higher in urban areas than in rural areas, a program effect might have been expected in rural areas, but the separate estimates did not differ substantially from those presented in this article. Therefore, an aggregation bias from pooling the urban and rural samples appears not to be driving the results.

parameters are allowed to differ across urban and rural areas in the fixed effects models.

Nonsignificant coefficients in a population survey may mask a response if beneficiaries are a small share of the population. However, the program covers nearly half the population and two-thirds of the poor. Thus, more than 8,000 individuals in the sample are beneficiaries of the program. Nonsignificant coefficients may also reflect imprecision in measurement. In this case, the confidence intervals—while narrow—do cross over into positive values. Still, at the largest positive value for the 95 percent confidence interval for the impact of 37 soles (\$11.50) of program expenditures per capita in, say, model 2, the program would increase z -scores by 0.09.

V. CONCLUSION

This article looked for evidence that the politically popular focus on the provision of milk to young children can further nutritional objectives. It studied expenditures on the Vaso de Leche feeding program in Peru, a program that is reasonably progressive in its distributional impact relative to international experience (Grosh 1994; Coady, Grosh, and Hoddinott 2004). The article was motivated by the paucity of evidence on the nutritional impact of similar programs worldwide despite the millions of dollars spent on them each year. It illustrates a methodology for linking public expenditure data with household survey data first to substantiate the targeting and then to model the determinants of nutritional outcomes of children to see whether Vaso de Leche program interventions have an impact on nutrition.

In the models of standardized child heights, the magnitude of program expenditures provided to the community rather than household participation is used as an explanatory variable, solving the issue of endogenous household choice. Even when accounting for endogenous program placement with fixed effects models, and further with alternative approaches to instrumental variables, Vaso de Leche program expenditures are found to have no impact on the nutritional outcomes of young children—the group to whom the program is targeted.

The results do confirm that the Vaso de Leche program is reasonably well targeted to poor households and to households with low nutritional status: some 50 percent of the poor received program benefits, while less than 20 percent of the nonpoor did. In value terms, more than 60 percent (possibly up to 75 percent) of the allocated Vaso de Leche budget goes to the poor. Therefore, the absence of a measurable impact on child growth is not likely explained by mistargeting. Indeed, given that the program expenditure has no observed impact on nutritional status other than that through any increase in household expenditures, further improvements in targeting are not by themselves likely to affect nutritional outcomes. To gauge this, the

effect of redistributing all of the Vaso de Leche benefits received by the nonpoor evenly among the poor was simulated using the household expenditure coefficient from the district fixed effects models. Malnutrition rates decreased only 0.28 percentage points.

One possible reason why the impact of food subsidies beyond their value as income transfers is limited may be the degree to which the commodity transfers are *inframarginal*. The 2002 Public Expenditure Tracking Survey data show that transfers of milk and milk substitutes from the Vaso de Leche program are *inframarginal* for approximately half the households that receive them. This can be only a partial explanation, however, since the results do not change when the estimates are confined to the poorest 40 percent, a group for which milk is less likely to be *inframarginal*.

There are, however, other means by which a subsidy to milk might achieve an improvement in child nutrition. For example, milk fortification may be a promising way to address anemia; programs in Argentina, Chile, and Mexico currently fortify milk with iron to achieve this objective. Randomization of fortification in two dozen communities in Mexico verified the efficacy of this approach (J. Rivera, pers. comm.). However, this was not an objective of the Vaso de Leche program in Peru during the period being studied.

Another possible approach to milk subsidies is to include them as part of a broader program for improving nutrition, as with the Women, Infants, and Children program in the United States. The review by Rivera, Hotz, and others (2003) and an earlier article by Garcia and Pinstrip-Andersen (1987) indicate the efficacy of supplementation programs that include nutrition education. Again, the Vaso de Leche program is not directly embedded in other interventions aimed at improving nutrition, so the current study cannot assess this possibility. Still, as this study (and the few similar studies in the literature) fails to find an impact on child growth from a subsidy on commercial milk products, the results reinforce the view that without additional measures dairy subsidies do not address the efficiency objectives of a transfer program in addressing malnutrition. Despite being reasonably well targeted to the poor and malnourished, the Vaso de Leche program fails to improve the nutritional status of young children. The implication for Peru and for other countries is that where costly *in-kind* transfers are also largely *inframarginal* and not fortified with micronutrients, they offer little if any efficiency gains as measured by improved child nutritional outcomes. Under such circumstances, cash transfers could be less costly and potentially an equally effective means of achieving nutritional and other distributional objectives.

APPENDIX TABLE A.1. Data Sources and Their Use in the Analysis

| Data source | Type | Year | Anthropo- metrics | Welfarist metric | Information on Vaso de Leche participation | Use in analysis |
|--|---|-----------|----------------------|---------------------|---|--|
| National Household Survey (Encuesta Nacional de Hogares, ENAHO) | Household survey | 1998 | No | Income | Indicators | Income targeting |
| | | 1999 | No | Income | Indicators | Income targeting |
| | | 2000 | No | Income | Indicators | Income targeting |
| National Living Standards Survey (Encuesta Nacional de Hogares sobre Medicion de Niveles de Vida, ENNIV) | Household survey | 1994 | Yes | Consumption | Indicators | Income and nutrition targeting |
| | | 1997" | Yes | Consumption | Values | Income and nutrition targeting |
| Demographic and Household Survey | Household survey | 1996 | Yes | Asset index | Indicators | Nutrition targeting and child nutrition models |
| Public Expenditure Tracking Survey | Expenditure tracking/household survey | 2000 | Yes | Asset index | No | Child nutrition models |
| | | 2001102 | No | Asset index | All households surveyed are participants | Inframarginality analysis |
| Vaso de Leche program expenditure data | District-level records | 1994–2000 | | | | Nutrition impact regressions |

"Data set used by Ruggeri Laderchi (2001).

REFERENCES

- Alderman, H., P. Chiappori, L. Haddad, J. Hoddinott, and R. Kanbur. 1995. "Unitary Versus Collective Models of the Household: Time to Shift the Burden of Proof?" *World Bank Research Observer* 10(1):1–20.
- Alderman, H., and C. del Ninno. 1999. "Poverty Issues for Zero Rating VAT in South Africa." *Journal of African Economies* 8(2):182–208.
- Becker, G. 1981. *A Treatise on the Family*. Cambridge, Mass.: Harvard University Press.
- Behrman, J. R., and J. Hoddinott. 2005. "Program Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican *Progresa* Impact on Child Nutrition." *Oxford Bulletin of Economics and Statistics* 67(4):547–69
- Breusch, T., and A. Pagan. 1979. "A Simple Test for Heteroskedasticity and Random Coefficient Variation." *Econometrica* 47(5):1287–94.
- Carlson, A., and B. Senauer. 2003. "The Impact of the Special Supplemental Nutrition Program for Women, Infants, and Children on Child Health." *American Journal of Agricultural Economics* 85(2):479–91.
- Coady, D., M. Grosh, and J. Hoddinott. 2004. *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*. Washington, D.C.: World Bank.
- Das, J., Q.-T. Do, and B. Ozler. 2005. "Reassessing Conditional Cash Transfer Programs." *World Bank Research Observer* 20(1):57–80.
- El Peruano. 2001. "Normas Legales." *Diario Oficial El Peruano* 7650:203875–76.
- Esanu, C., and K. Lindert. 1996. *An Analysis of Consumer Food Price and Subsidy Policy in Romania*. Washington, D.C.: World Bank.
- Filmer, D., and L. Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data - or Tears: An Application of Educational Enrollment in States of India." *Demography* 38(1):115–32.
- Garcia, M., and P. Pinstrip-Andersen. 1987. "The Pilot Food Price Subsidy Scheme in the Philippines: Its Impact on Income, Food Consumption, and Nutritional Status." Research Report 61. International Food Policy Research Institute, Washington, D.C.
- Grosh, M. 1994. *Administering Targeted Social Programs in Latin America: From Platitudes to Practice*. Washington, D.C.: World Bank.
- Gundersen, C., M. Yañez, C. Valdez, and B. Kuhn. 2000. "A Comparison of Food Assistance Programs in Mexico and the United States." Economic Research Service Food Assistance and Nutrition Research Report 6. United States Department of Agriculture, Washington, D.C.
- Haddad, L., H. Alderman, S. Appleton, L. Song, and Y. Yohannes. 2003. "Reducing Child Malnutrition: How Far Does Income Growth Take Us?" *World Bank Economic Review* 17(1):107–31.
- Heckman, J., R. Lalonde, and J. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3A. Amsterdam: Elsevier Science.
- Instituto Apoyo and World Bank. 2002. "Central Government Transfers to Municipalities in Peru: A Detailed Look at the Vaso de Leche Program." Lima.
- Kennedy, E., and H. Alderman. 1987. "Comparative Analyses of Nutritional Effectiveness of Food Subsidies and Other Food-Related Interventions." International Food Policy Research Institute, Washington, D.C. and World Health Organization–United Nations Children's Fund, Geneva, Switzerland.
- Lanjouw, P., and M. Ravallion. 1999. "Benefit Incidence, Public Spending Reforms, and the Timing of Program Capture." *World Bank Economic Review* 13(2):257–73.
- Levinger, B. 1986. "School Feeding Programmes in Developing Countries: An Analysis of Actual and Potential Impact." AID Evaluation Special Study 30. United States Agency for International Development, Washington, D.C.

- Lewbel, A. 2004. "Identification of Heteroskedastic Endogenous or Mismeasured Regressor Models." Unpublished paper. Boston College, Boston, Mass.
- Martorell, R., and J. Habicht. 1986. "Growth in Early Childhood in Developing Countries." In F. Falkner and J. Tanner, eds., *Human Growth: A Comprehensive Treatise*, Vol. 3. New York: Plenum Press.
- Moffitt, R. 1991. "Program Evaluation with Nonexperimental Data." *Evaluation Review* 15(3):291–314.
- Powell, C., S. Grantham-McGregor, S. Walker, and S. Chang. 1998. "School Breakfast Benefits Children's Nutritional Status and School Performance." *American Journal of Clinical Nutrition* 68(4):873–79.
- Rigobon, R. 2003. "Identification through Heteroskedasticity." *Review of Economics and Statistics* 85(4):777–92.
- Rivera, J., C. Hotz, T. González-Cossío, L. Neufeld, and A. Garcia-Guerra. 2003. "The Effect of Micronutrient Deficiencies on Child Growth: A Review of Results from Community-Based Supplementation Trials." *Journal of Nutrition* 133(November):4010S–20S.
- Rivera, J., D. Sotres-Alvarez, J.-P. Habicht, T. Shamah, and S. Villalpando. 2003. "Impact of the Mexican Program for Education, Health, and Nutrition (Progresa) on Rates of Growth and Anemia in Infants and Young Children: A Randomized Effectiveness Study." *Journal of the American Medical Association* 291(4):2563–70.
- Rosenzweig, M., and K. Wolpin. 1986. "Evaluation of the Effects of Optimally Distributed Programs: Child Health and Family Planning Initiatives." *American Economic Review* 76(3):470–82.
- Rouse, C. 1998. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *The Quarterly Journal of Economics* 113(2):553–602.
- Ruggeri Laderchi, C. 2001. "Killing Two Birds with the Same Stone? The Effectiveness of Food Transfers on Nutrition and Monetary Policy." Paper presented at the Latin American Economic Association Conference, October 18–20, Montevideo, Uruguay.
- Rush, D., J. Leighton, N. Sloan, J. Alvir, D. Horvitz, W. Seaver, G. Garbowski, S. Johnson, R. Kulka, and J. Devore. 1988. "The National WIC Evaluation: Evaluation of the Special Supplemental Food Program for Women, Infants, and Children. VI. Study of Infants and Children." *American Journal of Clinical Nutrition* 48(2 Suppl.):484–511.
- Sahn, D., and D. Stifel. 2002. "Parental Preferences for Nutrition of Boys and Girls: Evidence from Africa." *Journal of Development Studies* 39(1):21–45.
- 2003. "Exploring Alternative Measures of Welfare in the Absence of Expenditure Data." *Review of Income and Wealth* 49(4):463–89.
- Schady, N. 1998. *Picking the Poor: Indicators for Geographic Targeting in Peru*. Washington, D.C.: World Bank.
- Stifel, D., and H. Alderman. 2005. "Targeting at the Margin: The 'Glass of Milk' Subsidy Program in Peru." *Journal of Development Studies* 41(5):839–64.
- Strauss, J., and D. Thomas. 1995. "Human Resources: Empirical Modeling of Household and Family Decisions." In J. Behrman and T. Srinivasan, eds., *Handbook of Development Economics*, Vol. 3A. Amsterdam: Elsevier Science.
- Tuck, L., and K. Lindert. 1996. "From Universal Food Subsidies to a Self-Targeted Program: A Case Study of Tunisian Reform." Discussion Paper 351. World Bank, Washington, D.C.
- World Bank and Inter-American Development Bank. 2002. "Peru, Restoring Fiscal Discipline for Poverty Reduction: Public Expenditure Review." Report 24286-PE. Washington, D.C.: World Bank.
- World Bank. 1999. *Poverty and Social Development in Peru, 1994–1997*. Washington, D.C.: World Bank.
- World Bank. 2003. "Brazil, Growth and Poverty Reduction in Rio Grande do Norte: A State Economic Memorandum." Report 24891-BR. World Bank, Washington, D.C.
- Younger, S. 2002. "Public Social Sector Expenditures and Poverty in Peru." In C. Morrison, ed., *Education, Health Expenditure, and Development: The Cases of Indonesia and Peru*. Paris: OECD Development Centre.

How Endowments, Accumulations, and Choice Determine the Geography of Agricultural Productivity in Ecuador

Donald F. Larson and Mauricio León

Spatial disparity in incomes and productivity is apparent across and within countries. Most studies of the determinants of such differences focus on cross-country comparisons or location choice among firms. Less studied are the large differences in agricultural productivity within countries related to concentrations of rural poverty. For policy, understanding the determinants of this geography of agricultural productivity is important, because strategies to reduce poverty often feature components designed to boost regional agricultural incomes. Census and endowment data for Ecuador are used to estimate a model of endogenous technology choice to explain large regional differences in agricultural output and factor productivity. A composite-error estimation technique is used to separate systemic determinants from idiosyncratic differences. Simulations are employed to explore policy avenues. The findings suggest a differentiation between the types of policies that promote growth in agriculture generally and those that are more likely to assist the rural poor.

Regional differences in incomes within countries are often striking and are similar in many ways to differences in average incomes among countries. Why inequality takes on spatially identifiable forms is thought to be related to the characteristics of place that constrain and influence current economic choice and to the historical influences of geography on accumulations of assets by families and communities. To a large degree, spatial differences in economic opportunity are locally recognized and motivate the movement of labor among economic sectors and the migration of households that are a ubiquitous aspect of economic development. Most noticeably, such movements take the form of

Donald F. Larson is a senior economist in the Development Research Group at the World Bank; his email address is dlarson@worldbank.org. Mauricio León is coordinator of the Sistema Integrado de Indicadores Sociales del Ecuador, in Ecuador's Ministry of Social Welfare; his email address is mleon@frentesocial.gov.ec. The authors would like to thank Jamie de Melo, Daniel Lederman, Rinku Murgai, Yair Mundlak, Norbert Schady, J. Edward Taylor, and two anonymous referees for helpful suggestions on earlier drafts. The authors thank the journal technical editor for suggesting improvements when preparing the article for publication. A supplemental appendix to this article is available at <http://wber.oxfordjournals.org/>.

THE WORLD BANK ECONOMIC REVIEW, VOL. 20, NO. 3, pp. 449–471

doi:10.1093/wber/lhl003

Advance Access publication July 25, 2006

© The Author 2006. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development/ THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.

an out-migration from agriculture and a movement from rural to urban space (Larson and Mundlak 1997).

At the same time, not everyone is positioned to take advantage of opportunities in other sectors or locations. Moreover, barriers such as up-front costs, risks, and asymmetries in information further constrain migration. When this is so, circumstances related to household and location characteristics can work in persistent and reinforcing ways to impoverish communities and regions. Commonly, this gives rise to lagging regions with high concentrations of rural poor who depend substantially on agriculture for food and income.

For policymakers, a key question is whether the aspects of disadvantaged areas can be changed through policies in a way that creates greater economic opportunity for poor households. Related to this is the question of how quickly policy can affect the underlying determinants of regional inequality. In particular, the short-term efficacy of policy and the mutability of regional differences are expected to hinge on the degree to which geographic disparities in income depend on unchanging natural endowments, on quasi-fixed accumulations of public and private factors and institutions, and on policy-related incentives and constraints that shape how current resources are used.

To some extent, the study of what makes some regions within a country less prosperous than others is related to the question of what determines international differences in growth and productivity. In this area of research, evidence from cross-country macroeconomic data suggests that growth and productivity are driven by broad state variables that determine productivity directly and also influence choices that affect accumulations over time. A similar conclusion is reached in studies that look at what determines productivity differences among manufacturing firms.

This article takes another perspective and focuses on the determinants of spatial differences in agricultural productivity. It does so chiefly because lagging regions in poor developing countries usually have few linkages to sectors outside agriculture. Consequently, it is important to know whether policy instruments can be identified that bring about significant improvements in agriculture incomes. However, also of interest is the related question of whether results from cross-country and firm studies have counterparts in a study of what determines spatial income differences within a geographically pervasive sector of a single country.

There are several advantages to studying both issues in the context of agriculture in Ecuador. First, because disparities in agricultural productivity and incomes are large in Ecuador, the determinants of spatial inequality can be examined in the context of a sector that is important to Ecuador and in a framework that is different from that of other studies of regional inequality. Second, because census data are available, techniques can be employed that provide more focused measures of productivity and its determinants. Third, the role of endowments is more easily studied in the context of agriculture,

where soil and climate endowments are well measured and their potential contribution well identified.

I. BACKGROUND

Most models of production, including related growth models, start with the assumption that observed income levels arise from a common technology.¹ The assumption is rooted in the firm-theory conditions required to derive a well-defined production function. However, the notion that all firms use the same method to produce creates logical tensions because very heterogeneous levels of productivity are often observed in practice.² For studies based on farm data, differences in productivity can arguably be assigned to idiosyncratic differences in farms and farmers. However, in studies based on aggregate country or sector data, differences in the capability to apply available technology are attributed instead to identifiable country or sector traits. Even so, a common underlying technology is expected to prevail in the long run, and this notion has consequences for modeling approaches.

One view, associated with the endogenous growth literature, suggests that output and productivity differences among nations and among firms are driven largely by differences in adopted technology.³ Because, in this view, the cost of technology diffusion is low and the benefits are high, countries and firms will converge to a common technology in the long run. Trade and open investment channels are expected to facilitate technology diffusion. Consequently, this literature generally sees initial conditions and the speed with which new technologies can be adopted as keys to development. Empirically, studies of this type focus on short-run growth rates or rates of productivity convergence.

A related literature sees technology adoption as less automatic and focuses on barriers (see Parente and Prescott 1994 and references therein). In a similar way, the economic geography literature argues that the costs of adopting new production methods are often specific to location, potentially creating a range of barriers to adoption that vary geographically (see the survey by Henderson, Shalizi, and Venables 2001). In particular, local information about techniques and local-factor markets that support particular forms of production is seen as the basis for spillovers that contribute to productivity and become a force concentrating on economic activity. Transaction costs, the location of endowments, and history play a role as well, creating a shifting set of incentives for centers of economic activity.

1. Griliches (1996) provides a history of early efforts to measure productivity, including a review of early structural models. Mundlak (2001) reviews agricultural production modeling.

2. See related criticisms raised by Stigler (1976).

3. See early studies by Mankiw, Romer, and Weil (1992) and Barro and Sala-i-Martin (1992). Klenow and Rodriguez-Clare (1997) and Brock and Durlauf (2001) provide reviews.

The relationships among transaction costs, markets, and the capacity of governments to protect property rights and enforce contracts are the focus of another related set of studies that emphasize the role of institutions (North 1994; Hall and Jones 1997). Although expensive to build and maintain, institutions are expected to contribute to growth in several ways. Institutions are expected to reduce the risk of diversion or expropriation and to facilitate capital, insurance, hedging, and other related markets that allow risks to be shared. Reducing and sharing risks allow for anonymous exchange, increasing competition, and reducing marginal transaction costs. Dynamically, the workings of institutions allow for faster rates of accumulation of human and physical capital, which contribute directly to greater output.

Ideas related to institutions and economic geography are integrated with growth modeling in several studies designed to identify the deep-seated determinants of growth (Sachs and Warner 1997; Bhattacharyya 2004; Rodrik, Subramanian, and Trebbi 2004). In these studies, factor inputs are viewed as proximate endogenous variables because investment and other choices related to factor accumulation are influenced by the same conditioning variables that determine productivity. For this reason, these studies sometimes take an empirical approach that excludes factors as a determinant of production growth and instead rely on reduced-form applied models expressed in terms of the conditioning variables alone.

With this literature in mind, a model is developed in the following section in which production techniques are endogenous, although potentially constrained by available technology. The applied model includes many variables related to endowments, institutions, communities, and households that are expected to describe the decision environment that determines which of the available technologies is applied. Because some of these determinants are related to geographic features that cannot be affected by policy, the relative importance of these features in explaining regional income disparities is examined.

To some extent, the variables used to describe the decision environment are related to the types of variables viewed as the fundamental or deep determinants in the growth literature. This makes possible an exploration of whether the same basic notions about what determines income differences among countries hold in explaining regional differences in agricultural incomes. Because cross-sectional data are used, the study does not look at the dynamic relationship between factor accumulation and broad determinants of productivity. However, it does focus on the relative roles of factor use and technology choice to draw inferences about how quickly changes in these conditioning variables affect incomes.

This has relevance for policy because policies tend to work through the aspects of the conditioning environment. To the extent that current output is determined primarily by factors of production, the role of policy will be limited to its effect on the rates of accumulation. In this case, the benefits of improved policy will accrue slowly, and previous policies, embodied in current stocks of

accumulated factors, will determine outcomes largely in the short run. Alternatively, if policy directly affects the choice of technique, changes in policy can affect growth through productivity increases that are immediate and additional to long-run effects.

II. THEORY AND THE APPLIED MODEL

The starting point for the applied model is Mundlak's (1988), 1993 model of endogenous implemented technology. The model accommodates heterogeneous production technologies, based on the assumption that the choices that producers make regarding which technology to apply, and therefore which inputs to use, are conditioned by earlier decisions, manifested in quasi-fixed factors, and by the decision environment in which the producers operate. Because the aspects of the decision environment vary among producers, a set of microeconomic production solutions results, each potentially characterized by a different technology.

Using the vectors y_n^* to denote production, s to denote state variables that characterize the decision environment, and $x^* = x(s)$ to denote inputs, the aggregation of output can be written as a function of s alone, where $\sum y_n^* \equiv F(x^*, s) \equiv \Phi(s)$. In general, the production function is not identified. It is however possible to find an approximating aggregate function, $F(x, s)$, based on the assumption that observed differences in input allocations are associated with different implemented technologies conditioned on s . Operationally, the result is an approximating empirical representative model of production, where elasticities are functions of the state variables and possibly of the inputs. This is written as $\ln y = \Gamma(s) + B(x, s)$, where y is output, $B(x, s)$ represents a production technology that depends jointly on production factors and state variables, and $\Gamma(s)$ represents a vector of additional productivity effects that depend on state variables alone. In this framework, productivity and the marginal contributions of inputs to output are endogenous and arise in response to the changing decision environment.

The applied model provides functional form to the conceptual model. Because the production technology function $B(x, s)$ potentially depends on the input and state variables, it is modeled as a flexible combination of the factors and exogenous state variables:

$$(1) \quad B(x, s) \equiv \sum_i b_i \ln x_i + \frac{1}{2} \sum_{ij} b_{ij} \ln x_i \ln x_j + \frac{1}{2} \sum_{i,n} a_{in} \ln x_i \ln s_n.$$

The additional systemic state-specific productivity effects are modeled as a linear combination of the state variables, $\Gamma(s) \equiv \sum_n a_n \ln s_n$.⁴ The model is

4. Although, in principle, the state variables are measured as log transformations, most state variables used to estimate the model are either discrete or expressed as a proportion.

completed by adding a farm-specific idiosyncratic productivity term, e . The applied model is therefore given by:

$$(2) \ln y = \sum_n a_n \ln s_n + \sum_i b_i \ln x_i + \frac{1}{2} \sum_{ij} b_{ij} \ln x_i \ln x_j + \frac{1}{2} \sum_{i,n} a_{in} \ln x_i \ln s_n + e.$$

In anticipation of the discussion of the data used to estimate the applied model, additional comments about the model are in order. Because cross-sectional data are used to estimate the model, farm-specific subscripts are implicit in equation (2). Some state variables relate to location or to ecological measures that are repeated over subsets of households, and thus it is possible to denote these using location-specific subscripts instead. Later this relationship is used to distinguish farm and household effects that are independent of effects related to geography.

The idiosyncratic term in equation (2) is given a specific form that is motivated by a potential constraint imposed by the set of available technologies from which the endogenous applied technologies are chosen. To see this, consider the stochastic productivity measure:

$$(3) \quad P(x, s) \equiv \ln y - B(x, s) \equiv \Gamma(s) + e.$$

Implicit in the endogenous applied technology framework is the notion that some states will result, through technology choice, in higher levels of total factor productivity than others. This is reflected in the deterministic component, $\Gamma(s)$. However, productivity may be additionally affected by an idiosyncratic component related to unobservable characteristics of the farm or farmer. Without the loss of generality, it is possible to rank the deterministic component of productivity to say something about the idiosyncratic term.

The output outcome associated with the highest level of productivity can be labeled $P_0^* = \Gamma(s_0)$, and a conditioned measure of inefficiency with deterministic and idiosyncratic components $u_n^* = P_0^* - P_n^*$ can be calculated for each observation. If the conditional technology that produces P_0^* is binding in the sense that no greater output is feasible, the expected value of the inefficiency term will be non-negative. For a given set of state variables, producers might be expected to make the best of their available resources, so that the productivity outcomes cluster near the limiting technological frontier. However, relatively large inefficiencies are possible, in which case the distribution of the inefficiency term may be skewed as well as truncated.

The notion that stochastic productivity is constrained has motivated a series of applied stochastic frontier models. Generally, applied frontier models treat stochastic departures from the frontier as inefficiencies in the application of a single technology. This differs conceptually from the model developed here, where applied technologies are endogenous. Nevertheless, in a way that is similar to statistical frontier models, productivity in the applied model is stochastic and potentially constrained in a manner that would result in

idiosyncratic terms that have a skewed distribution. Consequently, the estimation techniques associated with frontier models can be used to potentially improve the estimation of the proposed model.

Specifically, the stochastic component of the model can be more fully specified as $e \equiv v - u$, so that the error is composed of a symmetric normally distributed error term, v , and a non-negative random term, u . By convention, the composite error is expressed as a difference between the two components. Consequently, all other things equal, lower values of u are associated with higher levels of output.

To estimate the composite-error model, it is necessary to assign a specific underlying distributional form to the unobserved distance term, u , to separate it from the also unobserved random component, v . For reasons that are developed later, a normal-truncated-normal composite error is specified, where $v \sim \text{iid } N(0, \sigma_v^2)$ and where $u \sim \text{iid } N^+[\mu(z), \sigma_u^2]$, where $\mu = \sum_k \delta_k z_k$. Using this specification, the distribution of the idiosyncratic productivity term can be modeled as conditional on additional variables. This feature is used later to include related endogenous variables that are expected to influence productivity.

Less complex error structures are nested within the normal-truncated-normal distribution given above. First, when $\mu = 0$, the two-parameter truncated-normal distribution collapses to the single-parameter half-normal, that is, $u \sim \text{iid } N^+[0, \sigma_u^2]$ (Stevenson 1980). Additionally, when $E(u) = 0$, the composite error can be represented by a single symmetric distribution, and simpler estimation techniques can be used. A variety of tests have been devised to test for the composite-error structure, as discussed after the following review of data used to estimate the model.

III. DATA

Farm and household data used in the analysis are taken from the 2000 Third Agricultural Census of Ecuador. The data are generated by a complete census of large-scale farms and a large representative sample of smaller farms. The complete survey contains observations on more than 128,000 farms and is representative at the canton level. This study used data on the nearly 108,000 farms in Ecuador that produce field crops. The census contains information about physical output, land use, labor, and production methods, as well as key information related to marketing. Information about farming households is collected as well. Output prices are not part of the survey, although detailed spatial data on farm products are available from ongoing producer price surveys by the National Institute of Statistics and Census. How these data were matched with the physical output data is explained in a supplemental appendix (available at <http://wber.oxfordjournals.org/>). The census data were also supplemented with environmental and climate data from the Sistema de Monitoreo Socioambiental Ecuatoriano (Ecociencia 2002). These data were matched with the census data

at the canton level. After the data sets were matched, observations with inconsistencies were dropped. Large-scale plantations were also excluded, leaving a sample of 107,269 farms. The variables used to estimate the model are described briefly below.

Production is measured as valued crop output and is calculated for each farm by matching spatial price data with production quantities from the census. This measure does not include livestock production, although livestock production enters into the analysis as a conditioning variable. The factors of production—variables related to land, labor, and capital—are taken directly from the census and measured as quantities.

The census distinguishes between irrigated and rainfed croplands, and the share of cropland irrigated is included as an explanatory variable. The census data also reveal whether additional inputs, including fertilizers, pesticides, and improved seeds, were used in combination with irrigation or applied to rainfed land. The census does not indicate quantities of inputs applied but reports the surface area receiving inputs. In practice there is less variation in the data than the questionnaire might suggest, as the share of irrigated or rainfed land receiving additional inputs is frequently either zero or one, especially for small- and medium-scale farms. Moreover, when additional inputs are used, they are generally used in combination. Consequently, the share of rainfed land and the share of irrigated land receiving fertilizer are included as separate explanatory variables and taken as an indicator of additional input bundles. Specified in this way, the marginal effects of shifting into irrigation land that is already in production and the separate marginal effects of applying additional inputs to rainfed or irrigated lands can be identified.

Labor is given as the number of workers. The data distinguish between family members who work on the farm and hired workers. The census also notes the number of seasonal workers for farms that employ them. The census provides data on farm machinery used on each farm but does not provide sufficient information to calculate a standard representation of on-farm capital. Therefore, the number of vehicles—tractors, trucks, and related machinery (thrashers, plows)—used on each farm is a proxy for on-farm capital.

The state variables fall into three broad categories: farmer characteristics and social capital, markets and institutions, and nature and risk. The farmer characteristics and social networks include three household measures. One is a measure of the farmer's formal education, the second is of the level of agricultural education, both measured in years, and the third is a discrete variable indicating the gender of the primary farmer. Two variables capture social capital. One is a discrete variable that indicates whether the farmer has received assistance from a *gremio*, a type of voluntary producer association common in Ecuador. The other variable indicates whether an indigenous language other than Spanish is spoken at home and is meant to capture the effects of belonging to an indigenous ethnic group.

The second set of state variables captures the influences of markets and government services. Included are private markets for credit and technical assistance, intermediate buying arrangements, and participation in output markets. An additional variable signals whether the farm is isolated from markets and is set to one when the nearest market is 90 minutes or more away. Consequently, the variable captures both distance to market and the quality of the transportation system. Three variables capture differences in government services. Two are discrete variables indicating whether the farm has received technical assistance from the government and whether the government has provided credit. A third, a continuous variable, gives the share of farmland that is titled.

The third set of state variables captures nature and risk. The indicators of nature include climate and topology measures. The climate measure is the ratio of average precipitation to moisture lost from the soil due to evaporation and transpiration at the canton level. The measure is used to classify canton climates as arid/dry, moist, humid, or wet, following the classifications used by the United Nations Convention to Combat Desertification.⁵ The topology measure is related to steeply sloped land in Ecuador and is reported as the percentage of canton area at risk of eroding. Two variables are related to production and income risk. One is the historic coefficient of variation in rainfall. The other is the share of farmland devoted to uses other than crops, a measure of diversification, a common risk mitigation strategy.

The census data reveal a wide range of scale for agriculture in Ecuador. Production technology choices likely vary in ways related to scale, as indicated by the descriptive statistics reported in table 1. The table contains sample averages and median values for three quantiles, which are based on farm area under field crops. First-quantile farms are very small; crops typically cover about two-thirds of a hectare and range up 1.5 hectares. Middle-quantile farms are typically 3 hectares in size, and the largest farms in the group are under 5 hectares. The third quantile contains a wide range of farms, including large-scale commercial farms of nearly 400 hectares.

Revenue per hectare from crops in Ecuador averaged \$676 for the census year and increased with scale, from \$472 for small-scale farms to \$1,031 for the largest farms. The number of workers per farm increased with scale as well, but not proportionally. Small farms had more than five full-time workers for every hectare of land, whereas large farms had more than 3 hectares of land for every full-time worker. The decline in labor with scale was matched by an increase in capital.⁶ Differences among the remaining variables in table 1 are small. The average share of cropland irrigated is slightly higher on small farms than on medium- and large-scale farms. The rates of fertilizer application on irrigated

5. Details of the climate classifications are available at www.unccd.entico.com/english/glossary.htm.

6. The movement of labor out of agriculture and capital into agriculture is a pervasive pattern associated with economic growth; see Mundlak, Larson, and Crego (1998).

TABLE 1. Descriptive Statistics, by Farm Size

| | Small Scale | Medium Scale | Large Scale | All Farms |
|---|-------------|--------------|-------------|-----------|
| Number of farms | 36,428 | 35,085 | 35,756 | 107,269 |
| Median size of cropland | 0.64 | 2.70 | 10.00 | 2.60 |
| Farm averages | | | | |
| Crop revenue per hectare (US\$) | 472 | 525 | 1,031 | 676 |
| Cropland (hectares) | 0.66 | 2.78 | 21.40 | 8.26 |
| Family workers | 3.34 | 3.50 | 3.19 | 3.34 |
| Hired workers | 0.11 | 0.29 | 2.79 | 1.06 |
| Capital index | 35 | 51 | 215 | 100 |
| Average shares | | | | |
| Share of full-time workers who are family members | 0.96 | 0.94 | 0.79 | 0.90 |
| Share of hired workers who are seasonal | 0.10 | 0.21 | 0.67 | 0.33 |
| Cropland irrigated | 0.25 | 0.20 | 0.21 | 0.22 |
| Irrigated cropland with fertilizer | 0.17 | 0.15 | 0.20 | 0.17 |
| Rainfed cropland with fertilizer | 0.27 | 0.24 | 0.27 | 0.26 |
| Landholdings titled | 0.73 | 0.70 | 0.77 | 0.73 |

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

lands are relatively low, but similar across scale; fertilizer use rates were higher on rainfed land.

IV. ESTIMATION RESULTS

The model was estimated using the data described above. In some cases, input variables such as hired labor take on zero values. To handle this in a log-based functional form, we used a set of corresponding dummy variables, based on an approach suggested by Battese (1997). To address possible differences in the distribution of the composite error due to scale, we also added dummy variables to the set of variables determining u . With these modifications, the final model contains 118 parameters. Of these, 68 percent are individually statistically significant. To keep the discussion of the estimation results manageable, we calculated mean-value elasticities from the estimated parameters, and these are discussed below. The full set of estimated parameter results are given in tables S.1 and S.2 of the supplemental appendix.

Before the derived elasticities are discussed, results related to the composite-error specifications should be mentioned. The composed form of the stochastic term may arise when the distribution of the farm-specific idiosyncratic components of productivity is skewed and truncated by a binding technology. The results from two tests are reported, and both are consistent with this characterization.

TABLE 2. Tests of the Composite-Error Structure

| Test | Score |
|------------------------------------|---------------|
| Parameter estimate, $\phi(\gamma)$ | 1.76 (0.019)* |
| Coelli's test statistic | -141.97* |

Note: Numbers in parentheses are standard errors.

*Results are significant at the 1 percent level.

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

The first test is based on a statistic constructed from the variances of the two composite-error terms, $\gamma = \sigma_u^2 / (\sigma_v^2 + \sigma_u^2)$. When $y = 0$, the composite-error model is indistinguishable from a model with a symmetrically distributed error term, and thus a test of the statistical significance of y constitutes a test of the composite-error specification. The test is easy to perform, because a related parameter, $\phi(\gamma)$, appears as a parameter in the model's likelihood function. The estimated value of the parameter is statistically different from zero, lending support for the assumed composite form (table 2).

Even so, because y is tied to estimates of the composite-term variances, the test is conditional and potentially sensitive to how the composite error is specified. For this reason, a second test was applied, based on an approach developed by Schmidt and Lin (1984) and modified by Coelli (1995). The related test statistic is based on least squares residuals and consequently is independent of prior assumptions about the form of the composite error. The calculated statistic is also significant (table 2). Moreover, the test-statistic value is negative, which is consistent with a clustering of production technologies around a binding technological frontier (Coelli 1995, p. 253).

Elasticity Estimates

Elasticities and standard errors, calculated from estimated model coefficients and sample averages, are reported in table 3 for the four factors of production, for qualitative differences in the factors, and for state variables.

The factor elasticities relate to the slope function, $B(x,s)$. Because the elasticities are functions of both state and input variables, each farm is potentially associated with a different set of factor elasticities. In this sense, the elasticities reported in table 3 represent average effects across technologies.

The mean-valued factor elasticities in table 3 are all significantly different from zero. They sum to 1.159, suggesting increasing returns to scale for a typical farm in Ecuador.⁷ The elasticity for family labor is positive, but quantitatively

7. It is also the case that the underlying parameters used to calculate the elasticities are collectively different from zero at standard confidence levels and that average returns to scale are different from 1. The elasticities for family and hired labor are statistically different from one another as well. See supplemental appendix table S.4.

TABLE 3. Elasticities Calculated at Mean Values

| | Elasticity | Standard Error |
|---|------------|----------------|
| Production factors | | |
| Family labor | 0.120 | 0.018** |
| Hired labor | 0.264 | 0.025** |
| Cropland | 0.656 | 0.008** |
| Capital | 0.119 | 0.048* |
| Returns to scale | 1.159 | 0.056** |
| Factor characteristics | | |
| Irrigation | 0.209 | 0.007** |
| Additional inputs on irrigated land | 0.020 | 0.004** |
| Additional inputs on rainfed land | 0.131 | 0.006** |
| Seasonal labor | 0.069 | 0.003** |
| State variables | | |
| Formal education | 0.014 | 0.002** |
| Agricultural education | 0.017 | 0.004** |
| Female head of house ^a | -0.125 | 0.015** |
| Indigenous ethnicity ^a | 0.033 | 0.022 |
| Participates in output markets ^a | 0.562 | 0.044** |
| Sells to intermediate buyer ^a | 0.019 | 0.020 |
| Isolated from markets ^a | 0.013 | 0.020 |
| Assistance from the gremio ^a | 0.055 | 0.026** |
| Precipitation variation | -0.047 | 0.010** |
| Land diversification | -0.039 | 0.009** |
| Moist climate ^a | 0.161 | 0.028** |
| Humid climate ^a | -0.055 | 0.024* |
| Wet climate ^a | 0.157 | 0.030** |
| Risk of erosion ^a | -0.027 | 0.004** |

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

^aDiscrete variable.

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

less than half that of hired labor.⁸ The land elasticity is about as large as the combined elasticities of the other production factors. The elasticity on capital is lower than might be expected, although the estimated value is consistent with cross-country studies that have relied on proxy measures of capital (Mundlak, Larson, and Butzer 1999).

Estimates related to qualitative differences in the factors suggest that bringing an additional 1 percent of existing cropland under irrigation increases output by 0.2 percent. Applying chemical inputs to irrigated land brings about a slight increase in output. The elasticity for chemical inputs on rainfed land is significantly higher and in line with fertilizer elasticities estimated from cross-country

8. As an anonymous referee pointed out, the marginal value product for family labor, given by the average product times the elasticity, is less than 15 percent that of hired labor at sample averages.

data. Supplementing full-time labor with seasonal workers has a small but statistically significant effect on output.

The elasticities reported in the bottom block of table 3 relate to the systemic state-specific productivity effects, $\Gamma(s)$. Because the state variables also affect the factor elasticities related to $B(x,s)$, the reported elasticities are partial and capture only the direct effect of the state variables on output and productivity. The role that the state variables play in determining factor elasticities is discussed later.

Although many of the state-variable elasticities are statistically significant, most are quantitatively small. Both general education and education related to agriculture have a small but positive effect on productivity. Productivity among households headed by a woman is lower, perhaps because this group includes a disproportionate number of single-parent households. Productivity does not differ significantly by ethnic group. The dummy variable that distinguishes between farmers who produce for market and those who do not is significant and quantitatively large, but because more than 90 percent of farmers in Ecuador already participate in output markets, the variable is less interesting for policy. Productivity is higher for farmers who participate in a marketing cooperative, but other variables related to marketing are not statistically significant or quantitatively large. Productivity is less for farms subject to greater rain variability and also for farms that diversify out of crops, a result consistent with the notion that some costs associated with risk take the form of forgone opportunities. The effects of climate and erosion risk are significant and similar in size.

As discussed, measured productivity includes both the systemic component, $\Gamma(s)$, and a stochastic term, ϵ , which includes the non-negative idiosyncratic term, u . To a degree, the idiosyncratic component of productivity relates to farm or farmer characteristics that cannot be observed. However, it is likely that some variables that ultimately affect technology choice and productivity are determined by a process that links these observed variables to the unobserved idiosyncratic characteristics of the farm or farmer. The most discussed example is access to credit, with the likelihood of receiving credit expected to be related to a borrower's unobserved entrepreneurial and cognitive abilities (McKernan 2002; Khandker and Faruqee 2003). However, related arguments have been made about the provision of government technical assistance (Godtland and others 2004) and land titling (Deininger and Chamorro 2004). For these reasons, some state-like variables are endogenous and therefore stochastic and are related to the idiosyncratic component of productivity. Consequently, a set of cotermined variables are included in the specification of e . These correspond to z in the model presented in Section III.⁹

9. This approach, believed to be novel, retains information about the stochastic component of the related endogenous variable. This information is stripped away in traditional approaches designed to construct nonstochastic proxies that can be treated as deterministic explanatory variables.

TABLE 4. Z-Variable Coefficients and Calculated Elasticities

| | Elasticity | Coefficient | Standard Error |
|--|------------|-------------|----------------------|
| Technical assistance, private ^a | 0.045 | -0.501 | 0.069 ^{***} |
| Credit, private ^a | 0.075 | -0.848 | 0.080 ^{***} |
| Share of land titled | 0.034 | -0.363 | 0.034 ^{***} |
| Technical assistance, public ^a | 0.007 | -0.077 | 0.086 |
| Credit, public ^a | 0.025 | -0.270 | 0.105 [*] |
| Medium scale ^a | 0.048 | -0.525 | 0.038 ^{***} |
| Large scale ^a | 0.106 | -1.213 | 0.055 ^{***} |

^aSignificant at the 5 percent level; **significant at the 1 percent level.

Note: Elasticities are calculated as discrete changes.

^aDiscrete variable.

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

The five endogenous variables included as determinants relate to private market access for credit and technical assistance and access to public programs for credit, technical assistance, and land titling services. Additionally, two dummy variables are included, which correspond to the second (medium) and third (large) cropland quantiles, to account for possible differences in the distribution of the stochastic productivity measure related to scale.¹⁰

Related coefficients from the estimated model are summarized in table 4. As discussed, frontier production models conventionally represent the non-negative component of the composite error as a departure from the efficiency frontier. Consequently, the negative coefficients in table 4 indicate that these variables are associated with a reduction in the mode of the inefficiency distribution and, all other things equal, an increase in output. As table 4 summarizes, all coefficients are negative and all but the coefficient on public technical assistance are statistically significant.

Although indicative, the coefficients themselves are a very rough measure of how shifts in the distributional mode of u affect output. This is because the average effect of the stochastic productivity component is determined not only by the mode of the distribution but also by the overall shape of the distribution, especially the point of truncation. For this reason, discrete changes in the predicted value of u are used to calculate output elasticities, rather than relying on an evaluation of the elasticities at a single point along the distribution.

Because for each farm observation f the estimated model provides predictions of u_f conditional on v_f , the average effect of a discrete change in the z variables can be calculated. This is done by using the estimated model to simulate the

10. Larger farms and larger endowments of managerial skill are potentially related. However, heteroskedasticity related to variance of the strictly positive efficiency term will also affect its mean, so there are mechanical reasons for including the terms as well.

effect on each u_f of switching in sequence each z_k from zero to one, holding other values constant, so that $\Delta \ln y / \Delta z_k = -\Delta u|_{dz_k=0,1}$. Average effects, $\Delta \bar{u}_k$, can be calculated over all observations or any subgroup of farms. Given the specification in equation (2), the calculated difference is similar to the coefficient on a dummy variable in a semilogarithmic equation; consequently, mean elasticities are given by $\varepsilon_k = \exp(\Delta \bar{u}_k) - 1$ (Halvorsen and Palquist 1980).

Elasticity estimates based on the procedure outline above are also reported in table 4. Among these, elasticities associated with private market access to credit and technical assistance are large relative to public program elasticities and to some state-variable elasticities. This is noteworthy, because these effects are additional to past accumulations of physical or educational capital. Among the public programs, land titling has the largest effect on output. Although the elasticity on access to credit through public programs is smaller than the elasticity associated with private credit markets, the results suggest that credit programs have a measurable impact. In contrast, the measured effect of public technical assistance programs is quantitatively small and statistically insignificant.

Because the implicit small-farm dummy variable is suppressed in the estimated model, the elasticities associated with the remaining quantile dummy variables can be interpreted as differences from the efficiencies found on small-scale farms. The results indicate that as the amount of land brought under crops increases, inefficiencies decline in a way that is separate from the returns to scale or the related effects of increasing factor use. That is to say, for a given set of factors and conditioning state variables, larger farms tend to cluster more closely to a frontier that is presumably limited by technology. This has policy relevance because it suggests that innovations in technology will especially benefit larger farms that are currently constrained by available technology. The results also suggest a potential to improve agricultural incomes by identifying constraints that lead smaller farms to choose less efficient technologies.¹¹

Choice and Factor Elasticities

In the conceptual model, technological choice is expected to affect both total factor productivity and the marginal productivity of factors. In the applied model, state variables are used to measure these influences. As discussed, one consequence is that measured values of total factor productivity take on geographic patterns because some determining state variables are location specific. To a degree, this is also true for measured factor elasticities. Nevertheless, factor elasticities depend on combined factor levels in addition to state variables, so that geography may play a relatively smaller role.

As a practical matter, it is possible to use parameters from the applied model to quantify the relative contributions of factors and state variables in

11. Key results are robust to alternative specifications. See tables S.4–S.6 in the supplemental appendix.

determining factor elasticities. To illustrate, consider the effect on output of a change in the first input:

$$(4) \quad \frac{\partial \ln y}{\partial \ln x_1} \equiv \frac{\partial B(x, s)}{\partial \ln x_1} = b_1 + \frac{1}{2} \sum_j b_{1j} \ln x_j + \frac{1}{2} \sum_n a_{1n} \ln s_n.$$

Of the three terms on the right side of equation (4), the first two are the portion of the factor elasticity that is due to input levels, whereas the last term captures changes due to state variables.

Using the approach illustrated in equation (4), factor elasticities were calculated based on variable averages from each of the three farm scale quantiles. The elasticities were further decomposed into factor and state-variable effects. Overall, the results suggest that factor elasticities are determined largely by input levels (table 5). This can be seen in the returns to scale elasticity, which summarizes the sometimes offsetting changes in the underlying composition of elasticities. The returns to scale elasticity falls slightly as scale increases, and both the absolute values of the elasticities and the direction of change are driven by factor effects. Moreover, for medium- and large-scale farms, the state variables largely account for differences from constant returns to scale.

For small farms, measured differences between the elasticities of family and hired labor are large, but the gap closes as more land is brought under production, operating mostly through factor effects. However, for hired labor, the complementary effects of state variables on the elasticity of labor increases

TABLE 5. Decomposition of Production Elasticities, by Farm Size

| | Small Scale | Medium Scale | Large Scale |
|------------------------|-------------|--------------|-------------|
| Elasticity | | | |
| Family labor | 0.099 | 0.114 | 0.151 |
| Due to factors | 0.043 | 0.062 | 0.106 |
| Due to state variables | 0.056 | 0.052 | 0.045 |
| Hired labor | 0.343 | 0.239 | 0.192 |
| Due to factors | 0.273 | 0.127 | 0.022 |
| Due to state variables | 0.070 | 0.112 | 0.169 |
| Cropland | 0.693 | 0.680 | 0.642 |
| Due to factors | 0.724 | 0.717 | 0.685 |
| Due to state variables | -0.031 | -0.037 | -0.043 |
| Capital | 0.102 | 0.153 | 0.122 |
| Due to factors | 0.066 | 0.118 | 0.119 |
| Due to state variables | 0.036 | 0.035 | 0.003 |
| Returns to scale | 1.238 | 1.186 | 1.107 |
| Due to factors | 1.106 | 1.024 | 0.932 |
| Due to state variables | 0.132 | 0.161 | 0.175 |
| Technical efficiency | 0.398 | 0.425 | 0.476 |

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

with scale, preventing the gap from closing further.¹² Differences in cropland elasticities are not large among the three quantiles and are driven by a slight decline due to factor effects. The elasticity of capital increases rapidly between the first and second quantiles, whereas the state-variable effect falls between the second and third quantiles.

As discussed, the estimated parameters suggest that the distribution of u shifts with scale. There are additional determinants of the distribution relating to markets and government programs, and related average participation rates vary among the three scale quantiles as well. The combined effect is reflected in the summary measure of technical efficiency, T_q , which is a quantile average of the farm-specific measures of technical efficiency, $E[\exp(-u_f|v_f)]$ (table 5). Generally, access to markets and government programs increases with scale, although the resulting elasticity differences are small. Consequently, the distributional determinants combine to produce rates of technical efficiency that increase with scale.¹³

V. DETERMINANTS OF REGIONAL DIFFERENCES

This section turns to a key motivation of the article, explaining observed regional differences in output and productivity. The analysis relies on the estimated model, which is used to map observed differences in factor use and the conditioning state variables to observed differences in output. This allows a decomposition of differences in revenue among typical farms of each region into factor and productivity effects. In addition, spatial differences in factor productivity are further decomposed into effects that correspond to the following classes of the state variable: household characteristics and social capital, nature and risk, and markets and institutions. From a policy perspective, this approach complements the elasticity discussion, which focused on which factors are most important in determining agricultural output and productivity. This analysis explores how differences in the spatial array of factors and state variables combine to generate observed spatial differences in output and productivity.

Broadly, the simulation strategy is to construct regional and national representations of farm output and to use those representations to measure the relative importance of factor use and productivity in explaining average revenue differences among regions. Specifically, for each classification of farm scale (small, medium, and large), national averages are calculated for the vector of production factors (\bar{x}_q), state variables (\bar{s}_q), z -variables (\bar{z}_q), and also provincial averages (\bar{x}_q^R , \bar{s}_q^R , and \bar{z}_q^R). To take into account the indirect role of state variables

12. Calculated marginal value products follow the same pattern, but in a more dramatic fashion. Estimated marginal value product for family labor is roughly 1 percent that of hired labor for small farms and close to 70 percent for large farms.

13. Details about participation rates are summarized in table S.6 in the supplemental appendix.

in the factor marginal products, we calculated elasticities are calculated based on the regional averages and coefficients from the model.

Let $\varepsilon_{iq}^R(\beta)$ represent input factor elasticities evaluated at $\beta(\bar{x}_q^R, \bar{s}_q^R)$, and let $\varepsilon_{nq}^R(\Gamma)$ represent the systemic productivity elasticities evaluated at $\Gamma(\bar{s}_q^R)$. Following the approach noted earlier, regional idiosyncratic productivity elasticities, $\varepsilon_{kq}^R(u)$, are calculated based on regional quantile averages of $u(z)$. Consequently, the calculated productivity measure includes systemic differences due to differences in the level of state variables plus regional differences in the distribution of idiosyncratic efficiencies.

Expected differences in regional (R) output for each farm classification (q) can then be expressed as

$$(5) \quad \Delta y_q^R \equiv \Delta F_q^R + \Delta P_q^R,$$

where the change in factor use is given by

$$(6) \quad \Delta F_q^R \equiv \sum_i \varepsilon_{iq}^R(\beta) (\ln \bar{x}_{iq}^R - \ln \bar{x}_{iq})$$

and the change in total factor productivity is given by¹⁴

$$(7) \quad \Delta P_q^R \equiv \sum_n \varepsilon_{nq}^R(\Gamma) (\ln \bar{s}_{nq}^R - \ln \bar{s}_{nq}) + \sum_k \varepsilon_q^R(u) (\bar{z}_{kq}^R - \bar{z}_{kq}).$$

For each class of farm, the exercise explains how differences in the distribution of factors and state variables result in differences in agricultural output. The simulation results are detailed and cover each of the 21 provincial areas of Ecuador. To conserve space, we reported only summary results here.¹⁵

The results illustrate the differences in regional output and productivity in Ecuador (table 6). Across all sizes of farms, regional differences average about 17 percent in absolute terms. In a mechanical way, output differences can be attributed primarily to factors generally and to land specifically. However, this is partly because the farm-size classes are based on area under crops. The first and third quantiles contain the tails of the underlying distribution of land among farms and therefore contain greater variation. After adjusting for land, output differences related to more intensive use of capital and labor inputs and those related to productivity are similar on average.

Still, the averages mask differences related to scale. With increases in scale, the role of productivity in explaining regional differences diminishes, whereas effects tied to factor intensification become more pronounced. This can be seen

14. The z variables and some state variables are expressed as shares (for example, share of farm land titled) and are therefore not converted into logs.

15. Detailed results are summarized in tables S.7-S.12 in the supplemental appendix.

TABLE 6. Average Absolute Differences in Regional Output and Related Determinants, by Farm Size

| | Small Scale | Medium Scale | Large Scale | All Farms |
|----------------------|-------------|--------------|-------------|-----------|
| Output | 0.167 | 0.090 | 0.256 | 0.171 |
| Factors | 0.124 | 0.059 | 0.243 | 0.142 |
| Land | 0.117 | 0.035 | 0.180 | 0.111 |
| Capital and labor | 0.051 | 0.046 | 0.091 | 0.063 |
| Productivity | 0.073 | 0.053 | 0.034 | 0.053 |
| Public programs | 0.004 | 0.004 | 0.004 | 0.004 |
| Households | 0.021 | 0.017 | 0.009 | 0.016 |
| Markets and networks | 0.045 | 0.025 | 0.009 | 0.027 |
| Risk | 0.008 | 0.008 | 0.006 | 0.008 |
| Nature | 0.045 | 0.040 | 0.034 | 0.040 |

Note: Differences are expressed as shares of national average output value by scale.

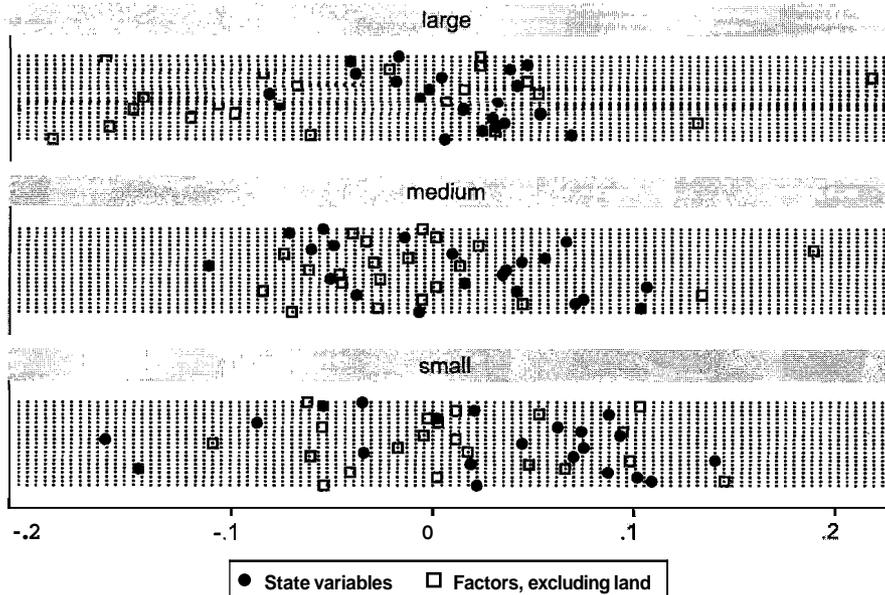
Source: Authors' analysis based on data described in text.

graphically in figure 1, where the circles mark simulated differences due to capital and labor and the squares mark productivity effects for each of the 21 geographic areas in the data. For small-scale farms, the factor and productivity effects are commingled. In moving along the figure from the small- to medium- to large-scale farms, it can be seen that the state-variable effects begin to shift closer to zero, whereas the spread in capital-plus-labor effects increases.

The productivity effects can be further decomposed by the underlying determining state variables (table 6). The results suggest that the three public programs included in the model explain little of the observed regional differences in output, regardless of scale, even though the elasticities on the credit and titling programs were of moderate size. For credit, this is because program penetration is limited—only 2 percent of farms receive credit through this program. For titling, there is little regional variation, because most farmland in most provinces is titled. The effect of differences in household characteristics accounts for about 2 percent of regional output for small farms and less than 1 percent for large farms. Differences in the use of markets and networks accounts for about 4.5 percent of regional output among small farms, 2.5 percent among medium-scale farms, and less than 1 percent among large farms. The effects of risk, as given by the variation in rainfall and the effects of diversifying land use, have a limited role in explaining regional differences. In contrast, nature, as measured by differences in climate, soil conditions, and slope characteristics, is an important determinant of output differences.

To summarize, the simulation results suggest that differences in applied technologies, conditioned by the state variables and associated with productivity, diminish as farms become larger in scale and more intensive in other inputs. Of the differences attributable to productivity, nature is an important

FIGURE 1. Simulated Differences in Output Attributable to Factor Intensification and Productivity, by Farm Scale



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

determinant on all farms. For small farms, household characteristics and access to markets play important roles as well.

VI. SUMMARY AND CONCLUSIONS

Because farmers face different circumstances with different resources, they choose different approaches to farming. This means that economic data about agricultural production span a variety of applied technologies. This article applies a flexible-form model to measure the interrelated effects on production of inputs and the state variables that condition this endogenous choice of technology.

Working through the decision environment, state variables are expected to influence total factor productivity directly and to influence the elasticities of production inputs. Statistically significant evidence of both effects is found. Even so, the results suggest that some of the ways factors and state variable interact to determine different levels of output and productivity vary with scale, and this has implications for policy.

To a large degree, output differences among large- and medium-size farms are explained by differences in factor use. Factors are important for small-scale farmers as well; however, significant differences in productivity outcomes remain among small-scale farms even after factor differences are taken into account. These remaining differences are explained largely by differences in the conditioning state variables. This is consistent with the notion that technology choice is more constrained among smallholder farms and that productivity is therefore more sensitive to differences in the decision environment.

As farm scale increases, differences in productivity are reduced, and this comes about by way of two complementary effects. First, productivity contains a component related to the unobservable idiosyncratic characteristics of farms and farm managers. In the applied model, this gives rise to a composite-error stochastic term from which a measure of technical efficiency can be derived. Evidence is found in favor of the composite-error structure, and further evidence suggests that the derived measure of technical efficiency increases with farm size alone.

Second, there is also evidence that market and program participation outcomes, potentially related to these idiosyncratic farm characteristics, influence the technical efficiency measure. Access to such programs is more common among larger farms, which also boosts measured efficiency. Consequently, the two effects work in reinforcing ways to provide higher estimates of technical efficiency on larger farms. This is taken as evidence that larger farms are better able to, and frequently do, choose farming approaches that are closer to a binding technological frontier.

Taken together, the findings suggest a differentiation between the types of policy that promote growth in agriculture generally and those that are more likely to assist the rural poor. Because most agricultural output is produced on larger farms that operate close to the technological frontier, programs that promote relevant new technologies can be expected to spur sectorwide growth. For regions that depend primarily on agriculture, growth may have additional spillover effects on incomes through markets for related goods and services. At the same time, smallholder households in Ecuador that depend on agriculture are disproportionately poor and tend to use a ranging set of technologies. Consequently, policies most likely to benefit the poor are those that change the constraints and incentives that lead some households to choose less efficient technologies over more productive alternatives.

Still, the results suggest that building effective strategies for reducing rural poverty is no easy task. Many of the state variables that explain productivity differences among smallholders are related to the aspects of geography that are not easily changed. This limits the range of available policy instruments and the scope for policy-led increases in productivity. Among the remaining policy avenues, simulation results point to the importance of investments in infrastructure and institutions that support markets, especially credit markets. These

markets can lead to the adoption of more productive technologies in the short run and can facilitate the buildup of productive assets in the long run.

Accumulated assets can be lost, and the results suggest that farmers forgo more productive technologies to take ex ante precautions against such loss. This, together with evidence of the central role that productive assets play in determining incomes, suggests the importance of policies that promote formal and informal insurance markets and provide for safety nets when these markets prove inadequate.

Studies of cross-country growth experiences find important roles for geography and for market-enhancing institutions. In a similar way, the results of this study suggest that climate and institution-dependent markets influence regional differences in agricultural productivity. The study results also indicate the importance of accumulated factors to short-run output. With time, the same conditioning factors that influence short-run productivity will likely influence stocks of these variables as well. However, this process may take generations to complete. Consequently, care should be exercised in drawing policy conclusions about the pace of growth from studies that methodologically set aside the influence of accumulated factors. This is particularly the case for many developing countries, where agriculture remains an important component of national income.

REFERENCES

- Barro, Robert, and Xavier Sala-I-Martin. 1992. "Convergence." *Journal of Political Economy* 100(2):223–51.
- Battese, George E. 1997. "A Note on the Estimation of Cobb-Douglas Production Functions When Some Explanatory Variables Have Zero Values." *Journal of Agricultural Economics* 48(2):250–52.
- Bhattacharyya, Sambit. 2004. "Deep Determinants of Economic Growth." *Applied Economics Letters* 11(9):587–90.
- Brock, William A., and Steven N. Durlauf. 2001. "What Have We Learned from a Decade of Empirical Research on Growth? Growth Empirics and Reality." *World Bank Economic Review* 15(2):229–72.
- Coelli, Timothy J. 1995. "Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis." *Journal of Productivity Analysis* 6(4):247–68.
- Deininger, Klaus, and Juan Sebastian Chamorro. 2004. "Investment and Equity Effects of Land Regularisation: The Case of Nicaragua." *Agricultural Economics* 30(2):101–16.
- Ecociencia 2002. *Sistema de Monitoreo Socioambiental Ecuatoriano*. Quito, Ecuador.
- Godtland, Erin, Elisabeth Sadoulet, Alain de Janvry, Rinku Murgai, and Oscar Ortiz. 2004. "The Impact of Farmer-Field Schools on Knowledge and Productivity: A Study of Potato Farmers in the Peruvian Andes." *Economic Development and Cultural Change* 53(1):63–92.
- Griliches, Zvi. 1996. "The Discovery of the Residual: A Historical Note." *Journal of Economic Literature* 34(3):1324–30.
- Hall, Robert E., and Charles Jones. 1997. "Levels of Economic Activity Across Countries." *American Economic Review* 87(2):173–77.
- Halvorsen, Robert, and Raymond Palquist. 1980. "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review* 70(3):474–5.

- Henderson, Vernon, Zmarak Shalizi, and Anthony Venables. 2001. "Geography and Development." *Journal of Economic Geography* 1(1):81–105.
- Khandker, Shahidur, and Rashid Faruquee. 2003. "The Impact of Farm Credit in Pakistan." *Agricultural Economics* 28(3):197–213.
- Klenow, Peter J., and Andrés Rodríguez-Clare. 1997. "Economic Growth: A Review Essay." *Journal of Monetary Economics* 40(3):597–617.
- Larson, Donald F., and Yair Mundlak. 1997. "On the Intersectoral Migration of Agricultural Labor." *Economic Development and Cultural Change* 45(2):295–319.
- Mankiw, N. Gregory, David Romer, and David Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107(2):407–37.
- McKernan, Signe-Mary. 2002. "The Impact of Microcredit Programs on Self-employment Profits: Do Noncredit Program Aspects Matter?" *Review of Economics and Statistics* 84(1):93–115.
- Mundlak, Yair. 1988. "Endogenous Technology and the Measurement of Productivity." In S. M. Capalbo and J. M. Antle, eds., *Agricultural Productivity: Measurement and Explanation*. Washington: Resources for the Future.
- . 1993. "On the Empirical Aspects of Economic Growth Theory." *American Economic Review* 83(2):415–20.
- . 2001. "Production and Supply." In B. Gardner and G. Rausser, eds., *Handbook of Agricultural Economics*. Volume 1A: *Agricultural Production*, *Handbooks in Economics*, Vol.18. Amsterdam, London and New York: Elsevier Science, North-Holland.
- Mundlak, Yair, Donald F. Larson, and Al Crego. 1998. "Agricultural Development: Issues, Evidence and Consequences." In Y. Mundlak, ed., *Contemporary Economic Issues: Proceedings of the Eleventh World Congress of the International Economic Association*, Volume 2, Labour, Food, and Poverty, Conference. New York: St. Martin's Press; London: Macmillan Press.
- Mundlak, Yair, Donald F. Larson, and Rita Butzer. 1999. "Rethinking Within and Between Regressions: The Case of Agricultural Production Functions." *Annales d'Economie et de Statistique* 55–56: 475–501.
- North, Douglas. 1994. "The Evolution of Efficient Markets." In John James and Mark Thomas, eds., *History in Capitalism in Context: Essays on Economic Development and Cultural Change in Honor of R. M Hartwell*. Chicago and London: University of Chicago Press.
- Parente, Stephen, and Edward Prescott. 1994. "Barriers to Technology Adoption and Development." *Journal of Political Economy* 102(2):298–321.
- Rodrik Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9(2):131–65.
- Sachs, Jeffrey, and Andrew Warner. 1997. "Fundamental Sources of Long-Run Growth." *American Economic Review* 87(2):184–85.
- Schmidt, Peter, and Tsai-Fen Lin. 1984. "Simple Tests of Alternative Specifications in Stochastic Frontier Models." *Journal of Econometrics* 24(March):349–61.
- Stevenson, Rodney. 1980. "Likelihood Functions for Generalized Stochastic Frontier Estimation." *Journal of Econometrics* 13(May):57–66.
- Stigler, George. 1976. "The Xistence of X-Efficiency." *American Economic Review* 66(1):213–16.

"a must-have online resource"

—*Library Journal netConnect*

OXFORD

Scholarship Online

Now Available for Purchase or Subscription

NEW FEATURES! Over 1,150 complete works • Automatic Export Citations • Printer Friendly Format Throughout Site

OXFORD'S VERY BEST IN KEY AREAS INCLUDING:

Philosophy

- Metaphysics and Epistemology
- Moral Philosophy and Ethics
- Ancient Philosophy

Political Science

- Political Economy
- International Relations
- US, UK, EU, and Russian Politics

Religion

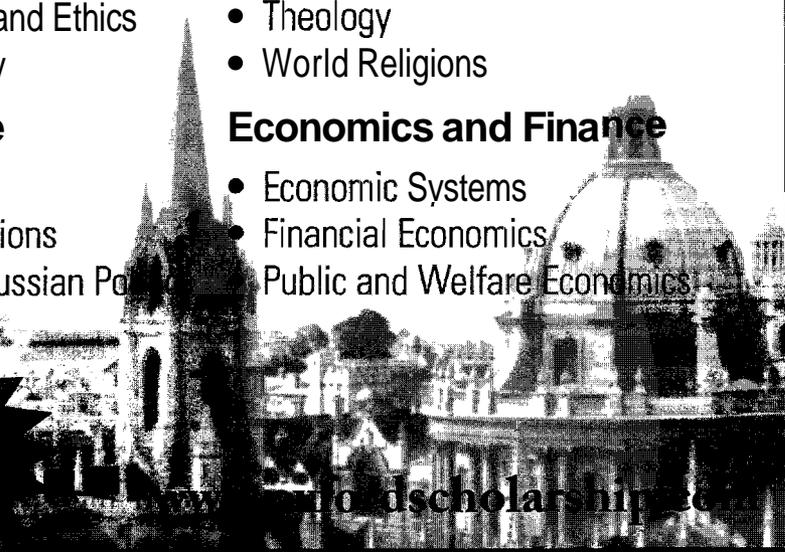
- Biblical and Early Christian Studies
- Theology
- World Religions

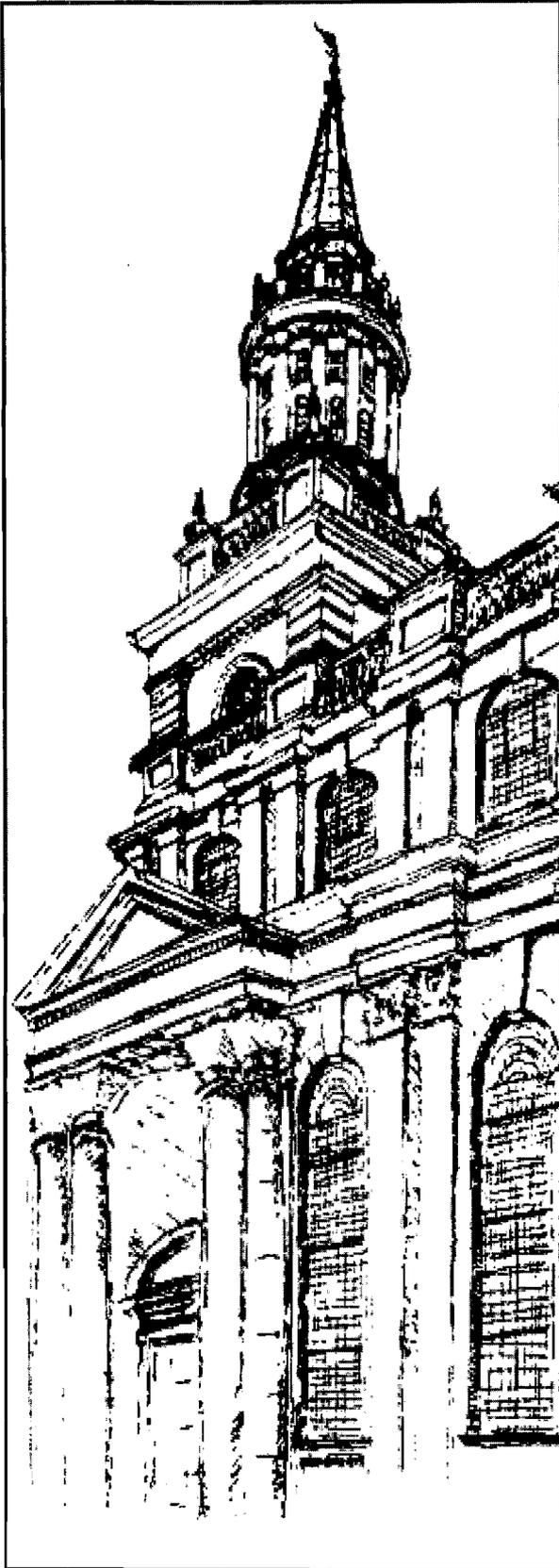
Economics and Finance

- Economic Systems
- Financial Economics
- Public and Welfare Economics

FREE TRIAL
available
online

OXFORD
UNIVERSITY PRESS





Oxford Journals Advertising

For information on
advertising in this journal
please contact:

Helen Pearson
Oxford Journals Advertising
PO Box 347
Abingdon, OX14 1GJ, UK

Tel: +44 (0) 1235 201904
Fax: +44 (0) 8704 296864

e-mail: helen@oxfordads.com

Information available online
Visit www.oxfordads.com

Stay alert to the very latest research

Design your own email alerts to track journal content that is important to you

CiteTrack will alert you by email whenever new content is published that matches criteria based on the topics, authors and articles you want to track

Topics

Tell CiteTrack which words or subjects to watch for in new content

Authors

Be alerted whenever key authors you are following publish a new paper

Articles

Know whenever a paper of interest to you is referenced by another paper or letter

CiteTrack checks for content matching your criteria in new issues of the journal you've asked it to monitor, and emails you each time new content is found.

You will receive an email every time matching content is found in new issues, one email for each new issue of each journal you are tracking.

With over 340 journals currently available through this service, you can easily monitor a wealth of quality research with one simple service.

Sign up for CiteTrack today – go to www.highwire.org, and click on 'My Email Alerts' for more information.



www.oupjournals.org

OXFORD
UNIVERSITY PRESS

Forthcoming papers in

**THE WORLD BANK
ECONOMIC REVIEW**

Volume 21, Number 1, 2007

- **Protecting the vulnerable: the trade-off between risk reduction and public insurance**
Shanta Devarajan and William Jack
- **Growth and Risk: Methodology and Micro Evidence**
Chris Elbers, Jan Willem Gunning and Bill Kinsey
- **Did the Health Card Program Ensure Access to Medical Care for the Poor during Indonesia's Economic Crisis?**
Menno Pradhan, Fadia Saadah, and Robert Sparrow
- **Incremental Reform and Distortions in China's Product and Factor Markets**
Xiaobo Zhang and Kong-Yan Tan



THE WORLD BANK

1818 H Street, NW

Washington, DC 20433, USA

World Wide Web: <http://mv.worldbank.org/>

E-mail: wber@worldbank.org

ISBN 0-19-920470-5



9 780199 204700