

26452

THE WORLD BANK ECONOMIC REVIEW

Volume 16 • 2002 • Number 2

FINANCIAL CRISES, CREDIT RATINGS, AND BANK FAILURES

An Introduction

Carmen M. Reinhart

Default, Currency Crises, and Sovereign Credit Ratings

Carmen M. Reinhart

Emerging Market Instability

Do Sovereign Ratings Affect Country Risk and Stock Returns?

Graciela Kaminsky and Sergio L. Schmukler

On the Use of Portfolio Risk Models and Capital Requirements
in Emerging Markets: The Case of Argentina

Veronica Balzarotti, Michael Ealkenhem, and Andrew Powell

IMPACT EVALUATION OF SOCIAL FUNDS

An Introduction

Laura B. Rawlings and Norbert R. Schady

Supporting Communities in Transition:

The Impact of the Armenian Social Investment Fund

Robert S. Chase

An Impact Evaluation of Education, Health, and Water Supply
Investments by the Bolivian Social Investment Fund

*John Neuman, Menno Pradhan, Laura B. Rawlings, Geert Ridder,
Ramiro Coa, and Jose Luis Etia*

The Impact and Targeting of Social Infrastructure Investments:
Lessons from the Nicaraguan Social Fund

Menno Pradhan and Laura B. Rawlings

The Allocation and Impact of Social Funds: Spending on School
Infrastructure in Peru

Christina Paxson and Norbert R. Schady

www.iber.cupjournal.org

OXFORD

ISSN 0258-6770

FILE COPY

THE WORLD BANK ECONOMIC REVIEW

EDITOR

François Bourguignon, *World Bank*

EDITORIAL BOARD

- | | |
|--|---|
| Abhijit Banerjee, <i>Massachusetts Institute of Technology, USA</i> | Ravi Kanbur, <i>Cornell University, USA</i> |
| Kaushik Basu, <i>Cornell University, USA</i> | Elizabeth M. King, <i>World Bank</i> |
| Tim Besley, <i>London School of Economics, UK</i> | Justin Yifu Lin, <i>China Center for Economic Research, Peking University, China</i> |
| Anne Case, <i>Princeton University, USA</i> | Mustapha Kamel Nabli, <i>World Bank</i> |
| Stijn Claessens, <i>University of Amsterdam, The Netherlands</i> | Juan Pablo Nicolini, <i>Universidad di Tella, Argentina</i> |
| Paul Collier, <i>World Bank</i> | Howard Pack, <i>University of Pennsylvania, USA</i> |
| David R. Dollar, <i>World Bank</i> | Jean-Philippe Platteau, <i>Facultés Universitaires Notre-Dame de la Paix, Belgium</i> |
| Antonio Estache, <i>World Bank</i> | Boris Pleskovic, <i>World Bank</i> |
| Augustin Kwasi Fosu, <i>African Economic Research Council, Kenya</i> | Martin Ravallion, <i>World Bank</i> |
| Mark Gersovitz, <i>The Johns Hopkins University, USA</i> | Carmen Reinhart, <i>University of Maryland, USA</i> |
| Jan Willem Gunning, <i>Free University, Amsterdam, The Netherlands</i> | Mark R. Rosenzweig, <i>University of Pennsylvania, USA</i> |
| Jeffrey S. Hammer, <i>World Bank</i> | Joseph E. Stiglitz, <i>Columbia University, USA</i> |
| Karla Hoff, <i>World Bank</i> | Moshe Syrquin, <i>University of Miami, USA</i> |
| Gregory K. Ingram, <i>World Bank</i> | Vinod Thomas, <i>World Bank</i> |
| | L. Alan Winters, <i>University of Sussex, UK</i> |

The World Bank Economic Review is a professional journal for the dissemination of World Bank-sponsored and outside research that may inform policy analyses and choices. It is directed to an international readership among economists and social scientists in government, business, and international agencies, as well as in universities and development research institutions. The *Review* emphasizes 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 will illustrate how professional research can shed light on policy choices. Inconsistency with Bank policy will not be grounds for rejection of an article.

Articles will be drawn from work conducted by World Bank staff and consultants and from papers submitted by outside researchers. Before being accepted for publication, all articles will be reviewed by two referees who are not members of the Bank's staff and one World Bank staff member. Articles must also be recommended by a member of the Editorial Board. Non-Bank contributors are requested to submit a proposal of not more than two pages in length to the Editor or a member of the Editorial Board before sending in their paper.

Comments or brief notes responding to *Review* articles are welcome and will be considered for publication to the extent that space permits. Please direct all editorial correspondence to the Editor, *The World Bank Economic Review*, The World Bank, 1818 H Street, Washington, DC 20433, USA, or wber@worldbank.org.

For more information, please visit the Web sites of the *Economic Review* at www.wber.oupjournals.org, the World Bank at www.worldbank.org, and Oxford University Press at www.oup-usa.org.

THE WORLD BANK ECONOMIC REVIEW

Volume 16 • 2002 • Number 2

FINANCIAL CRISES, CREDIT RATINGS, AND BANK FAILURES

An Introduction	149
<i>Carmen M. Reinhart</i>	
Default, Currency Crises, and Sovereign Credit Ratings	151
<i>Carmen M. Reinhart</i>	
Emerging Market Instability: Do Sovereign Ratings Affect Country Risk and Stock Returns?	171
<i>Graciela Kaminsky and Sergio L. Schmukler</i>	
On the Use of Portfolio Risk Models and Capital Requirements in Emerging Markets: The Case of Argentina	197
<i>Veronica Balzarotti, Michael Falkenheim, and Andrew Powell</i>	

IMPACT EVALUATION OF SOCIAL FUNDS

An Introduction	213
<i>Laura B. Rawlings and Norbert R. Schady</i>	
Supporting Communities in Transition: The Impact of the Armenian Social Investment Fund	219
<i>Robert S. Chase</i>	
An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund	241
<i>John Newman, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia</i>	
The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund	275
<i>Menno Pradhan and Laura B. Rawlings</i>	
The Allocation and Impact of Social Funds: Spending on School Infrastructure in Peru	297
<i>Christina Paxson and Norbert R. Schady</i>	

The World Bank Economic Review (ISSN 0258-6770) is published three times a year by Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009 for The International Bank for Reconstruction and Development / THE WORLD BANK. Communications regarding original articles and editorial management should be addressed to The Editor, *The World Bank Economic Review*, 66, avenue d'Iéna, 75116 Paris, France. E-mail: wber@worldbank.org.

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide.

SUBSCRIPTIONS: Subscription is on a yearly basis. The annual rates are US\$40 (£30 in UK and Europe) for individuals; US\$101 (£71 in UK and Europe) for academic libraries; US\$123 (£84 in UK and Europe) for corporations. Single issues are available for US\$16 (£12 in UK and Europe) for individuals; US\$43 (£30 in UK and Europe) for academic libraries; US\$52 (£35 in UK and Europe) for corporations. All prices include postage. Individual rates are applicable only when a subscription is for individual use and are not available if delivery is made to a corporate address. *Subscriptions are provided free of charge to non-OECD countries.* All subscription requests, single issue orders, changes of address, and claims for missing issues should be sent to:

North America: Oxford University Press, Journals Customer Service, 2001 Evans Road, Cary, NC 27513-2009, USA. Toll-free in the USA and Canada: 800-852-7323, or 919-677-0977. Fax: 919-677-1714. E-mail: jnlorders@oup-usa.org.

Elsewhere: Oxford University Press, Journals Customer Service, Great Clarendon Street, Oxford OX2 6DP, UK. Tel: +44 1865 353907. Fax: +44 1865 353485. E-mail: jnl.cust.serv@oup.co.uk.

ADVERTISING: Helen Pearson, Oxford Journals Advertising, P.O. Box 347, Abingdon SO, OX14 1GJ, UK. Tel/Fax: +44 1235 201904. E-mail: helen@oxfordads.com.

BACK ISSUES: The current plus all back volumes (from 1997) are available from Oxford University Press at the *North America* contact information listed above.

REQUESTS FOR PERMISSIONS, REPRINTS, AND PHOTOCOPIES: 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 either prior written permission of the publisher (Oxford University Press, Journals Rights and Permissions, Great Clarendon Street, Oxford OX2 6DP, UK; tel: +44 1865 354490; fax: +44 1865 353485) or a license permitting restricted copying issued in the USA by the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 (fax: 978-750-4470), or in the UK by the Copyright Licensing Agency Ltd., 90 Tottenham Court Road, London W1P 9HE, UK. Reprints of individual articles are available only from the authors.

COPYRIGHT: Copyright © 2002 The International Bank for Reconstruction and Development / THE WORLD BANK. It is a condition of publication in the journal that authors assign copyright to The International Bank for Reconstruction and Development / THE WORLD BANK. However, requests for permission to reprint material found in the journal should come to Oxford University Press. This ensures that requests from third parties to reproduce articles are handled efficiently and consistently and will also allow the article to be disseminated as widely as possible. Authors may use their own material in other publications provided that the journal is acknowledged as the original place of publication and Oxford University Press is notified in writing and in advance.

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 Sciences Citation Index*. The microform edition is available through ProQuest (formerly UMI), 300 North Zeeb Road, Ann Arbor, MI 48106, USA.

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).

POSTAL INFORMATION: *The World Bank Economic Review* (ISSN 0258-6770) is published three times a year by Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009. Send address changes to *The World Bank Economic Review*, Journals Customer Service Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513-2009.

Financial Crises, Credit Ratings, and Bank Failures

An Introduction

Carmen M. Reinhart

Financial crises of every variety rocked emerging markets in the second half of the 1990s. Nearly every region experienced currency crashes, banking crises were both numerous and severe, and a few countries, facing extreme duress, defaulted on their sovereign debt. Not surprisingly, then, there is considerable interest in policy and academic circles and within the investment community in gaining a better understanding of financial and economic distress.

An aspect that has received particular attention in the growing empirical literature on financial crises is crisis prediction or, more broadly, assessment of imminent downside risks—for banks, currency, or sovereign debt. Much of this literature focuses on selecting the “right” set of economic and occasionally political fundamentals. Rating agencies, responsible for monitoring credit risks, came under considerable scrutiny following their lackluster performance in anticipating the Mexican peso crisis (which nearly resulted in default) and the Asian crises of 1997–98.

The impact of changes in sovereign credit ratings on financial markets also became a matter for debate, with some evidence that credit rating agencies could be fueling a boom-bust cycle in international capital markets. Central banks across the globe and multilateral lending institutions in particular became painfully aware of the pressing need to develop and implement more robust ways of “stress testing” financial systems, to ensure adequate capital requirements and provisioning and limit the risk of bank failure.

Three articles in this issue address various aspects of these topics—in particular, the behavior and impact of sovereign credit ratings and the development of a portfolio model for testing the vulnerability of banks. Two of the articles address the usefulness and impact of sovereign credit ratings. Reinhart evaluates the performance of credit ratings in anticipating currency crises and defaults, finding that ratings systematically fail to anticipate currency crises but do somewhat better in anticipating defaults. However, downgrades in credit ratings usually follow currency crises, suggesting both that currency instability increases the risk of default and that credit ratings tend to be procyclical. Using a very

Carmen M. Reinhart is a professor of economics with the University of Maryland, on leave at the International Monetary Fund. Her e-mail address is creinhart@imf.org.

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

different sample, data frequency, and analytical approach from Reinhart, Kaminsky, and Schmukler also find evidence of procyclicality in sovereign credit ratings. They show that credit upgrades usually take place during market rallies and downgrades during bond and equity market slumps.

In the third article, Balzarotti, Falkenheim, and Powell focus on bank defaults rather than sovereign defaults, exploring how capital and provisioning measures can limit bank failures. Their approach stresses a portfolio-based model for estimating the potential losses of banks under a variety of adverse but plausible scenarios. An important implication of their study is that for the many emerging markets that have developed public credit registries, the information collected by these agencies can be used to develop better measures of banks' credit risks.

Financial Crises, Credit Ratings, and Bank Failures

Default, Currency Crises, and Sovereign Credit Ratings

Carmen M. Reinhart

Sovereign credit ratings play an important part in determining countries' access to international capital markets and the terms of that access. In principle, there is no reason to expect that sovereign credit ratings should systematically predict currency crises. In practice, in emerging market economies there is a strong link between currency crises and default. Hence if credit ratings are forward-looking and currency crises in emerging market economies are linked to defaults, it follows that downgrades in credit ratings should systematically precede currency crises. This article presents results suggesting that sovereign credit ratings systematically fail to predict currency crises but do considerably better in predicting defaults. Downgrades in credit ratings usually follow currency crises, possibly suggesting that currency instability increases the risk of default.

Sovereign credit ratings play an important part in determining countries' access to international capital markets and the terms of that access. As more countries are added to the list of rated sovereigns, the information content of ratings becomes even more important.¹ Credit ratings have been shown to have a significant impact on the yield spreads of sovereign bonds.² Indeed, sovereign credit ratings are taken as summary measures of the likelihood that a country will default. It is hardly surprising that the countries with the lowest ratings are those that are unable to borrow from international capital markets and are dependent on official loans from multilateral institutions or governments. In a cross-sectional setting, sovereign credit ratings do well in distinguishing across borrowers.

Developed countries take access to international capital markets for granted. At the other end of the spectrum, many low-income countries, mired in debt,

Carmen M. Reinhart is Professor of Economics at the University of Maryland, on leave at the International Monetary Fund. Her e-mail address is creinhart@imf.org. This article was prepared while the author was professor at the University of Maryland, for the New York University–University of Maryland project “The Role of Credit Rating Agencies in the International Economy.” The author thanks François Bourguignon, Stijn Claessens, Ricardo Hausmann, Peter Kenen, and three anonymous referees for helpful comments; Facundo Martin, Ioannis Tokatlidis, and Juan Trevino for superb research assistance; and the Center for International Political Economy for financial support. The article represents the views of the author and not necessarily those of the institution with which she is affiliated.

1. To cite a recent example, Standard and Poor's added Guatemala to the list of rated sovereigns in November 2001.

2. See, for example, Larraín and others (1997), who find evidence that ratings “Granger cause” the yield spreads of sovereign bonds.

have no access to international lending even under relatively favorable states of nature.³ For emerging market economies, access to international capital markets is precarious and highly variable over time. For these economies sovereign credit ratings are most critical.

In principle, there is no reason to expect that sovereign credit ratings should systematically predict currency or banking crises. After all, developed countries have had their share of currency crises (such as the 1992–93 Exchange Rate Mechanism crisis in France and the United Kingdom), banking crises (such as that in Japan in the 1990s and the savings and loan crisis in the United States in the early 1980s), and simultaneous currency and banking crises (such as the twin crises in the Nordic countries in the early 1990s and in Spain in the late 1970s and early 1980s).⁴ These crises did not lead to systematic markdowns in credit ratings.

In practice, however, in emerging market economies there is a strong link between currency crises and default.⁵ Without the colossal bailout packages put together by the international community, there is little doubt that the currency crises in Mexico, the Republic of Korea, Thailand, and more recently Turkey would all have produced a sovereign debt default.⁶ Hence if credit ratings are forward-looking and currency crises in emerging market economies are linked to defaults, it follows that downgrades should systematically precede currency crises. It is this temporal (rather than cross-sectional) behavior of credit ratings that this article investigates.

Contrary to logic, recent anecdotal evidence suggests that downgrades in credit ratings have not preceded financial crises. Downgrades appear to have followed, not preceded, the crises in Asia in 1997 (table 1). A natural question, then, is whether this failure by the rating agencies to anticipate debt servicing difficulties is systematic. Goldstein and others (2000) examine the links between currency and banking crises and changes in sovereign credit ratings by Institutional Investor and Moody's for 20 countries. They find mixed evidence on the ability of the rating agencies to anticipate financial crises. Neither rating agency predicted banking crises, but there is evidence that the Moody's sovereign ratings have some (very low) predictive power for currency crises.

This article casts a wider net, examining the links among crises, default, and rating changes for 46–62 economies, depending on the rating agency. It also extends the analysis to Standard and Poor's sovereign ratings and different approaches to dating the currency crises.

3. Favorable states of nature include both shocks that are idiosyncratic to a country, such as an increase in its terms of trade, and common shocks, such as a decline in world interest rates.

4. For an analysis of twin crises see Kaminsky and Reinhart (1999).

5. Furthermore, some of the indicators useful in predicting currency crises are also useful in predicting debt crises. See Detragiache and Spilimbergo (2001).

6. Even if the government itself has little outstanding debt, history has shown that, time after time, governments assume private-sector debt during currency crises.

TABLE 1. Performance of Rating Agencies before the Asian Crisis: Long-Term Debt Ratings, 1996–97

Rating agency and country	Jan. 15, 1996		Dec. 2, 1996		June 24, 1997		Dec. 12, 1997		
	Rating	Outlook	Rating	Outlook	Rating	Outlook	Rating	Out	
<i>Moody's foreign currency debt</i>									
Indonesia	Baa3		Baa3		Baa3		Baa3		
Korea, Rep. of	A1		A1	Stable			Baa2	Negative	
Malaysia	A1		A1		A1		A1		
Mexico	Ba2		Ba2		Ba2		Ba2		
Philippines	Ba2		Ba2		Ba2		Ba2		
Thailand	A2		A2		A2		Baa1	Negative	
Rating agency and country	Jan. 15, 1996		Dec. 2, 1996		June 24, 1997		Oct. 1997		
	Rating	Outlook	Rating	Outlook	Rating	Outlook	Rating	Out	
<i>Standard and Poor's</i>									
Indonesia	Foreign currency debt	BBB	Stable	BBB	Stable	BBB	Stable	BBB	Negative
	Domestic currency debt			A+		A+		A-	Negative
Korea, Rep. of	Foreign currency debt	AA-	Stable	AA-	Stable				
	Domestic currency debt								
Malaysia	Foreign currency debt	A+	Stable	A+	Stable	A+	Positive	A+	Negative
	Domestic currency debt	AA+		AA+		AA+		AA+	Negative
Mexico	Foreign currency debt	BB	Negative	BB		BB			
	Domestic currency debt	BBB+		BBB+	Stable	BBB+	Positive		
Philippines	Foreign currency debt	BB	Positive	BB	Positive	BB+	Positive	BB+	Stable
	Domestic currency debt	BBB+		BBB+		A-		A-	Stable
Thailand	Foreign currency debt	A	Stable	A	Stable	A	Stable	BBB	Negative
	Domestic currency debt			AA		AA		A	Negative

Note: The rating system for Moody's is as follows (from highest to lowest): Aaa, Aa1, Aa2, Aa3, A1, A2, A3, Baa1, Baa2, Baa3, Ba1, Ba2, Ba3. The rating system for Standard and Poor's is as follows (from highest to lowest): AAA, AA+, AA, AA-, A+, A, A-, BBB+, BBB, BBB-, BB+, BB, BB-. Blank cells denote no action or change at that time.

Source: Radelet and Sachs (1998).

I. CRISES, DEFAULT, AND THE AFTERMATH

Calvo and Reinhart (2000) have suggested that one reason emerging market economies may fear devaluations (or large depreciations) is that the devaluations are associated with a loss of access to international credit, which in turn is associated with severe recessions.⁷ To examine this issue, the article begins by assessing the temporal links between episodes of sovereign default and currency crises. In response to the recent anecdotal evidence suggesting that downward adjustments in sovereign credit ratings have come after currency crises were well under way, it also reviews the behavior of credit ratings in the aftermath of crises. It also examines the extent to which currency crises lead to a reassessment of the risk of sovereign default and whether distinct patterns emerge for developed and emerging market economies.

Data and Definitions

The analysis covers the sovereign credit ratings issued by Institutional Investor, Moody's Investors Service, and Standard and Poor's. The Institutional Investor sample begins in 1979 and runs through 1999. The panels for the Moody's and Standard and Poor's ratings are unbalanced (that is, they do not have the same number of observations for all economies).

The Institutional Investor ratings run from 0 (least creditworthy) to 100 (most creditworthy). The ratings are reported twice a year and changed frequently. Moody's and Standard and Poor's use multiple letters to rate a sovereign's creditworthiness. For the purposes of the analysis the letter ratings are mapped into 17 categories, with 16 corresponding to the highest rating and 0 to the lowest (for an illustration see table 2).⁸ Moody's and Standard and Poor's may change their ratings at any time (though they do so much less often than Institutional Investor), so the samples for these rating agencies include, for each economy, the months in which any rating changes took place.

With 62 economies, the Institutional Investor sample is the largest (table 3). The Standard and Poor's sample is the smallest, with 46 economies, but is nonetheless more than twice the size of the 20-country sample used by Goldstein and others (2000). The sample for the analysis of the links between currency crises and default includes 58 economies and spans the period 1970–99.

Methodological Issues

To assess the interaction among currency crises, default, and sovereign credit ratings, the crises need to be dated. Two indexes of currency crises are constructed, to assess whether the results are sensitive to the definition of crises used. The first index is that used by Kaminsky and Reinhart (1999) for 20 countries now

7. Calvo and Reinhart (2000) present evidence that the recessions following devaluations are deeper in emerging market economies than in developed economies.

8. This approach follows the procedure adopted in Cantor and Packer (1996a, 1996b).

TABLE 2. Scale for Moody's
Foreign Currency Debt Rating

Rating	Assigned value
Aaa	16
Aa1	15
Aa2	14
Aa3	13
A1	12
A2	11
A3	10
Baa1	9
Baa2	8
Baa3	7
Ba1	6
Ba2	5
Ba3	4
B1	3
B2	2
B3	1
C	0

Source: Compiled by the author based on data from Moody's.

extended to the larger sample. The second is the definition of crises employed by Frankel and Rose (1996).⁹

Kaminsky and Reinhart's (1999) crisis index, (KR) , I , is a weighted average of the rate of change of the exchange rate, $\Delta e/e$, and the rate of change of reserves, $\Delta R/R$, with weights such that the two components of the index have equal sample volatilities:

$$(1) \quad I = (\Delta e/e) - (\sigma_e/\sigma_R) * (\Delta R/R),$$

where σ_e is the standard deviation of the rate of change of the exchange rate and σ_R is the standard deviation of the rate of change of reserves. Because changes in the exchange rate enter with a positive weight and changes in reserves with a negative weight, readings of this index that are three standard deviations or more above the mean are catalogued as crises.

Construction of the index is modified for economies in the sample that have experienced hyperinflation. Although a 100 percent devaluation may be traumatic for a country with low to moderate inflation, a devaluation of that size is common during episodes of hyperinflation. For countries that have had such an episode, a single index would miss sizable devaluations and reserve losses in the periods of moderate inflation, since the high-inflation episode distorts the his-

9. An earlier version of this article included a modified version of Frankel and Rose's index that dates "milder" crises. See Reinhart (2002).

TABLE 3. The Samples

Institutional Investors: biannual observations for 62 economies, 1979–99

Algeria, Argentina, Australia, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Czech Republic, Denmark, Dominican Republic, Ecuador, Arab Republic of Egypt, El Salvador, Ethiopia, Finland, Ghana, Greece, Hong Kong (China), Hungary, India, Indonesia, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Republic of Korea, Malaysia, Mali, Mexico, Morocco, Nepal, New Zealand, Nigeria, Norway, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, the Philippines, Poland, Portugal, Romania, Saudi Arabia, Singapore, South Africa, Spain, Sri Lanka, Swaziland, Sweden, Tanzania, Thailand, Turkey, United States, Uruguay, Venezuela, and Zimbabwe.

Moody's Investors Service: monthly observations for 48 economies, unbalanced panel, 1979–99

Argentina, Australia, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Czech Republic, Denmark, Dominican Republic, Ecuador, Arab Republic of Egypt, El Salvador, Finland, Greece, Hong Kong (China), Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Republic of Korea, Malaysia, Mexico, Morocco, New Zealand, Norway, Pakistan, Panama, Paraguay, the Philippines, Poland, Portugal, Romania, Saudi Arabia, Singapore, South Africa, Spain, Sweden, Thailand, Turkey, United States, Uruguay, and Venezuela.

Standard and Poor's: monthly observations for 46 economies, unbalanced panel, 1979–99

Argentina, Australia, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Czech Republic, Denmark, Dominican Republic, Arab Republic of Egypt, El Salvador, Finland, Greece, Hong Kong (China), Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Republic of Korea, Malaysia, Mexico, Morocco, New Zealand, Norway, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, the Philippines, Poland, Portugal, Romania, Singapore, Spain, Sweden, Thailand, Turkey, United States, and Uruguay.

Sample for the interaction between currency crises and defaults: 58 economies, 1970–99, 151 currency crises, 113 defaults

Algeria, Argentina, Bolivia, Brazil, Burkina Faso, Cameroon, the Central African Republic, Chile, Colombia, the Democratic Republic of Congo, Costa Rica, Côte d'Ivoire, Denmark, Dominican Republic, Ecuador, Arab Republic of Egypt, Ethiopia, Finland, Gabon, Gambia, Ghana, Guatemala, Guinea, Guinea-Bissau, Guyana, Honduras, India, Indonesia, Israel, Jamaica, Jordan, Madagascar, Malawi, Mexico, Morocco, Nicaragua, Niger, Nigeria, Norway, Paraguay, Peru, the Philippines, Romania, São Tomé and Príncipe, Senegal, Sierra Leone, Spain, Sudan, Sweden, Tanzania, Togo, Trinidad and Tobago, Turkey, Uganda, Uruguay, Venezuela, Zambia, and Zimbabwe.

torical mean. To avoid this, the sample economies are sorted into two groups according to whether the inflation rate in the previous six months exceeded 150 percent, and an index is then constructed for each subsample.¹⁰ As earlier studies (see Frankel and Rose 1996) have noted, the dates of crises obtained using this method map well onto the dates obtained when crises are defined exclusively on the basis of events, such as the closing of the exchange markets or a change in the exchange rate regime.

10. Similar results are obtained by using significant departures in inflation from 6- and 12-month moving averages.

The Frankel and Rose (*FR*) definition of a currency crisis is a 25 percent or greater devaluation in a given month that is also at least 10 percent greater than the devaluation in the preceding month (Frankel and Rose 1996).¹¹

Episodes of default are dated using Beers and Bhatia (1999), who provide dates of default from 1824 to 1999 (the analysis here focuses on 1970–99); Beim and Calomiris (2001); the World Bank's *Global Development Finance* (various years); and the dates of debt crises provided by Detragiache and Spilimbergo (2001).¹² In some cases these sources pinpoint the exact month in which a default was announced. The sample includes defaults on both foreign currency bank debt and foreign currency bond debt but not on local currency instruments. It includes defaults for both rated and unrated sovereigns.

A Sketch of the Signals Approach

The “signals” approach developed by Kaminsky and Reinhart (1999) is used to compare the performance of the ratings—and of some of the economic indicators on which rating agencies focus—with the performance of some of the other (and better) predictors of financial crises. In a nutshell, the signals approach involves a set of possible outcomes, as presented in the following two-by-two matrix (matrix 1).¹³

MATRIX 1. Crisis and Signals

	Crisis occurs in the following 24 months	No crisis occurs in the following 24 months
Signal	A	B
No signal	C	D

A perfect indicator would have entries only in cells A and D.

This matrix permits the definition of several useful concepts employed to evaluate the performance of each indicator. If no information were available on the performance of the indicators, it would still be possible to calculate, for a given sample, the unconditional probability of crisis:

$$(2) \quad P(C) = (A + C)/(A + B + C + D).$$

11. The modified *FR* index (*MFR*), used in the earlier version of this article (Reinhart 2002), classifies as a crisis a devaluation in a given month that is 20 percent or greater and at least 5 percent greater than the devaluation in the preceding month.

12. Detragiache and Spilimbergo (2001) classify an observation as a debt crisis if either or both of the following conditions occur: there are arrears of principal or interest on external obligations to commercial creditors (banks or bondholders) exceeding 5 percent of total commercial debt outstanding, or there is a rescheduling or debt restructuring agreement with commercial creditors as listed in the World Bank's *Global Development Finance*.

13. For a more detailed description of the signals approach see Kaminsky and others (1998) and Goldstein and others (2000).

If an indicator sends a signal and that indicator has a reliable track record, the probability of a crisis, conditional on a signal, $P(C | S)$, can be expected to be greater than the unconditional probability where

$$(3) \quad P(C | S) = A/(A + B).$$

Formally,

$$(4) \quad P(C | S) - P(C) > 0.$$

The intuition is clear: if the indicator is not “noisy” (prone to sending false alarms), there will be relatively few entries in cell B , and $P(C | S) \approx 1$. This is one of the criteria used to rank the indicators. The noise-to-signal ratio can be defined as

$$(5) \quad N/S = [B/(B + D)]/[A/(A + C)].$$

If an indicator has a track record of relatively few false alarms, this may mean that the indicator issues signals relatively rarely and that there is a danger that it misses crises altogether (that is, that it does not signal before a crisis). Hence the proportion of crises accurately signaled is also calculated for each indicator:

$$(6) \quad PC = C/(A + C).$$

For credit ratings, a downgrade in the 24 months before a crisis would be considered a signal.

Currency Crises and Default: The Links

To analyze the interaction between defaults and currency crises, the two-by-two matrix (matrix 2) can be recast as follows if defaults signal currency crises.

MATRIX 2. Defaults Lead Currency Crisis

	Currency Crisis occurs in the following 24 months	No crisis occurs in the following 24 months
Default	A	B
No default	C	D

If the converse is true, the matrix can be recast in this way (matrix 3):

MATRIX 3. Currency Crises Lead Default

	Default occurs in the following 24 months	No default occurs in the following 24 months
Currency crisis	A	B
No currency crisis	C	D

To simply look at the joint occurrence of default and currency crises, the 24-month window would be extended so that it is two-sided around the default date.

The sample for the analysis of the links between defaults and currency crises includes 113 defaults and 151 currency crises, 135 of them in emerging market economies. The unconditional probability of defaulting is 13.3 percent if devel-

oped economies (for which the probability of default is zero) are excluded from the sample (table 4).¹⁴ The unconditional probability of a currency crisis is about 17 percent. This probability changes little when developed economies are excluded from the sample, highlighting the fact that the key difference between developed and emerging market economies is debt problems—although debt problems are tightly linked to currency problems in emerging market economies, as we shall see. The probability of having a currency crisis within 24 months of defaulting (with the crisis either before or after the default) is about 84 percent. Because defaults are somewhat rarer than currency crises, the probability of having a default within 24 months of a currency crisis is lower: about 58 percent for the entire sample and 66 percent for emerging market economies. This second subset is the relevant group, because it accounts for all the episodes of default in the sample. This exercise points to the strong association between debt events and currency crises in emerging market economies.

What temporal pattern do the results reveal? The probability of having a currency crisis conditional on having defaulted is about 69 percent, whereas the probability of defaulting conditional on having had a currency crisis is somewhat lower, at around 46 percent for emerging market economies. What inference is to be drawn from these results? Not so much that there is any obvious causal pattern—although currency crises conditional on having defaulted are more common—but that currency crises are more frequent and in about half the cases (even in emerging market economies) do not necessarily lead to default. Indeed, this stylized fact may help explain why credit ratings do poorly in predicting currency crises, an issue taken up later. Of course, as discussed, there are a few cases in the sample in which a currency crisis would have precipitated a default in the absence of a major intervention by the financial community.

Sovereign Credit Ratings in the Aftermath of Crises

Further evidence that devaluations (or large depreciations) are followed more often than not by debt servicing difficulties can be gleaned from studying the behavior of sovereign credit ratings in the aftermath of such events. Results from analysis of Institutional Investor ratings around currency crises show no significant difference between developed and emerging market economies in the probability of a downgrade (or multiple downgrades) following a currency crisis (table 5). But this is where the similarities between the two groups of economies end. The average rating for emerging market economies at the time of a crisis is 37.6, slightly less than half the average score for developed countries. This suggests, of course, that even in the absence of a crisis access to international lending is far from even for the two groups. Moreover, that vast gap widens further in the aftermath of de-

14. This is the probability of a new default, not the probability of being in a state of default, which is larger. For example, Sierra Leone was in default on its foreign currency bank debt during 1983–84 and 1986–95. This is treated as a single episode of default beginning in 1983, just as in Beim and Calomiris (2001).

TABLE 4. The Twin D's: Devaluation and Default, 1970–99 (percent)

Probability	All sample economies	Emerging market economies only
Unconditional probability of a default occurring in the next 24 months	12.0 ^a	13.3
Unconditional probability of a currency crisis occurring in the next 24 months	17.3	16.9
Probability of a currency crisis occurring within 24 months before or after a default	84.0 ^b	84.0 ^b
Probability of a default occurring within 24 months before or after a currency crisis	57.5	65.7
Probability of a currency crisis occurring within 24 months of having defaulted	69.3	69.3
Probability of a default occurring within 24 months of having had a currency crisis	39.4	46.0

^aThe probability of default for the developed economies in the sample is 0.

^bThe probabilities for the entire sample and the subset of emerging market economies are the same because although developed economies had currency crises, none defaulted within the sample period.

Source: Author's calculations.

valuations associated with currency crises. In the 12 months following a currency crisis, the sovereign rating index for emerging market economies falls by 10.8 percent on average, a downgrade about five times that for developed economies. The difference between the postcrisis downgrades for emerging market and developed economies is significant at standard confidence levels.

A comparable exercise performed for the Moody's ratings shows an even greater gulf between emerging market and developed economies. Like the Institutional Investor ratings, the Moody's ratings at the outset of a currency crisis are significantly lower for emerging market economies—on average, about a third that for developed economies (table 6). Again, the downgrade for emerging market economies (about 9 percent) is far greater than that for developed economies (less than 1 percent). But the probability of a downgrade—and the probability of multiple downgrades—in the 12 months following a crisis are significantly higher for emerging market economies in the case of Moody's sovereign ratings. Consistent with the findings on the interaction between defaults and currency crises, the behavior of sovereign credit ratings in the aftermath of currency crises suggests that such crises increase the probability of default—but not necessarily that currency crises equal default.

A useful analysis complementing the preceding one examines whether knowing that there was a currency crisis indeed helps predict downgrades in sovereign credit ratings for emerging market and developed economies. For the Institutional Investor sample, for which there is a continuous time series, the six-month change in the credit rating index is regressed on a dummy variable that takes the value of one when there was a crisis six months earlier, and zero otherwise. The

TABLE 5. Probability and Size of Downgrades in Institutional Investor Sovereign Credit Ratings around Currency Crises, 1979–99

	Probability (percent)		
	Of downgrade in the 6 months following the crisis	Of downgrade in the 12 months following the crisis	Of more than one downgrade in the 12 months following the crisis
Emerging market economies	39.0	79.3	31.7
Developed economies	38.4	73.1	30.8
Difference	0.6	6.2	0.9
	Index level		
	At time of crisis	In the next 6 months	12 months later
Emerging market economies	37.6	36.0	33.5
Developed economies	76.0	74.9	74.5
Difference	-38.4**	-38.9**	-41.0**
	Size of downgrade (percentage change)		
	In the 6 months following the crisis	In the next 6 months	In the 12 months following the crisis
Emerging market economies	4.3	6.9	10.9
Developed economies	1.4	0.5	1.9
Difference	2.8*	6.4**	8.9**

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Calvo and Reinhart (2000); author's calculations.

TABLE 6. Probability and Size of Downgrades in Moody's Sovereign Credit Ratings around Currency Crises, 1979-99

	Probability (percent)		
	Of downgrade in the 6 months following the crisis	Of downgrade in the 12 months following the crisis	Of more than one downgrade in the 12 months following the crisis
Emerging market economies	20.0	26.7	6.7
Developed economies	10.0	10.0	0.0
Difference	10.0**	16.7**	6.7*

	Index level		
	At time of crisis	In the next 6 months	12 months later
Emerging market economies	4.9	4.5	4.3
Developed economies	15.0	14.9	14.9
Difference	-10.1**	-10.4**	-10.6**

	Size of downgrade (percentage change)		
	In the 6 months following the crisis	In the next 6 months	In the 12 months following the crisis
Emerging market economies	8.2	4.4	12.2
Developed economies	0.7	0.0	0.7
Difference	7.5**	4.4**	11.5**

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Calvo and Reinhart 2000; author's calculations.

method of estimation is generalized least squares, correcting for both generalized forms of heteroscedasticity and serial correlation in the residuals. For the Moody's sample the dependent variable is the three-month change in the rating, and the explanatory variable is the dummy variable for currency crises three months earlier. This specification makes it possible to determine more precisely whether downgrades follow rapidly after crises occur. The dependent variable assumes the value -1 (if there was a downgrade), 0 (no change), or 1 (an upgrade). The parameters of interest are estimated using an ordered probit technique that permits correction for heteroscedastic disturbances.

The results of the estimation show that for emerging market economies, currency crises help predict downgrades regardless of which rating index is used (table 7). For developed economies, however, there is no conclusive evidence that ratings respond to currency crises in a systematic and significant way, at least after 1970. This finding is perfectly consistent with the probability assessment showing no links between currency crises and default. For emerging market economies the coefficients are significant at standard confidence levels, but the marginal predictive contribution of currency crises to predicting default remains small. For example, a currency crisis increases the likelihood of a downgrade in the Moody's ratings by 5 percent. The results from this exercise reinforce the view that large devaluations or depreciations in emerging market economies increase the likelihood of default, as evidenced by the downgrades in sovereign ratings. That the predictive ability of currency crises is so low suggests that other economic fundamentals are important in explaining changes in sovereign credit ratings (see Cantor and Packer 1996a, 1996b).

The results are also consistent with the conclusions of Larraín and others (1997), who find evidence of two-way causality between sovereign ratings and

TABLE 7. Sovereign Credit Ratings following Currency Crises in Developed and Emerging Market Economies

	Developed economies	Emerging market economies
<i>Institutional investor</i> ^a		
Coefficient	-0.007	-0.08**
Standard error	0.023	0.011
R ²	0.01	0.07
<i>Moody's</i> ^b		
Coefficient	-0.08	-0.31**
Standard error	0.76	0.11
Pseudo-R ²	0.000	0.060

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Note: Independent variable is a dummy variable for currency crises.

^aEstimation method is ordinary least squares with robust standard errors. Dependent variable is the 6-month change in the sovereign credit rating.

^bEstimation method is ordered probit. Dependent variable is the 3-month change in the sovereign credit rating.

Source: Author's calculations.

market spreads. Hence, not only do international capital markets react to changes in the ratings, but the ratings systematically react (with a lag) to market conditions, as reflected in the yield spreads of sovereign bonds.

II. SOVEREIGN CREDIT RATINGS BEFORE CRISES

The analysis has shown that there is a strong connection between currency crises and default in emerging market economies—about 84 percent of defaults in these economies are associated with a currency crisis in the months before or after the default. It has also shown that slightly more than half the currency crises in emerging market economies are not linked with a subsequent default. Nonetheless, it is evident from the preceding exercise that currency crises do affect the probability of default, as sovereign credit ratings for emerging market economies are systematically downgraded after currency crises. Hence, although it is critical to assess how well the ratings predict default, it is also useful to assess how well changes in sovereign credit ratings predict crises, given the close connection between the two in emerging market economies.

Probit estimation is used to assess the predictive ability of ratings, with the Institutional Investor, Moody's, and Standard and Poor's ratings as regressors for currency crises and defaults. The dependent variable is a dummy variable for currency crises; the independent variable is the 12-month change in the credit rating lagged 1 year. A comparable exercise is performed for episodes of default. Alternative specifications, such as the 6-month change in the credit rating lagged 6, 18, and 24 months, are also considered.¹⁵ The method of estimation corrects for serial correlation and for heteroscedasticity in the residuals.

The basic premise underpinning the simple postulated model is as follows. If the credit rating agencies use all available information on the economic "fundamentals" to form their rating decisions, then credit ratings should help predict defaults and may (or may not) predict currency crises—if the macroeconomic indicators on which the ratings are based have some predictive power. Moreover, the simple model should not be misspecified—that is, other indicators should not be statistically significant, because that information would presumably already be reflected in the ratings. Hence the state of the macroeconomic fundamentals would be captured in a single indicator: the ratings.

Recent studies that have examined the determinants of credit ratings provide support for the basic premise that ratings are significantly linked with certain economic fundamentals (see Cantor and Packer 1996a, Lee 1993). For example, Cantor and Packer (1996a) find that per capita GDP, inflation, external debt, and indicators of default history and economic development are all significant determinants of sovereign ratings.

15. The alternative time horizons, ranging from 6-month changes to 18- and 24-month changes at a variety of lag lengths, produce very similar results. A subset of these results are reported in Reinhart (2002), the rest are available from the author.

In the results of the probit estimation for Institutional Investor ratings, the coefficients of the credit ratings have the expected negative sign for any of the two definitions of currency crises—that is, an upgrade reduces the probability of a crisis (table 8). But for the two definitions of currency crises the coefficient is significant at the 10 percent level. Moreover, as in Reinhart (2002), this result is not robust to other specifications. For example, if the six-month change in the credit rating six months before the crisis is used as a regressor, none of the coefficients are statistically significant. For defaults, the coefficients of the Institutional Investor ratings are significant—but only at the 10 percent confidence level.

For the Moody's sample the coefficients on the ratings variable are statistically insignificant for both the definitions of currency crises, and for the *FR* definition the coefficient has the wrong sign (table 9). Hence for this larger, 48-country sample the Goldstein, Kaminsky, and Reinhart (2000) results do not hold.¹⁶ Interestingly, for defaults, the Institutional Investor ratings are significant at the 5 percent confidence level. If the potential cases of default in the 1990s (the countries that received massive bailout packages, without which default would have been certain) are included, however, ratings are significant only at the 10 percent level and are very sensitive to the lag structure used.

The results for the Standard and Poor's sample are in line with those for the Moody's sample (table 10). Regardless of the definition of crises or specification of lag structure used, none of the coefficients on the changes in credit rating is statistically significant at standard confidence levels. Moreover, the coefficients often have the wrong sign for the dates of crises, though they are much better for the dates of default.¹⁷

These results appear to be at odds with those of Larraín and others (1997), who find evidence that ratings “cause” interest rate spreads. The interpretation here, however, is that although ratings may systematically lead yield spreads (Larraín and colleagues present evidence of two-way causality), yield spreads are poor predictors of crises but better predictors of default. The reason is that, as shown, not all currency crises lead to default. Hence the inability of ratings to predict currency crises is not inconsistent with their ability to influence spreads.

Sovereign Credit Ratings and Macroeconomic Indicators of Crises

A comparison of the performance of credit ratings and of some of the economic indicators on which rating agencies focus with the performance of some of the better predictors of financial crises produces results underscoring the preceding ones. Performance is assessed on the basis of the basic descriptive statistics used

16. For the Moody's sovereign ratings Goldstein and others (2000) find a statistically significant coefficient for their 20-country sample. Even so, the marginal contribution of the ratings variable was very small.

17. As for the Moody's sample, the results for the Standard and Poor's sample are sensitive to the inclusion of potential cases of default and to the lag structure used (that is, the predictive performance was much worse with longer lead times for the ratings).

TABLE 8. Do Changes in Sovereign Credit Ratings Predict Currency Crises or Default? Institutional Investor Ratings (2,195 biannual observations)

	Coefficient	Standard error	Significance level	Pseudo R ²
<i>Currency crises</i>				
Kaminsky and Reinhart definition	-0.435	0.540	0.072	0.005
Frankel and Rose definition	-0.288	0.015	0.059	0.007
Defaults	-0.214	0.015	0.063	0.011
Defaults and potential defaults	-0.161	0.021	0.141	0.008

Note: Estimation method is probit with robust standard errors. Dependent variable is a dummy variable for currency crises. Independent variable is the 12-month change in the sovereign credit rating one year earlier.

Source: Author's calculations.

in the signals approach to gauge an indicator's ability to predict crises: the noise-to-signal ratio, the percentage of crises accurately called, and the marginal predictive power (the difference between the conditional and unconditional probabilities). The basic story that emerges is that the Institutional Investor credit ratings perform much worse in predicting both currency and banking crises than do the better indicators of economic fundamentals (table 11). For the credit ratings the noise-to-signal ratio is higher than one for both types of crises, suggesting a similar incidence of good signals and false alarms. Hence, not surprisingly, the marginal contribution to predicting a crisis is small relative to that of the top indicators; for banking crises the marginal contribution is nil. Moreover, the credit ratings call a much smaller percentage of crises than do the top indicators. Indeed, the Institutional Investor ratings compare unfavorably with even the worst indicators (see Goldstein and others 2000 for details). The results for the Institutional Investor ratings for the larger sample considered here are even worse than those shown in Goldstein and others (2000).

TABLE 9. Do Changes in Sovereign Credit Ratings Predict Currency Crises or Default? Moody's Ratings (4,774 monthly observations)

	Coefficient	Standard error	Significance level	Pseudo R ²
<i>Currency crises</i>				
Kaminsky and Reinhart definition	-0.217	0.761	0.412	0.001
Frankel and Rose definition	0.014	1.582	0.975	0.000
Defaults	-0.197	0.102	0.048	0.010
Defaults and potential defaults	-0.204	0.180	0.099	0.007

Note: Estimation method is probit with robust standard errors. Dependent variable is a dummy variable for currency crises. Independent variable is the 12-month change in the sovereign credit rating 1 year earlier.

Source: Author's calculations.

TABLE 10. Do Changes in Sovereign Credit Ratings Predict Currency Crises or Defaults in Emerging Market Economies? Standard and Poor's Ratings (3,742 monthly observations)

	Coefficient	Standard error	Significance level	Pseudo R ²
<i>Currency crises</i>				
Kaminsky and Reinhart definition	-0.080	0.091	0.772	0.001
Frankel and Rose definition	-0.014	0.076	0.721	0.001
Defaults	-0.120	0.076	0.054	0.011
Defaults and potential defaults	-0.356	0.170	0.117	0.007

Note: Estimation method is probit with robust standard errors. Dependent variable is a dummy variable for currency crises. Independent variable is the 12-month change in the sovereign credit rating 1 year earlier.

Source: Author's calculations.

Why Don't Sovereign Credit Ratings Do Better in Predicting Financial Distress?

Financial crises are generally difficult to predict—witness the poor performance of international interest rate spreads and currency forecasts.¹⁸ Moreover, though the overwhelming majority of defaults in the sample are associated with currency crises, the converse is not true. The results presented here offer a tentative (though partial) answer to the question of why sovereign credit ratings don't do better in predicting financial distress: rating agencies appear to have focused on a set of fundamentals that are not the most reliable in predicting currency crises. For example, they have given much weight to the debt-to-export ratio, yet this indicator has tended to be a poor predictor of financial stress (see table 11). As in Reinhart (2002), rating agencies have attached little weight to indicators of liquidity, currency misalignments, and asset price behavior, which are more reliable leading indicators of the kind of financial stress that can lead to both currency crises and default.

Detragiache and Spilimbergo (2001) present evidence that liquidity indicators, such as short-term debt and debt repayments due, perform particularly well in explaining subsequent debt servicing difficulties. Openness and measures of currency overvaluation score high marks in their study.

III. RESULTS AND IMPLICATIONS

This article has addressed several questions. What is the interaction between currency crises and defaults? The overwhelming majority of the defaults (84 percent) in emerging market economies in the sample are associated with currency crises. But the converse is not true—only slightly less than half the currency crises

18. See Kaminsky and others (1998) on the performance of interest rate spreads, and Goldfajn and Valdés (1998) on the performance of currency forecasts.

TABLE 11. Performance of Institutional Investor Sovereign Credit Ratings and Economic Fundamentals in Predicting Crises

Type of crisis and indicator	Noise-to-signal ratio	Percentage of crises accurately called	Difference between conditional and unconditional probability (percent)
<i>Currency crises</i>			
Institutional Investor sovereign rating	1.07	29	5.2
Average of the top 5 monthly indicators	0.45	70	19.1
Average of the top 3 annual indicators	0.49	36	15.4
Debt-to-export ratio	0.91	53	6.1
<i>Banking crises</i>			
Institutional Investor sovereign rating	1.62	22	0.9
Average of the top 5 monthly indicators	0.50	72	9.1
Average of the top 3 annual indicators	0.41	44	16.3
Debt-to-export ratio	1.04	56	0.9

Note: The top 5 monthly indicators for currency crises are: the real exchange rate, banking crises, stock returns, exports, and M2/reserves. As for currency crises, for banking crises the top 5 monthly indicators include the real exchange rate, stock returns, and exports, but output and the M2 multipliers completes the list. As for the annual indicators, the current account balance and a percent of investment and the overall budget deficit as a percent of GDP make the top 3 for both currency and banking crises. For currency crises the current account as a percent of GDP completes the top 3, whereas for banking crises short-term capital inflows as a percent of GDP rates highly.

Source: Author's calculations; Goldstein and others (2000).

in such economies are linked to default. For developed economies there is no evidence of any connection between currency crashes and default.

How do credit ratings behave following a currency crisis? Are there important differences between developed and emerging market economies in the behavior of ratings? There is evidence that sovereign credit ratings tend to be reactive, particularly those for emerging market economies. Both the probability and the size of a downgrade are significantly greater for emerging market economies. Taken together, these findings point to a procyclicality in the ratings. Perhaps a more instructive interpretation, however, is that currency crises in emerging market economies increase the likelihood of a default. The economic intuition is straightforward. Much of the debt of emerging market economies is denominated in dollars, so devaluations can have significant balance sheet effects. Moreover, most of the empirical evidence suggests that devaluations are contractionary. Calvo and Reinhart (2000), for example, ask how the differences between developed and emerging economies in access to international capital markets influence the outcomes of a currency crisis, particularly with respect to output. They present evidence that in emerging market economies devaluations

(or large depreciations) are contractionary, with the adjustments in the current and capital accounts far more acute and abrupt. Hence currency crises often become credit crises as sovereign credit ratings collapse following the currency collapse, and the economy loses access to international credit.

Do sovereign credit ratings systematically help predict currency crises and default? The results of the empirical tests presented here suggest that sovereign credit ratings systematically fail to predict currency crises but do considerably better in predicting defaults. Even so, ratings would not have predicted the nearly certain defaults that would have occurred in several recent crises had the international community not provided large-scale bailouts. These results appear to be robust across different definitions of crises, model specifications, and approaches.

Finally, why are sovereign ratings such poor predictors of currency crises? Financial crises are generally difficult to predict; international interest rate spreads and currency forecasts also perform poorly in predicting such crises. Yet ratings do better in predicting defaults than they do in predicting currency crises, although these results are less robust across different model specifications. Nonetheless, the results presented here suggest that rating agencies would do well to incorporate many indicators of vulnerability that have received high marks from the literature on the antecedents of currency crises. For example, rating agencies have given much weight to debt-to-export ratios, which have proved to be poor predictors of financial stress, but they have given little to indicators of liquidity, currency misalignments, and asset price behavior. Many of these indicators have been shown to be useful in predicting not only currency crises but also debt crises (Detragiache and Spilimbergo 2001). As noted, much can be learned about the antecedents and incidence of default from the literature on currency crises. This should not come as a surprise because after all, about one half of the currency crises are not associated with default, but an equal share of currency crises are linked in one way or another to a sovereign default incident.

REFERENCES

- Beers, David T., and Ashok Bhatia. 1999. "Sovereign Defaults: History." *Standard and Poor's Credit Week* December 22.
- Beim, David O., and Charles W. Calomiris. 2001. *Emerging Financial Markets*. New York: McGraw-Hill/Irwin.
- Calvo, Guillermo A., and Carmen M. Reinhart. 2000. "Fixing for Your Life." In Susan Collins and Dani Rodrik, eds., *Brookings Trade Forum 2000: Policy Challenges in the Next Millennium*. Washington, D.C.: Brookings Institution.
- Cantor, Richard, and Frank Packer. 1996a. "Determinants and Impact of Sovereign Credit Ratings." *Federal Reserve Bank of New York Economic Policy Review* (October):1-15.
- . 1996b. "Sovereign Risk Assessment and Agency Credit Ratings." *European Financial Management* 2:247-56.
- Detragiache, Enrica, and Antonio Spilimbergo. 2001. "Short-Term Debt and Crises." International Monetary Fund, Washington, D.C.

- Frankel, Jeffrey A., and Andrew K. Rose. 1996. "Exchange Rate Crises in Emerging Markets." *Journal of International Economics* 41(3/4):351–68.
- Goldfajn, Ilan, and Rodrigo O. Valdés. 1998. "Are Currency Crises Predictable?" *European Economic Review* 42(May):887–95.
- Goldstein, Morris, Graciela L. Kaminsky, and Carmen M. Reinhart. 2000. *Assessing Financial Vulnerability: An Early Warning System for Emerging Markets*. Washington, D.C.: Institute for International Economics.
- Kaminsky, Graciela L., and Carmen M. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems." *American Economic Review* 89(3):473–500.
- Kaminsky, Graciela L., Saul Lizondo, and Carmen M. Reinhart. 1998. "Leading Indicators of Currency Crises." *IMF Staff Papers* 45(1):1–48.
- Larraín, Guillermo, Helmut Reisen, and Julia von Maltzan. 1997. "Emerging Market Risk and Sovereign Credit Ratings." OECD Development Centre Technical Paper 124. Organisation for Economic Co-operation and Development, Paris.
- Lee, Suk Hun. 1993. "Are the Credit Ratings Assigned by Bankers Based on the Willingness of Borrowers to Repay?" *Journal of Development Economics* 40(April):349–59.
- Radelet, Steven, and Jeffrey Sachs. 1998. "The East Asian Financial Crisis: Diagnosis, Remedies, Prospects." *Brookings Papers on Economic Activity* 1.
- Reinhart, Carmen M. 2002. "Sovereign Credit Ratings before and after Financial Crises." In R. Levich, Giovanni Majnoni (ed.), and C. M. Reinhart (ed.), *Ratings, Rating Agencies and the Global Financial System*. Boston: Kluwer Academic Press.
- World Bank. Various years. *Global Development Finance*. Washington, D.C.: World Bank.

Financial Crises, Credit Ratings, and Bank Failures

Emerging Market Instability: Do Sovereign Ratings Affect Country Risk and Stock Returns?

Graciela Kaminsky and Sergio L. Schmukler

Changes in sovereign debt ratings and outlooks affect financial markets in emerging economies. They affect not only the instrument being rated (bonds) but also stocks. They directly impact the markets of the countries rated and generate cross-country contagion. The effects of rating and outlook changes are stronger during crises, in nontransparent economies, and in neighboring countries. Upgrades tend to take place during market rallies, whereas downgrades occur during downturns, providing support to the idea that credit rating agencies contribute to the instability in emerging financial markets.

Worldwide financial market instability has been the focus of attention in both academic and policy circles. Following the series of currency crashes in the past decade, most of the discussion has centered on balance of payments crises. This attention on crises is not going to fade any time soon, with the financial crashes in Argentina and Turkey in 2001 surely fueling an avid interest in crises well into the new millennium. But currency collapses are not the only crises to have attracted attention. The daily volatility of stock and bond markets during normal periods has also stirred interest, with, for example, the vagaries of the NASDAQ index in the United States making the daily headlines.

Many have argued that globalization is at the heart of this volatility, with highly diversified investors paying little attention to economic fundamentals and following the herd in the presence of asymmetric information.¹ Policies that can lead to moral hazard, including bailouts by both international institutions and governments, have also been blamed for financial volatility and financial excesses (see, for example, Dooley 1998, McKinnon and Pill 1997).

Graciela Kaminsky is with George Washington University. Her e-mail address is graciela@gwu.edu. Sergio Schmukler is with the Development Research Group at the World Bank. His e-mail address is sschmukler@worldbank.org. We are grateful to Eduardo Borensztein, François Bourguignon, Hali Edison, Cam Harvey, Richard Levich, Rick Mishkin, Carmen Reinhart, three anonymous referees, and two members of the *World Bank Economic Review* editorial board, as well as participants at the New York University and University of Maryland World Bank conferences and workshops for helpful comments and suggestions. We thank Gloria Alonso, Tatiana Didier, and Chris van Klaveren for excellent research assistance. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily represent the views of the World Bank.

1. See, for example, Calvo and Mendoza (2000). This argument has provided ammunition to those who have supported the reintroduction of capital controls, including Krugman (1998) and Stiglitz (2000).

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

The list of culprits does not stop here. Rating agencies have recently come under scrutiny as promoters of financial excesses. As Ferri and others (1999) suggest, their procyclical behavior (upgrading countries in good times and downgrading them in bad times) may have magnified the boom-bust pattern in stock markets.² Rating changes may also reveal new (private) information about a country, fueling rallies or downturns. This effect is likely to be stronger in emerging markets, where problems of asymmetric information and transparency are more severe. Changes in ratings may also act as a wake-up call, with upgrades or downgrades in one country affecting other, similar economies.

Even if rating agencies do not behave procyclically, their announcements may still trigger market jitters because many institutional investors can hold only investment-grade instruments. Downgrading (or upgrading) sovereign debt below (or above) investment grade may thus have a drastic impact on prices because these rating changes can affect the pool of investors. These effects are not confined to the pool of investors acquiring sovereign debt. When a credit rating agency downgrades a country's sovereign debt, all debt instruments in that country may have to be downgraded accordingly because of the sovereign ceiling doctrine. Commercial banks downgraded to subinvestment grade will find it costly to issue internationally recognized letters of credit for domestic exporters and importers, isolating the country from international capital markets. Downgrading corporate debt to subinvestment grade means that firms will face difficulties issuing debt on international capital markets.

Research on the effects of rating changes flourished in the 1990s. Most of this work focused on the effects of ratings on the instruments being rated or on the instruments of the institutions being rated. Cantor and Packer (1996), Larrain and others (1997), and Reisen and von Maztlan (1999), for example, examine the effects of sovereign ratings on emerging market bond yield spreads. Other researchers have focused on ratings of banks and nonfinancial firms. Hand and others (1992) estimate the effects of ratings of corporate firms on the securities they issue. Using bank-level data from emerging markets, Richards and Deddouche (1999) examine the impact of bank ratings on bank stock prices.

Research has not examined whether changes in ratings of assets from one country trigger contagious fluctuations in other countries, and it has largely neglected whether changes in ratings of one type of security affect other asset markets.³ These two possible spillover effects of credit ratings are important to analyze for several reasons. First, cross-country contagion effects can be large, as spillover effects of the Russian default on industrial and developing econo-

2. Mora (2001) extends these results. She agrees that ratings are procyclical but questions the notion that changes in ratings increased the cost of borrowing and decreased the supply of international credit during the East Asian crisis.

3. To our knowledge, the only article that examines the contagious role of credit ratings is Kaminsky and Schmukler (1999). Erb and others (1996a, 1996b, 1996c) study how the effect of changes in ratings of one type of security affect other asset markets, studying the link between expected stock returns and future fixed-income returns with different measures of country risk.

mies showed.⁴ Rating agencies may contribute to this comovement in financial markets around the world. Second, news about one type of security can affect yields of other securities, through various channels. For example, stock markets can be adversely affected by the downgrading of sovereign bonds because governments may raise taxes on firms (reducing firms' future stream of profits) to neutralize the adverse budget effect of higher interest rates on government bonds triggered by the downgrade. These cross-asset effects can be large, heightening financial instability.

Another line of research on emerging market instability has focused largely on the effects of changes in monetary policy in financial centers. The results have been conflicting. Eichengreen and Mody (1998) and Kamin and von Kleist (1999) find that U.S. interest rate shocks do not affect sovereign bond spreads, whereas Herrera and Perry (2000) find that they do. The Eichengreen and Mody (1998) and Kamin and von Kleist (1999) studies do not include episodes of crises, and the Herrera and Perry (2000) work does. These conflicting results may be reconciled if economic fragility makes countries more sensitive to changes in international financial markets. The degree of economic fragility can be captured by country ratings. Thus, we are able to link the research on the effects of monetary shocks in financial centers on emerging market instability to the research on credit ratings.⁵

This article complements earlier research on rating agencies by examining the cross-country and cross-security spillover effects of rating changes. It contributes to the literature on contagion and international transmission of shocks by examining the effect of domestic vulnerability, as measured by the ratings of credit agencies, on the extent of international spillovers.

The article is organized as follows. Section I describes the institutional features of rating agencies. Section II presents the data. Section III describes the methodology. Section IV discusses the results. Section V summarizes the conclusions.

I. INSTITUTIONAL FEATURES OF RATING AGENCIES

Three major international agencies—Moody's, Standard and Poor's (S&P), and Fitch-IBCA—rate debt.⁶ These agencies assign ratings to different types of borrowers and financial instruments. We study sovereign ratings (also known as country ratings), the ratings of both domestic and foreign currency-denominated sovereign debt.

4. The word *contagion* is used in a broad sense to denote cross-country spillover effects, regardless of the nature of the shock. For alternative definitions and related articles, see <http://www.worldbank.org/contagion>.

5. Another factor that can influence the transmission of international shocks is the exchange rate regime. Frankel and others (2000), for example, find that world interest rates shocks have a stronger effect on countries under pegs.

6. Another important agency is Institutional Investors. Unlike the other three agencies, Institutional Investors reports ratings only twice a year at a predetermined date. It also tends to change its ratings more often than the other agencies. Because of these differences, we excluded Institutional Investors from the sample.

Rating agencies assess the capacity of sovereign borrowers to service their debt. Each of the three agencies has its own rating scale (see appendix table 1). Moody's scale, for example, ranges from Aaa to C. Rating agencies also provide an outlook, or watchlist, that includes prospective changes in ratings. The outlook is typically positive, stable, or negative. A positive (negative) outlook means that a rating may be revised upward (downward).

Moody's, S&P, and Fitch-IBCA upgrade or downgrade particular countries almost simultaneously (figure 1). All three agencies downgraded the East Asian countries immediately following the start of the crisis in July 1997; all three simultaneously upgraded the same countries once the crisis faded.⁷

The number of upgrades and downgrades rose after the Mexican crisis (figure 2). Downgrades increased considerably after the devaluation of the Thai baht, the Korean crisis, and the Russian default, with a peak of 25 downgrades in December 1997. After November 1998 many countries started to be upgraded, but downgrades were also announced in the midst of the Brazilian crisis in January 1999.

A large proportion of changes in outlook are followed by a change in rating (table 1). Between 1990 and 2000, 78 percent of changes in S&P outlook were followed by changes in ratings. Rating changes followed outlook changes 69 percent of the time at Moody's and 50 percent of the time at Fitch-IBCA.

The time interval between changes in outlook and changes in rating varies across agencies. Most of the changes in rating occurred within two months for Moody's and Fitch-IBCA. For S&P most of the upgrades took place five or more months after the change in outlook was announced.

II. DATA

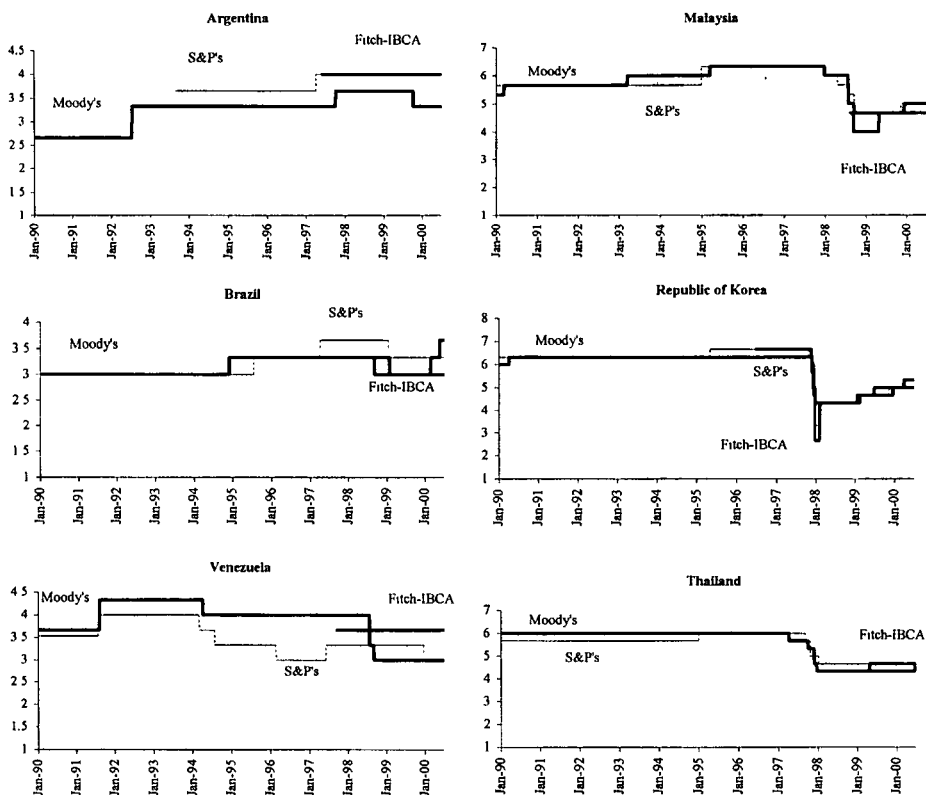
We examine data from 16 emerging markets: Argentina, Brazil, Chile, Colombia, Indonesia, Malaysia, Mexico, Peru, the Philippines, Poland, the Republic of Korea, the Russian Federation, Taiwan (China), Thailand, Turkey, and Venezuela. The data cover the period January 1990–June 2000. We chose countries in the three regions (East Asia, Eastern Europe, and Latin America) that suffered crises and contagion during the 1990s and for which data were available. (Appendix table 2 reports the time periods for which data were available for each country.)

The sample includes 244 changes in ratings and outlooks, 99 upgrades, and 145 downgrades (tables 2 and 3). Most of these changes were changes in ratings rather than changes in outlooks. Countries with currency collapses during the 1990s—such as Brazil, Indonesia, Malaysia, the Republic of Korea, the Russian Federation, and Thailand—were frequently reevaluated by rating agencies.

Sovereign bond yield spreads were obtained from JP Morgan's Emerging Markets Bond Index (EMBI). The yield spread index for each country is either the EMBI or the EMBI+, based on availability. The two indexes track foreign currency-denominated government bond yields for several emerging market econo-

7. For a detailed study of how ratings are changed, see Cruces (2001).

FIGURE 1. Ratings of Foreign-Currency Sovereign Debt for Selected Countries



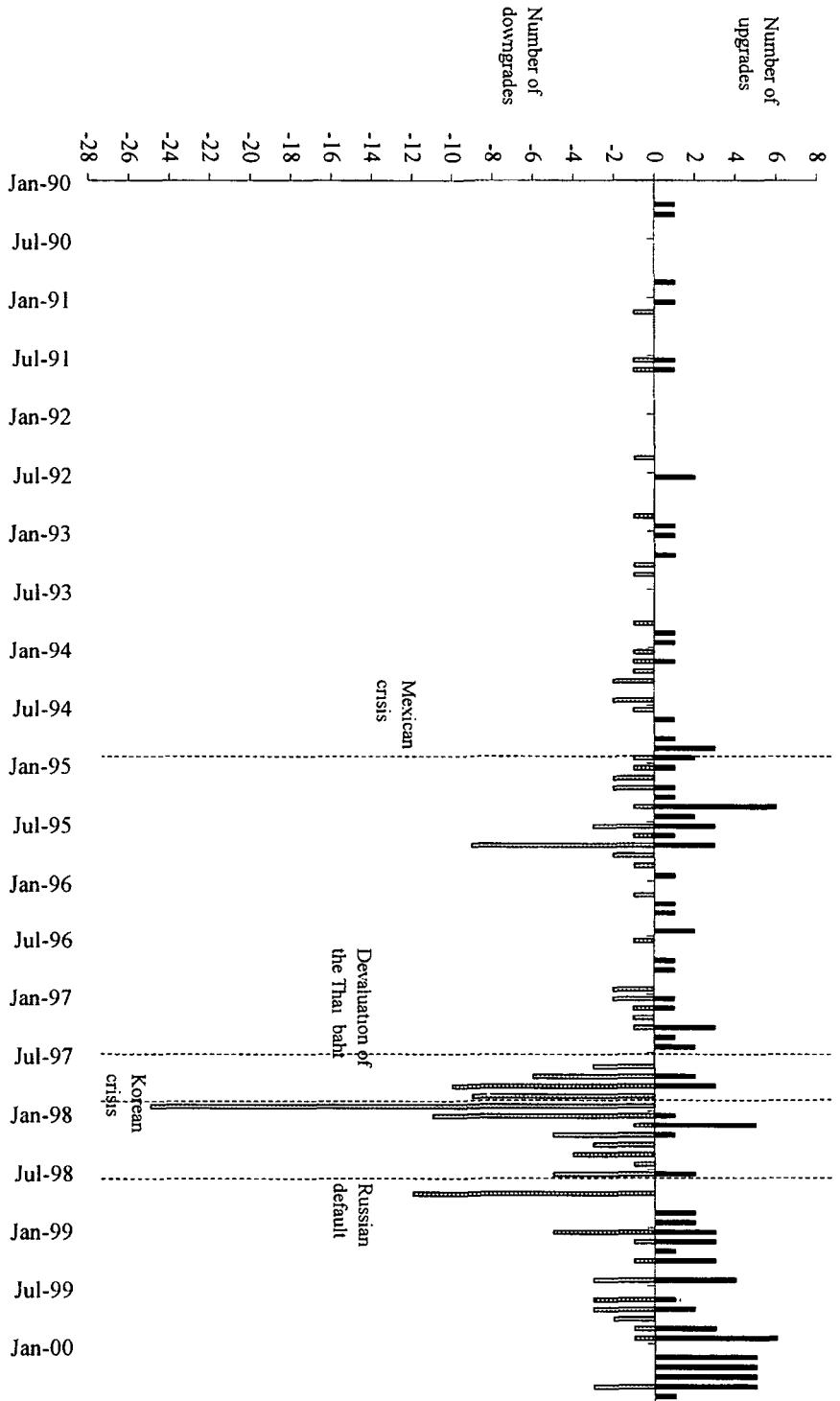
Source: Bloomberg.

mies and compare them with the yields of benchmark instruments issued by industrial countries. The securities included in the EMBI index are Brady bonds, which are traded internationally in highly liquid markets. The EMBI+ is a more comprehensive index and includes benchmark Eurobonds, loans, and Argentine domestic debt. EMBI and EMBI+ (henceforth EMBI) spreads are commonly used as measures of country premia, country risk, or default risk. When the probability of a sovereign default increases, bond prices decrease and yield spreads increase.

Data on stock prices, U.S. interest rates, and credit ratings come from Bloomberg and Datastream. Stock market price indexes for each country are measured in U.S. dollars to be able to compare returns across countries in the same unit of account. Returns in dollars are the ones relevant for international investors. The U.S. interest rate is the one-month interbank offer rate.

Daily changes (in absolute values) in bond and stock markets oscillate about 2.5 percentage points for sovereign spreads and about 1.6 percentage points for stock prices (table 4). The number of observations is high (about 11,000 for bond spreads and 22,000 for stock prices).

FIGURE 2. Changes in Ratings and Outlooks 1990–2000



Source: Authors' calculations.

TABLE 1. Number of Changes in Ratings Following Change in Outlook

Items	Moody's		s&P		Fitch-IBCA	
	Upgrades	Downgrades	Upgrades	Downgrades	Upgrades	Downgrades
Total number of changes in outlooks	13	16	13	23	5	3
Total number of changes in ratings	9	11	13	15	3	1
<i>Within</i>						
1 month	0	2	1	4	1	1
2 months	6	7	0	4	1	0
3 months	1	1	0	4	1	0
4 months	2	1	1	1	0	0
More than 4 months	0	0	11	2	0	0

Source: Authors' calculations.

TABLE 2. Number of Upgrades and Downgrades, by Rating Agency

Agency	Total changes	Ratings		Outlooks	
		Upgrades	Downgrades	Upgrades	Downgrades
<i>Moody's</i>	77	19	29	13	16
Foreign-currency debt	37	14	23		
Domestic-currency debt	11	5	6		
<i>S&P</i>	112	28	48	13	23
Foreign-currency debt	45	19	26		
Domestic-currency debt	31	9	22		
<i>Fitch-IBCA</i>	55	21	26	5	3
Foreign-currency debt	30	15	15		
Domestic-currency debt	17	6	11		
Total	244	68	103	31	42

Source: Authors' calculations.

III. METHODOLOGY

To study the effects of ratings and outlooks on financial markets, we estimate panel regressions and perform event studies. The panel regressions focus on the immediate response of financial markets to rating and outlook changes. The event studies examine the dynamic response of financial markets around the time of important events.

Panel Regressions

The panel estimations study the daily reactions of country premia and stock returns to changes in ratings, outlooks, and U.S. interest rates. The fact that we use daily data does not allow us to control for country fundamentals, which are typically reported on a monthly or quarterly basis, but we do control for past changes of the explanatory variables. We use only one lag, because additional lags appear to be insignificant.

We estimate different regressions for both country premia and stock prices. We start with a benchmark regression, which we then modify to examine to test various hypotheses:

$$(1) \quad \Delta Y_{i,t} = \alpha + \delta \Delta Y_{i,t-1} + \beta \Delta R_t + \gamma \Delta i_t^{US} + \varepsilon_{i,t}$$

such that $i = 1, \dots, N$ and $t = 1, \dots, T$. $\Delta Y_{i,t}$ represents the log change in spreads and the log change in stock market prices. The subindexes i and t stand for country and time. ΔR_t stands for the change in ratings and outlooks. It is equal to 1 (–1) if there is an upgrade (downgrade) in rating or outlook at time t by any agency of any type of debt (denominated in foreign or domestic currency) from any country in the sample; otherwise it is equal to zero. Δi_t^{US} stands for the change in U.S. interest rates; strictly speaking, the interest rate is $100 \times \log(1 + i_t^{US})$.

TABLE 3. Number of Upgrades and Downgrades, by Country

Agency	Total changes	Ratings		Outlooks	
		Upgrades	Downgrades	Upgrades	Downgrades
Argentina	14	4	3	3	4
Brazil	19	9	6	3	1
Chile	5	3	1	1	0
Colombia	11	0	7	1	3
Indonesia	22	1	20	1	0
Korea, Rep. of	40	14	17	2	7
Malaysia	28	5	12	3	8
Mexico	17	6	5	5	1
Peru	2	1	0	1	0
Philippines	8	4	0	2	2
Poland	10	9	0	1	0
Russian Federation	26	7	15	2	2
Taiwan (China)	0	0	0	0	0
Thailand	22	2	11	1	8
Turkey	10	1	3	3	3
Venezuela	10	2	3	2	3
Total	244	68	103	31	42

Source: Authors' calculations.

The second regression is

$$(2) \quad \Delta Y_{i,t} = \alpha + \delta \Delta Y_{i,t-1} + \beta^r \Delta R_{i,t}^r + \beta^o \Delta R_{i,t}^o + \gamma \Delta i_t^{US} + \varepsilon_{i,t}.$$

The variable $\Delta R_{i,t}^r$ is equal to 1 (-1) if there is a change in rating (upgrade or downgrade) at time t by any agency on any type of debt from any country in the sample. The variable is equal to zero otherwise. The variable $\Delta R_{i,t}^o$ is similar to $\Delta R_{i,t}^r$ but takes the value 1 (-1) when there is a change in outlook (upgrade or downgrade) in any country in the sample. This specification tries to disentangle the effects of changes in ratings from those of changes in outlooks.

TABLE 4. Sovereign Yield Spreads and Stock Prices

	Mean	Median	Standard deviation	Minimum	Maximum	Number of observations
Log change in EMBI spreads	-0.0004	-0.0012	0.0379	-0.4986	0.4652	11,122
Log change in absolute value of EMBI spreads	0.0243	0.0160	0.0291	0.0000	0.4986	11,122
Log change in stock prices	-0.0001	0.0000	0.0257	-0.3947	0.3171	21,788
Log change in absolute value of stock prices	0.0158	0.0095	0.0203	0.0000	0.3947	21,788

Source: Authors' calculations.

The third regression is

$$(3) \quad \Delta Y_{i,t} = \alpha + \delta \Delta Y_{i,t-1} + \beta^i \Delta R_{i,t} + \beta^r \Delta R_{r,t} + \beta^{nr} \Delta R_{nr,t} + \gamma \Delta i_t^{US} + \varepsilon_{i,t}.$$

$\Delta R_{i,t}$ is equal to 1 (–1) if there is an upgrade (downgrade) in rating or outlook at time t by any agency on any type of debt from country i . It is equal to zero otherwise. The variable $\Delta R_{r,t}$ is similar to $\Delta R_{i,t}$ but takes the value 1 (–1) when there is an upgrade (downgrade) in country r for $r \neq i$. The variable r represents a country that belongs to the same geographic region (East Asia, Eastern Europe, or Latin America) as i . The variable $\Delta R_{nr,t}$ is used for countries in other regions; nr represents a country that belongs to a geographic region other than i . The last two variables are related to changes in foreign country ratings and outlooks. This specification tries to examine whether there is a contagion effect of credit ratings and whether these effects are of a regional or nonregional nature.

We estimate different versions of the third regression. The first divides the sample into crisis and noncrisis periods to test whether markets react more strongly to changes in domestic and foreign ratings in good and bad times. This difference can arise in models with multiple equilibria. In this type of model, a signal can coordinate investors' expectations, shifting them from a good to a bad equilibrium in both the domestic economy and other economies (see, for example, Masson 1998). In our case the signal can be provided by a rating downgrade.

A second version of the third regression divides the sample between transparent and nontransparent countries, based on the data used by Mehrez and Kaufmann (2000). If rating changes provide any information to markets, they should do so more in nontransparent countries than in transparent ones.

As a fourth regression, we estimate

$$(4) \quad \Delta Y_{i,t} = \alpha + \delta \Delta Y_{i,t-1} + \beta^i \Delta R_{i,t} + \beta^r \Delta R_{r,t} + \beta^{nr} \Delta R_{nr,t} + \gamma^h h_{i,t} \Delta i_t^{US} + \gamma^l l_{i,t} \Delta i_t^{US} + \varepsilon_{i,t}.$$

This regression allows for different responses (γ^h and γ^l) of the dependent variable to changes in U.S. interest rates depending on the state of vulnerability of the domestic economy, as captured by the assessment of rating agencies. In particular, we divide the observations into two groups, those with above-average and below-average ratings relative to the mean rating of all the countries in our sample. Those observations are divided using two dummy variables, $h_{i,t}$ and $l_{i,t}$, which capture high and low ratings.

We estimate all of these specifications using pooled panels in which the error term $\varepsilon_{i,t}$ can be characterized by an independently distributed random variable with mean zero and variance $\sigma_{i,t}^2$. We estimate the equations using least squares, allowing for heteroscedastic residuals.

The least squares specifications assume a zero correlation between the error term and the explanatory variables. This correlation may arise if an explanatory variable is endogenously determined. However, we do not expect changes in U.S. interest rates or changes in ratings to respond to contemporaneous daily changes

in emerging market spreads or stock prices. Still, a correlation between the lagged dependent variable and the error term is possible. This correlation can arise if, for example, the true original model were in levels and then first differenced. In that case, the error term in our equations would be in first differences and correlated with the lagged endogenous variable by construction. To correct for potential biased coefficients, we estimate the more complete specification, equation (4), using instrumental variables or two-stage least squares. As instruments, we use lagged values of the lagged dependent variable, as proposed by Anderson and Hsiao (1982).

We expect certain signs for the estimated coefficients. If changes in ratings convey new information to market participants, we expect $\hat{\beta} < 0$ in the regressions for country premia; that is, rating upgrades (downgrades) lead to decreases (increases) in bond spreads. Analogously, in the regressions for stock returns, we expect $\hat{\beta} > 0$ for the coefficients of both ratings and outlooks and those for domestic- and foreign-country ratings.

If increases in U.S. rates lead to higher country premia, we expect $\hat{\gamma} > 0$ in the equation for country premia. As Kamin and von Kleist (1999) argue, there are different channels through which changes in U.S. interest rates might positively affect country premia. First, if there is a positive probability that a government will not pay its debt, increases in U.S. rates will prompt a larger rise in the interest rate of the government's debt. These higher rates compensate for the probability of no repayment. Second, increases in U.S. interest rates increase the burden of the debt, decreasing a country's repayment capacity. Third, increases in U.S. rates can decrease investors' appetite for risk, reducing the demand for risky assets from emerging countries and thereby increasing the country premia.

A similar argument can be used to explain stock returns. Governments can levy taxes on corporations if they face higher debt payments. Therefore, we expect that U.S. interest rates negatively affect stock returns ($\hat{\gamma} < 0$ in the equation for stock returns).

We expect countries with healthy economies to be less affected by changes in U.S. rates ($|\hat{\gamma}^l| > |\hat{\gamma}^h|$), for several reasons. First, given that higher ratings mean a lower probability of default, changes in U.S. interest rates will have a greater effect on spreads in countries with lower ratings. Second, countries with higher ratings tend to have lower levels of debt, so the burden of the debt will increase less in countries with high ratings when U.S. interest rates increase. Third, if there is a flight to quality when U.S. rates increase, sovereign yield spreads of countries with lower ratings should react more strongly. Similar arguments can be made for the quantitative responses of stock returns to changes in U.S. interest rates in more vulnerable and less vulnerable countries.

The coefficient on the lagged dependent variable $\hat{\delta}$ is expected to be positive if returns are autocorrelated. In efficient capital markets $\hat{\delta}$ should be zero, because returns are unpredictable. However, recent research has shown that returns

are to some degree predictable and are positively autocorrelated (see, for example, Richards 1996, Rouwenhorst 1998, and Kaminsky and others 2000).⁸

Event Studies

The regressions presented focus on the contemporaneous effect of ratings on bond spreads and stock returns. To capture the dynamic effects around the time of changes in outlooks or ratings, we use event studies. Event studies can provide evidence on whether rating agencies act procyclically, downgrading countries during bad times and upgrading them during good times. They can also help determine whether the actions of rating agencies have sustained or merely transitory effects on financial markets.

The event studies examine the evolution of country premia (sovereign bond yield spreads) and stock market spreads (domestic stock market prices relative to the S&P 500 index) during a ± 10 -day window around an upgrade or downgrade of a rating or outlook. We use stock market spreads because we want to measure the evolution of local stock prices relative to a benchmark.

The event study methodology allows us to study the effect of an upgrade or downgrade on the evolution of spreads around the event. Of course, other events that affect spreads may take place at the same time. We do not control for those factors and assume that on average there is no particular bias in the event studies. That is, we expect that other factors influence spreads both positively and negatively in a random way. If, however, rating changes are serially correlated, the event studies will be biased. To control for this effect, we work with “clean events,” that is, upgrades and downgrades that do not overlap during the 10-day window. In this manner, we ensure that we are studying the effect of only one upgrade or downgrade in each event.

IV. RESULTS

We first examine the contemporaneous impact of changes in ratings and outlooks. We then we report on the dynamic aspects of market responses to these changes.

Panel Regressions

The panel regression results for EMBI spreads show that the coefficient for the lagged dependent variable is positive and statistically significant (table 5). As found in previous research, this result suggests that returns are somewhat predictable, so that trading strategies (such as momentum trading) may be profitable. This result holds in several specifications.

The coefficient for rating and outlook changes (both domestic and foreign) is negative and statistically significant. The negative sign of the coefficient is as

8. For other alternative specifications, including those that look at ratings on domestic and foreign currency-denominated debt, see Kaminsky and Schmukler (2001).

TABLE 5. Panel-Regression Estimates
(dependent variable: log change in EMBI spreads)

Explanatory variable	Alternative specifications								Independent variable 9
	1	2	3	Crisis periods 4	Noncrisis periods 5	Transparent countries 6	Nontransparent countries 7	8	
Lagged dependent variable	0.039* (1.863)	0.039* (1.829)	0.039* (1.844)	0.049 (1.167)	0.017 (1.080)	0.051 (1.287)	0.032 (1.307)	0.040* (1.874)	-0.510 (1.204)
<i>Changes in ratings and outlooks</i>									
All countries (ratings and outlooks)	-0.006*** (4.673)								
All countries (ratings)		-0.004*** (3.011)							
All countries (outlooks)		-0.007*** (3.857)							
Domestic country (ratings and outlooks)			-0.021*** (3.447)	-0.028* (1.889)	-0.015*** (2.654)	-0.022** (2.504)	-0.020** (2.557)	-0.021*** (3.448)	-0.020*** (3.040)
Regional countries (ratings and outlooks)			-0.007*** (3.129)	-0.028*** (4.047)	0.001 (0.355)	-0.006** (2.226)	-0.007** (2.186)	-0.007*** (3.126)	-0.010*** (2.770)
Nonregional countries (ratings and outlooks)			-0.004*** (2.915)	-0.007** (2.231)	-0.001 (0.945)	-0.002 (0.999)	-0.005*** (2.883)	-0.004*** (2.755)	-0.006** (2.323)
<i>Changes in U.S. interest rates</i>									
Change in U.S. interest rates	0.029*** (2.714)	0.029*** (2.730)	0.028*** (2.699)	0.168*** (3.378)	0.015 (1.541)	0.028* (1.791)	0.029** (2.037)		
Change in U.S. interest rates: high ratings*								0.023 (1.534)	0.043 (1.613)
Change in U.S. interest rates: low ratings*								0.034** (2.328)	0.067** (2.445)
Number of observations	11,122	11,122	11,122	1,948	9,206	4,481	6,641	10,923	10,408
R-squared	0.005	0.005	0.006	0.021	0.002	0.007	0.006	0.006	0.006

*Significant at the 10 percent level.

**Significant at the 5 percent level.

***Significant at the 1 percent level.

Note: Table reports panel estimates with robust standard errors, using the White correction for heteroscedasticity. A constant is estimated but not reported. The instrumental variables estimation (specification 9) uses a third lag of the dependent variable as an instrument. The crisis periods are from December 1, 1994, to March 30, 1995; July 1, 1997, to January 30, 1998; August 1 to October 30, 1998; and January 1 to February 28, 1999. Countries are classified transparent or nontransparent countries based on the Mehrez and Kaufmann (2000) data. Countries with ratings above the median (Brazil, Chile, Malaysia, Mexico, Peru, and Taiwan [China]) are considered transparent. Countries with ratings below the median (Argentina, Colombia, Indonesia, the Philippines, Poland, the Republic of Korea, the Russian Federation, Thailand, Turkey, and Venezuela) are considered nontransparent. *t*-statistics are in parentheses.

Source: Authors' calculations.

expected: a rating or outlook upgrade (downgrade) decreases (increases) bond spreads. This result holds in all specifications. Though significant, the coefficient is small, with spreads changing about 0.6 percentage points following a rating or outlook announcement. The average absolute change of spreads in our sample is about 2 percentage points.

The coefficient on U.S. interest rates is statistically significant. The sign is positive, as expected. A hike in U.S. interest rates increases bond spreads. That is, higher U.S. rates increase domestic interest rates more than proportionally to compensate for the higher expected default risk, among other things. This result holds in almost all specifications.

We investigate separately whether changes in ratings have different effects from changes in outlooks, finding that both coefficients are statistically significant and with a negative sign (column 2). The coefficient on outlooks is significantly larger than the coefficient on ratings, suggesting that investors may anticipate rating changes, perhaps because countries are put on a watchlist before being downgraded.

We separate the effect of domestic- and foreign-country changes in ratings and outlooks (column 3). We use both changes in ratings and outlooks in the same variable to avoid studying the effects of a small number of changes, because there are relatively few changes in outlooks. We find that changes in ratings and outlooks have substantially stronger effects on the country being assessed than on other countries, with own-country effects averaging 2.1 percentage points. Still, rating and outlook changes do contribute to contagion, with ratings of foreign-country debt spilling over to domestic financial markets. These spillover effects range from 0.4 to 0.7 percentage points.

The results also provide evidence on a widely discussed issue in the contagion literature: whether contagion is regional or global. The crises of the 1990s and the speed with which a crisis in one country was transmitted throughout the region and even to other regions have spawned a still growing literature on contagion. Much of the research has centered on the role of financial links versus trade links.⁹ But there is a growing interest in the geography of contagion. The Tequila crisis was confined largely to Latin America, and the crisis in Thailand spread mostly to Asian economies.¹⁰

We examine whether these regional spillovers are also present following rating and outlook changes. Our results shows that regional effects are stronger than those from nonregional countries: Within-region upgrades and downgrades led to an average increase in yields of 0.7 percentage points, whereas nonregional

9. Kaminsky and Reinhart (2000b) and Kaminsky and others (2000) point to the role of financial links and focus on the behavior of international banks and mutual funds. In contrast, Corsetti and others (2000) focus on the role of trade links.

10. Kaminsky and Reinhart (2000a) analyze why some crises become systemic whereas others are confined within national borders or are at most regional.

upgrades and downgrades triggered an average change in spreads of about 0.4 percentage points.

We use the last specification to examine the effect of rating and outlook changes during crisis and noncrisis periods (columns 4 and 5). Our results reveal that these changes have stronger effects during crises, with changes in domestic ratings of 2.8 percentage points during crises and 1.5 during noncrisis periods. Moreover, some of the variables are significant only in crises. Cross-country spillover effects are statistically significant only during crises, a result that is consistent with the evidence on contagion. Changes in U.S. interest rates are significant only during crises.

Our results also show that rating and outlook changes have different effects in transparent and nontransparent countries (columns 6 and 7). Nontransparent countries are affected by nonregional ratings and outlooks, whereas transparent countries are not.

We are also interested in the effect of changes in international financial markets on emerging economies. This topic has generated many articles following Calvo and others (1993), who brought to the limelight the close relation between the capital inflows episode to emerging markets during the early 1990s to monetary policy in the United States. Many have focused on the relation between capital flows or foreign exchange reserves and interest rates in financial centers. Some have focused on the links between returns in emerging markets and returns in financial centers. Others have focused on the effects of interest rate hikes on interest rates and bond spreads. These links were strong in the early 1990s, weakened somewhat in the mid-1990s, and reappeared in the late 1990s.

The changing relation between financial markets in emerging economies and financial centers is particularly clear in the research on the determinants of country premia, with some articles finding a positive relation and others finding no significant relation. Although understanding the determinants of this time-varying relation is beyond the scope of this article, we examine whether hikes in interest rates in financial centers are transmitted more strongly to vulnerable economies. We divide the sample into two equal parts based on sovereign credit ratings. The point estimates (column 8 of table 5) indicate that fluctuations in U.S. interest rates have about a 50 percent greater effect on more vulnerable economies (those with worse ratings) than on less vulnerable economies. Interestingly, countries with higher credit ratings are not affected in a statistically significant way by changes in U.S. interest rates, but economies with lower credit ratings are.

We use instrumental variables to try to control for potentially biased estimates (last column of table 5). Using the same specification reported in column 8, we find that the results on credit rating and outlook changes and those on changes in U.S. interest rates hold when estimating the equation with two-stage least squares.

The results of estimations of the same specifications using stock market returns as the dependent variable are very similar to those obtained for EMBI spreads, with some interesting differences (table 6). First, stock returns display more persistence than EMBI spreads, as shown in the estimates of the lagged dependent variable. Second, the magnitude of the point estimates for the other variables tends to be smaller, suggesting that ratings have stronger effects on the prices of the instrument they are assessing. Third, domestic ratings are significant only in nontransparent economies, suggesting that rating agencies do provide valuable signals in countries in which information is lacking.

Event Studies

In the panel estimations, we focused on the instantaneous response of bond and stock markets in emerging economies to changes in credit ratings and outlooks. To capture whether these changes persistently affect investors' mood, we rely on event study methods commonly used in the finance literature. The event study methodology also allows us to examine the claim that rating agencies behave procyclically, upgrading countries in good times and downgrading them during crises.

We examine the behavior of asset markets around the time of rating and outlook changes (10-day windows before and after changes). We look only at clean events, examining 103 domestic-country rating and outlook changes (56 upgrades and 47 downgrades) (table 7). Including foreign-country changes increases the number of changes to 653. Standard event study methodology requires linking rating events to abnormal returns. For this reason, we base the event study on the yield spreads between sovereign government debt and the benchmark instruments from industrial countries. For stocks we use the dollar "stock spreads" between emerging markets stock prices and the S&P 500 U.S. stock market index.

The evidence supports the hypothesis that rating agencies may have exacerbated the boom-bust pattern in emerging markets (figures 3 and 4). Upgrades tend to occur when markets are rallying and downgrades when emerging markets are collapsing. Bond spreads, for example, rose by as much as 7 percent in the 10 days before downgrades, and stock market spreads increased by as much as 4 percent. In both cases, the effect is statistically significant. Rallies were more muted in the days leading up to rating upgrades, with bond spreads barely declining and stock spreads increasing about 2 percent.

Similar results hold for changes in foreign-country ratings and outlooks. The results suggest that upgrades of other countries' debt trigger important declines in yield spreads and substantial increases in stock market prices. Likewise, foreign downgrades are followed by increases in EMBI spreads and declines in the domestic stock market relative to that of the U.S. market. As expected, the change in spreads is smaller in this case; domestic-country rating and outlook changes have larger effects on financial markets than foreign-country changes. Relative

TABLE 6. Panel Regression Estimates
(dependent variable: log change in stock prices)

Explanatory variable	Alternative specifications								
	1	2	3	Crisis periods 4	Noncrisis periods 5	Transparent countries 6	Nontransparent countries 7	8	Independent variable 9
Lagged dependent variable	0.088*** (4.458)	0.088*** (4.448)	0.087*** (4.406)	0.098*** (3.047)	0.074*** (2.946)	0.022 (0.557)	0.126*** (6.127)	0.088*** (4.399)	0.333 (1.565)
<i>Changes in ratings and outlooks</i>									
All countries (ratings and outlooks)	0.003*** (5.119)								
All countries (ratings)		0.002*** (2.970)							
All countries (outlooks)		0.004*** (4.882)							
Domestic country (ratings and outlooks)			0.009*** (2.837)	0.017** (2.323)	0.002 (1.010)	0.008 (1.310)	0.009*** (2.582)	0.009*** (2.836)	0.006 (1.517)
Regional countries (ratings and outlooks)			0.004*** (4.029)	0.010*** (4.187)	0.000 (0.295)	0.004*** (2.954)	0.004*** (2.803)	0.004*** (4.051)	0.003*** (2.713)
Nonregional countries (ratings and outlooks)			0.001** (2.382)	0.001 (0.950)	0.001 (1.495)	0.000 (0.202)	0.002*** (2.682)	0.002** (2.444)	0.001* (1.884)
<i>Changes in U.S. interest rates</i>									
Change in U.S. interest rates	-0.009** (2.481)	-0.009** (2.521)	-0.009** (2.481)	-0.005 (0.232)	-0.011*** (3.030)	-0.011** (1.969)	-0.008 (1.638)		
Change in U.S. interest rates: high ratings*								-0.007 (1.218)	-0.006 (0.920)
Change in U.S. interest rates: low ratings*								-0.012** (2.184)	-0.011* (1.750)
Number of observations	21,788	21,788	21,788	4,330	17,521	8,898	12,890	21,247	20,508
R-squared	0.010	0.010	0.010	0.019	0.006	0.002	0.020	0.011	0.011

*Significant at the 10 percent level.

**Significant at the 5 percent level.

***Significant at the 1 percent level.

Note: Table reports panel estimates with robust standard errors, using the White correction for heteroskedasticity. A constant is estimated but not reported. The instrumental variables estimation (specification 9) uses a third lag of the dependent variable as an instrument. The crisis periods are from December 1, 1994, to March 30, 1995; July 1, 1997, to January 30, 1998; August 1 to October 30, 1998; and January 1 to February 28, 1999. Countries are classified as transparent or nontransparent countries based on the Mehrez and Kaufmann (2000) data. Countries with ratings above the median (Brazil, Chile, Malaysia, Mexico, Peru, and Taiwan [China]) are considered transparent. Countries with ratings below the median (Argentina, Colombia, Indonesia, the Philippines, Poland, the Republic of Korea, the Russian Federation, Thailand, Turkey, and Venezuela) are considered nontransparent. T-statistics are in parentheses.

TABLE 7. Number of Clean Events, by Country

Country	Total events	Upgrades	Downgrades
<i>Latin America</i>			
Argentina	4	1	3
Brazil	8	6	2
Chile	3	2	1
Colombia	7	1	6
Mexico	5	3	2
Peru	2	2	0
Venezuela	7	3	4
Total	36	18	18
<i>East Asia</i>			
Indonesia	6	2	4
Korea, Rep. of	10	9	1
Malaysia	12	5	7
Philippines	5	4	1
Taiwan (China)	0	0	0
Thailand	8	1	7
Total	41	21	20
<i>Eastern Europe</i>			
Poland	6	6	0
Russian Federation	14	8	6
Turkey	6	3	3
Total	26	17	9
Grand total	103	56	47

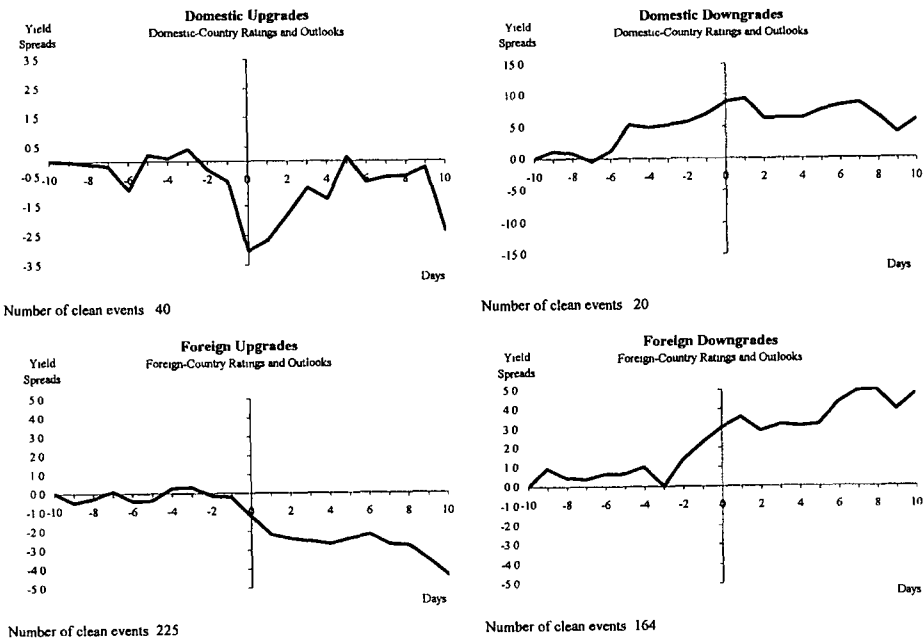
Source: Author calculations.

to domestic-country changes, foreign-country rating and outlook changes appear to have more persistent effects, as if market participants had anticipated these changes to a lesser extent than the changes in domestic-country ratings. Overall, these event studies suggest important spillover effects of changes in ratings, with financial markets in emerging economies jointly rallying or collapsing following rating changes.

These results could be interpreted as indicating that rating agencies are behaving procyclically. Rating agencies decide to downgrade (upgrade) a country when the prices of its financial instruments go down (up). Alternatively, the behavior of prices in the days preceding rating and outlook changes could reflect an anticipation effect. Market participants anticipate the behavior of rating and outlook changes, so markets discount those events.

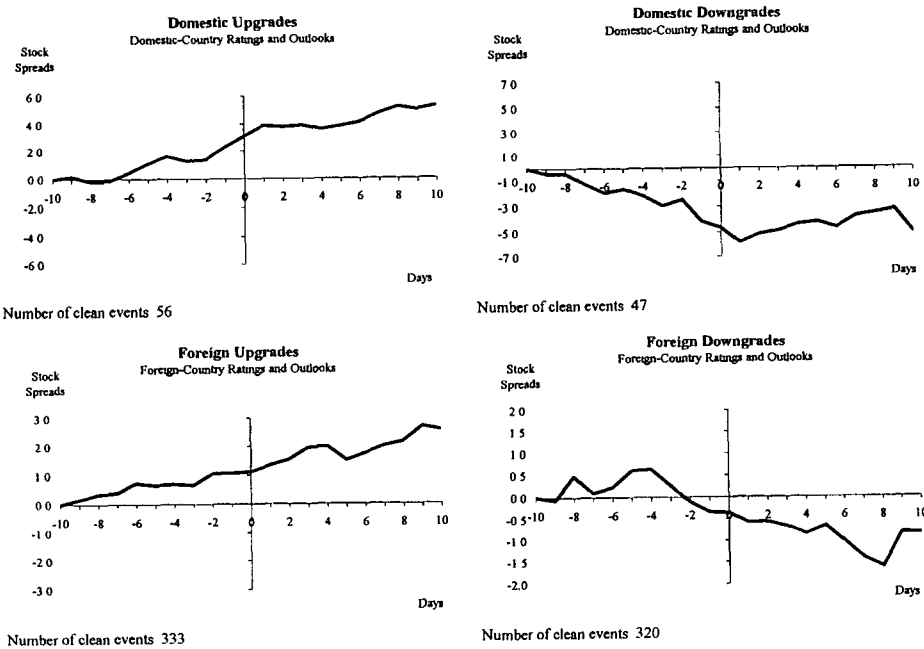
We are inclined to interpret the results as evidence of procyclical behavior by rating agencies. Anecdotal evidence suggests that market participants do not try to anticipate the actions of rating agencies but that these agencies follow market sentiment closely. Moreover, our results are consistent with the findings in Reinhart (2001), who examines whether rating agencies actions anticipated the

FIGURE 3. Even Studies of EMBI spreads



Note: Figure displays spreads between local sovereign debt yields and benchmark instruments from industrial countries (in logs, normalized to 0 on day -10).
 Source: Authors' calculations.

FIGURE 4. Event Studies of Stock Market Indices



Note: Figure displays spreads between local stock market index and the U.S. S&P 500 (in logs, normalized to 0 on day -10).
 Source: Authors' calculations.

crises of the 1990s. With a large sample of countries and crises, she concludes that far from being leading indicators of crises, rating changes are lagging indicators of financial collapses. In contrast, the aftermath of rating changes is uneventful, with sovereign bond yield spreads and stock spreads remaining largely unchanged after announcements and both spreads maintaining the gains or losses observed in the days preceding the rating changes.

V. CONCLUSIONS

Most of the research on the effects of credit ratings on financial markets has focused on quantifying the effects of changes in sovereign ratings on sovereign risk, as measured by the yield spread of domestic instruments relative to benchmark instruments in industrial countries. In this article, we expand the data set not only to update previous tests but to test new hypotheses about the effects of changes in sovereign rating and outlook on financial markets in emerging economies. The data set we assembled enabled us to test the spillover effects across securities and countries, among other things, and to provide a more complete characterization of the relation between credit ratings and financial markets.

We draw six conclusions about the effect of credit ratings on financial markets. First, changes in ratings and outlook significantly affect bond and stock markets, with average yield spreads increasing 2 percentage points and average stock returns declining about 1 percentage point in response to a domestic downgrade. Outlook changes appear to be at least as important as rating changes.

Second, rating changes contribute to contagion or spillover effects, with rating changes of bonds in one emerging market triggering changes in both yield spreads and stock returns in other emerging economies. As in the case of contagious crises, the spillover effects of rating changes are stronger at the regional level (see, for example, Kaminsky and Reinhart 2000b).

Third, changes in credit ratings and outlooks have a stronger effect on both domestic markets and foreign financial markets during crises. Spillover effects are also stronger during crises. This evidence supports crisis-contingent theories of how shocks are transmitted internationally. Masson (1998) shows how a crisis in one country may coordinate investors' expectations, shifting them from a good equilibrium to a bad equilibrium and thus triggering a crash in the other economy's financial markets. Rating and outlook changes could provide this coordinating signal.

Fourth, as expected, rating changes have a stronger effect on more non-transparent economies than on transparent ones, as these changes reveal more information about nontransparent countries.

Fifth, domestic-country rating upgrades do take place following market rallies, whereas downgrades occur after market downturns. This evidence is consistent

with the notion that rating agencies may be contributing to the instability of financial markets in emerging economies. Our results may explain why the effects of upgrades and downgrades do not appear to be large in economic terms, although they are statistically significant. Rating agencies provide bad news in bad times and good news in good times, reinforcing investors' expectations. Rigobon (1997), among others, note that this type of news is not very informative to investors, so markets do not react very strongly to it.

Finally, fragile economies, as measured by low credit ratings, are more severely affected by changes in U.S. interest rates than other economies. In fact, interest rate hikes in financial centers fuel increases in sovereign risk by 50 percent more in vulnerable economies than in countries with higher ratings.

Several potential extensions to this research would improve the understanding of the effects of credit ratings and outlooks. It would be interesting to study whether different ratings agencies affect markets differently. To do so, researchers may need to collect more data to run tests that are statistically meaningful. Another important issue to examine is whether coordinated rating changes across agencies convey stronger signals about a country's health than isolated rating changes and thus trigger more dramatic reactions in financial markets.

An additional extension would be to build better models with which to explain the movements of financial markets in emerging economies. We are still far from explaining daily volatility in either developing countries or mature markets, with the R^2 in most studies tending to be very low.

It is also important to examine the effects of sovereign rating changes on a broader set of securities. For example, sovereign ratings may have a stronger effect on firms with large foreign exposure because sovereign default and currency crises are closely associated (Reinhart 2001). Several researchers have suggested that instability due to "liability dollarization" can be reduced by granting access to security markets in mature markets. Stulz (1999), for example, claims that when firms in emerging market list on stock exchanges in industrial economies they become more accountable and transparent, reducing adverse selection and moral hazard and alleviating liquidity problems that firms in emerging markets often face. One way of testing this hypothesis would be to examine whether sovereign ratings have less of an effect on firms listed on industrial country stock markets.

Regarding the procyclicality of rating upgrades and downgrades, it would be interesting to understand how rating agencies behave beyond the 10-day window analyzed here. This would shed light on how lasting their effects persist.

TABLE A-1. Range of Possible Ratings Assigned by Each Rating Agency to Sovereign Debt

Moody's			s&P			Fitch-IBCA		
Rating	Number	Outlook	Rating	Number	Outlook	Rating	Number	Outlook
Aaa	8.00	Positive	AAA	8.00	Positive	AAA	8.00	Positive
Aa1	7.33	Negative	AA+	7.33	Negative	AA+	7.33	Negative
Aa2	7.00	Stable	AA	7.00	Stable	AA	7.00	Stable
Aa3	6.66		AA-	6.66		AA-	6.66	
A1	6.33		A+	6.33		A+	6.33	
A2	6.00		A	6.00		A	6.00	
A3	5.66		A-	5.66		A-	5.66	
Baa1	5.33		BBB+	5.33		BBB+	5.33	
Baa2	5.00		BBB	5.00		BBB	5.00	
Baa3	4.66		BBB-	4.66		BBB-	4.66	
Ba1	4.33		BB+	4.33		BB+	4.33	
Ba2	4.00		BB	4.00		BB	4.00	
Ba3	3.66		BB-	3.66		BB-	3.66	
B1	3.33		B+	3.33		B+	3.33	
B2	3.00		B	3.00		B	3.00	
B3	2.66		B-	2.66		B-	2.66	
Caa1	2.33		CCC+	2.33		CCC+	2.33	
Caa2	2.00		CCC	2.00		CCC	2.00	
Caa3	1.66		CCC-	1.66		CCC-	1.66	
Ca	1.33		CC	1.33		CC	1.33	
C	1.00		SD	1.00		C	1.00	

Note: The numbers assigned are the ones used to construct figure 2.

Source: Bloomberg.

TABLE A-2. Time Span of EMBI Spreads, Stock Returns, and Sovereign Ratings, by Country

Country	EMBI Spreads		Stock Returns		Sovereign Ratings	
	Initial date	End date	Initial date	End date	Initial date	End date
Argentina	April 30, 1993	June 30, 2000	Jan. 3, 1992	June 30, 2000	Jan. 1, 1990	June 30, 2000
Brazil	Dec. 31, 1991	June 30, 2000	Jan. 23, 1992	June 30, 2000	Jan. 1, 1990	June 30, 2000
Chile	—	—	Jan. 2, 1992	June 30, 2000	Dec. 7, 1992	June 30, 2000
Colombia	—	—	Jan. 2, 1996	June 30, 2000	Jan. 1, 1990	June 30, 2000
Indonesia	—	—	Nov. 5, 1991	June 30, 2000	Dec. 7, 1992	June 30, 2000
Korea, Rep. of	April 30, 1998	June 30, 2000	June 30, 1995	June 30, 2000	Jan. 1, 1990	June 30, 2000
Malaysia	—	—	June 30, 1995	June 30, 2000	Jan. 1, 1990	June 30, 2000
Mexico	Dec. 31, 1991	June 30, 2000	Jan. 2, 1995	June 30, 2000	Dec. 18, 1990	June 30, 2000
Peru	May 30, 1997	June 30, 2000	Jan. 2, 1996	June 30, 2000	Feb. 5, 1996	June 30, 2000
Philippines	Jan. 4, 1993	Jan. 30, 1997	Jan. 4, 1993	June 30, 2000	June 30, 1993	June 30, 2000
Poland	Jan. 17, 1995	June 30, 2000	April 3, 1996	June 30, 2000	June 1, 1995	June 30, 2000
Russian Fed.	Dec. 31, 1997	June 30, 2000	Dec. 1, 1993	June 30, 2000	April 11, 1994	June 30, 2000
Taiwan (China)	—	—	Jan. 2, 1996	June 30, 2000	Jan. 1, 1990	June 30, 2000
Thailand	—	—	Jan. 2, 1996	June 30, 2000	Jan. 1, 1990	June 30, 2000
Turkey	—	—	June 30, 1995	Dec. 30, 1999	May 5, 1992	June 30, 2000
Venezuela	Dec. 31, 1991	June 30, 2000	April 23, 1996	June 30, 2000	Jan. 1, 1990	June 30, 2000

—Not applicable.

Source: Bloomberg and JP Morgan.

REFERENCES

- Anderson, T., and C. Hsiao. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18(1):47–82.
- Calvo, G., and E. Mendoza. 2000. "Rational Contagion and the Globalization of Securities Markets." *Journal of International Economics* 51(1):79–114.
- Calvo, G., L. Leiderman, and C. Reinhart. 1993. "Capital Inflows and Real Exchange Rate Appreciation in Latin America: The Role of External Factors." *Staff Papers International Monetary Fund* 40(1):108–51.
- Cantor, R., and F. Packer. 1996. "Determinants and Impact of Sovereign Credit Ratings." Federal Reserve Bank of New York. *Economic Policy Review* 2(2):37–53.
- Corsetti, G., P. Pesenti, and N. Roubini. 2000. "Competitive Devaluations: A Welfare-Based Approach." *Journal of International Economics* 51(1):217–41.
- Cruces, J. 2001. "Statistical Properties of Sovereign Credit Ratings." Universidad de San Andrés, Buenos Aires, Argentina.
- Dooley, M. 1998. "A Model of Crises in Emerging Markets." NBER Working Paper 6300. National Bureau of Economic Research, Cambridge, Mass.
- Eichengreen, B., and A. Mody. 1998. "What Explains Changing Spreads on Emerging Market Debt: Fundamentals or Market Sentiment?" NBER Working Paper 6408. National Bureau of Economic Research, Cambridge, Mass.
- Erb, C., C. Harvey, and T. Viskanta. 1996a. "Expected Returns and Volatility in 135 Countries." *Journal of Portfolio Management* 22(3):46–58.
- . 1996b. "The Influence of Political, Economic and Financial Risk on Expected Fixed Income Returns." *Journal of Fixed Income* 6(1):7–31.
- . 1996c. "Political Risk, Financial Risk and Economic Risk." *Financial Analysts Journal* 52(6):28–46.
- Ferri, G., G. Liu, and J. Stiglitz. 1999. "The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis." *Economic Notes* 28(3):335–55.
- Frankel, J., S. Schmukler, and L. Servén. 2000. "Global Transmission of Interest Rates: Monetary Independence and the Currency Regime." World Bank Working Paper 2424. World Bank, Washington, D.C.
- Hand, J., R. Holthausen, and R. Leftwich. 1992. "The Effect of Bond Rating Agency Announcements on Bond and Stock Prices." *Journal of Finance* 157(2):733–52.
- Herrera, S., and G. Perry. 2000. "Determinants of Latin Spreads in the New Economy Era: The Role of U.S. Interest Rates and Other External Variables." Latin American and the Caribbean Region, World Bank, Washington, D.C.
- Kamin, S., and K. von Kleist. 1999. "The Evolution and Determinants of Emerging Market Credit Spreads in the 1990s." International Finance Discussion Paper 1999-653. Federal Reserve Board, Washington, D.C.
- Kaminsky, G., and C. Reinhart. 2000a. "The Center and the Periphery: The Globalization of Financial Turmoil." IMF Working Paper, International Monetary Fund, Washington, D.C., forthcoming.
- . 2000b. "On Crises, Contagion, and Confusion." *Journal of International Economics* 51(1):145–68.
- Kaminsky, G., and S. Schmukler. 1999. "What Triggers Market Jitters?" *Journal of International Money and Finance* 18(4):537–60.

- . 2001. "Emerging Markets Instability: Do Sovereign Ratings Affect Country Risk and Stock Returns?" World Bank Policy Research Working Paper 2678. World Bank, Washington, D.C. Available online at <http://www.worldbank.org/research>.
- Kaminsky, G., R. Lyons, and S. Schmukler. 2000. "Managers, Investors, and Crisis: Mutual Fund Strategies in Emerging Markets." NBER Working Paper 7855. National Bureau of Economic Research, Cambridge, Mass.
- Krugman, P. 1998. "Saving Asia: It Is Time to Get Radical." *Fortune* (September 7):74–80.
- Larrain, G., H. Reisen, and J. von Maltzan. 1997. "Emerging Market Risk and Sovereign Credit Rating." OECD Development Centre Technical Paper 124. Organisation for Economic Co-operation and Development, Paris.
- Masson, P. 1998. "Contagion: Monsoonal Effects, Spillovers, and Jumps between Multiple Equilibria." International Monetary Fund Working Paper WP/98/142. IMF, Washington, D.C.
- Mora, N. 2001. "Sovereign Credit Ratings: Guilty Beyond Reasonable Doubt?" Department of Economics, MIT, Cambridge, Mass.
- McKinnon, R., and H. Pill. 1997. "Credible Economic Liberalizations and Overborrowing." *American Economic Review* 87(2):189–93.
- Mehrez, G., and D. Kaufmann. 2000. "Transparency, Liberalization, and Banking Crises." World Bank Policy Research Working Paper 2286. World Bank, Governance, Regulation, and Finance, World Bank Institute, Washington, D.C.
- Reinhart, C. 2001. "Do Sovereign Credit Ratings Anticipate Financial Crises? Evidence from Emerging Markets." School of Public Affairs, University of Maryland, College Park.
- Reisen, H., and J. von Maltzan. 1999. "Boom and Bust and Sovereign Ratings." OECD Development Centre Technical Paper 148. Organisation for Economic Co-operation and Development, Paris.
- Richards, A. 1996. "Winner-Loser Reversals in National Stock Market Indices: Can They Be Explained?" *Journal of Finance* 52(5):2129–44.
- Richards, A., and D. Deddouché. 1999. "Bank Rating Changes and Bank Stock Returns: Puzzling Evidence from Emerging Markets." IMF Working Paper. International Monetary Fund, Washington, D.C.
- Rigobon, R. 1997. "Informational Speculative Attacks: Good News Is No News." Sloan School of Management, MIT, Cambridge, Mass.
- Rouwenhorst, G. 1998. "International Momentum Strategies." *Journal of Finance* 53(1):267–84.
- Stiglitz, J. 2000. "Capital Market Liberalization, Economic Growth, and Instability." *World Development* 28(6):1075–86.
- Stulz, R. 1999. "Globalization of Capital Markets and the Cost of Capital: The Case of Nestle." *Journal of Applied Corporate Finance* 8(3):30–38.

Financial Crises, Credit Ratings, and Bank Failures

On the Use of Portfolio Risk Models and Capital Requirements in Emerging Markets: The Case of Argentina

Veronica Balzarotti, Michael Falkenheim, and Andrew Powell

A portfolio-based model (CreditRisk+ of Credit Suisse First Boston) and recent Central Bank of Argentina credit bureau data are used to estimate whether current capital and provisioning regulations match actual risks. Arguing that provisions should cover expected losses and that capital requirements should cover potential losses beyond expected losses subject to some statistical level of tolerance, the article assesses how well actual capital and provisioning requirements match the estimated requirements given by the model. Actual provisioning requirements were found to be close to implied levels of expected losses. The estimate of potential losses was found to be highly sensitive to the assumptions of the model, especially the parameter relating the volatility of a loan's rate of default to its mean value. This volatility parameter cannot be estimated accurately with the credit bureau data because of the short time span covered, so proxy data were used to estimate it, and two values around that estimate were tried. The difficulty of estimating this critical parameter implies that the results should only be regarded as suggestive. Moreover, the methodology only seeks to estimate credit risk and not interest rate risk or exchange rate risk, nor does it fully take into account the indirect effects of interest rates and exchange rates on credit risk. As recent events in Argentina have demonstrated, estimating credit risk along these lines should be thought of as just one tool in attempting to assess the appropriate level of bank provisions and capital.

Recent literature stresses the need for capital requirements and provisions to maintain a healthy financial system by limiting the risk of bank failures. In the past, the requirements reflected rules of thumb or were the outcome of complex political negotiations. More recently greater efforts have been made to quantify

Veronica Balzarotti is manager of the Regulatory Policy Department with the Central Bank of Argentina. Michael Falkenheim is a financial economist with the Office of Management and Budget of the United States and a former Research Economist in the Research Department of the Central Bank of Argentina. Andrew Powell is Professor at Universidad Torcuato Di Tella and former Chief Economist at the Central Bank of Argentina. Their e-mail addresses are vbalzarotti@bcra.gov.ar, Michael_C._Falkenheim@omb.eop.gov, and apowell@utdt.edu, respectively. The authors would like to thank George McAndless and Guillermo Escudé for comments, Christian Castro and Matías Gutierrez Girault for assistance, and two anonymous referees for invaluable comments. All remaining errors remain their own. The opinions expressed in this article are entirely those of the authors and do not necessarily reflect those of the Central Bank of Argentina or any other institution with which they are affiliated.

appropriate levels of regulatory capital. The methodology usually applied is to consider what stock of capital would cover potential losses in all but a small percentage of scenarios that could prevail in the time needed to take risk-mitigating actions, such as selling risky loans or replenishing capital.

The credit bureau of the Central Bank of Argentina provides a rare tool for quantifying provisions and capital requirements in this way. This credit bureau was established in 1991 to collect information on the larger debtors of the financial system and help elucidate how those debtors posed risks for the financial system and for individual banks.

The credit bureau has served other needs, including the need for more accurate information on the credit history of debtors. At an early stage the database was given to virtually all financial institutions at low cost. The power of the data to address "willingness to pay" issues prompted both an extension of the database and its wider distribution. Today the database covers virtually all loans in the financial system (more than 6 million entries). It is updated monthly and available free of charge on the Internet through the Web site of the Central Bank of Argentina at <http://www.bcra.gov.ar>.

In this article we describe a study that used the Central Bank of Argentina's credit data to estimate the potential losses of a portfolio of Argentine loans. We use estimates of potential losses from a portfolio-based model, CreditRisk+ of Credit Suisse First Boston, to determine whether current capital and provisioning regulations match actual risks. We argue that provisions should cover expected losses and capital requirements should cover potential losses beyond expected losses, subject to some statistical level of tolerance. We then assess how actual capital and provisioning requirements match the estimated requirements produced by our application of the model and calibrated using a particular sample of recent data.

The results are subject to several caveats, especially considering recent events in Argentina. First, we employ a limited database, which includes barely two years of data. The period covered was one of recession, but it ended before the real crisis began. Probabilities estimated on the basis of such a database may not cover particular infrequent events. Second, we attempt to measure only credit risk; we do not analyze the direct effect of interest rate or foreign exchange risk on banks' balance sheets. Indeed, given the limited nature of the database, it is unlikely that we fully capture the indirect effect of these risks on banks' clients that then feed through to credit risk. Finally, we do not consider the severe effects on bank solvency of the default of the public sector in Argentina and the forced revaluation of loans and deposits at different exchange rates. These caveats underline the fact that value at risk models should be thought of as partial tools, to be used in conjunction with scenario analysis or other methods rather than as sole estimators of bank risk.

This article is organized as follows. In section I we briefly describe the role of provisioning and capital requirements and suggest that capital requirements may need to reflect portfolio considerations. In section II we describe the source of

our data, the Central Bank of Argentina's credit bureau. In section III we present our application of CreditRisk+, explaining our methodology for choosing its input parameters. In section IV we compare the model's estimates of expected and potential losses to actual provisioning and capital requirements. Section V summarizes our conclusions.

I. PROVISIONS, CAPITAL REQUIREMENTS, AND CREDIT RISK

Both provisions and capital requirements attempt to control credit risk by creating a buffer against credit losses (Basel Committee on Banking Supervision 1999a, 1999b). In practice it is sometimes difficult to differentiate them. In the Basel 1988 Accord, for example, it was agreed that a general provisioning requirement might be recorded as Tier II capital against requirements.¹

We believe that provisions and capital serve two distinct purposes. In our view, provisions alone should protect banks against ordinary levels of credit loss, whereas capital requirements should protect banks against unforeseen losses. In statistical terms, provisions should reflect the expected value of credit losses, and capital requirements should protect against unexpected losses, subject to some level of statistical tolerance. This means that in theory, both provisioning and capital requirements may be specified from the same distribution (the distribution of potential credit losses), but they reflect different statistics of that distribution.

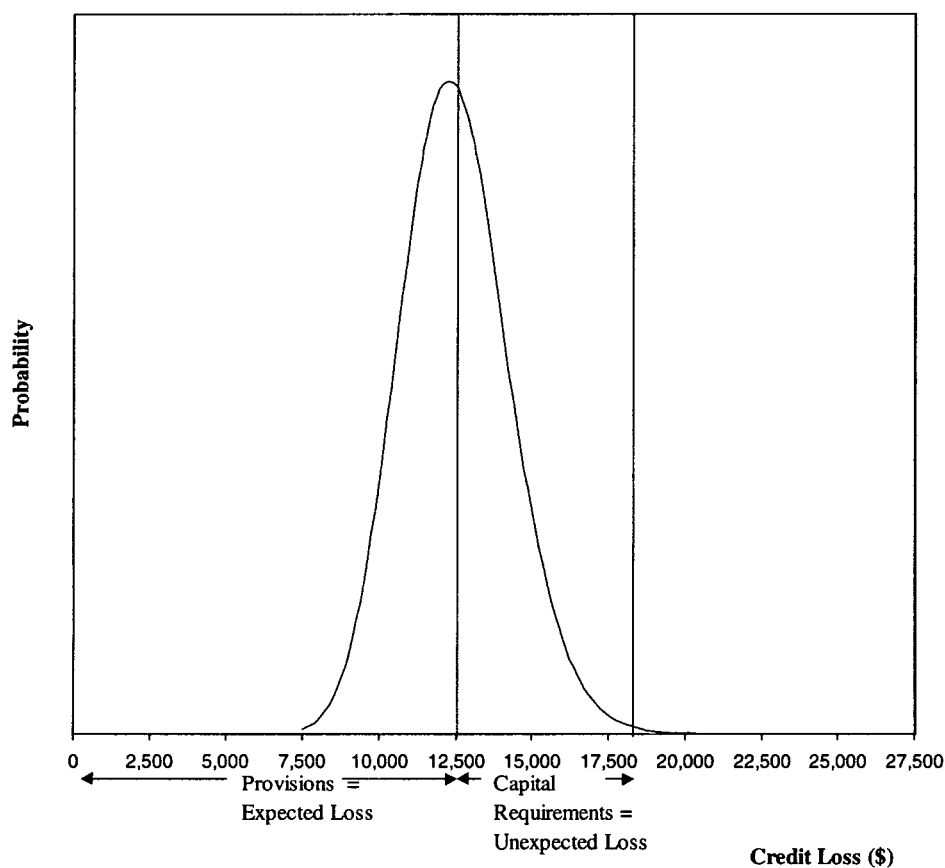
This theoretical concept of provisions and capital allows for clear comparisons between their actual levels and an estimated probability distribution. Moreover, it reflects what appears to be the emerging consensus in the regulatory community and the banking industry.

The Basel Committee on Banking Supervision (1998) stated that the "aggregate amount of specific and general [provisions] should be adequate to absorb estimated credit losses associated with the loan portfolio" (p. 23). The same committee (Basel Committee on Banking Supervision 1999a) suggests that most sophisticated financial institutions view economic capital as covering unexpected as opposed to expected losses. Part of our analysis compares the total coverage provided by both capital and provisions against the 99.9th percentile of the loss distribution, thus making no assumptions about the division of labor between capital and provisions and recognizing that capital and provisions work in tandem.

Figure 1, which plots a distribution of potential credit losses (for a single loan or a loan portfolio), illustrates our view of the appropriate level of provisions and capital requirements. In this example the appropriate level of provisions is

1. The Capital Accord defined tier I capital as "core capital," including equity capital and published reserves from posttax retained earnings. It defined tier II capital as "supplementary capital," including hybrid capital instruments, subordinated debt, and provisions. The Capital Accord required that at least 50 percent of capital used to meet minimum levels be tier I capital.

FIGURE 1. An Illustration of the Probability Distribution of Credit Losses



the expected loss, US\$12,500 (for the moment we disregard the role of net interest). In contrast, capital requirements should reflect unexpected losses, usually defined as the difference between a given percentile level and the expected loss. In our example we calculate appropriate capital requirements as equal to the difference between the 99.9th percentile of credit losses and the expected value. The 99th percentile defines a line that places just 0.1 percent of the distribution to the right (in our example the value is \$18,000). Appropriate capital requirements in our example are $\$18,000 - \$12,500 = \$5,500$. Note that credit losses should exceed the 99.9th percentile in only 1 out of 1,000 possible economic scenarios (just over once every millennium if annual drawings are made from this distribution).

Interest complicates this picture somewhat. Provisions and capital requirements are supplemented by the interest that banks charge borrowers to compensate them for the expected cost of default and as a premium for the systematic risk of their loan portfolio (see Escudé 1999, Rochet 1992). The Rochet and Escudé models suggest that the net interest charged by banks should cover the losses expected

by the lender at the moment of origination and the nondiversifiable portion of unexpected losses. One argument is that interest compensates the financial institution for expected losses and systematic risk, partly duplicating the roles of provisions and capital. However, the Basel Committee on Banking Supervision (1998) recommended that provisions be adequate to protect against expected losses, making no mention of interest margins.² Moreover, net interest is available as protection against credit losses only if it is not first paid as a dividend to bank shareholders or owners. In normal times it might be argued that shareholders will expect a normal return; provisions should hence cover normal losses without including net interest earnings. That said, it is likely that a bank will reduce or even suspend payments of dividends if its portfolio begins to deteriorate seriously and increase its provisions.

In our view, deducting future interest margin from the calculation of capital requirements would imply a much less conservative treatment of capital than that included in the current (and future) Basel Capital Accord. The accord's definition of capital admits earnings at tier I level only if retained earnings have been appropriated into a disclosed reserve; otherwise the accumulated after-tax surplus of retained profits may be included as tier II capital, with the approval of the national supervisor. This means that these profits must be observed, not merely expected, and may be subject to the limits on tier II capital if they have not been appropriated into a reserve.

For these complications it remains controversial whether expected net interest should be counted along with capital and provisions as a buffer against credit losses. We remain agnostic on this point. We therefore make two comparisons: one that takes interest into account and another that does not. First, we compare capital requirements and provisions with expected and potential losses. Second, we add an estimate of expected net interest to provisions and capital requirements and then compare them with expected and potential losses.

In Argentina, as elsewhere, for each loan the Central Bank requires a minimum level of provisions, which depends on the economic classification of the debtor. For commercial loans financial institutions are required to rate the debtor on a scale of one to five, depending on its expected cash flow. For consumer and housing loans, financial institutions must base the classification of debtors on their current payment status. Current regulations allow commercial loans of less than \$200,000 to be treated as consumer loans in terms of these requirements.

2. The Basel Committee on Banking Supervision (2001) notes that its proposed Internal Ratings Based (IRB) approach has been criticized for not taking interest margins into account. It recommends that regulators allow some technically sophisticated institutions to take interest margins into account for retail portfolios under the IRB: "For the retail portfolio, allow (by definition advanced) IRB banks to use their own internally generated estimates of [Expected Loss]-related capital charges based on a comparison of expected future credit losses and future margin income. . . . For non-retail portfolios sticking to the one year ahead [Probability of Default] times [Loss Given Default] without any [Future Margin Income] recognition would seem to be an acceptable approximation for [Expected Loss]" (p. 4).

One of the central purposes of this study is to evaluate the current system of provisioning requirements. We use the credit bureau data to assess whether the level of provisions is adequate given observed loss probabilities.

Performing loans are classified as 1 or 2, whereas loans rated 3, 4, or 5 are considered nonperforming loans. Each of the first two categories represents a broad range of risk and would cover several ratings classes in a private ratings system, such as Moody's or Standard and Poor's. In developing our credit risk model we use econometric techniques to distinguish between different risk levels within each category, and we give a more precise estimate of risk than that conveyed by the rating alone.

The Basel 1988 Capital Accord formally establishes the current form of capital requirements for "internationally active" banks in G-10 countries, but more than 100 countries, including Argentina, have explicitly adopted the accord in their own banking regulations or in law. In many countries the rules are applied not only to internationally active banks but also to domestic banks. Moreover, in some countries that have adopted the general Basel methodology, requirements have been stricter than the minimum 8 percent of assets at risk recommended by the accord. Countries have adopted their own limits within the general methodology depending on the perceived level of credit risk.

In Argentina the Central Bank sets capital requirements for the banking system. Since the end of 1994 it has required banks to set aside 11.5 percent of risk-weighted assets for counterparty risk. The Basel Accord defined risk weights for different assets in an attempt to capture the different levels of risk in their returns. These weights are used in Argentina, but they are complemented with a risk indicator based on the interest rate charged on each loan. This additional risk indicator is a factor by which the base capital requirement is multiplied. Under this system loans with higher interest rates have higher capital requirements because they are presumed to have a higher level of risk. Argentine capital requirements also include a factor that depends on the CAMELS (capital, assets, management, earnings, liquidity, and sensitivity) rating assigned to each financial institution by the Superintendency of Financial Institutions.

Unlike the original Basel Accord, Argentina's regulations do not allow general provisions to be included as tier II capital. In addition to counterparty risk capital requirements, the Central Bank imposed capital requirements for market risk and interest rate risk. The current Argentine capital requirement was then specified according to the following formula:

$$(1) \text{ Overall Capital Requirement} = 11.5\% * IR * W * K * A + \text{Market Risk} \\ + \text{Interest Rate Risk}$$

where IR is the interest rate factor; W is the average Basel risk weight for assets, varying between 0 and 1; and K is the CAMELS factor. K ranges from 0.97, for banks with a rating of 1, to 1.15 for banks with a rating of 5. A is the outstanding value of the asset.

One shortcoming of Basel-style capital requirements is that they do not take into account how individual exposures are combined in the loan portfolio. One \$100 million loan to a single company has the same capital requirement as 1,000 loans of \$100,000 each to 1,000 different companies if these loans are in the same risk category. However, maintaining a diversified portfolio usually reduces the total credit risk of an institution. Moreover, requirements do not differ according to the level of correlation of asset returns in a portfolio. A portfolio of loans exclusively to companies in the textile industry would have the same capital requirement as a portfolio of loans spread across various industries, assuming they are all in the same risk category. If the standard rules do not reflect the actual risks of financial institutions' credit portfolios, capital allocation decisions may be distorted.

A recent proposal to remedy this problem is the use of internal models to assess capital adequacy. Under this system financial institutions would apply to use their own measures of credit risk to determine the capital requirement. As Jones and Mingo (1998) suggest, regulators would define a minimum permitted probability of insolvency, and financial institutions would develop models to estimate the probability distribution of credit losses. Capital adequacy would be calculated based on that probability. Regulators would then decide which models deserve authorization based on their technical merit and historical performance.

Although senior regulators in the United States and the United Kingdom view the use of internal models as very promising, such models remain at an early stage of development. Indeed, the recent proposal to modify the Basel 1988 Accord, though including many ideas to improve credit risk assessment of individual debtors, shied away from methods of analyzing portfolio risk based on internal models. In using a credit risk model to evaluate provision and capital requirements, it is therefore important to recognize the limitations of this approach. The Basel Committee and other institutions studying credit risk models all concluded that these models are not yet sufficiently well developed to use in a capital requirement system. According to the Basel Committee on Banking Supervision (1999a, 1999b) and others (see Jackson and others 1999), important issues, such as the correct shape of the loss distribution, have not been resolved, and the short span of historical data makes it impossible to properly validate credit risk models. Furthermore, other risk factors, such as operational risk, have not been adequately studied. The basic ratio established in the Capital Accord might also provide a needed hedge against operational and other risks.

The problems associated with implementing models for credit risk regulatory capital may be magnified in emerging markets, where assumptions about structure and parameters are likely to be less stable and technical and human resources are likely to be more constrained. Faced with this reality, in the exercise that follows we consider using such models only as a check to see whether current regulations broadly match implied theoretical levels. Our concern is with the total

level of provisions and capital available to an institution or to the financial system as a whole. We do not consider how this capital is distributed across the loan portfolio. We thus address questions of overall prudential standards rather than questions of efficiency.

Another possible use of credit risk models is in supervision. Capital requirements in Argentina depend on the CAMELS rating that the Superintendency of Financial Institutions assigns to each institution. Supervisors in Argentina give institutions a rating of one to five based on the level of risk of their assets, among other things. This rating translates into a lower or higher capital requirement, because each rating causes a different multiplier to be applied to the global credit risk capital requirement (table 1). Credit risk models may help supervisors quantify the credit risk of institutions and perhaps become an explicit part of their (CAMELS) ratings decisions.

II. DATA

Our source of data for studying credit risk is the Central de Deudores del Sistema Financiero (CDSF) data set, which includes information on virtually every loan in the Argentine financial system. The CDSF originated in January 1991 when the Central Bank of Argentina began collecting and disclosing information about the largest debtors in the financial system (those with debts of more than \$200,000). Financial institutions provide the information, which is then validated and redistributed to all contributors.

Information on debtors originally included only the classification assigned to each debtor by each financial firm. Later the Central Bank required banks to report more detailed information, including the principal activity of the debtor, its links (if any) to the lending institution, the business group to which it belonged, debts by currency denomination, collateral, provisions, and net worth. In September 1994 the Central Bank of Argentina decided to make the information available to the public, for a modest fee, through an agency named the Risk Center (Diaz 1998, Roisenzvit 1997).

In 1995 the Central Bank decided to extend the range of debtors by creating the Credit Information Center (CIC), which began operations in January 1996.

TABLE 1. Bank of Argentina Loan Classifications and Provision Requirements

Classification	Definition	Current minimum provision (percent)
1	Normal	1
2	Potential risk	5
3	Substandard	25
4	Doubtful	50
5	Loss	100

Source: Central Bank of Argentina.

This new register includes information on debtors from the nonfinancial sector with debts greater than \$50, thus covering virtually the entire range of borrowers. For each debtor the CIC collects information on principal activity, total debt, collateral, and the financial institution's classification of the debtor. In October 1996 the Central Bank decided to disclose the information collected by the CIC for an annual fee. In July 1997 both the Risk Center and the CIC were unified in the CDSF.

Using the CDSF to study credit risk entails several practical problems. The database does not contain information on individual debts, instead grouping together all debts by an individual at a financial institution. The CDSF does not provide information about the structure of debts and only recently began to provide information on loan duration and debt interest rates.

III. METHODOLOGY

To quantify the credit risk of Argentine financial institutions, we use the CreditRisk+ model, a portfolio-based credit risk model developed by Credit Suisse Financial Products. Portfolio-based credit risk models estimate the probability distribution of total credit losses for a portfolio of loans. These models generate a probability distribution of loan losses based on certain specifications and parameters supplied by the user.

Portfolio-based credit risk models capture the two factors that create the potential for unexpected losses: the concentration of exposures in large borrowers and the correlation of changes in credit quality of separate borrowers. If exposures are concentrated in a few large borrowers their default can create larger than expected total losses. Correlation between the defaults of different borrowers creates a situation in which defaults tend to come in bunches, causing a higher than expected default rate.

We chose CreditRisk+ after considering several alternative models. Crouhy and others (2000) and Koyluoglu and Hickman (1998) make excellent comparisons of the different models. Although the presentations in the models' technical documents give the impression that they are quite different from one another, Koyluoglu and Hickman (1998) and Gordy (2000) point out that their statistical structure is fairly similar. The differences have to do largely with calculation methods and assumptions about the correlation of loan defaults. Some models use Monte Carlo techniques, whereas others, including the CreditRisk+ model, derive analytical formulas for the probability distribution of total portfolio losses. Different models also make different assumptions to capture the correlation between the defaults of different debtors.

The CreditRisk+ technical document gives a full description of the model (Credit Suisse First Boston 1997). To illustrate our application of CreditRisk+, we highlight its main features.

The model calculates the probability distribution of total portfolio losses for a fixed time horizon—one year, for example—although it can be extended to

calculate losses over many periods or under the assumption that all loans will be held to maturity. For each loan the model requires as inputs the size of the exposure, the probability of default, the volatility of that probability over time, and the loan's loss given default (one minus the recovery rate), assumed to be a constant. One or more stochastic factors drives the probability of default of the loans in the portfolio. These stochastic factors, which are assumed to have a gamma distribution, capture the correlation between loan defaults. The correlation of default between each pair of debtors depends on how much common risk factors drive their probabilities of default.

In applying CreditRisk+ the user needs first to define the basic design of the model, including its time horizon and the number of independent risk factors that will affect the probability of default of individual loans. The user then needs to supply parameters for each loan: the exposure, the probability of default, the volatility of that probability, and the loss given default. The formulas of the CreditRisk+ model aggregate the individual probabilities of default and volatilities to derive an analytical expression for the probability distribution of total losses.

The main benefit of CreditRisk+ for our application is its simplicity. It requires a minimal amount of data, all of which are available in the CDSF. Its analytical (as opposed to Monte Carlo) method for calculating the loss distribution of a portfolio allows us to calculate results for large portfolios relatively quickly, a key feature given that one of our goals is to incorporate its results in the supervision process. In addition, using CreditRisk+ facilitates analysis of the sensitivity of our results to key modeling assumptions. In comparing CreditRisk+ and CreditMetrics™, Gordy (2000) notes that estimates of implied capital derived from CreditRisk+ can vary greatly with the kurtosis of the default rate's distribution, which in turn is related to the parameters of the underlying gamma distribution.

We define the time horizon as one year. Most analysts have viewed one year as an appropriate time horizon for measuring capital adequacy because they believe that financial institutions can take risk-mitigating action or replenish their capital levels within that time period.

For the sake of simplicity we assume a single risk factor. The formulas for the probability distribution of total portfolio losses become substantially simpler when a single factor is assumed, and the model's processing time is shorter. Assuming a single risk factor implies that the default probabilities of all loans are driven by a common factor, which one could think of as the overall macroeconomic climate. Assuming more than one risk factor potentially could have allowed us to better model the risk of institutions with portfolios concentrated in single industries by assigning a risk factor to each major industry. But assuming many risk factors would have complicated the model substantially and would have required the estimation of numerous additional parameters, many of which could not have been estimated with any accuracy. Given our main goal—to determine whether current regulations broadly match implied theoretical levels—we considered it best to avoid these complications.

Having specified a one-year time horizon and one risk factor for each loan, i , we need to identify the size of the loss given default, μ_i ; the mean probability of default over the one-year horizon; and the volatility of that probability, σ_i . We assume that the exposure is the loan balance recorded in the CDSF database and that the loss given default is equal to the exposure minus 50 percent of the value of the collateral covering the credit. This assumption is consistent with the Central Bank's provisioning requirements, which oblige each bank to allocate provisions for 100 percent of the value of an irrecoverable loan minus 50 percent of the value of the collateral guaranteeing that loan.

Using historical data from the CDSF and an econometric model, we estimate each loan's probability of default as a function of its classification and other characteristics. We specify the model as an ordered probit. The ordered probit estimates the probability that at the end of the year a loan will have each possible classification, given its current classification and other characteristics. The estimated probability that a loan ends the year with a classification of 5 (defined as loss) gives the estimated probability of default. Using an ordered probit—which estimates the probability of obtaining any classification—as opposed to a simple probit—which would estimate only the probability of loss—we were able to take into account all the observed changes in classification, not just instances when a loan changed its classification to “loss.” Under the ordered probit the probability that the classification at the end of the year c'_{t+1} is K is given by the following expression:

$$(2) \quad \Pr(c'_{t+1} = K | \beta; X^i) = \begin{cases} F[\nu_1 + \beta' X^i] & K = 1 \\ F[\nu_K + \beta' X^i] - F[\nu_{K-1} + \beta' X^i] & K = 2, 3, 4 \\ 1 - F[\nu_K + \beta' X^i] & K = 5 \end{cases}$$

In this equation, $F[\cdot]$ is the standard cumulative normal distribution, and ν_1, \dots, ν_5 are estimated parameters that define the cutoffs between each pair of adjacent classifications. The value X^i is a vector of characteristics for the individual loan, including the following variables:

- *The borrower's classification.* Naturally borrowers with a better (lower) classification are less likely to default. For each classification, we estimate a dummy variable measuring the relative risk of default of that classification.
- *The borrower's activity.* The database includes a three-digit code identifying the borrower's industry or intended use of the loan. For each category defined by that code, we estimate a dummy variable to measure the relative propensity of borrowers in that industry to default.
- *The size of the exposure.* We conjecture that larger loans have on average a lower probability of default than smaller loans, because lenders may screen them more carefully.
- *The CAMELS rating of the lender.* Because the CAMELS rating should reflect the quality of the financial institution, a better rating should tend to be associated with lower probabilities of default. For each CAMELS rating we

estimate a dummy variable measuring the relative risk of default for loans made by an institution with that rating.

- *The percent of the debt backed by collateral.* Collateral may affect the probability of default for two offsetting reasons. On the one hand, borrowers who post collateral may work harder to avoid default to prevent losing their collateral. On the other hand, lending institutions may relax their lending requirements for borrowers with collateral.

The parameter β is the vector of coefficients that multiply the variables in the vector X . We use maximum likelihood estimation to estimate the parameters ν and β using data on the actual one-year changes in classification for all loans in the credit bureau starting in September 1999. We then use the estimate $\hat{\beta}$ to generate estimated probabilities of default for each loan, given by the following expression.

$$(3) \quad \hat{\mu}_i = \Pr(c_{i,t+1} = 5 \mid \hat{\beta}, X_i^t) = 1 - F[\nu_5 + \hat{\beta}'X_i^t].$$

To estimate the econometric equation we use data from the CDSF on borrower characteristics in March 1998 and one-year changes in classification between that month and March 1999, a period in which there were about 4.5 million loans in the database.

The volatility of the probability of default reflects its sensitivity to economic and other factors that affect different loans simultaneously. High volatility indicates that the probability of default will tend to go up in difficult times, increasing the loan's contribution to the portfolio level of unexpected risk. To simplify the estimation of the volatility parameter for each loan, we adopt a suggestion from the Technical Document and assume that the volatility to the probability of default is given by the expression $\sigma_i = \delta\mu_i$. In other words, the ratio of the volatility of the probability of default to its mean is the same for all loans and equal to δ . Because of the short time span, we cannot depend on our data to estimate this ratio. Had we had a longer time series of data, it would have been possible to estimate a sequence of one-year default probabilities and calculate their standard deviation in order to estimate δ . Lacking such a time series, we relied on external data to generate an estimate of δ and analyzed the sensitivity of our results to this estimate by reestimating the CreditRisk+ model with alternative values of δ around that estimate.

The U.S. Federal Reserve collects data on the delinquency rate of bank loans in a variety of categories. Between 1985 and 2000 the loan delinquency rate averaged 3.86 percent, with a standard deviation of 1.44 percent, suggesting that $\delta \cong 0.38$. In our estimation, we consider two possibilities around this level, $\delta = 0.3$ and $\delta = 0.5$.

IV. IMPLICIT PROVISIONS AND CAPITAL REQUIREMENTS

To compare the implied capital requirements from our application of CreditRisk+ with actual capital requirements and capital levels, we examine the five largest

private financial institutions in Argentina. We estimate the probability distribution of credit losses of their combined portfolio (figure 2).

The figure shows that the shape of the loss distribution is highly sensitive to the assumption about the parameter δ , which represents the ratio of the volatility of the default rate to its mean. Under the assumption $\delta = 0.5$, the distribution is much more skewed to the right than under the assumption $\delta = 0.3$, resulting in higher estimates of potential losses.

We compare both required and actual levels of capital and provisions held by financial institutions against the probability distribution of losses under each assumption about the parameter δ (table 2). We also make another calculation that takes interest into account. To calculate the amount of interest to protect financial institutions against credit losses, we note that in recent years the interest margin of Argentine financial institutions has been about 4 percent (see Raffin 1999), whereas administrative costs and noninterest income have almost exactly offset each other. This means that for every dollar of assets, a typical institution has available to it a net income of 4 cents to cover credit losses. Given that total assets of the five institutions were equal to \$25 billion, about \$1 billion in net interest is expected on their loan portfolios. This interest creates a buffer against credit losses. We also compare the distribution of credit losses against the total of provisions, capital, and net interest income available to the five institutions (table 3).

The results of both comparisons indicate that, subject to the caveats noted earlier, Argentina's provisions provide adequate protection against the expected value of credit losses. The capital requirement is adequate at a 99.9 percent level

FIGURE 2. The Estimated Probability Distribution of Credit Losses for the Five Largest Private Banks

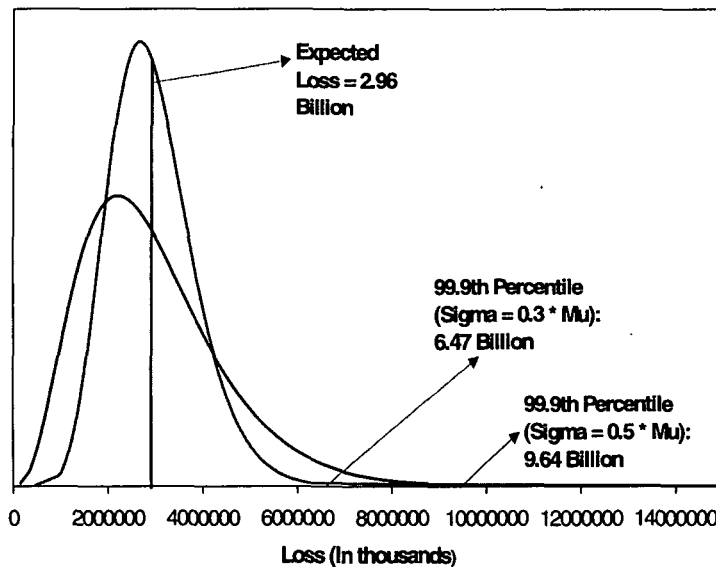


TABLE 2. Estimated Distribution of Losses
(billions of pesos)

Item	CreditRisk+ ($\delta = 0.3$)	CreditRisk+ ($\delta = 0.5$)	Current minimum requirement	Actual level	Actual level including interest
Expected loss/provisions/ provisions + annual net interest income	2.96	2.96	2.98	3.19	4.19
Unexpected loss 99.9th percentile minus expected loss/capital requirement	3.51	6.68	3.72	4.88	4.88
Sum (99.9th percentile)	6.47	9.64	6.70	8.07	9.07

Source: Authors' calculations.

of tolerance under the assumption that $\delta = 0.3$ but not under the assumption that $\delta = 0.5$. Total protection against credit losses offered by provision and capital requirements is adequate under the assumption of $\delta = 0.3$ but not under the assumption that $\delta = 0.5$. Under the more conservative assumption, there is a 2 percent chance in any year that actual credit losses will exceed the required level of capital and provisions. The probability falls to 0.5 percent when the actual level of capital and provisions is considered and to 0.2 percent when the protection offered by net interest income is included.

V. CONCLUSIONS

The extensive credit bureau data of the Central Bank of Argentina can be used to evaluate the provisions and capital requirements of the Argentine financial system. Although there are significant limitations on using these data in credit risk models—a purpose for which the database was not originally intended—the data can be used to estimate implicit capital requirements for Argentine financial institutions.

Using a portfolio model, CreditRisk+, and the credit bureau data, we find actual provisioning requirements to be close to implied levels of expected losses. The estimate of unexpected losses is highly sensitive to the assumptions of the model,

TABLE 3. Estimated Probability That Credit Losses Exceed Indicated Values
(percent)

Level of capital, provisions, and net interest	CreditRisk+ ($\delta = 0.3$)	CreditRisk+ ($\delta = 0.5$)
Current required level of capital and provisions	Less than 0.1	2.0
Actual level of capital and provisions	Less than 0.1	0.5
Actual level of capital, provisions, and net interest income	Less than 0.1	0.2

Source: Authors' calculations.

especially to the parameter relating the volatility of a loan's rate of default to its mean value. This implies that it is more problematic to calibrate capital requirements. However, this volatility parameter cannot be estimated accurately with the credit bureau data given its short time span. We use proxy data to estimate this parameter and tried two values around that estimate. To ensure a 99.9 percent probability of solvency, the capital levels appear to be adequate under one value of that parameter and inadequate under a more conservative assumption.

This work represents a first attempt to use a simple portfolio model of credit risk and credit bureau data to assess regulatory capital requirements. Many restrictions and caveats must accompany this type of analysis. The theoretical assumptions are strong, and the data cover only a short time span. The results reflect these limitations and should therefore be read as suggestive rather than authoritative. Given recent events in Argentina, the credit risk analyzed in this model is clearly only a partial measure of the risk on banks' balance sheets. Moreover, we do not seek to estimate other risks, such as interest rate risk or exchange rate risk. Our results are nevertheless important because they suggest that for the many emerging countries that have developed public credit registries (see Miller 2002 for a review), the information collected may be used to develop measures of credit risk that can help assess appropriate levels of bank provisioning and capital.

REFERENCES

- Basel Committee on Banking Supervision. 1998. "Sound Practices for Loan Accounting, Credit Risk Disclosure and Related Matters." Basel Committee Publication 43. Basel, Switzerland.
- . 1999a. "Credit Risk Modeling Current Practices and Applications." Basel Committee Publication 49. Basel, Switzerland.
- . 1999b. "A New Capital Adequacy Framework." Basel Committee Publication 50. Basel, Switzerland.
- . 2001. "Working Paper on the IRB Treatment of Expected Losses and Future Margin Income." Basel Committee Working Paper. Basel, Switzerland.
- Credit Suisse First Boston. 1997. "CreditRisk+, A Credit Risk Management Framework." London.
- Crouhy, Michel, Dan Galai, and Robert Mark. 2000. "A Comparative Analysis of Current Credit Risk Models." *Journal of Banking and Finance* 24(1–2):59–117.
- Diaz, Julio. 1998. "Credit Information: Conceptual Issues and a Description of the Argentine Case." Central Bank of Argentina, Buenos Aires.
- Escudé, Guillermo. 1999. "El Indicador de Riesgo Crediticio de Argentina dentro de un enfoque de teoría de carteras de la exigencia de capital por riesgo crediticio." Working Paper No. 8. Central Bank of Argentina, Buenos Aires.
- Gordy, Michael. 2000. "A Comparative Anatomy of Credit Risk Models." *Journal of Banking and Finance* 24(1–2):119–49.
- Jackson, Patricia, Pamela Nickell, and William Perraudin. 1999. "Credit Risk Modeling." *Financial Stability Review* 6(June):94–102.

- Jones, David, and John Mingo. 1998. "Industry Practices for Credit Risk Modeling and International Capital Allocations: Implications for a Models Based Regulatory Capital Standard." Federal Reserve Bank of New York *Economic Policy Review* 4(3):53–60.
- Koyluoglu, H. U., and A. Hickman. 1998. "Reconcilable Differences." *Risk* 11(10):56–62.
- Miller, M. 2002. "Credit Reporting Systems around the Globe: The State of the Art in Public and Private Credit Registries." In Margaret J. Miller, ed., *Credit Reporting Systems and the International Economy*. Cambridge, Mass.: MIT Press.
- Raffin, Marcelo. 1999. "Una nota sobre la Rentabilidad de los Bancos Extranjeros en la Argentina." Technical Note No. 6. Central Bank of Argentina, Buenos Aires.
- Roisenzvit, Alfredo. 1997. "Los institutos de Información Crediticia en Argentina." Central Bank of Argentina, Buenos Aires.
- Rochet, Jean Charles. 1992. "Capital Requirements and the Behaviour of Commercial Banks." *European Economic Review* 36(5):1137–70.

Impact Evaluation of Social Funds

An Introduction

Laura B. Rawlings and Norbert R. Schady

I. WHY CARRY OUT IMPACT EVALUATIONS?

Despite the importance of knowing whether development programs achieve their objectives, impact evaluations remain rare in developing economies. This is unfortunate. With the growing use of results-based management by governments, determining whether goals have been attained and convincingly linking changes to specific programs has become increasingly critical. Tracking such outcomes as gains in school enrollment or reductions in infant mortality is indispensable. But simply gathering good data on outcomes sheds little light on why objectives have or have not been met. For this reason, impact evaluations should be a key instrument in policymakers' monitoring and evaluation toolbox.

Impact evaluations rely on the construction of a counterfactual—an attempt to estimate what a given outcome would have been for the beneficiaries of a program if the program had not been implemented. Impact evaluations thus address causality and allow results to be attributed to specific interventions. The challenge of evaluation research arises from the fact that the counterfactual outcome is inherently unobservable, because people cannot simultaneously participate and not participate in a program. The four social fund evaluation studies in this issue illustrate that establishing a counterfactual is usually a matter of using statistical or econometric techniques to construct a control or comparison group.

II. WHY EVALUATE SOCIAL FUNDS?

Social funds have become popular vehicles for channeling development assistance, with a reputation for implementing community-based development projects quickly and with broad participation. That reputation led to their rapid expansion after the creation of the first social fund in Bolivia in 1987.

By May 2001 the World Bank had financed more than 98 social fund projects in 58 countries. Almost all countries in Latin America and the Caribbean have social funds or development projects that embody many of their operational

Laura B. Rawlings is Senior Monitoring and Evaluation Specialist and Norbert R. Schady is Senior Economist, both in the World Bank's Latin America and the Caribbean Region. Their e-mail addresses are lrawlings@worldbank.org and nschady@worldbank.org, respectively.

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

characteristics. Many countries in Africa, Asia, Europe, and the Middle East have also established social funds.¹ These social funds have absorbed more than US\$8 billion in investments by the World Bank, other international agencies, and governments.² Nonetheless, at the national level social fund spending remains relatively limited, with total expenditures usually accounting for no more than 1 percent of gross domestic product (Rawlings and others 2002).

The scope and financial scale of the global portfolio of social funds, along with the renewed interest in community-based development models, sparked demand for the evaluation of social funds among national governments and development institutions alike.

III. WHAT CHALLENGES DO SOCIAL FUNDS POSE FOR IMPACT EVALUATION?

Social funds pose particular challenges for impact evaluation. Perhaps the two most important ones relate to the nonrandom placement of projects. Most social funds use a poverty map and other tools to target their investments. But allocation rules may rely at least in part on characteristics of communities that are observed by the social fund administrator but not by the evaluator. If these characteristics affect outcomes, the evaluator's inability to adequately control for them can bias the results. For example, if a social fund spends more in poor communities and poverty is correlated with poor health status and imperfectly observed by the evaluator, estimates of the social fund's impact on health outcomes may be biased downward.

The second challenge arises because communities generally self-select into social fund projects. This, too, can bias the results. For example, if a social fund spends more on school infrastructure in communities that have greater (unobserved) organizational capacity and that in the absence of the project would have been more likely to find other solutions for decaying school infrastructure, estimates of the social fund's impact on educational attainment may be biased upward.

Also complicating the impact evaluation of social funds is the range of objectives they address. This makes selecting valid outcome indicators difficult and comparing outcomes across countries even more so. Despite these challenges, the four evaluation studies in this issue shed light on the ability of social funds to reach poor communities and households and the impact of social fund investments on a number of outcomes.

1. For periodically updated information on social funds, go to the World Bank Social Fund database Web site at <http://worldbank.org/sp> (click on the link for "social funds"). Questions about the social fund Web site or database can be directed to the Social Protection Advisory Service, 1818 H Street NW, Room G8-138, Washington, D.C. 20433 (phone 202-458-5267; fax 202-614-0471; e-mail socialprotection@worldbank.org).

2. Of the \$8 billion in investments in social funds, the World Bank accounts for about \$3.5 billion. This total excludes social funds that do not receive World Bank financing and are instead financed by other multilateral and bilateral sources or solely through domestic resources.

IV. WHAT METHODS DO THE FOUR EVALUATIONS USE?

The articles that follow present four different approaches to the same basic challenge: determining whether social fund investments have led to changes in the well-being of beneficiaries.³ The evaluations have used a range of methods, including randomization, propensity score matching, and instrumental variables. Some have relied on several approaches, which can provide a useful check on the robustness of the assumptions underlying different estimates. In each case the choice of evaluation methods reflects available data, time, and resources as well as the particular focus of the different evaluations.

Randomization

Randomization assigns the “treatment”—in this case a social fund intervention—through some sort of lottery, allowing researchers to construct treatment and control groups. In a study of the Bolivian social fund, Newman and others use randomization of the offer to participate in a social fund project to evaluate the impact of improvements in school infrastructure on a variety of school outcomes in the rural Chaco region.

Randomization is immensely appealing because if the sample is large enough, this method controls for all differences, observable and unobservable, between the treatment and control groups. Simple differences in mean outcomes between the two groups, or differences in changes in outcomes, can then be credibly interpreted as the impact of the treatment on the treated.

But the evaluation of the Bolivian social fund also shows some of the limitations of randomization and some of the challenges that social-sector programs pose for evaluation. The evaluation had to deal with changes that occurred after the allocation of the offer to participate, as a result of which some schools not selected for treatment ended up receiving the social fund intervention. Newman and others use bounds estimates to correct for these changes.

Techniques Matching

Propensity score matching in its simplest form involves predicting the probability of treatment on the basis of observed covariates for both the treatment and the control group samples. This probability, the propensity score, is then used to match treated and untreated observations—for example, through nearest-neighbor matching. Under some conditions the difference in mean outcomes between the two groups is then a reasonable estimate of impact.

Propensity score matching is often fairly simple to carry out. For example, if a national household survey has recently been administered, a separate survey of beneficiaries can be fielded using the same questionnaire. The results of this survey can then be combined with those of the national survey to construct treatment and comparison groups through propensity score matching.

3. A good summary of evaluation techniques can also be found in Ravallion (2001).

Still, propensity score matching requires an exhaustive questionnaire to accurately match treated and untreated populations based on their observable characteristics. It demands careful consideration of the extent to which unobserved differences remain between the two samples. Three of the articles in this issue—Chase on Armenia, Newman and others on Bolivia, and Pradhan and Rawlings on Nicaragua—use propensity score matching to estimate impact. In addition, the Armenia evaluation uses a “pipeline” method to minimize possible bias arising from any residual unobserved differences. The logic of this approach is appealing: If there are unmeasured characteristics that make some groups more likely to apply for or receive a social fund project, it should be possible to match outcomes in communities that have already been “treated” with others in the “pipeline”—that is, communities that have self-selected and been preapproved for the next round of interventions.

Instrumental Variables

Two-stage least squares estimation attempts to mimic an experimental design. It relies on a variable, the instrument, which is assumed to be correlated with the probability of treatment but uncorrelated with unobserved determinants of outcomes. Under these circumstances instrumental variables can purge the estimation of treatment effects of such problems as measurement error, reverse causality, or nonrandom program placement. But because instrumental variable estimates are predicated entirely on the validity of the instrument, any correlation between the instrument and unobserved determinants of treatment effects can result in serious biases.

Paxson and Schady, in an evaluation of the Peruvian social fund, use two-stage least squares to estimate the impact of the fund’s investments in school infrastructure on attendance rates. They use the distribution of the progovernment vote as an instrument for social fund spending, building on earlier work showing that changes in the distribution of the vote between 1990 and 1993 affected the distribution of social fund resources in Peru.

V. WHAT ARE THE LESSONS FOR EVALUATION AND DIRECTIONS FOR FUTURE RESEARCH?

The four social fund evaluations in this issue provide plausible estimates of the impact of social fund investments on a number of outcomes, including the availability of infrastructure in communities and the education and health status of beneficiary households. Equally important, they use a range of evaluation techniques to produce those estimates. What lessons do these evaluations of social funds offer for the impact evaluation of social-sector interventions more broadly?

First, when little is known about the likely impact of an intervention and accurately estimating that impact would provide an important public good, randomization is still the most convincing choice of evaluation technique. The random

selection of treatment and control groups provides a solid basis for an impact evaluation even when randomization is on project promotion.

Second, when randomization is not an option, because of resistance to randomization or a desire to reach all eligible beneficiaries, researchers and policymakers should be opportunistic. Credible instrumental variables are hard to come by, but when available they can provide convincing estimates of impact. Propensity score matching is a promising approach to constructing a counterfactual based on similarities between treatment comparison groups. All the evaluations in this issue except the Bolivian case applied propensity score matching to existing data sets (such as household surveys), often supplemented with data collected on social fund beneficiaries—a practical way of reducing the cost of an impact evaluation.

For social funds, two areas of future research seem particularly fertile. An important one is to estimate whether social funds are more (or less) cost-effective than comparable interventions in achieving a particular impact, such as raising enrollment or reducing infant mortality. Such information is indispensable in making the kinds of tradeoffs that policymakers face daily. The evaluations in this issue provide a benchmark against which other interventions could be measured. A second useful exercise would be to estimate the impact of social funds on other outcomes, such as the organizational capacity and social capital of beneficiary communities.

REFERENCES

- Ravallion, Martin. 2001. "The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation." *World Bank Economic Review* 15(1):115–40.
- Rawlings, Laura B., Lynne Sherburne-Benz, and Julie van Domelen. 2002. "Evaluating Social Fund Performance: A Cross-Country Analysis of Community Investments." World Bank, Social Protection Network, Washington, D.C.

Impact Evaluation of Social Funds

Supporting Communities in Transition:
The Impact of the Armenian
Social Investment Fund

Robert S. Chase

The Armenian Social Investment Fund supports communities' efforts to improve local infrastructure during Armenia's economic transition away from central planning, financing community-designed and -implemented projects to rehabilitate primary schools, water systems, and other infrastructure. This article considers the targeting, household impact, and community effects of the social fund's activities. It relies on a nationally representative household survey, oversampled in areas where the social fund was active. Using propensity and pipeline matching techniques to control for community self-selection into the social fund, it evaluates the household effects of rehabilitating schools and water systems. The results show that the social fund reached poor households, particularly in rural areas. Education projects increased households' spending on education significantly and had mild effects on school attendance. Potable water projects increased household access to water and had mild positive effects on health. Communities that completed a social fund project were less likely than the comparison group to complete other local infrastructure projects, suggesting that social capital was expended in these early projects. By contrast, communities that joined the social fund later and had not yet completed their projects took more initiatives not supported by the social fund.

In centrally planned economies, national governments exerted tremendous economic control. This control extended to investment in local infrastructure, including building and maintaining roads, schools, and water systems. When these economies collapsed, governments became bereft of resources. Systems for maintaining local infrastructure began to fail; as deep economic recession took hold, schools and water systems fell into disrepair. Local public services deteriorated, compounding other hardships for people living in postcommunist conditions. But because communities were accustomed to relying on central authorities to meet local needs, they often were unable to address their problems.

Robert S. Chase is an assistant professor of International Economics at the School of Advanced International Studies of Johns Hopkins University; Senior Economist and Social Capital Coordinator in the World Bank's Social Development Family; and Research Associate of the William Davidson Institute. His e-mail address is rchase@worldbank.org. The research underlying this article progressed thanks to support from the Armenian Social Investment Fund, the State Department of Statistics of Armenia, the Social Funds 2000 study, Caroline Mascarell, and Lynne Sherburne-Benz. The author benefited from extensive assistance in data collection and analysis from Ghislaine Delaine, Julia Grutzner, Kalpana Mehra, Mirvat Sewadeh, and Sylvia Zucchini, and from helpful comments from Julia Magluchants and Amalia Poghosian.

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

The World Bank supported social funds in postcommunist transition economies as part of a strategy to improve this situation. These flexible financing instruments provide resources for community initiatives to improve local infrastructure. In the former Eastern Bloc countries social funds sought to provide temporary employment, alleviate local public service hardships, and mobilize communities to address local needs. This article investigates the household and community impact of one such project, the Armenian Social Investment Fund (ASIF). It seeks to ascertain whether ASIF resources reached poorer households, whether the social fund altered the behavior or welfare of those in ASIF communities, and how the fund related to communities' ability to act collectively.

There have been many investigations of social funds since their introduction in Latin America at the end of the 1980s (see, for example, Goodman and others 1997, Khadiagala 1995, and Marc and others 1993). These have adopted many different evaluative approaches, including analyses of fund disbursement patterns, institutional studies of operating procedures, and sociological studies of beneficiaries' attitudes. But by the mid-1990s household data and quantitative techniques had been used only in evaluating the Bolivian social fund. To provide deeper, more diverse evidence on the impact of social funds, in 1997 the World Bank initiated a multicountry analysis, *Social Funds 2000*. This article comes out of that research program. *Social Funds 2000* used household survey data to isolate statistically significant social fund effects in Armenia, Bolivia (Newman and others 2002), Honduras (Walker and others 1999), Nicaragua (Pradhan and Rawlings 2002), Peru (Paxson and Schady 2002), and Zambia (Chase and Sherburne-Benz 2001). This article on Armenia is the first quantitative impact evaluation to consider a social fund in a postcommunist economy.

I. THE ASIF

Though social funds share common characteristics, each is designed to fit the objectives and institutions of the country operating it. The ASIF began as a pilot project, funded by the U.S. Agency for International Development, to provide employment, support community initiatives, and enhance civil society during the postcommunist transition. In January 1996 the first World Bank loan for ASIF became effective, providing US\$12 million in concessional financing for a \$20 million project. Although some of this funding helped develop institutions necessary to administer the social fund and monitor household welfare, most went to support projects that communities designed and implemented.

Between 1996 and December 2000, when the first World Bank loan for ASIF closed, the social fund received proposals for 726 projects, of which 334 were approved and 259 completed. Over the course of the loan the average project size was \$50,000. As a direct result of World Bank financing, 178 contractors implemented projects, providing jobs to 5,000 people. The infrastructure improvements reached an estimated 640,000 beneficiaries.

Like other social funds, ASIF offered a menu of projects from which communities chose. It specified types of projects that would meet the country's pressing needs in local infrastructure and would likely be interesting only to communities in difficult circumstances, thus self-targeting ASIF resources to the poor. Of the 259 completed projects in 1996–2000, 35 percent were small-scale school rehabilitations, 32 percent potable water projects, 11 percent minor irrigation works, and 5 percent rehabilitations of health facilities. The other 17 percent included initiatives for community centers, pension homes, roads and landscaping, and sewage and waste management. The focus here is on primary school rehabilitations and water projects. These two groups not only are the largest but also consist of projects that are relatively homogeneous, allowing easier comparisons within each group.

Though communities stepped forward to participate in ASIF, the social fund administration also targeted resources to areas of the country with the most pressing need for small-scale infrastructure improvements. From 1996 to 2000, 38 percent of projects were in *marzes* (regions) where the 1988 earthquake caused the most devastation—Aragosotn, Lori, and Shirak. Marzes that suffered most from the Karabakh conflict—Sunik and Tavush—implemented 21 percent of the projects. Yerevan received 25 percent. Fifteen percent was spread among the remaining five marzes. As will be discussed, though this regional distribution of projects focused resources on marzes in the most difficult circumstances, it created technical challenges for evaluation. Notably, the targeting of resources to specific areas makes it difficult to identify control communities that did not participate in the ASIF project but otherwise had characteristics similar to those of participating communities.

II. DATA AND METHODOLOGY

To analyze changes in household behavior and outcomes, the article relies primarily on an integrated household survey. This comprehensive, nationally representative data source allows in-depth analysis of the welfare of the Armenian population. Among other topics, the core survey instrument includes information on household composition, income, expenditures, education, and health. Conducted from July 1, 1998, to June 30, 1999, the survey includes roughly 3,600 households in its basic sample. Enumerators visited 20 households per sample cluster.

To allow impact evaluation, ASIF staff and the State Directorate of Statistics added a module to the survey instrument that posed questions about ASIF activities and community organizations and initiatives. It asked households to report changes to community infrastructure that had taken place in the previous five years. It also asked whether they had taken part in the effort to repair or upgrade the infrastructure and what their attitudes were toward the resulting infrastructure.

To ensure adequate coverage of ASIF treatment areas, the survey oversampled households in areas where the social fund was active. Beyond the base sample

representative of the Armenian population, survey enumerators visited an additional 2,260 households in 113 clusters where the social fund was active. Within this group of oversampled communities, the survey collected data on two groups of households: those where projects had been completed and those where ASIF had approved a project but the project had not yet been completed.

As the evaluation literature has long emphasized, it is difficult to isolate the effects of an intervention, particularly when potential participants chose to involve themselves in the intervention (for an overview of key evaluation issues, see Moffitt 1991). Fundamentally, impact evaluation compares outcome indicators for a group that completed a project—the treatment group—with those for a comparison group. If the comparison group is correctly identified, the difference between the treatment and comparison groups isolates the effect of the intervention. But in many cases, including that of social funds, identifying appropriate comparison groups can be difficult.

Because of the way social funds operate, communities wanting to participate must organize themselves to earn funding for their initiatives. Before the social fund has disbursed any resources, a treatment community distinguishes itself from its neighbors by assembling a project committee and proposal. Thus a simple “with-and-without” comparison for social fund participants and nonparticipants is biased. It mistakenly attributes to the social fund the community selection effects that encourage participation.

Randomized control design avoids these selection problems in creating a comparison group by randomly selecting parts of the country where the social fund can and cannot operate. But like most other social funds (the Bolivian social fund being the exception), ASIF did not randomly choose where it would operate, precluding this robust evaluative approach. But alternative techniques allow the impact evaluation to generate treatment and comparison groups for ASIF.

Information collected about households in communities where ASIF had approved a project but the project had not yet been completed makes it possible to establish a pipeline comparison group. These communities have demonstrated that they can organize themselves for social fund projects, so there is no selection bias. But they have not yet gained the benefits of the projects. So a comparison of these pipeline communities with those that have completed projects offers insight into the effects of those projects, abstracting away from characteristics that led communities to participate in ASIF.

In addition to pipeline matching, this evaluation also uses propensity score matching to correct for selection biases.¹ To create a comparison group, the traits of communities that participated in ASIF are analyzed, and then a propensity func-

1. Several studies have used the propensity score matching approach to evaluate impact. Some have used it for individual-level interventions (Heckman and others 1997, 1998). Among those applying it to community-level interventions are several contributions to the Social Funds 2000 study: Newman and others (2002) look at Bolivia, Pradhan and Rawlings (2002) consider Nicaragua, and Chase and Sherburne-Benz (2000) analyze Zambia.

tion is generated that links the characteristics of a community to the likelihood that it will submit a successful proposal for a social fund project.

In Armenia geography was a crucial determinant of which communities participated in the social fund. Because ASIF focused resources on the earthquake and conflict zones, communities in these areas were more likely to participate in ASIF, and the basis for their inclusion in the program differed from that in other areas. For these reasons three separate propensity functions are estimated to stratify the sample by the earthquake zone, the conflict zone, and nontargeted zones. The propensity functions isolate the effects of community means for household expenditures, share spent on food, female headship, and education levels in each of these zones.

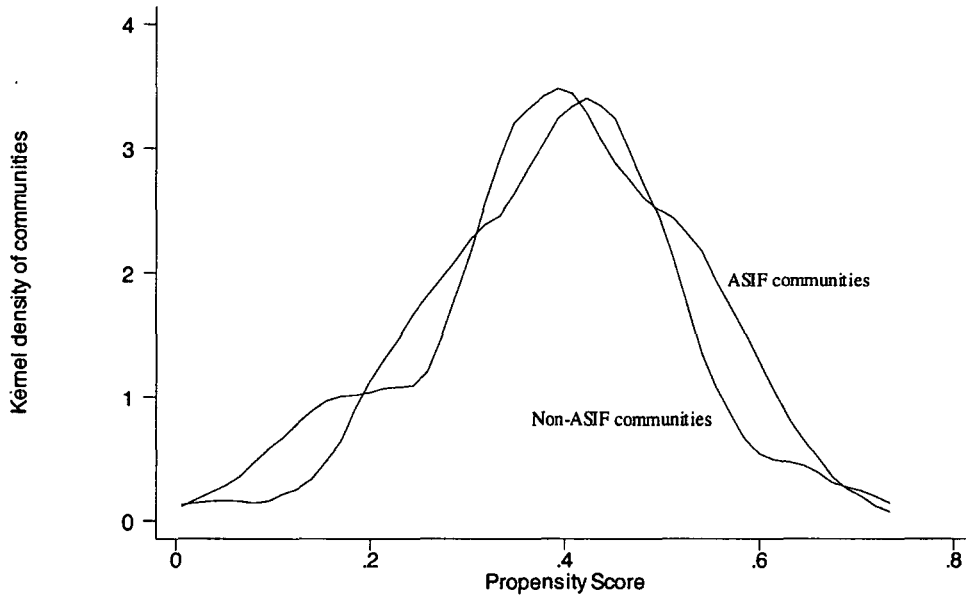
Estimates of propensity function parameters for each of the subsamples are used to predict the probability of program participation for all community clusters, by pooling those that participated and those that did not (see appendix). These probabilities are propensity scores.

To create a comparison group of communities whose propensity to participate in the social fund was comparable to that of treatment communities, each community that completed an ASIF project is matched with a community in the same zone that did not participate but that had an equivalent propensity score. This procedure creates a comparison group of communities just as likely to participate in the social fund as the treatment group. The difference between treatment and comparison thus isolates the effect of implementing the social fund project, abstracting away from traits that led communities to work with the social fund.

Although the idea of the propensity score matching procedure is clear, its application in Armenia is challenging. Notably, for treatment communities with very high propensity scores, it is not always possible to find a control community equally likely to participate in the social fund. The distribution of propensity scores before matching for ASIF communities (those that participated) and comparison communities (those that did not) shows that proportionately more ASIF communities had high propensity scores in nontargeted zones as well as in the conflict and earthquake zones (figures 1–3). Thus, treatment communities differ from randomly selected communities in their likelihood to participate in ASIF. Because of this difference, randomly selected communities are an inadequate comparison group.

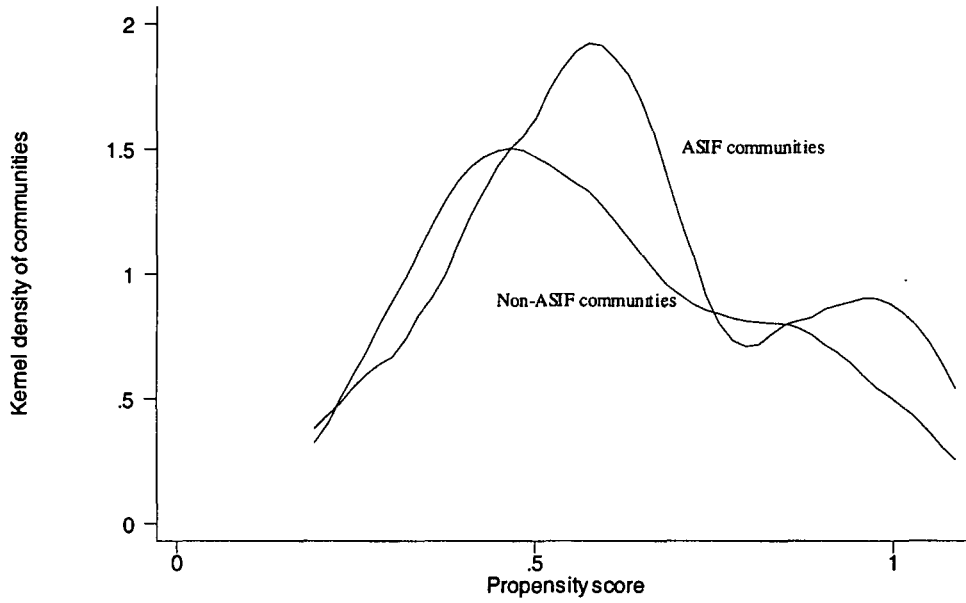
After communities are matched by propensity score, the treatment and comparison groups would have the same distributions of propensity scores if the matching procedures had worked perfectly. Although the distributions do become more similar after the matching procedure, in each case they are still distinct (figures 4–6). For many of the communities that completed ASIF projects estimated propensity scores are very close to one, signifying that the communities were almost sure to participate. For these communities there are no matches in the same zone—that is, communities that almost assuredly should have participated but did not. A comparison group created from each of these zone-specific propensity matches does not adequately correct for selection bias.

FIGURE 1. Pre-Match Propensity Score Distributions for ASIF versus Non-ASIF Communities in Non-Targeted Zones



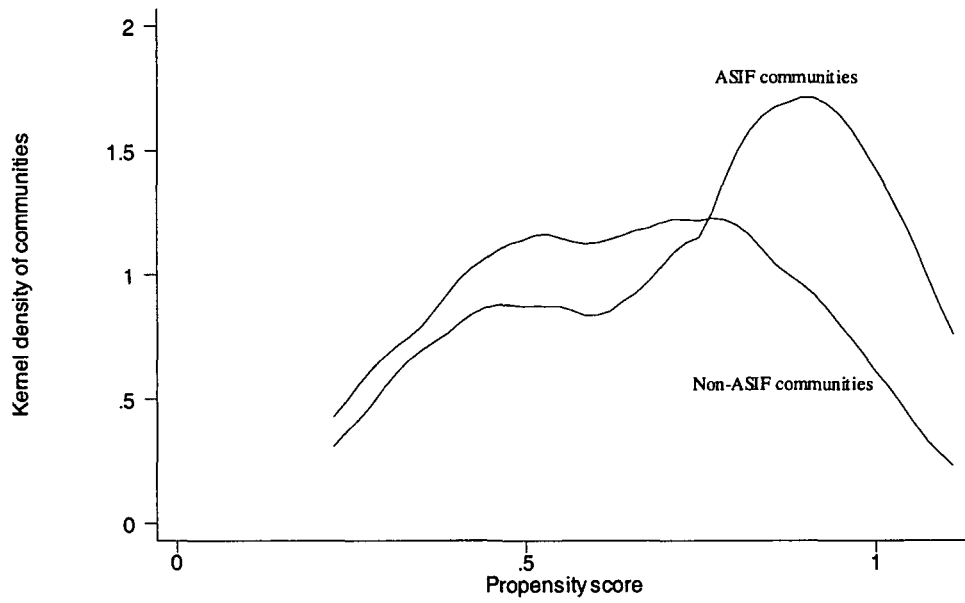
Source: Author's calculations.

FIGURE 2. Pre-Match Propensity Score Distributions for ASIF versus Non-ASIF Communities in Conflict Zones



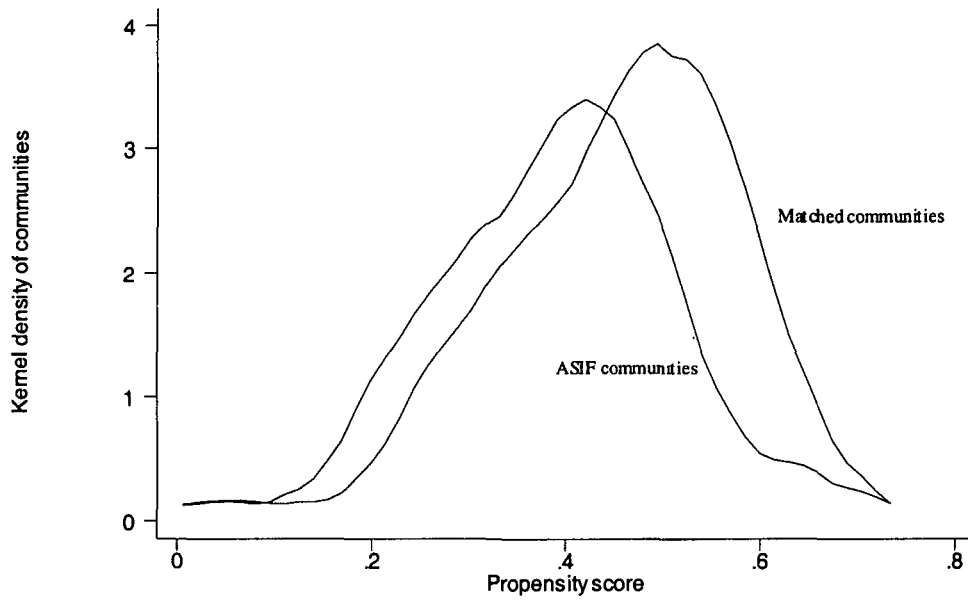
Source: Author's calculations.

FIGURE 3. Pre-Match Propensity Score Distributions for ASIF versus Non-ASIF Communities in Earthquake Zones



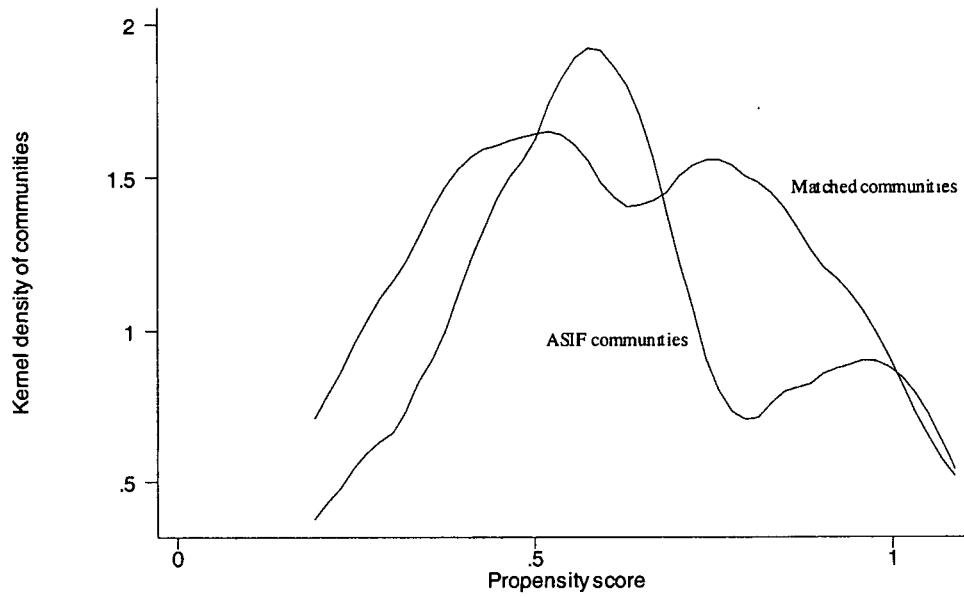
Source: Author's calculations.

FIGURE 4. Post-Match Propensity Score Distributions for ASIF versus Matched Communities in Non-Targeted Zones



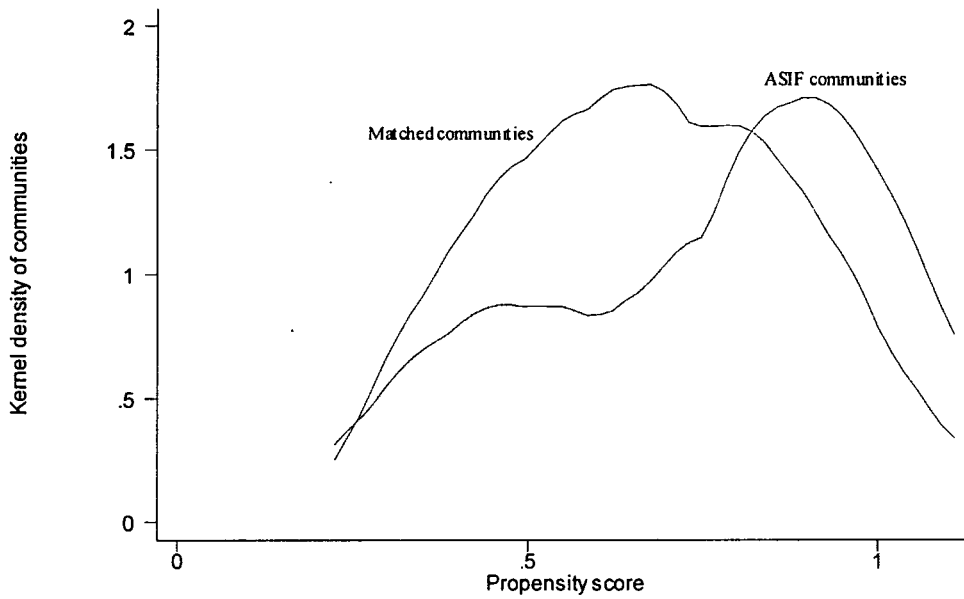
Source: Author's calculations.

FIGURE 5. Post-Match Propensity Score Distributions for ASIF versus Matched Communities in Conflict Zones



Source: Author's calculations.

FIGURE 6. Post-Match Propensity Score Distributions for ASIF versus Matched Communities in Earthquake Zones



Source: Author's calculations.

III. EFFECTIVENESS OF TARGETING

Social funds support project options that more less well-off communities find attractive. They also use administrative efforts to target poorer areas. Through a combination of these two strategies, social funds purport to focus resources on a country's poorer communities. Household expenditure data from communities where ASIF was active and from randomly selected communities give insights into whether ASIF resources reached relatively less-well-off communities.

The household expenditure data show that people in ASIF communities are poorer (on average) than other Armenians (table 1). To a statistically significant degree, ASIF households across the country spent less per capita than randomly selected households. Furthermore, households in ASIF communities devoted a larger share of their expenditures to food (82 percent) than did non-ASIF households (80 percent), a robust indicator of higher relative poverty.

These differences remain when the country is divided into urban and rural areas. In urban areas ASIF households spent significantly less per capita (11,800 drams/month) than non-ASIF households (13,100 drams) and directed a significantly larger share of their spending to food. Similarly, in rural areas ASIF households spent less than non-ASIF households and allocated significantly more of their money to food (85 percent, compared with 83 percent for non-ASIF households).

Thus households in social fund communities are poorer on average than other Armenian households. But did the social fund reach the poorest Armenians? Concentration curves showing the distribution of household per capita expenditures in ASIF communities and in the entire Armenian population offer an answer to this question. If the distribution of poverty among ASIF households was the same as that among the rest of the Armenian population, ASIF targeting would be neutral with regard to poverty, and the concentration curve would correspond to the 45° line. But if, say, the 20th percentile of ASIF households had the same income as the 10th percentile of all Armenian households, that would show that ASIF resources were being allocated progressively, targeting the relatively poor. Concentration curves above the 45° line indicate propoor targeting.

Concentration curves for the country as whole, for urban areas, and for rural areas are all fairly close to the 45° line, showing that ASIF targeting is relatively neutral (figure 7). Urban spending appears slightly progressive because the concentration curve is above the 45° line for all parts of the household expenditure distribution. By contrast, at lower parts of the distribution rural spending is slightly regressive.

These findings are notable and somewhat surprising. Studies of social funds in other countries have generally found rural spending to be more progressive and urban spending generally more regressive (see Chase and Sherburne-Benz 2001, Newman and others 2000, and Pradhan and Rawlings 1999). ASIF targeted areas with poor infrastructure, such as the conflict and earthquake zones, where rehabilitating schools and water systems could have a large direct effect. But it did not

TABLE 1. Monthly Household Expenditures in ASIF and Non- ASIF Communities, 1998–99

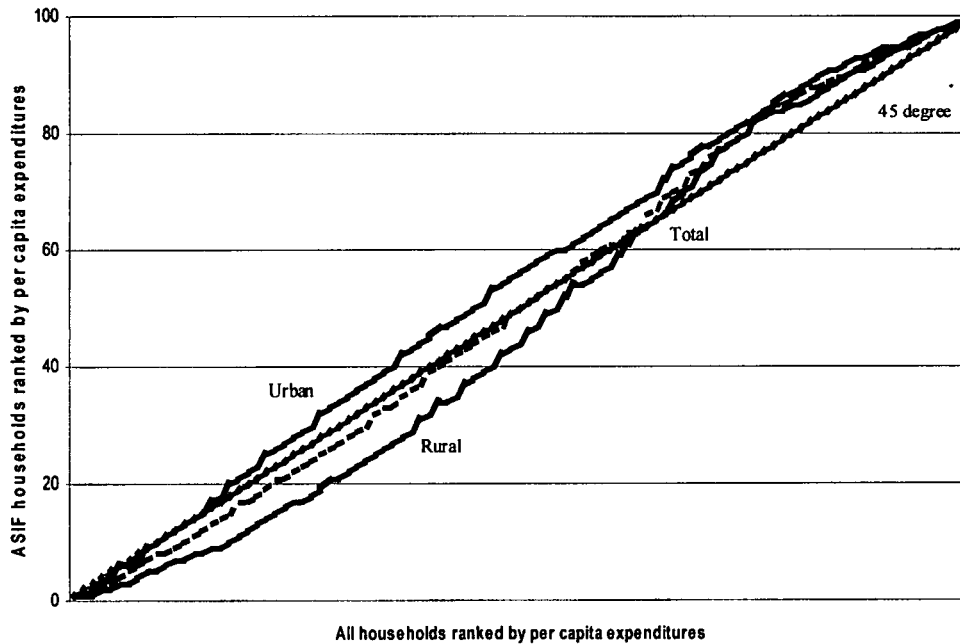
Indicator	Total			Urban			Rural		
	Non-ASIF communities	ASIF communities	<i>t</i> -statistic	Non-ASIF communities	ASIF communities	<i>t</i> -statistic	Non-ASIF communities	ASIF communities	<i>t</i> -statistic
Total expenditures (drams)	51,654	48,814	1.40	47,934	46,233	0.67	57,364	51,259*	1.86
Per capita expenditures (drams)	13,268	12,554*	1.66	13,144	11,762**	2.33	13,495	13,305	0.25
Share spent on food (%)	79.8	82.1**	6.03	77.8	78.8*	1.93	82.9	85.2**	4.45

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

FIGURE 7. Concentration Curves of ASIF Targeting in Relation to Armenian Population as a Whole



Source: Author's calculations from household data.

explicitly target areas with low household expenditure. In Armenia poor infrastructure does not appear where household expenditure is lowest.

The regressive rural targeting may result from the 10 percent contribution that ASIF requires of communities. According to ASIF staff, in rural areas this requirement selects against the poorest communities. Households there are unwilling or unable to contribute for community public goods, such as schools or improved water systems.

Though relatively progressive urban targeting is also unusual for social funds, ASIF focused its activity on the capital city, Yerevan, one of the least well-off areas of the country. Here, ASIF activities reached the poorest communities. With many ASIF projects in Yerevan, progressive targeting there implies that ASIF reached Armenia's poorest households.

IV. IMPACT OF EDUCATION AND WATER PROJECTS

To investigate the household impact of ASIF school rehabilitation projects, the impact evaluation compares education outcomes for households in communities where ASIF school projects were completed, households in propensity matched communities, households in pipeline communities, and an unmatched set of randomly selected households. The results show that across the country as a whole,

ASIF generated few significant differences in how much households with primary-school-age children spent on schooling (table 2). In earthquake zones, however, households in ASIF communities spent 22 percent more than those in the matched comparison group and 27 percent more than those in the pipeline control group. The opposite appears to be the case in conflict zones, where ASIF households spent significantly less than either of the control groups. Expenditures by the treatment group averaged 2,125 drams/month compared with 4,062 drams in the matched comparison group and 3,600 drams in the pipeline comparison group. These findings point to different social fund effects in earthquake and conflict zones.

There is also evidence that school enrollments were higher in communities in which ASIF supported school rehabilitation. In treatment communities 87 percent of primary school-age children were in school. This is significantly higher than the 79 percent in school in communities where an ASIF project had been approved but not yet completed.

In earthquake zones households near ASIF-supported school rehabilitation projects not only spent more on schooling but also were more likely to have their children attend school. Together, these pieces of evidence suggest that ASIF increased demand for education. If the quality of school facilities in the earthquake zones had been low, renovations financed by ASIF would make the schools more attractive, increasing demand for primary education.

Subjective measures of changes in the quality of school services provide interesting insight into the impact of ASIF. Although 8 percent of households in ASIF communities reported that school services had improved in the previous 12 months, fewer (6 percent) in matched comparison communities said that schools had recently improved, suggesting that people in ASIF communities had more positive impressions of their schools. In pipeline comparison communities, however, 12 percent reported improvements, significantly more than in the treatment communities. What explains this inconsistency? It is possible that ASIF projects had been completed some time ago, so that impressions of improvement were remote, and there had been recent discussion of school improvements in pipeline comparison communities. Overall, however, few households in Armenia believed that schools were improving.

Besides school rehabilitations, ASIF also supported community projects to improve local water supply. For the impact analysis of these projects, the primary indicators of interest are household access to water and sanitation. Household data offer evidence that the projects improved access. For example, in ASIF treatment communities 93 percent of households had access to cold running water, compared with 85 percent in matched communities and 72 percent in pipeline communities (table 3). Further, 92 percent of households in ASIF communities had central water systems, significantly more than in both matched communities (83 percent) and pipeline communities (68 percent). Finally, compared with pipeline communities, more households had an indoor tap in ASIF commu-

TABLE 2. Household Effects of ASIF-Supported School Rehabilitation Projects

Indicator	Treatment communities	Matched communities	<i>t</i> -statistic	Pipeline communities	<i>t</i> -statistic	Unmatched communities	<i>t</i> -statistic
Monthly household expenditures on schooling (drams)	3,105	3,627	1.32	2,640	1.44	3,967**	2.06
Earthquake zones	3,873	3,391	0.69	2,808**	1.98	3,790	0.12
Conflict zones	2,125	4,062**	2.13	3,600**	2.18	5,105	1.56
Log monthly household expenditures on schooling	7.84	7.90	0.91	7.76	0.92	7.95*	1.71
Earthquake zones	8.02	7.80*	1.94	7.75**	2.04	7.84	1.65
Conflict zones	7.62	8.18**	3.42	7.98	1.36	8.20**	2.90
Proportion of 7- to 12-year-olds attending school	0.87	0.83	1.52	0.79**	2.30	0.83	1.59
Earthquake zones	0.86	0.80	1.65	0.76**	2.15	0.80*	1.64
Conflict zones	0.93	0.84	1.04	0.83	0.75	0.83	1.24
Proportion of households reporting that school service improved in previous 12 months	0.08	0.06**	2.17	0.12*	1.84	0.07	1.44
Earthquake zones	0.11	0.10	0.30	0.21**	3.16	0.08	1.28
Conflict zones	0.13	0.01**	4.52	0.00*	1.66	0.03**	3.46
Number of observations	232	646		148		1,298	
Earthquake zones	113	150		80		247	
Conflict zones	20	87		5		208	

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

TABLE 3. Household Effects of ASIF-Supported Potable Water Projects

Indicator	Treatment communities	Matched communities	<i>t</i> -statistic	Pipeline communities	<i>t</i> -statistic	Unmatched communities	<i>t</i> -statistic
<i>Proportion of households with</i>							
Indoor water tap	0.68	0.69	0.46	0.43**	7.12	0.67	0.32
Earthquake zones	0.81	0.71**	2.53	0.90*	1.75	0.69**	3.20
Conflict zones	0.90	0.38**	7.92	0.28**	9.74	0.50**	6.10
Central water system	0.92	0.83**	4.17	0.68**	8.28	0.84**	4.05
Earthquake zones	0.94	0.74**	5.32	0.99*	1.75	0.72**	5.93
Conflict zones	0.85	0.58**	4.04	0.49**	4.93	0.71**	2.31
Flush toilet	0.52	0.58**	2.22	0.33**	5.29	0.58**	2.20
Earthquake zones	0.65	0.55**	2.08	0.71	0.97	0.55**	2.21
Conflict zones	0.50	0.25**	3.94	0.21**	4.17	0.38*	1.88
Cold running water	0.93	0.85**	3.95	0.72**	7.67	0.78**	2.33
Earthquake zones	0.93	0.80**	3.75	0.96	0.97	0.76**	4.91
Conflict zones	0.92	0.78**	2.49	0.53**	5.63	0.82*	1.83

<i>Proportion of households reporting (for previous 12 months)</i>							
Water service improvements	0.34	0.22**	4.78	0.28*	1.81	0.21**	5.83
Earthquake zones	0.52	0.30**	5.13	0.60	1.19	0.24**	7.25
Conflict zones	0.083	0.14	1.20	0.15	1.26	0.15	1.40
Sanitation improvements	0.09	0.09	0.16	0.06	1.57	0.08	0.31
Earthquake zones	0.18	0.13	1.46	0.26	1.46	0.11**	2.48
Conflict zones	0.02	0.00**	2.01	0.00	1.42	0.03	0.50
Illness	0.13	0.18**	2.68	0.17**	1.99	0.20**	4.07
Earthquake zones	0.16	0.17	0.45	0.06**	2.86	0.20	1.57
Conflict zones	0.09	0.20**	2.38	0.24**	2.81	0.20**	2.52
Inactivity due to illness	0.15	0.17	0.69	0.25**	2.80	0.20*	1.87
Earthquake zones	0.15	0.15	0.00	0.18	0.42	0.20	1.49
Conflict zones	0.19	0.27	0.76	0.34	1.33	0.26	0.66
Ill children	0.03	0.03	0.06	0.03	0.13	0.04	0.95
Earthquake zones	0.05	0.03	0.12	0.01	1.49	0.04	0.61
Conflict zones	0.02	0.02	0.07	0.03	0.52	0.04	0.59
Number of observations	340	1,740		380		3,600	
Earthquake zones	160	420		80		700	
Conflict zones	60	240		120		580	

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

nities. Thus, households in ASIF communities had greater access to water by several measures—across Armenia and in its earthquake and conflict zones.

Subjective measures of changes in water service indicate that households in ASIF communities were more likely to report improvements in the previous 12 months than either of the comparison groups. Within the treatment group 34 percent said that water service had improved, compared with 22 percent in the matched comparison group and 28 percent in the pipeline comparison group. But ASIF communities were no more likely to report improvements in sanitation.

Although this impact evaluation focuses on output indicators, comparing changes in facilities and their availability to households, ultimately ASIF projects seek to improve the welfare of people. Few indicators are available to measure these final outcomes, though there is some weak evidence that ASIF potable water interventions improved beneficiaries' health. Fewer households (13 percent) in treatment communities reported illness than in both matched (18 percent) and pipeline (17 percent) comparison groups. ASIF households were less likely to report inactivity due to illness (15 percent) than were those in the pipeline comparison group (25 percent). But there was no statistically significant effect on the proportion of households reporting ill children in ASIF communities compared with either of the comparison groups. Thus, although ASIF water supply interventions may have had some impact on these health indicators, the data show few strong effects.

V. COMMUNITY PARTICIPATION

Social fund support to local initiatives is often described as having important community impacts that may not be reflected in the welfare of individual households. In post-Soviet societies, where many look reflexively to central governments to solve local problems, a program demonstrating that the community can act collectively might have a positive impact on that institution. Indeed, ASIF beneficiary assessments show that community participation in the social fund can change community attitudes, encouraging people to solve local problems through their own efforts.

Further, as the literature on social capital suggests (see, for example, Grootaert and van Bastelaer 2001), communities that act collectively may be better endowed with positive structural or cognitive social capital. Though there is extensive debate about how to define and measure social capital (see, for example, Grootaert 1997, Knack and Keefer 1997, and Woolcock 1998), this analysis focuses on collective action, looking at the likelihood that communities that completed an ASIF-supported school rehabilitation or potable water project also undertook other recent community initiatives.

Information about ASIF education activities is not universally shared. In communities where ASIF financed the rehabilitation of a school, 49 percent of households reported that this change had recently taken place (table 4). In those where ASIF financed the rehabilitation of the water system, 46 percent were aware of it (table 5). These results could mean that the rehabilitations had taken place some

TABLE 4. Likelihood That Communities Completing ASIF-Supported School Rehabilitation Projects Undertake Other Community Infrastructure Projects

Variable	Treatment communities	Matched communities	<i>t</i> -statistic	Pipeline communities	<i>t</i> -statistic	Unmatched communities	<i>t</i> -statistic
Build new school	0.21	0.15**	3.48	0.38**	6.15	0.11**	7.46
Build health facility	0.08	0.09	1.05	0.18**	4.92	0.06**	2.04
Rehabilitate health facility	0.11	0.14*	1.75	0.24**	5.79	0.11	0.09
Build new road	0.02	0.04**	2.07	0.05**	2.79	0.04**	2.10
Rehabilitate road	0.24	0.29**	2.46	0.35**	4.15	0.25	0.93
Build or rehabilitate piped water	0.07	0.15**	5.12	0.30**	10.37	0.17**	6.47
Build or rehabilitate reservoir	0.04	0.03	1.24	0.04	0.46	0.03	0.50
Build or rehabilitate sanitation	0.03	0.01**	3.80	0.01	-0.07	0.03**	3.86
Proportion of households reporting school rehabilitation	0.49	0.25**	11.64	0.51	0.49	0.22**	14.26
Number of observations	620	1,740		439		3,600	

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

TABLE 5. Likelihood that Communities Completing ASIF-Supported Potable Water Projects Undertake Other Community Infrastructure Projects

Variable	Treatment communities	Matched communities	<i>t</i> -statistic	Pipeline communities	<i>t</i> -statistic	Unmatched communities	<i>t</i> -statistic
Build new school	0.11	0.15**	2.28	0.21**	3.93	0.11**	0.14
Rehabilitate school	0.28	0.25	1.44	0.36**	2.32	0.22	2.47
Build health facility	0.07	0.10*	1.77	0.06	0.53	0.06	0.41
Rehabilitate health facility	0.18	0.14*	1.79	0.19	0.36	0.11**	3.67
Build new road	0.08	0.04**	3.10	0.05*	1.80	0.04**	3.44
Rehabilitate road	0.16	0.29**	5.08	0.36**	6.46	4.05	0.93
Build or rehabilitate reservoir	0.20	0.03**	13.3	0.18	0.44	0.032**	14.28
Build or rehabilitate sanitation	0.03	0.03	0.04	0.07**	2.45	0.033	0.04
Proportion of households reporting building or rehabilitation of piped water	0.46	0.15**	13.5	0.72**	7.51	0.17**	12.91
Number of observations	340	1,740		380		3,600	

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

time in the past, so that households were unaware of them or did not think of the changes as recent.

There is evidence that the community effort required to complete an ASIF school rehabilitation project displaces effort on other local infrastructure projects. To a statistically significant degree, communities that rehabilitated a school are less likely to have also built or rehabilitated a road or piped water system or rehabilitated a health facility. Further, communities in the ASIF pipeline to do an education project were also more likely to have carried out other types of infrastructure projects than were those that had completed an ASIF education project. Communities in the pipeline for an ASIF education project were more likely than propensity-matched communities to take other initiatives. This suggests that communities whose participation in ASIF came later had greater social capital.

Where ASIF water projects were completed, collective action does not appear to be as uniformly weakened as with education projects. Completing an ASIF water project reduced the likelihood that a community would build a health facility or school, rehabilitate a road, or build or rehabilitate a reservoir. But it increased the likelihood that a community would rehabilitate a health facility or build a new road. As with education projects, communities that had not yet completed their ASIF-supported water initiatives were more likely to have undertaken other infrastructure projects, such as building or rehabilitating a school.

In general, if undertaking other community initiatives is used as an indicator of social capital, the communities that completed ASIF projects had less social capital than propensity score-matched comparators. But the ASIF communities that had not yet completed their projects—those whose participation began after the social fund had been in place for some time—had significantly higher social capital.

VI. CONCLUSION

This article offers several insights into how the Armenian social fund affected households and communities during the postcommunist transition. It provides evidence about the degree to which ASIF reached poor Armenian households, the effects of the infrastructure projects on households within the projects' catchment area, and the effects on community collective action.

ASIF was not specifically designed to reach poorer communities. Instead, it sought to reach areas with poor infrastructure, where primary schools and water systems were in particular disrepair. Nonetheless, by some robust measures, ASIF resources reached less well-off parts of the population. Across Armenia and in both urban and rural areas, ASIF households are on average less well off than other Armenian households. But when concentration curves are used to consider the entire distribution of household expenditures, the story becomes less clear. The targeting of ASIF resources was relatively neutral with regard to poverty—slightly progressive in urban areas and slightly regressive in rural areas. One explanation for the progressive urban targeting is ASIF's focus on Yerevan, whose population suffered acutely from economic dislocation. The regressive rural tar-

getting may result from the difficulties rural communities faced in coming up with the required 10 percent community contribution, which could have excluded poorer communities.

Using propensity score and pipeline matching of household data, the analysis demonstrated several impacts of ASIF-supported projects to rehabilitate schools and water systems. In the earthquake zones household spending on primary education and primary school attendance both rose in communities that had completed school projects, suggesting increased demand for education. In communities that had undertaken water projects, households reported improvements in access to water and in water services. But there were few robust indicators of improvements in health in these communities.

One of the central objectives of the ASIF was to increase community involvement. The evidence suggests that the communities that completed a social fund project were less likely than comparison groups to complete other local infrastructure projects, suggesting that social capital was expended in these early projects. By contrast, communities that joined ASIF later and had not yet completed their projects reported more collective action.

Although further research is needed to improve on evaluative approaches, this analysis provides substantial evidence from Armenia that social funds do reach communities in difficult economic circumstances. Furthermore, social funds affect the services available to households. Finally, at least in the communities that became involved more recently, the social fund bolsters their ability to address local needs. These effects suggest that social funds are a useful tool for improving public services in Armenia. As social funds begin to operate in an increasing number of countries undergoing transition from central planning, more opportunities will emerge to learn whether it is appropriate to generalize from the Armenian evidence analyzed here to social funds supporting the transition in other countries.

APPENDIX. COMMUNITY DETERMINANTS OF LIKELIHOOD OF PARTICIPATING IN SOCIAL FUND

TABLE A-1. Propensity Score Probits for Nontargeted Zones

Variable	Coefficient	z-statistic
Community mean per capita expenditure	-0.00009**	2.70
Community mean share of food in expenditure	0.021	0.97
Community share of female household heads	-0.0093	0.01
Community mean of household head's education	0.25*	1.65
Constant	-2.07	0.98
Number of observations	145	
Chi-squared (4)	11.7	

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

TABLE A-2. Propensity Score Probits for Earthquake Zones

Variable	Coefficient	z-statistic
Community mean per capita expenditure	0.0002**	2.21
Community mean share of food in expenditure	-0.076	1.21
Community share of female household heads	-4.77*	0.01
Community mean of household head's education	0.79**	2.08
Constant	2.15	0.37
Number of observations	41	
Chi-squared (4)	15.9	

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

TABLE A-3. Propensity Score Probits for Conflict Zones

Variable	Coefficient	z-statistic
Community mean per capita expenditure	-0.00003	0.52
Community mean share of food in expenditure	0.162**	2.61
Community share of female household heads	6.06**	1.97
Community mean of household head's education	-0.217*	0.63
Constant	-14.6	2.30
Number of observations	43	
Chi-squared (4)	15.0	

*Significant at the 10 percent level.

**Significant at the 5 percent level.

Source: Author's calculations.

REFERENCES

- Chase, Robert, and Lynne Sherburne-Benz. 2001. "Impact Evaluation of the Zambia Social Fund." World Bank, Social Protection, Washington, D.C.
- Goodman, Margaret, Samuel Morley, Gabriel Siri, and Elaine Zuckerman. 1997. "Social Investment Funds in Latin America: Past Performance and Future Role." Inter-American Development Bank, Social Programs and Sustainable Development Department, Evaluation Office, Washington, D.C.
- Grootaert, Christiaan. 1997. "Social Capital: The Missing Link?" In *Expanding the Measure of Wealth: Indicators of Environmentally Sustainable Development*. Environmentally Sustainable Development Studies and Monographs Series No. 17. Washington, D.C.: World Bank.
- Grootaert, Christiaan, and Thierry van Bastelaer. 2001. "Understanding and Measuring Social Capital: A Synthesis of Findings and Recommendations from the Social Capital Initiative." Social Capital Initiative Working Paper No. 24. World Bank, Washington, D.C.
- Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64(4):605-54.

- . 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65(2):261–94.
- Khadiagala, Lynn. 1995. "Social Funds: Strengths, Weaknesses and Conditions for Success." World Bank, Environment and Social Protection Department, Washington, D.C.
- Knack, Stephen, and Philip Keefer. 1997. "Does Social Capital Have an Economic Pay-off? A Cross-Country Investigation." *Quarterly Journal of Economics* 112:1251–88.
- Marc, Alexandre, Carol Graham, Mark Schachter, and Mary Schmidt. 1993. "Social Action Programs and Social Funds: A Review of Design and Implementation in Sub-Saharan Africa." World Bank Discussion Paper 274. World Bank, Washington, D.C.
- Moffitt, Robert. 1991. "Program Evaluation with Nonexperimental Data." *Evaluation Review* 15(3):291–314.
- Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia. 2002. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* 16(2):241–274.
- Paxson, Christina, and Norbert R. Schady. 2002. "The Allocation and Impact of Social Funds: Spending on School Infrastructure in Peru." *World Bank Economic Review* 16(2):297–319.
- Pradhan, Menno, and Laura B Rawlings. 2002. "The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund." *World Bank Economic Review* 16(2):275–295.
- Walker, Ian, Roberto del Cid, Francisco Ordoñez, and Francisco Rodríguez. 1999. *Ex-Post Evaluation of the Honduran Social Investment Fund (FHIS 2)*. Produced by ESA Consultants, Honduras, for the World Bank, Latin American and Caribbean Region.
- Woolcock, Michael. 1998. "Social Capital and Economic Development: Toward a Theoretical Synthesis and Policy Framework." *Theory and Society* 27:151–208.

Impact Evaluation of Social Funds

An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund

*John Newman, Menno Pradhan, Laura B. Rawlings, Geert Ridder,
Ramiro Coa, and Jose Luis Evia*

This article reviews the results of an impact evaluation of small-scale rural infrastructure projects in health, water, and education financed by the Bolivian Social Investment Fund. The impact evaluation used panel data on project beneficiaries and control or comparison groups and applied several evaluation methodologies. An experimental design based on randomization of the offer to participate in a social fund project was successful in estimating impact when combined with bounds estimates to address noncompliance issues. Propensity score matching was applied to baseline data to reduce observable preprogram differences between treatment and comparison groups. Results for education projects suggest that although they improved school infrastructure, they had little impact on education outcomes. In contrast, interventions in health clinics, perhaps because they went beyond simply improving infrastructure, raised utilization rates and were associated with substantial declines in under-age-five mortality. Investments in small community water systems had no major impact on water quality until combined with community-level training, though they did increase the access to and the quantity of water. This increase in quantity appears to have been sufficient to generate declines in under-age-five mortality similar in size to those associated with the health interventions.

This article provides an overview of the results of an impact evaluation study of the Bolivian Social Investment Fund (SIF) and the methodological choices and

John Newman is Resident Representative with the World Bank in Bolivia; Menno Pradhan is with the Nutritional Science Department at Cornell University and the Economics Department at the Free University in Amsterdam; Laura Rawlings is with the Latin America and the Caribbean Region at the World Bank; Geert Ridder is with the Economics Department at the University of Southern California; Ramiro Coa is with the Statistics Department at the Pontificia Universidad Catolica de Chile at Universidad de Belo Horizonte; and Jose Luis Evia is a researcher at the Fundación Milenium. Their e-mail addresses are jnewman@worldbank.org, mpradhan@feweb.vu.nl, lrawlings@worldbank.org, ridder@usc.edu, rcoa@mat.puc.cl, and jlaevia@hotmail.com, respectively. Financial support for the impact evaluation was provided by the World Bank Research Committee and the development assistance agencies of Germany, Sweden, Switzerland, and Denmark. Data were collected by the Bolivian National Statistical Institute. The authors would like to thank Connie Corbett, Amando Godinez, Kye Woo Lee, Lynne Sherburne-Benz, Jacques van der Gaag, and Julie van Domelen for support and helpful suggestions. Cynthia Lopez of the World Bank country office in La Paz and staff of the SIF, particularly Jose Duran and Rolando Cadina, provided valuable assistance in carrying out the study. The research was part of a larger cross-country study in the World Bank, Social Funds 2000.

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

constraints in designing and implementing the evaluation. The study used each of the main evaluation designs generally applied to estimate the impact of projects.¹ These include an experimental design applied to assess the impact of education projects in Chaco, a poor rural region of Bolivia, where eligibility for a project financed by the social fund was randomly assigned to communities.² Through the results from the randomization of eligibility in this case and those from statistical matching procedures using propensity scores in others, this article contributes to the body of empirical evidence on the effectiveness of improving infrastructure quality in education (Hanushek 1995, Kremer 1995), health (Alderman and Lavy 1996, Lavy and others 1996, Mwabu and others 1993), and drinking water (Brockerhoff and Derose 1996, Lee and others 1997).

The main conclusions of the study are as follows. Although the social fund improved the quality of school infrastructure (measured some three years after the intervention), it had little effect on education outcomes. In contrast, the social fund's interventions in health clinics, perhaps because they went beyond simply improving the physical infrastructure, raised utilization rates and were associated with substantial declines in under-age-five mortality. Its investments in small community water systems had no major effect on the quality of the water but did increase the access to and the quantity of water. This increase in quantity appears to have been sufficient to generate declines in under-age-five mortality similar in size to those associated with the health interventions. How the study came to these conclusions is the subject of this article.

I. THE BOLIVIAN SIF

Bolivia introduced the first social investment fund when it established the Emergency Social Fund in 1986. Program staff and international donors soon recognized the potential of the social fund as a channel for social investments in rural areas of Bolivia and as an international model for community-led development. In 1991 a permanent institution, the SIF, was created to replace the Emergency Social Fund, and the social fund began concentrating on delivering social infrastructure to historically underserved areas, moving away from emergency-driven employment-generation projects.

The Bolivian social fund proved that social funds could operate to scale, bringing small infrastructure investments to vast areas of rural Bolivia that line ministries had been unable to reach because of their weak capacity to execute projects.

1. Impact evaluations of World Bank-financed projects continue to be rare even where knowledge about development outcomes is at a premium, such as in new initiatives about which little is known or in projects with large sums of money at stake. A recent study by Subbarao and others (1999) found that only 5.4 percent of all World Bank projects in fiscal year 1998 included elements necessary for a solid impact evaluation: outcome indicators, baseline data, and a comparison group.

2. In the evaluation literature the random assignment of potential beneficiaries to treatment and control groups is widely considered to be the most robust evaluation design because the assignment process itself ensures comparability (Grossman 1994, Holland 1986, Newman and others 1994).

Providing financing to communities rather than implementing projects itself, the social fund introduced a new way of doing business that rapidly absorbed a large share of public investment. Between 1994 and 1998 (roughly the period between the baseline and the follow-up of the impact evaluation study) the SIF disbursed more than US\$160 million, primarily for projects in education (\$82 million), health (\$23 million), and water and sanitation (\$47 million).

The World Bank project that helped finance the SIF built in an impact evaluation at the outset. The design for the evaluation was developed in 1992; baseline data were collected in 1993. The Bolivian social fund is the only one for which there are both baseline and follow-up data and an experimental evaluation design, adding robustness to the results not found in other impact evaluations.³

II. EVALUATION DESIGN

Impact evaluations seek to establish whether a particular intervention (in this case a SIF investment) changes outcomes in the beneficiary population. The central issue for all impact evaluations is establishing what would have happened to the beneficiaries had they not received the intervention. Because this counterfactual state is never actually observed, comparison or control groups are used as a proxy for the state of the beneficiaries in the absence of the intervention. Several evaluation designs and statistical procedures have been developed to obtain the counterfactual, most of which were used in this evaluation. The average difference between the observed outcome for the beneficiary population and the counterfactual outcome is called the average treatment effect for the treated. This effect is the focus of this evaluation study and most others.

The evaluation used different methodologies for different types of projects (education, health, and water) in two regions, the Chaco region and the Resto Rural—an amalgamation of rural areas (table 1). The design of the SIF projects motivated the original choice of evaluation designs applied when setting up the treatment and control or comparison groups during the sample design and baseline data-collection phase. Similarly, changes in the way projects were implemented affected the choice of evaluation methodologies applied in the impact assessment stage.

Education: Random Assignment of Eligibility and Matched Comparison

The education case shows how two different evaluation designs were applied in the two regions: random assignment of eligibility in the Chaco region and matched comparison in the Resto Rural. The choice of evaluation design in each region was conditioned by resource constraints and the timing of the evaluation relative to the SIF investment decisions.

3. The impact evaluation cost about \$880,000, equal to 1.4 percent of the World Bank credit to help finance the SIF and 0.5 percent of the amount disbursed by the SIF between 1994 and 1998.

TABLE 1. Evaluation Designs by Type of Project and Region

	Education		Health	Water
	Chaco	Chaco and Resto Rural combined	Chaco and Resto Rural combined	Chaco and Resto Rural combined
Original evaluation design	Random assignment of eligibility	Matched comparison	Reflexive comparison	Matched comparison
Final evaluation design	Random assignment of eligibility	Matched comparison	Matched comparison	Matched comparison
Final control or comparison group	Nonbeneficiaries randomized out of eligibility for receiving project promotion	Nonbeneficiaries matched on observable 1992 characteristics before the baseline; further statistical matching on baseline characteristics	Nonbeneficiaries statistically matched on baseline characteristics, after determining which clinics did not receive intervention	Nonbeneficiaries from health subsample
Impact analysis methodology ^a	Bounds on treatment effect derived from randomly assigned eligibility	Difference in differences on matched comparisons	Difference in differences on matched comparisons	Difference in differences on matched comparisons

^aEstimations are of the average effects of the SIF interventions on community means, often assessed by aggregating household data.

RANDOM ASSIGNMENT OF ELIGIBILITY. In 1991 the German Institute for Reconstruction and Development earmarked funding for education interventions in Chaco. But the process for promoting SIF interventions in selected communities had not been initiated, and funding was insufficient to reach all schools in the region. This situation provided an opportunity to assess schools' needs and use a random selection process to determine which of a group of communities with equally eligible schools would receive active promotion of a SIF intervention.

To determine which communities would be eligible for active promotion, the SIF used a school quality index.⁴ Only schools with an index below a particular value were considered for SIF interventions, and the worst off were automatically designated for active promotion of SIF education investments.⁵ A total of 200 schools were included in the randomization, of which 86 were randomly assigned to be eligible for the intervention. Although not all eligible communities selected for active promotion ended up receiving a SIF education project, and though a few schools originally classified as ineligible did receive a SIF intervention, the randomization of eligibility was sufficient to measure all the impact indicators of interest.

MATCHED COMPARISON. In the Resto Rural schools had already been selected for SIF interventions, precluding randomization: Nonetheless, it was possible to collect baseline data from both the treatment group and a similar comparison group constructed in 1993 during the evaluation design and sample selection stage.

In the original evaluation design applied to education projects in the Resto Rural, treatment schools were randomly sampled from the list of all schools designated for SIF interventions. A comparison group of non-SIF schools was then constructed using a two-step matching process based on observable characteristics of communities (from a recent census) and schools (from administrative data). First, using the 1992 census, the study matched the cantons in which the treatment schools were located to cantons that were similar in population (size, age distribution, and gender composition), education level, infant mortality rate, language, and literacy rate. Second, it selected comparison schools from those cantons to match the treatment schools using the same school quality index applied in the Chaco region.

Once follow-up data were collected and the impact analysis conducted, the study refined the matching, using observed characteristics from the baseline preintervention data. It matched treatment group observations to comparison

4. This index for the Chaco region assigned each school a score from 0 to 9 based on the sum of five indicators of school infrastructure and equipment: electric lights (1 if present, 0 if not), sewage system (2 if present, 0 if not), a water source (4 if present, 0 if not), at least one desk per student (1 if so, 0 if not), and at least 1.05 m² of space per student (1 if so, 0 if not). Schools were ranked according to this index, with a higher value reflecting more resources.

5. Because the worst-off and best-off schools were excluded from the randomization and the sample, the study's findings on the impacts of the SIF cannot be generalized to all schools.

group observations on the basis of a constructed propensity score that estimates the probability of receiving an intervention.⁶ Following the approach set forth in Dehejia and Wahba (1999), the study matched the observations with replacement, meaning that one comparison group observation can be matched to more than one treatment group observation. This matching was based on variables measured in the treatment and comparison groups before the intervention. Preintervention outcome variables as well as other variables that affect outcomes in the propensity score were included.

In effect, the matching produced a reweighting of the original comparison group so as to more closely match the distribution of the treatment group before the intervention. These weights were then applied to the postintervention data to provide an estimate of the counterfactual—what the value in the treatment schools would have been in the absence of the intervention. The ability to match on preintervention values is one of the main advantages of having baseline data. This analysis combined Chaco and Resto Rural data to yield a larger sample.

Finally, the results were presented using a difference-in-difference estimator, which assumes that any remaining preintervention differences between the treatment schools and the (reweighted) comparison group schools would have remained constant over time if the SIF had not intervened. Thus the selection effect was corrected for in three rounds: first by constructing a match in the design stage, then by using propensity score matching, and finally by using a difference-in-difference estimator.

Health: Reflexive Comparison and Matched Comparison

The health case demonstrates how an evaluation design can evolve between the baseline and follow-up stages when interventions are not implemented as planned. It also underscores the value of flexibility and relatively large samples in impact evaluations.

A reflexive comparison evaluation design based solely on before and after measures was originally developed for assessing SIF-financed health projects. This type of evaluation design involves comparing values for a population at an earlier period with values observed for the same population in a later period. It is considered one of the least methodologically rigorous evaluation methods because isolating the impact of an intervention from the impact of other influences on observed outcomes is difficult without a comparison or control group that does not receive the intervention (Grossman 1994). The original evaluation design was chosen in the expectation that the SIF would invest in all the rural health clinics in the Chaco and Resto Rural.

At the time of the follow-up survey German financing had enabled the SIF to carry out most of its planned health investments in the Chaco region, but financial constraints had prevented it from investing in all the health centers in the

6. See Baker (2000) for a description of propensity score matching.

Resto Rural. This change in implementation allowed the application of a new evaluation design—matched comparison. The question remained, however, whether the SIF interventions had been assigned to health centers on the basis of observed variables and time-constant unobserved variables or on the basis of unobservable variables that changed between the baseline and follow-up surveys. In discussions with SIF management in 1999 it proved impossible to identify the criteria used to select which health centers that would receive the interventions.

An examination of the baseline data revealed significant differences in characteristics between health centers that received the interventions and those that did not. To adjust for these differences, a propensity-matching procedure similar to that used with the education data in the Resto Rural was carried out. The difference between the distribution of the propensity scores in the treatment and comparison groups before and after the matching narrowed considerably, pointing to the effectiveness of the propensity-score-matching method in eliminating observable differences between the treatment and comparison groups.

Once the propensity score matching was applied to the baseline data, a difference-in-difference estimation was performed to assess the impact of the SIF-financed health center investments in rural areas. As will be discussed in the section on results, a series of additional tests were also applied to confirm the robustness of the results on infant mortality.

Water Supply: Matched Comparison

The water case illustrates how impact evaluation estimates for a particular type of intervention can be generated by taking advantage of data from a larger evaluation. At the time of the baseline survey, 18 water projects were planned for the Chaco and Resto Rural. These projects consisted of water supply investments designed to benefit all households within each intervention area. Project sites were selected on the basis of two criteria: whether a water source was available and whether the beneficiary population would be concentrated enough to allow economies of scale.

No specific comparison group was constructed *ex ante*. Instead, it was expected that the comparison group could be constructed from the health subsample using a matched comparison technique to identify similar nonbeneficiaries. At the follow-up data collection and analysis stage it was determined that all 18 projects had been carried out as planned and that there were sufficient data from which to construct a comparison group using the health sample, as originally expected. Thus the water case is the only one of the three in which the evaluation design did not change between the baseline and follow-up stages of the evaluation.

III. RESULTS IN EDUCATION

SIF-financed education projects either repaired existing schools or constructed new ones and usually also provided new desks, blackboards, and playgrounds. In many cases new schools were constructed in the same location as the old

schools, which were then used for storage or in some cases adapted to provide housing for teachers.

Schools that received a SIF intervention benefited from significant improvements in infrastructure (the condition of classrooms and an increase in classroom space per student) and in the availability of bathrooms compared with schools that did not receive a SIF intervention. They also had an increase in textbooks per student and a reduction in the student-teacher ratio.⁷ But the improvements had little effect on enrollment, attendance, or academic achievement. Among student-level outcomes, only the dropout rate reflects any significant impact from the education investments.

Estimates Based on Randomization of Eligibility

The evaluation for the Chaco region was able to take advantage of the randomization of active promotion across eligible communities to arrive at reliable estimates of the average impact of the intervention (table 2). Because of the demand-driven nature of the SIF, not all communities selected for active promotion applied for and received a SIF-financed education project. This does not represent a departure from the original evaluation design, and randomization of eligibility (rather than the intervention) is sufficient to estimate all the impacts of interest (see appendix A).

But the fact that some communities not selected for active promotion nevertheless applied for and received a SIF-financed education project does represent a departure from the original evaluation design. This noncompliance in the control group (as it is known in the evaluation literature) can be handled by calculating lower and upper bounds for the estimated effects.⁸ Thus the cost of the noncompliance is a loss of precision in the impact estimate as compared with a case in which there is full compliance. In the case considered here, the differences between the lower and upper bounds of the estimates are typically small and the results are still useful for policy purposes (see table 2 for these bounds estimates and appendix A for an explanation).

Estimates Based on Matched Comparison

In the Resto Rural schools had already been selected for the SIF interventions and no randomization of eligibility took place, making it impossible to apply an

7. For all education and health results the Wilcoxon-Mann-Whitney nonparametric test was used to detect departures from the null hypothesis that the treatment and comparison cases came from the same distribution. The alternative hypothesis is that one distribution is shifted relative to the other by an unknown shift parameter. The p -values are exact and are derived by permuting the observed data to obtain the true distribution of the test statistic and then comparing what was actually observed with what might have been observed. In contrast, asymptotic p -values are obtained by evaluating the tail area of the limiting distribution. The software used for the exact nonparametric inference is StatXact 4 (<http://www.cytel.com>). Although the exact tests take account of potentially small sample bias, in practice there were no major differences between the exact and asymptotic p -values.

8. This approach of working with bounds follows in the spirit of Manski (1995).

TABLE 2. Average Impact of SIF Education Investments in Chaco, with Estimation Based on Randomization of Eligibility

Indicator ^a	Mean for all schools, 1993	Impact of intervention, 1997			
		Lower bound	<i>p</i> -value	Upper bound	<i>p</i> -value
<i>School-level outcomes</i>					
Blackboards	0.35	1.46	0.17	1.79	0.08**
Blackboards per classroom	0.08	0.40	0.03*	0.43	0.02*
Desks	33.32	9.20	0.70	29.44	0.11
Desks per student	0.52	0.57	0.15	0.65	0.10**
Classrooms in good condition	0.37	1.01	0.42	1.98	0.06**
Fraction of classrooms in good condition	0.11	0.34	0.07**	0.41	0.02*
Teachers' tables	0.42	1.12	0.31	1.67	0.11
Teachers' tables per classroom	0.18	0.54	0.00*	0.59	0.00*
Fraction of schools with sanitation facilities	0.39	0.47	0.02*	0.58	0.00*
Fraction of schools with electricity	0.06	-0.05	0.75	-0.07	0.69
Fraction of teachers with professional degrees	0.46	-0.09	0.65	-0.10	0.63
Textbooks	17.47	-25.72	0.64	1.79	0.97
Textbooks per student	0.32	0.41	0.87	0.05	0.98
Students per classroom	22.93	2.12	0.68	0.47	0.93
<i>Students' education outcomes</i>					
Repetition rate (percent)	12.65	-1.75	0.61	-5.45	0.17
Dropout rate based on household data (percent)	9.49	-3.90	0.26	-6.00	0.08**
Dropout rate based on administrative data (percent)	10.73	3.01	0.53	3.17	0.50*
Enrollment ratio (ages 5-12)	0.83	0.15	0.14	0.05	0.63
Fraction of days of school attended in past week	0.93	-0.02	0.38	-0.07	0.11

*Significant at the 5 percent level.

**Significant at the 10 percent level.

^aIn 1997 (but not in 1993) achievement tests in language and mathematics were administered to the treatment and control schools. No significant differences were found.

Source: SIF Evaluation Surveys

experimental design and calculate impact in the same way as in the Chaco region. Instead, a matching procedure based on propensity scores was used, as described in the section on evaluation design. This analysis combined the Chaco and Resto Rural samples. The first-stage probit estimations used to calculate the propensity scores employed only values for 1993, before the intervention, to ensure preintervention comparability between the treatment and comparison groups.

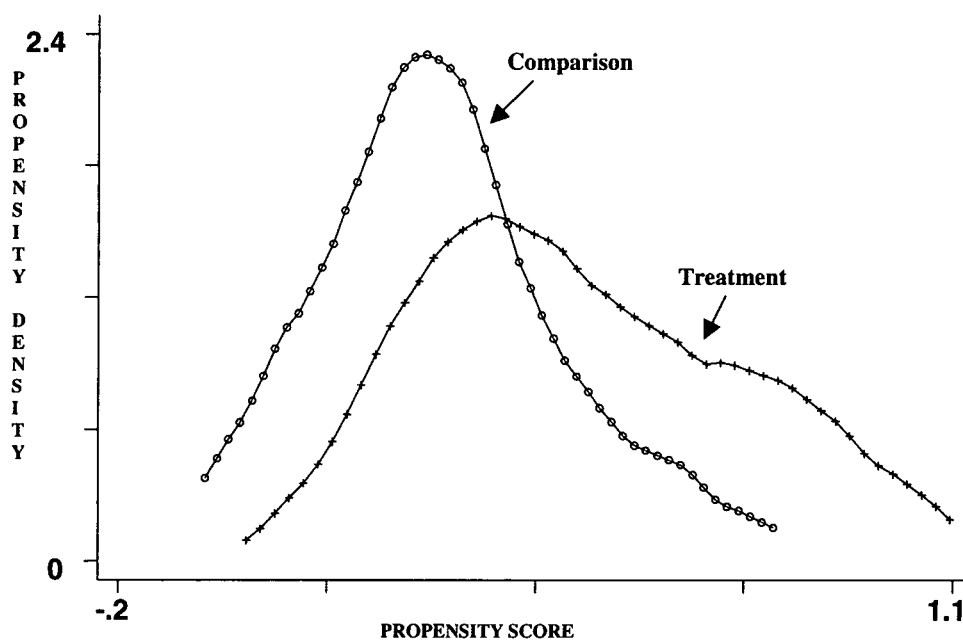
The kernel density estimates of the propensity scores for the treatment and comparison groups before propensity score matching indicate that differences

remained between the groups before the intervention took place (figure 1). The kernel density estimates of the propensity scores after matching, however, show that propensity matching does a relatively good job of eliminating preprogram differences between SIF and non-SIF schools (figure 2).

Even so, there is a range where the propensity scores do not overlap. In this range observations in the treatment group have propensity scores exceeding the highest values in the comparison group. For this group of treatment observations no comparable comparison group is available. The group consists of only five observations, however, and can be taken into account by setting bounds on the possible counterfactual values for these five. In practice, for each treatment school that cannot be matched to a comparison school, a comparison is constructed by matching the school with itself. That is, the comparison is an exact replica but with the intervention dummy variable set to 0. This is equivalent to assuming that for these schools the intervention has no effect. (For a discussion of the upper bound, see appendix A.)

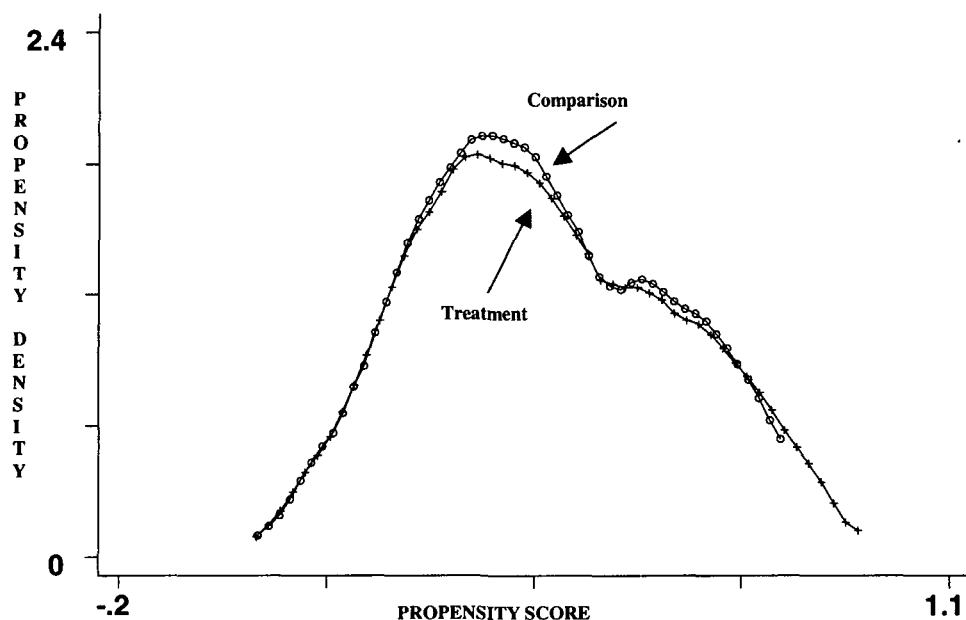
The results of a difference-in-difference estimation (intertemporal change in the treatment group minus intertemporal change in the comparison group) before and after the propensity score matching are not dramatically different from those based on randomization of eligibility (table 3). This indicates that the matching in the evaluation design stage, before the statistical propensity score matching, was rela-

FIGURE 1. Kernel Density Estimates of Treatment and Comparison Schools' Propensity Scores Before Matching



Source: Authors' calculations.

FIGURE 2. Kernel Density Estimates of Propensity Scores for Treatment and (Reweighted) Comparison Schools After Matching



Source: Authors' calculations.

tively effective. Only for a couple of variables were there preprogram differences, and these were eliminated with the propensity score matching.

The ability to eliminate the preintervention differences in means between treatment and comparison groups after matching increases confidence in the evaluation results, although it is by no means a guarantee that the estimates are unbiased. But the matching procedure did remove observable differences between treatment and comparison groups, and the difference-in-difference estimation also removed the time-constant unobservable differences. In presenting the impact estimates, one has to assume that the matching has also eliminated the preintervention differences in time-varying unobservable variables that affect outcomes.

Although initial differences in unobservable characteristics cannot be examined, baseline data make it possible to check whether differences in observable characteristics between the treatment and comparison groups have been addressed. Baseline data also make it possible to use difference-in-difference estimates to eliminate the effect of time-constant unobservables in estimating program impact. Most evaluations that have only postintervention data on beneficiaries and nonbeneficiaries rely on some type of statistical matching procedure to try to generate appropriate comparison groups for those receiving the intervention (Rosenbaum and Rubin 1983, Heckman and others 1998, Angrist and Krueger 1999).

TABLE 3. Difference-in-Difference Estimates of Average Impact of SIF Education Investments in Chaco and Resto Rural (intertemporal change in the treatment group minus intertemporal change in the comparison group)

Indicator	Before matching differences			After matching differences		
	Treatment group	Comparison group	<i>p</i> -value	Treatment group	Comparison group	<i>p</i> -value
<i>School-level outcomes</i>						
Fraction of schools with electricity	0.152	0.127	0.70	0.152	0.159	0.93
Fraction of schools with sanitation facilities	0.347	0.082	0.032*	0.341	-0.048	0.016*
Textbooks per student	3.78	3.05	0.219	3.78	1.97	0.027*
Square meters per student	1.87	0.47	0.004*	1.87	0.448	0.002*
Students per classroom	-7.53	1.22	0.006*	-7.53	3.01	0.002*
Fraction of classrooms in good condition	0.365	0.064	0.005*	0.365	0.019	0.015*
Students per desk	-1.30	-0.72	0.97	-1.30	0.30	0.74
Students per teacher	-5.05	-1.17	0.176	-5.05	-0.136	0.048*
<i>Students' education outcomes</i>						
Dropout rate	-0.028	0.006	0.010*	-0.028	-0.003	0.045*
Number of registered students per school	6.4	18.27	0.68	6.4	42.6	0.038*
Number of students attending classes regularly per school	8.76	17.2	0.68	8.76	3.84	0.042*
Number of students repeating classes	-2.36	1.09	0.417	-2.39	38.8	0.40

*Significant at the 5 percent level.

**Significant at the 10 percent level.

Source: Authors' calculations.

IV. RESULTS IN HEALTH

SIF-financed health projects repaired existing health centers and constructed new ones. The SIF worked with prototype designs that included a waiting room, a room for outpatient consultations, a room with several beds for inpatients, a space for a pharmacy, bathrooms, and a meeting room for presentations on health topics. The SIF also provided health centers with medicines, furniture, and medical equipment; a motorcycle to allow health personnel to conduct more home visits; and a radio to call for ambulances and to keep in contact with other health centers. Where centers lacked electricity, the SIF provided solar panels to power lights, a radio, and a refrigerator for storing medicines and vaccines. Finally, it made drinking water available and typically installed showers.

As explained, the SIF originally intended to make investments in all health clinics in the sample but was unable to do so mainly because of financial constraints. Thus by the time of the follow-up survey some clinics had received an interven-

tion and some had not. Thanks to the financing from the German bilateral aid agency, most clinics in the Chaco region received an intervention. Fewer did in the Resto Rural sample.

Kernel density estimates of the propensity scores for the treatment and comparison groups before matching reveal considerably greater differences than was the case for education (figure 3). This may reflect the inability to construct a comparison group before the intervention owing to the initial plans to reach all health clinics. Despite the initial differences, the matching procedure managed to eliminate virtually all the observable preprogram differences in the reported variables (figure 4).

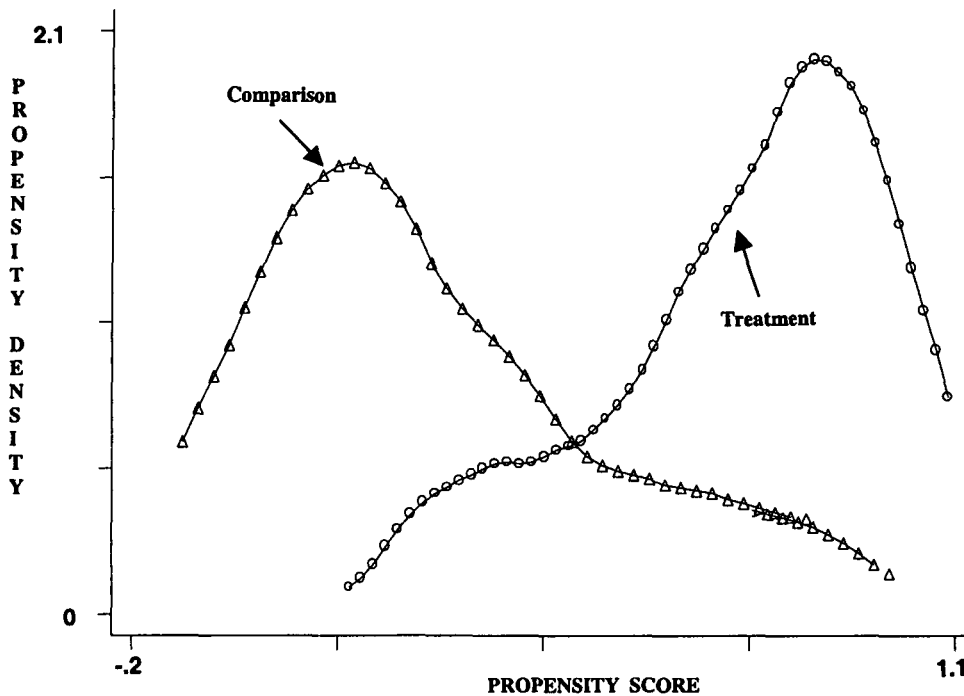
Infrastructure and Utilization Estimates

The SIF investments in health centers brought about significant improvements in their physical characteristics and in their utilization. Both the share of women's prenatal care and the share of births attended—two important factors affecting under-age-five mortality—increased significantly (table 4).

Under-Age-Five Mortality Estimates

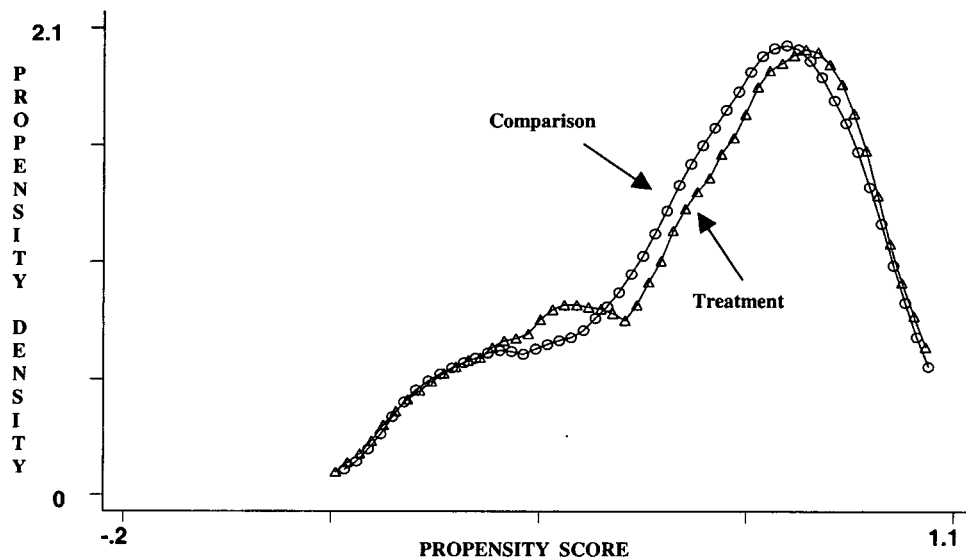
The impact evaluation drew on sufficiently large samples in the household surveys to allow assessment of the impact of SIF-financed investments in health

FIGURE 3. Kernel Density Estimates of Propensity Scores for Treatment and Comparison Health Clinics Before Matching



Source: Authors' calculations.

FIGURE 4. Kernel Density Estimates of Propensity Scores of Treatment and (Reweighted) Comparison Health Clinics After Matching



Source: Authors' calculations.

centers on under-age-five mortality. Using three different methods to assess this impact, the evaluation found consistent evidence of a significant reduction in under-age-five mortality in the areas served by health clinics receiving a SIF intervention.

The first method, using propensity score matching, uses recall data from the household surveys on deaths among children born 10 years before the survey. The results before propensity score matching show that the proportion of children dying was significantly higher in the treatment group than in the comparison group before the intervention, but significantly lower in the treatment group after the intervention (table 5). When matching, the study used the same procedure (and the same implicit weights) as it did when analyzing the effect of SIF investments on the infrastructure and utilization of health clinics. Just as with the variables for physical characteristics and utilization, the matching eliminates the preintervention differences. The postintervention differences remain, however: under-age-five mortality is lower in the treatment group.

The second method draws on life table estimates for the change in mortality using only the households for which survey data are available for both 1993 and 1997. For this reason the sample is smaller and no matching was done. The under-age-five mortality rates in this sample, covering the period 1988–93, are close to the rates reported in the 1994 National Demographic and Health Survey for the period 1989–94.

TABLE 4. Difference-in-Difference Estimates of Average Impact of SIF Health Investments in Chaco and Resto Rural
(intertemporal change in the treatment group minus intertemporal change in the comparison group)

Indicator	Before matching differences			After matching differences		
	Treatment group	Comparison group	<i>p</i> -value	Treatment group	Comparison group	<i>p</i> -value
<i>Health clinic characteristics</i>						
Number of beds	1.400	0.125	0.00*	1.39	0.71	0.003*
Fraction of clinics with electricity	0.077	0.050	0.81	0.078	0.098	0.89
Fraction of clinics with sanitation facilities	0.404	0.125	0.66	0.392	0.176	0.042*
Fraction of clinics with water	0.078	-0.025	0.58	0.08	0	0.64
Number of patient rooms	0.346	-0.205	0.07**	0.33	-0.54	0.00*
Index of availability of medical equipment in good condition***	0.252	0.109	0.24	0.25	0.22	0.40
Index of availability of medical supplies***	0.332	0.080	0.02*	0.33	0.07	0.00*
<i>Intermediate health outcomes</i>						
Use of public health service (unconditional)	0.002	-0.001	0.18	0.002	0.002	0.60
Use of public health service (conditional on illness)	0.011	-0.006	0.96	0.011	0.010	0.49
Fraction of women receiving any prenatal care	0.191	0.073	0.068**	0.207	0.007	0.001*
Fraction of births attended by trained personnel	0.068	0.020	0.60	0.063	0.050	0.58
Fraction of cases of diarrhea treated	0.006	0.069	0.92	0.006	-0.138	0.23
Fraction of cases of cough treated	0.030	0.053	0.18	0.031	0.133	0.08**
<i>Health outcomes</i>						
Incidence of diarrhea	-0.030	-0.079	0.17	-0.029	-0.013	0.84
Incidence of cough	-0.147	-0.089	0.64	-0.152	-0.178	0.34

*Significant at the 5 percent level.

**Significant at the 10 percent level.

***The index is calculated as the fraction of supplies that were found in a site inspection, relative to the norms for supplies specified by the Ministry of Health.

Source: Authors' calculations.

TABLE 5. Deaths among Children under Age Five among Children Born in Previous 10 Years in Chaco and Resto Rural, 1993 and 1997

Indicator	1993		1997	
	Treatment group	Comparison group	Treatment group	Comparison group
<i>Before matching</i>				
Percentage of children dying	10.6 (292)	8.4 (122)	6.1 (134)	9.8 (120)
Percentage of children surviving	89.4 (2,469)	91.6 (1,322)	93.9 (2,068)	90.2 (1,107)
Difference between comparison and treatment groups in percentage of children dying	-2.1 [0.076]**		3.7 [0.023]*	
<i>After matching</i>				
Percentage of children dying	10.3 (237)	10.2 (182)	6.0 (110)	10.7 (149)
Percentage of children surviving	89.7 (2,057)	89.8 (1,595)	94.0 (1,723)	89.3 (1,242)
Difference between comparison and treatment groups in percentage of children dying	-0.08 [0.96]**		4.7 [0.07]*	

*Significant at the 5 percent level.

**Significant at the 10 percent level.

Note: Figures in parentheses are number of deaths and survivors. Figures in square brackets are *p*-values. Results corrected for cluster sampling.

Source: Authors' calculations.

Again, the results show a significant reduction in mortality in the treatment group from 1993 to 1997 (table 6). In the comparison group mortality does not decline and, if anything, increases.

The third approach to measuring the change in mortality is based on estimations of a Cox proportional hazard function. The sample is first divided into a group of clinics that received a SIF intervention and a comparison group matched according to the propensity score, which takes into account characteristics of the health facility, the community and health outcomes, and characteristics of the households in the service area (see appendix C). Data on individual households residing in the service area of the two groups of clinics are used to estimate a hazard function and, based on the estimated hazard, an under-age-five mortality rate. The hazard function is written as

$$(1) \quad \lambda(\text{time}; X_j, i_j) = \lambda(\text{time})\exp(X_j\beta + \theta i_j)$$

where X is a vector of characteristics of child j and i denotes whether or not the clinic in the area received an intervention. The advantage of using a hazard model is that it allows one to easily deal with right censoring and thus to estimate an under-age-five mortality rate.

TABLE 6. Life Table Estimates of Infant and Under-Age-Five Mortality Rates in Chaco and Resto Rural, 1993 and 1997

	1993		1997	
	Treatment group	Comparison group	Treatment group	Comparison group
Infant mortality rate (per 1,000 live births)	61.5	59.8	30.8	67.2
Under-five mortality rate (per 1,000)	94.0	92.6	54.6	107.9
Number of observations	838	822	620	596
Cumulative failure at month				
0	0.029	0.027	0.016	0.032
1	0.038	0.038	0.020	0.044
3	0.050	0.050	0.025	0.053
6	0.062	0.061	0.031	0.067
12	0.072	0.074	0.040	0.081
24	0.091	0.090	0.055	0.107
60	0.091	0.090	0.055	0.107
Likelihood ratio test for homogeneity	0.007		10.04	
Chi2(1)	[0.932]		[0.002]*	

*Significant at the 5 percent level.

Note: Figures in square brackets are *p*-values.

Source: Authors' calculations.

The estimated coefficients of β and θ in table 7 represent results after matching, using the procedure described. Per capita consumption, age of mother at child's birth, and education of mother are expressed as deviations from the mean, with values of 2,600 (bolivianos), 27 (years), and 3 (years), respectively. The reported under-age-five mortality rates are derived from the estimated survival function evaluated at the mean values of X .

The results again show no significant differences in 1993 between the treatment and comparison groups (the intervention variable is not significant), but significantly lower under-age-five mortality in the treatment group after the intervention. The impact can be derived by using the differences in predicted under-age-five mortality rates with and without the intervention between the two years. Selection bias is addressed by using difference in differences.

Thus all three of the approaches show a similar pattern of declining under-age-five mortality in the treatment group receiving a SIF-financed health investment and no decline in the comparison group. The Cox proportional hazard estimates, the most accurate, show a decline in under-age-five mortality from 88.5 deaths per 1,000 to 65.8 among children living in the service area of a health center that received a SIF investment.

What are some possible explanations for the finding of lower mortality in the treatment group? One is that the treatment group might have received interventions not provided by the SIF that could have led to lower mortality, such as in water and sanitation.

TABLE 7. Cox Proportional Hazard Estimates of Under-Five Mortality in Chaco and Resto Rural, 1993 and 1997

Variable	1993			1997		
	Coefficient	Standard error	p-value	Coefficient	Standard error	p-value
Duration (year of birth-1992)	-.029	0.025	0.259	-.039	0.033	0.24
Intervention dummy variable (= 1 if living in area of influence of health clinic with intervention)	-.009	0.195	0.96	-.55	0.28	0.05*
Per capita household consumption	-.000012	0.00001	0.36	1.45e-07	4.40e-06	0.97
Age of mother at child's birth	.029	0.027	0.28	-.0007	0.01	0.95
Education of mother	.022	0.047	0.65	-.011	0.038	0.74
Number of observations	3,881			3,107		
Wald Chi2(5)	5.16			8.06		
Prob > Chi2	0.40			0.153		
<i>Estimated under-age-five mortality rate (per 1,000)</i>						
Treatment group	88.5			65.8		
Comparison group	89.3			111		

*Significant at the 5 percent level.

Source: Authors' calculations.

Between the baseline and follow-up surveys the comparison group received more non-SIF water interventions than the treatment group, though there was no significant difference in the non-SIF sanitation projects received (table 8). Although not reported here, regressions of the difference between 1997 and 1993 in availability of piped water, adequacy of water throughout the day and year, distance to water supply, and adequacy of sanitation facilities on the intervention dummy variable also revealed no significant differences between the treatment and comparison groups.

If the reduction in under-age-five mortality had something to do with the services provided in the clinics, greater reductions in mortality would be expected among those who used the clinics than among those who did not. Data show that under-age-five mortality among families in which the mother received at least one prenatal checkup before the last birth was significantly lower in the treatment group than in the comparison group in 1997 but not in 1993 (table 9). This result strongly suggests that something associated with the health clinic after the intervention accounts for the lower mortality observed.

V. RESULTS IN WATER SUPPLY

SIF water supply investments provided financing for small-scale potable water systems whose design varied depending on the geographic location. Initially, the investments in infrastructure were not accompanied by adequate training. But in later years greater effort was made to provide training through the World Bank-financed Rural Water and Sanitation Project (Prosabar).

Data from before and after the SIF water supply investments in Chaco and the Resto Rural show that the main changes were a reduction in the distance to

TABLE 8. Non-SIF Water and Sanitation Projects Benefiting Treatment and Comparison Groups in Chaco and Resto Rural, 1993–97

	Treatment group	Comparison group
<i>Non-SIF water projects</i>		
Percent of households who benefited from water projects not financed by the SIF	14.5 (656)	32.7 (457)
Percent of households who did not benefit	85.5 (3,863)	67.3 (941)
Design-based <i>F</i>	3.28 [0.073]	
<i>Non-sif sanitation projects</i>		
Percent of households who benefited from sanitation projects not financed by the SIF	8.5 (384)	6.2 (87)
Percent of households who did not benefit	91.5 (4,135)	93.8 (1,311)
Design-based <i>F</i>	0.144 [0.705]	

Note: Figures in parentheses are number of observations. Figures in square brackets are *p*-values. Results adjusted for cluster sampling.

Source: Authors' calculations.

TABLE 9. Deaths in Previous Five Years among Children under Age Five in Families with and without Prenatal Checkups in Chaco and Resto Rural, 1993 and 1997 (percent)

	1993		1997	
	Treatment group	Comparison group	Treatment group	Comparison group
<i>At least one prenatal checkup before last birth</i>				
Percentage of children dying	8.4 (57)	8.2 (23)	4.8 (37)	9.6 (31)
Percentage of children surviving	91.6 (620)	91.8 (258)	95.2 (728)	90.4 (293)
Design-based <i>F</i>	0.015 [0.90]		7.40 [0.01]*	
<i>No prenatal checkup before last birth</i>				
Percentage of children dying	7.8 (62)	7.7 (39)	9.6 (31)	8.4 (31)
Percentage of children surviving	92.2 (732)	92.3 (467)	90.4 (293)	91.6 (338)
Design-based <i>F</i>	0.003 [0.95]		0.267 [0.61]	

*Significant at the 5 percent level.

Note: Figures in parentheses are number of deaths and survivors. Figures in square brackets are *p*-values. Results corrected for cluster sampling.

Source: Authors' calculations.

the water source and, in the Resto Rural, a substantial improvement in sanitation facilities (table 10). Unfortunately, data on water consumption were collected only for 1997, making it impossible to measure the improvement in this important indicator.

A laboratory analysis of the quality of water from the old and new sources showed surprisingly little improvement in the 18 SIF water projects in the impact evaluation study.⁹ Results indicated fecal contamination in the old system for 9 of the 15 projects where samples could be taken, and in the new system for 7 of the 14 projects where samples were taken. Samples from both the old and the new systems showed a complete absence of residual chloride, suggesting that no chlorination had taken place. Interviews with beneficiaries pointed to the following explanations for the lack of improvement in water quality:

- The personnel designated by each community to maintain the water systems lacked training in procedures for cleaning the water tanks, repairing the water tubes, chlorinating the water supply, and managing the proceeds from user fees.

9. The testing followed recommended parameters defined by the World Health Organization. For more details see Coa (1997) and Damiani (2000).

TABLE 10. Impact of SIF Water Investments in Chaco and Resto Rural

Indicator	Chaco		Resto Rural	
	1993	1997	1993	1997
Incidence of diarrhea in past 24 hours among children less than 6 years old	0.11 (0.31)	0.09 (0.29)	0.09 (0.29)	0.09 (0.29)
Duration of diarrhea (days)	3.03 (2.26)	2.95 (2.71)	5.07 (5.79)	3.28 (2.76)
Fraction of diarrhea cases treated	0.34 (0.48)	0.37 (0.49)	0.53 (0.26)	0.36 (0.49)
Fraction of households with piped water	0.49 (0.50)	0.67 (0.47)	0.44 (0.50)	0.54 (0.50)
Fraction of households with sanitation facilities	0.58 (0.49)	0.61 (0.49)	0.27 (0.44)	0.71 (0.45)
Distance from house to principal water source (m)	211.47 (433.23)	57.95 (207.62)	92.48 (165.11)	41.11 (116.81)
Hours a day of water availability	21.95 (5.95)	19.38 (8.79)	18.49 (8.61)	21.15 (6.97)
Fraction of year with adequate water	0.79 (0.40)	0.89 (0.31)	0.87 (0.34)	0.91 (0.29)
Household water consumption (L/day)	—	23.73 (13.82)	—	20.51 (12.55)
Fraction of households boiling water before consumption	0.54 (0.50)	0.28 (0.45)	0.61 (0.49)	0.45 (0.50)
Fraction of households with knowledge of oral rehydration therapy	0.78 (0.41)	0.95 (0.21)	0.74 (0.44)	0.84 (0.36)
Fraction of households using oral rehydration therapy	0.52 (0.50)	0.55 (0.50)	0.33 (0.48)	0.44 (0.50)

— Not available.

Note: standard deviation in parentheses

Source: Authors' calculations.

- The systems lacked meters for measuring household water consumption, which would have made it easier to collect user fees adequate for providing the necessary maintenance of the system.
- In some cases inappropriate materials had been used (such as tubes designed for oil, not water) and the work was of poor quality (resulting in a rough finish for the water tanks, which made cleaning more difficult).

When the water quality results were presented to SIF representatives, they acknowledged that their initial water projects did have problems, mostly attributable to inadequate training. But they explained that this problem had been solved with the assistance of Prosabar. To test this explanation, a second water quality analysis was carried out using the same approach but covering more recent projects.

The second analysis found significant levels of fecal contamination in 10 of 18 old water sources but in only 2 of 15 new sources. In contrast with the first sample of projects, in which the beneficiaries received little training, in the second sample of projects all communities had received training through Prosabar.

A disturbing finding, however, was that no chlorination was taking place in any of the more recent projects. This could cause problems down the road if maintenance deteriorates or there is an external source of contamination.

In a review of several studies of the health impact of improvements in water supply and sanitation facilities, Esrey and others (1990) suggest that such improvements can be expected to reduce under-age-five mortality by about 55–60 percent. To maximize the health impacts of water projects, they indicate that the supply of water should be as close to the home as possible so as to increase the quantity available for hygiene. They conclude that safe excreta disposal and proper use of water for personal and domestic hygiene appear to be more important than the quality of drinking water in achieving broad health impacts.

The results from the SIF water and sanitation investments are consistent with these findings, showing a significant reduction in deaths among children under age five (table 11). Converting the results to under-age-five mortality rates by estimating a Cox proportional hazard model (as in the health case) shows a reduction from 105 deaths per 1,000 to 61, a decline of 42 percent.

VI. CONTRIBUTION OF THE SIF TO DECLINES IN DROPOUT RATES AND UNDER-AGE-FIVE MORTALITY

The results from the impact evaluation study can be scaled up to suggest the impact that the SIF had in the country as a whole. In Bolivia, as elsewhere, one of the important features of the social fund model is its ability to operate to scale. Between 1994 and 1998 the SIF financed investments in 1,041 of the roughly 3,900 rural primary schools in the country, benefiting roughly 185,000 students. The study estimated that these investments led to a reduction in dropout rates ranging from 3 percentage points (from the propensity score matching) to 3.8 percentage points (the lower bound from the randomization of eligibility). On the basis of these results it can be estimated that the SIF investments led to an additional 5,550–7,030 students remaining in school over the four-year period of the study.¹⁰ The average cost of the school interventions was about \$60,650.

SIF health and water investments accounted for roughly 25 percent of the reduction in deaths among children under age five in rural areas between 1994 and 1998. This finding is based on a scaling up of the estimated mortality effects of the sample of SIF investments in the evaluation compared with the change in the total number of deaths in the under-age-five population, in the five-year period before the survey. Data on total deaths are from Demographic and Health Surveys carried out in 1994 and 1998.

The estimate of the number of deaths averted as a result of the SIF health interventions (1,150) was obtained by multiplying the difference in the proportion of children dying between the treatment and comparison groups (0.04) by the

10. Of course, this says nothing about whether the additional students remaining in school stayed to graduate. More time and larger samples would be needed to determine how long lasting the effect is.

TABLE 11. Deaths in Previous Five Years among Children under Age Five in Households Benefiting from SIF Water Investment in Chaco and Resto Rural, 1993 and 1997

	1993	1997
Percentage of children dying	9.74 (167)	5.73 (77)
Percentage of children surviving	90.26 (1,547)	94.27 (1,247)
Pearson design-based $F(1,28)$	14.715 [0.0007]*	

*Significant at the 5 percent level.

Note: Figures in parentheses are number of survivors or deaths. Figure in square brackets is the p -value.

Source: Authors' calculations.

estimated number of children under age five served by the 473 SIF-financed health centers (28,853).¹¹ The estimate of the number of deaths averted because of the SIF water interventions (2,640) was similarly obtained by multiplying the difference in the proportion of children dying between the treatment and comparison groups (again, 0.04) by the estimated number of children under age five served by the 639 SIF-financed water projects (65,945).

Mortality data from the 1994 and 1998 Demographic and Health Surveys (which cover a period roughly coinciding with that covered by the baseline and follow-up surveys of the SIF evaluation) and rural population estimates from the National Statistical Institute indicate a decline of some 13,870 deaths between 1994 and 1998.¹² If not for the SIF interventions, there would have been a decline of only 10,080 deaths.

It is possible to arrive at a rough estimate of the cost per death averted for both the health and the water interventions. The average health intervention cost \$47,780, and the average water intervention \$62,905. Thus the cost per death averted was roughly \$20,000 for the health interventions and \$15,200 for the water interventions. This estimate refers only to the initial four years of SIF investments. As long as the investments are maintained, they can be expected to avert more deaths in the coming years. Moreover, the investments lead to benefits beyond the effects on under-age-five mortality.

VII. CONCLUSIONS

The main finding of the evaluation is that SIF-financed investments in health centers and water supply systems appear to have resulted in a significant reduction

11. The mean number of individuals served by health centers was 380, of which 16 percent (61) were under age five. The mean number of individuals benefiting from water projects was 645.

12. The calculations are based on an estimated under-age-five mortality rate of 115.6 per 1,000 in 1994 and 91.7 per 1,000 in 1998 and an estimated population of children under age five in rural areas of 505,510 in 1994 ($3,008,993 \times 0.168$ percent) and 485,984 in 1998 ($3,018,535 \times 0.161$ percent). The estimated number of deaths among children under age five in rural areas was 58,436 in the five-year period before 1994 and 44,564 in the five-year period before 1998.

in under-age-five mortality. By contrast, investments in school infrastructure led to little improvement in education outcomes apart from a decline in dropout rates. But in all three sectors the investments resulted in a demonstrable improvement in the physical facilities.

Why did the SIF investments in health facilities have a greater effect than those in schools? Part of the reason may be that the health investments went beyond simply providing infrastructure. They also provided medicines and medical supplies—and radios and motorcycles supporting outreach to patients and communication with regional health centers and hospitals. Moreover, the results suggest a link between an increase in the utilization of health centers—particularly for prenatal care—and the reduction in under-age-five mortality.

The finding that the investments in school infrastructure are insufficient to achieve the desired impact on education outcomes has implications as much for the education sector as it does for the SIF. Motivated in part by this finding, shared with the government of Bolivia in 1999, the SIF and the Ministry of Education have devoted much effort to changing the projects financed through the SIF. They now give more attention to the “software” of education, and where the SIF finances physical infrastructure, it does so as part of an integrated intervention.

The improvements in under-age-five mortality arising from the investments in water supply were accompanied by significant reductions in the distance of water sources from households and, in the Resto Rural, a substantial improvement in the adequacy of sanitation facilities but not by improvements in the quality of water. Water quality did not improve substantially until after training in operations and maintenance was provided to the communities receiving water projects.

From a methodological standpoint, the three cases highlight the variety of approaches available to evaluators, the benefit of having baseline data, and the need for flexibility in the face of changes in the implementation of interventions. Projects often are not carried out as planned, particularly when they are demand-driven.

Planning ahead for an evaluation and responding creatively to budgetary or administrative constraints can provide opportunities for randomization. In education randomization of eligibility for active promotion of projects was sufficient to obtain all the indicators of interest. This finding is especially useful for evaluations of demand-driven programs because people’s behavior can often result in changes to the original evaluation design, as it did for the SIF-financed education projects in the Chaco region. Noncompliance in the control group can be handled by working with bounds to estimate a range of impacts in cases where contamination is not too severe.

Where randomization was not possible, applying propensity score matching to baseline data was reasonably successful in eliminating preintervention differences between treatment and comparison groups and allowed difference-in-difference estimates to measure program impact. The baseline data collected from comparison and treatment groups were essential to this analysis. Preintervention

data can help form better statistical matches and also make it possible to check whether the statistical matching eliminates preintervention differences. If the statistical matching produces a treatment group and a comparison group that do not differ except for the effect of the intervention, there should be no differences in the average values of key characteristics before the intervention. Future impact evaluations should make a greater effort to collect preintervention data.

APPENDIX A. USING RANDOMIZATION OF ELIGIBILITY TO
ESTIMATE THE AVERAGE TREATMENT EFFECT ON THE
TREATED FOR SCHOOL INVESTMENTS IN CHACO

This appendix explains how the impact evaluation study derived an average impact estimate for the communities that received a SIF education intervention (the treated population) by taking advantage of the information that some communities were randomly assigned to be eligible to receive such an intervention.

The evaluation design for school investments in the Chaco region included two types of schools: those that were eligible to receive the SIF intervention and those that were not. In the implementation stage, however, the demand-driven nature of the SIF, combined with common difficulties in maintaining a planned evaluation design throughout a project's implementation, gave rise to four groups:

1. Schools that were eligible to receive a SIF intervention and did receive an intervention (compliers in the treatment group)
2. Schools that were eligible to receive a SIF intervention and did not receive an intervention (noncompliers in the treatment group)
3. Schools that were not eligible to receive a SIF intervention and did not receive an intervention (compliers in the control group)
4. Schools that were not eligible to receive a SIF intervention but did receive an intervention (noncompliers in the control group)

Consider first the situation with full compliance in both the treatment and the control group. Using a potential outcome notation, let $Y_i(1)$ denote the outcome for subject i under treatment and let $Y_i(0)$ denote the outcome for subject i without treatment. The average treatment effect on the treated (ATET) can be written as

$$(A-1) \quad ATET = E[Y(1) - Y(0) | se = 1] = E[Y(1) | se = 1] - E[Y(0) | se = 1]$$

where $se = 1$ denotes that treatment was received.

The first expectation in the last expression in equation (A-1) is just the average outcome for the treated, $E(Y | se = 1)$.

$$(A-2) \quad E[Y(1) | se = 1] = E[Y(1) | e = 1, se = 1]$$

where $e = 1$ denotes that the subject was eligible for the intervention. This expectation can be estimated by observing the mean outcomes for group 1.

The second expectation in equation (A-1) is counterfactual and not observed. However, it is possible to derive an expression for this counterfactual from the observed outcomes for other groups when there is randomization of eligibility and full compliance in the control group.

Note that the expected outcome without treatment for the eligible group can be expressed as the weighted average of the expected outcome without treatment for the subgroup that received treatment and that for the subgroup that did not:

$$(A-3) \quad E[Y(0) | e = 1] = E[Y(0) | se = 1, e = 1] \Pr(se = 1 | e = 1) \\ + E[Y(0) | se = 0, e = 1] \Pr(se = 0 | e = 1)$$

which, because of the randomization of eligibility, can be rewritten as

$$(A-4) \quad E[Y(0) | e = 0] = E[Y(0) | se = 1, e = 1] \Pr(se = 1 | e = 1) \\ + E[Y(0) | se = 0, e = 1] \Pr(se = 0 | e = 1)$$

The left-hand side of equation (A-4) is the average outcome for the noneligible group, which can be estimated as the average outcome in group 3 (with full compliance in the control group, group 4 is empty). The two probabilities on the right-hand side of the equation can be estimated using the fraction of the eligible group that received treatment and the fraction that did not. The last expectation can be estimated using the mean outcomes for group 2. Thus equation (A-4) can be solved for $E[Y(0) | se = 1, e = 1]$. With full compliance in the control group, this is equal to $E[Y(0) | se = 1]$. Substituting this solution into equation (A-1) yields an expression of the average treatment effect in terms of the observable expectations and probabilities:

$$(A-5) \quad ATET = E(Y | se = 1) - \frac{E(Y | e = 0)}{\Pr(se = 1 | e = 1)} + \frac{\Pr(se = 0 | e = 1) E(Y | se = 0, e = 1)}{\Pr(se = 1 | e = 1)}$$

Standard errors can be computed using the delta method.

In the data there is noncompliance in the control group: three schools received SIF funding even though they were not in the eligible group. Therefore, $E[Y(0) | e = 1]$ cannot be estimated directly. It is possible, however, to derive bounds for this expected value. Note that the original control group of schools consists of a group that did not receive SIF funding (compliers) and a group that did (noncompliers).

$$(A-6) \quad E[Y(0) | e = 0] = \Pr(se = 1 | e = 0) E[Y(0) | se = 1, e = 0] \\ + [1 - \Pr(se = 1 | e = 0)] E[Y(0) | se = 0, e = 0]$$

where $\Pr(se = 1 | e = 0)$ is the probability that a control group subject receives treatment. Note that $\Pr(se = 1 | e = 0) \neq \Pr(se = 1 | e = 1)$. That is, the probability of selecting for treatment depends on the assignment to the treatment or control group. It is usually easier to select for treatment if a subject is assigned to the eligible group. Thus

$$(A-7) \quad \Pr(se = 1 | e = 0) < \Pr(se = 1 | e = 1).$$

The last expectation on the right-hand side of equation (A-6) can be estimated directly using the mean observed outcomes for the subjects in the control group that did not receive an intervention (group 3). The probability can be estimated using the fraction of controls that received treatment. The first expectation on the right-hand side of the equation cannot be estimated directly. It is only possible to observe the expected outcome for the control group that received treatment. The following assumptions are used to derive an upper and lower bound for the unobserved expectation:

- Assumption 1: Treatment has no negative impact on outcomes.
- Assumption 2: The average outcome without treatment for control group members who received treatment is not less than the expected outcome before treatment for that same group plus 0.5 times the average change in outcome observed in the nontreated control group.

The first assumption needs little explanation. The second was chosen because education outcomes have generally improved in Bolivia, including for those groups not reached by the SIF. Stating that the expected improvement is more than half the trend observed for the nontreated population is therefore a mild assumption. The bound assumption can be written as

$$(A-8) \quad \begin{aligned} & E[Y_{t=0} | se = 1, e = 0] + 0.5(E[Y_{t=1} | se = 0, e = 0] \\ & - E[Y_{t=0} | se = 0, e = 0]) \leq E[Y_{t=1} (0) | se = 1, e = 0] \\ & \leq E[Y_{t=1} (1) | se = 1, e = 0] = E[Y_{t=1} | se = 1, e = 0] \end{aligned}$$

where the subscript t has been introduced to denote the period before the intervention (baseline, $t = 0$) or that after the intervention (follow-up, $t = 1$).

Using these bounds, one can estimate the upper and lower bounds of the treatment effect as defined in equation (A-1). Because $\Pr(se = 1 | e = 0)$ is small, these bounds will be relatively close. The three noneligible schools that managed to receive an intervention have relatively little impact on the final estimate of the treatment effect. Standard errors of the bounds can be computed using the delta method.

Rather than assumption 2, an alternative assumption could have been that the lower bound is 0 or that it is equal to $E(Y_{t=0} | se = 1, e = 0)$. These weaker assumptions give wider bounds. The bounds are reasonable if the degree of noncompliance in the control group is small. If there is substantial noncompliance, the local average treatment effect (LATE) estimator of Imbens and Angrist (1994) can be used. With full compliance in the control group, this estimator is the same as that derived. With noncompliance in the control group, the LATE estimator is not an estimator of the average treatment effect on the treated.

APPENDIX B. THE DATA

Data for the impact evaluation were collected through a baseline survey in 1993 and a follow-up survey that extended over late 1997 and early 1998 (the data from the follow-up survey are referred to as 1997 data in the article). Both sur-

veys collected data from 5 provinces in the Chaco region and 17 provinces in selected rural areas, referred to as the Resto Rural.

The data reflect the three types of investments considered: health care, education, and water supply. Five types of data collection instruments were used: household surveys, facilities surveys, community surveys, water quality samples, and student achievement tests. Each of these instruments was used in collecting data from both treatment and comparison groups.

Household Survey Data

The household survey consisted of three subsamples. The first was a random sample of all households in the Chaco and Resto Rural, which also served as the sample for the health component. The second was a sample of households living near the schools in the treatment or control group for the education component. The third consisted of households in the area of influence of water and sanitation projects.

The household survey gathered information on a variety of aspects, including household composition, household consumption (to generate a measure of poverty for an assessment of targeting), household involvement with the SIF project, individual household members' access to social services, and health and education outcome measures.

In 1997 an effort was made to interview the same households as in 1993, and 65 percent of the 1997 sample consisted of households that were also in the 1993 sample. A household that could not be traced was usually replaced by one nearby. The survey sample sizes are shown in table B-1.

Facilities Survey Data

SCHOOLS. The school survey used two questionnaires, one for the director and one for each teacher. It gathered information on infrastructure, equipment, teaching methods, and student dropout and repetition rates

In the Chaco region the sample of schools surveyed in 1993 was drawn from the group of primary and secondary schools that had been randomly selected as eligible (the treatment group) or not eligible (the control group) for active promotion of a SIF intervention (table B-2). In 1997 it appeared that half the eligible schools surveyed had succeeded in obtaining a SIF intervention.

TABLE B-1. Household Survey Samples
(number of households)

Type of investment	Chaco		Resto Rural	
	1993	1997	1993	1997
Health	2,029	1,941	2,138	1,901
Education	995	1,109	902	856
Water supply	666	594	569	540
Total	3,690	3,644	3,609	3,297

TABLE B-2. School Survey Sample in Chaco
(number of schools)

	1993	1997 intervention	1997 no intervention	Not surveyed in 1997 ^a
Eligible	36	17	18	1
Not eligible	35	3	31	1
Augmented sample (added in 1997)	n.a.	15	0	0
Total	71	35	49	2

n.a. = Not applicable.

^aSchools were not surveyed if key informants were absent at the time of the follow-up survey.

Because only a small number of schools from the original sample received an intervention, it was decided to augment the sample of treatment schools by selecting schools from the universe of those that had participated in the random assignment and had been selected for active promotion of a SIF intervention but had not been surveyed in 1993. The additional schools were randomly drawn from the set of schools that had obtained a SIF intervention by 1997. Because they had been subject to the original randomization process, it was assumed that their average characteristics would not differ significantly from those of the other schools included in the random assignment. The original design was contaminated by three schools that had been classified as noneligible for active promotion but had nevertheless obtained an intervention.

In the Resto Rural the sample consisted of a random sample of treatment schools along with a group of schools that had similar characteristics but did not receive investments (table B-3). Unlike in Chaco, schools were selected for the sample after decisions on interventions were made. Thus the sample included no schools that were eligible but did not receive an intervention.

WATER SUPPLY PROJECTS. At the time of the baseline survey the SIF was just beginning to invest in rural water projects. The 18 projects in the survey constitute the universe of projects considered for funding in 1992. (The SIF has greatly expanded its work in water and sanitation since then.) For these 18 projects, baseline and follow-up surveys were conducted to gather data on the character-

TABLE B-3. School Survey Sample in Resto Rural
(number of schools)

	1993	1997 intervention	1997 no intervention	Not surveyed in 1997 ^a
Treatment group	33	31	n.a.	2
Control group	35	n.a.	33	2
Total	68	31	33	4

n.a. = Not applicable.

^aSchools were not surveyed if key informants were absent at the time of the follow-up survey.

istics of households. A community survey was also conducted in the areas that received water projects, but there was no facility questionnaire like those for health and education.

In addition to these surveys, water quality tests were carried out, for both the old drinking water sources and the new, social fund-financed sources. These tests were carried out in situ using portable equipment and in one of the main water-quality testing laboratories in Bolivia.

HEALTH CENTERS. The health facility survey gathered information on staffing, visits to the center, and the quality of the infrastructure. Because the SIF had planned to invest in all health centers in Chaco and the Resto Rural, all were included in the survey (table B-4).

The survey distinguished between health centers at the sector, area, and district levels. Sector health centers are typically very small, providing basic health care. Area-level health centers provide more sophisticated care and serve a larger geographic region. District-level health centers are hospitals, the largest type of facility. The larger the health center, the more detailed the questionnaire administered. The questionnaires were nevertheless comparable and collected similar types of information.

Other Data

Mathematics and language achievement tests were applied in the follow-up survey in schools in the Chaco and Resto Rural. They could not be applied at the time of the baseline survey because they were developed and introduced as part of the Education Reform Program launched in 1994. However, the equivalency between the treatment and comparison groups established during the evaluation design stage in 1993, particularly through the application of the experimental design in the Chaco region, supports the assumption that the treatment and control groups would not have registered significantly different test scores in 1993.

A community survey collected data from community leaders on topics ranging from the quality of the infrastructure and distance to facilities to the presence of local organizations in both 1993 and 1997.

APPENDIX C. FIRST-STAGE ESTIMATES FOR PROPENSITY MATCHING

Education

Table C-1 presents estimates from the probit equation used to calculate the propensity score for the impact analysis of education investments. All variables are measured in 1993, before the intervention. Each variable was defined as the actual value if observed and 0 if missing. A dummy variable equal to 1 if any of the variables was missing for that observation was added to the probit equation. The number of observations was 119, and the pseudo- R^2 for the probit equation was 0.185.

TABLE B-4. Health Center Survey Sample
(number of health centers)

Type of health center	Chaco				Resto Rural			
	1993	1997 intervention	1997 no intervention	Not surveyed in 1997 ^a	1993	1997 intervention	1997 no intervention	Not surveyed in 1997 ^a
District	5	4	1	0	4	4	0	0
Area	16	9	6	1	22	4	17	1
Sector	62	47	5	10	84	22	58	4

^aHealth centers were not surveyed if key informants were absent at the time of the follow-up survey.

TABLE C-1. Results of First-Stage Probit for Impact Analysis of Education Investments

Variable	Coefficient	Standard error	t-statistic
<i>School characteristics</i>			
Number of classrooms	-.043	0.093	-0.457
Number of classrooms in adequate condition	.035	0.164	0.214
Have sanitation facilities	.090	0.289	0.312
Textbooks per student	-.204	0.412	-0.495
Cost of matriculation per student	-.218	0.09	-2.301
Number of students registered	.002	0.003	0.520
Repetition rate	-.002	0.019	-0.097
<i>Community characteristics</i>			
Number of nongovernmental organizations in community	.518	0.174	2.98
Population of community	.0003	0.0004	0.635
Functioning parents school association	-.007	0.357	-0.019
Whether respondents had heard of SIF	-.004	0.366	-0.012
<i>Household characteristics (average for community)</i>			
Father's education	-.15	0.10	-1.48
Mother's education	.22	0.16	1.32
Per capita household consumption	-.0002	0.0001	-1.08
Distance from household to school	.00004	0.00007	0.59
Dummy variable for missing observation	-.79	0.51	-1.57

Source: Authors' calculations.

Health

Table C-2 presents estimates from the probit equation used to calculate the propensity score for the impact analysis of health investments. All variables are measured in 1993, before the intervention. Each variable was defined as the actual value if observed and 0 if missing. A dummy variable equal to 1 if any of the variables was missing for that observation was added to the probit equation. The number of observations was 92, and the pseudo- R^2 for the probit equation was 0.403.

REFERENCES

- Alderman, Harold, and Victor Lavy. 1996. "Household Responses to Public Health Services: Cost and Quality Tradeoffs." *World Bank Research Observer* 11(1):3-22.
- Angrist, Joshua D., and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*. New York: Elsevier.
- Baker, Judy. 2000. *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Directions in Development Series. Washington, D.C.: World Bank.
- Brockerkhoff, Martin, and Laurie F. Derose. 1996. "Child Survival in East Africa: The Impact of Preventive Health Care." *World Development* 24(12):1841-57.

TABLE C-2. Results of First-Stage Probit for Impact Analysis of Health Investments

Variable	Coefficient	Standard error	t-statistic
<i>Health facility characteristics</i>			
Have electricity	-0.922	0.550	-1.676
Have sanitation facilities	-0.007	0.449	-0.017
Have water	0.412	0.478	0.863
Number of patient rooms	-0.576	0.411	-1.405
Index of availability of medical equipment in good condition	-0.017	0.016	-1.123
Index of availability of medical inputs	0.005	0.019	0.237
Number of beds	-0.328	0.237	-1.381
<i>Health outcomes before intervention</i>			
Fraction of women receiving any prenatal care	0.862	1.10	0.782
Fraction of women receiving at least one prenatal checkup who received at least four	2.513	0.975	2.577
Fraction of births attended by trained personnel	0.810	2.19	0.369
Incidence of cough	0.475	1.18	0.402
Fraction with cough receiving treatment	-0.767	0.887	-0.864
Incidence of diarrhea	-4.887	2.787	-1.753
Fraction with diarrhea receiving treatment	-0.448	0.757	-0.591
Fraction of children dying before age five	4.23	3.02	1.40
Fraction of entire community population using health center	10.85	6.325	1.716
<i>Community characteristics</i>			
Distance to nearest main road	0.003	0.012	0.251
Knowledge of SIF	0.558	0.451	1.239
Population of community	-0.001	0.0007	-0.810
Number of nongovernmental organizations in community	0.497	0.270	1.839
<i>Household characteristics (average for community)</i>			
Sum of father's and mother's education	-0.15	0.10	-1.48
Per capita household consumption	0.0002	0.0002	0.682
Dummy variable for missing observation	-1.45	0.51	-2.82
Constant	1.48	1.08	1.364

Source: Authors' calculations.

Coa, Ramiro. 1997. "Evaluación de la Calidad de Agua en una Muestra de Proyectos Financiados por el FIS." Mimeo.

Damiani, Ester. 2000. "Evaluación de la Calidad de Agua en una Muestra de Proyectos Financiados por el FIS." Mimeo.

Dehejia, Rajeev H., and Sadek Wahba. 1999. "Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94(448):1053-62.

Esrey, Steven A., James B. Potash, Leslie Roberts, and Clive Shiff. 1990. "Health Benefits from Improvements in Water Supply and Sanitation." WASH Reprint Technical Report 66. Water and Sanitation for Health Project, Arlington, Va. Available online at <http://www.sanicon.net/titles/title.php3?titleno=103>.

- Grossman, Jean. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *World Bank Research Observer* 9(2):159–80.
- Hanushek, Eric A. 1995. "Interpreting Recent Research on Schooling in Developing Countries." *World Bank Research Observer* 10(2):227–46.
- Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65(2):261–94.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–60.
- Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2):467–75.
- Kremer, Michael R. 1995. "Research on Schooling: What We Know and What We Don't. A Comment on Hanushek." *World Bank Research Observer* 10(2):247–54.
- Lavy, Victor, John Strauss, Duncan Thomas, and Philippe de Vreyer. 1996. "Quality of Health Care, Survival and Health Outcomes in Ghana." *Journal of Health Economics* 15(3):333–57.
- Lee, Lung-fei, Mark R. Rosenzweig, and Mark M. Pitt. 1997. "The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High-Mortality Populations." *Journal of Econometrics* 77(1):209–35.
- Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, Mass.: Harvard University Press.
- Mwabu, Germano, Martha Ainsworth, and Andrew Nyamete. 1993. "Quality of Medical Care and Choice of Medical Treatment in Kenya: An Empirical Analysis." *Journal of Human Resources* 28(4):838–62.
- Newman, John, Laura Rawlings, and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *World Bank Research Observer* 9(2):181–201.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.
- Subbarao, K., Kene Ezemerani, John Randa, and Gloria Rubio. 1999. "Impact Evaluation in FY98 Bank Projects: A Review." World Bank, Poverty Reduction and Economic Management Network, Washington, D.C.

Impact Evaluation of Social Funds

The Impact and Targeting of Social
Infrastructure Investments:
Lessons from the Nicaraguan Social Fund

Menno Pradhan and Laura B. Rawlings

The benefit incidence and impact of projects financed by the Nicaraguan Emergency Social Investment Fund are investigated using a sample of beneficiaries, a national household survey, and two distinct comparison groups. The first group is constructed on the basis of geographic proximity between similar facilities and their corresponding communities; the second is drawn from the national Living Standards Measurement Study survey sample using propensity score matching techniques. The analysis finds that the social fund investments in latrines, schools, and health posts are targeted to poor communities and households, whereas those in sewerage are targeted to the better-off. Investments in water systems are poverty-neutral. Education investments have a positive, significant impact on school outcomes regardless of the comparison group used. The results of health investments are less clear. Using one comparison group, the analysis finds that use of health clinics increased as a result of the investments; using both, it finds higher use of clinics for children under age six with diarrhea. With neither comparison group does it find improvements in health outcomes. Social fund investments in water and sanitation improve access to services but have no effect on health outcomes.

Social investment funds have quickly gained in popularity because of their capacity to carry out community development projects rapidly and with broad participation. An alternative to strategies led by central governments, social funds allow communities control in determining investment priorities. This model, widely implemented in a short period, has been the basis for the World Bank's first large-scale experience with small, community-led projects. The first social fund was created in Bolivia in 1987; today almost all countries in Latin America

Menno Pradhan is with the Nutritional Science Department at Cornell University and the Economics Department at Free University in Amsterdam, and Laura B. Rawlings is a Senior Monitoring and Evaluation Specialist in the World Bank's Latin America and the Caribbean Region. Their e-mail addresses are mpradhan@feweb.vu.nl and lrawlings@worldbank.org, respectively. The authors would like to thank the participants in the Nicaraguan Emergency Social Investment Fund (FISE) workshop held November 1999 at the World Bank for useful comments, particularly Florencia Castro-Leal, Carlos Lacayo, John Newman, Berk Ozler, Geert Ridder, and Steve Younger. They would also like to thank the team at the FISE for guidance and support and the team at the National Statistical Institute in Nicaragua for excellent data collection and processing. Funding for the research was provided by the governments of Nicaragua and Norway and the World Bank.

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

and the Caribbean have social funds or development projects that embody many of their operational characteristics. Social funds have also been established in Africa, Asia, Europe, and the Middle East.

Social funds finance small projects using a demand-driven process that allows the fund to appraise, finance, and supervise the implementation of social projects identified and executed by a range of local actors (Jorgensen and van Domelen 1999). In Latin America most social funds concentrate their investments in social infrastructure, particularly school construction projects (Goodman and others 1997). Although organizationally part of the central government, social funds generally operate outside the norms regulating public agencies, including those governing staff salaries. Some observers have praised social funds because as a result of these features, they can attract high-level professional staff and operate efficiently by separating the financing of investments from their provision; others have criticized them for providing a means of avoiding essential reforms in line ministries.

Despite the widespread use of social funds, until recently they have not been subjected to rigorous impact evaluations. As a result there is little knowledge and much debate about whether these demand-driven mechanisms can reach poor communities and households and whether the infrastructure investments they finance affect welfare outcomes. Given the scope of social funds and the national and international resources they have quickly absorbed, the need for serious evaluation of social funds is clear. This evaluation provides empirical data shedding light on the extent to which social funds have realized their goals.

This impact evaluation, one of the first on social investment funds, was motivated by the prominence of the Nicaraguan social fund and the interest in Nicaragua and beyond in assessing the ability of social funds to reach the poor and contribute to changes in their welfare. The Nicaraguan Emergency Social Investment Fund (Fondo de Inversión Social de Emergencia, or FISE) is the primary financier of health and education infrastructure in Nicaragua, with total operations accounting for more than 1 percent of the country's gross domestic product (GDP). It has grown remarkably since its establishment in 1990 and played a key role in expanding public infrastructure. In 1991–98 the FISE carried out 40 percent of the public investments in Nicaragua's social sector infrastructure (Bermudez and others 1999).

The FISE is patterned on the general model for social funds. Its central function is to finance infrastructure improvements in schools, health centers, water systems, and sanitation facilities at the request of local communities. It has also focused increasingly on combining its financial role with strengthening the planning and implementation capacity of local government.

This article examines whether FISE investments—in primary schools, rural health posts, latrines, and water and sewerage systems—are targeted to poor communities and poor households, improve access to basic social services, and help improve health and education outcomes. In doing so, the article contributes to the thin literature on the effects on household behavior and outcomes in

developing economies of quality improvements in health facilities (Alderman and Lavy 1996, Hotchkiss 1998, Lavy and others 1996), education facilities (Glewwe 1999, Hanushek 1995, Kremer 1995), and water and sanitation facilities (Brockhoff and Derose 1996, Lee and others 1997). Existing studies are based on cross-sectional variations in quality and do not take account of the endogeneity of the placement of government investments. By dealing explicitly with the endogeneity of investments, this article makes a new contribution. The article also contributes to the large literature on targeting, providing information on the outcomes achieved through the novel strategy of combining explicit targeting to poor areas with a demand-driven approach.

Because the locations of FISE interventions are determined through a non-random selection process, a simple comparison of health and education outcomes between areas that benefited from FISE investments and those that did not would not yield a valid estimate of the impact of the investments on beneficiaries. With only postintervention data available, the choice of evaluation techniques to address this selection issue is limited. This analysis applies a matched comparison technique in which each treatment subject is matched with a comparison subject that did not benefit from a FISE investment. Two comparison groups are used. One was drawn from similar schools and health posts in the proximity of the treatment facilities. The other was constructed using propensity score matching, a technique building on recent advances in the evaluation literature that has been applied mainly in labor market evaluations (Dehejia and Wahba 1998, Heckman and others 1998).

This household-level impact evaluation is part of a larger evaluation of the Nicaraguan social fund carried out by the World Bank in coordination with the FISE. The larger evaluation also includes an analysis of the quality and sustainability of FISE projects based on the results of a project-level survey, a review of the institutional evolution of the FISE, a comparison of the cost-effectiveness of FISE investments with that of similar projects carried out by another agency, and a contextual process evaluation of FISE projects in a subsample of 18 FISE communities selected from those surveyed for the impact evaluation. The results of all the studies are summarized in World Bank (2000).

I. THE FISE AND DATA SOURCES FOR THE IMPACT EVALUATION

The FISE was created in November 1990 to fund small-scale projects designed to meet the basic needs of the poor and create temporary employment, thereby contributing to the poor's economic and human capital and involving them in Nicaragua's economic and social development (Bermudez and others 1999). In 1991–98 the FISE invested US\$191 million, making it the largest social investment fund (as a percentage of GDP) in Latin America. On average, the FISE invested \$11.2 million a year in education, and the Ministry of Education invested \$11.7 million. The social fund's average yearly investment in health was \$5.8 million, and the Ministry of Health's was \$17.2 million. The FISE directed most

of its investments to infrastructure and equipment for primary and secondary schools (57 percent), health posts and health centers (8 percent), infrastructure for water and sanitation (9 percent), latrine facilities (7 percent), and public works (16 percent). This article considers all but the last category.

The FISE uses a poverty map to target investments to the poor. The poverty map used to guide the projects reviewed in this evaluation is based on the 1993 Living Standards Measurement Study (LSMS) survey—a nationwide household survey—and contains a poverty measure developed by the FISE for each municipality.¹ Estimates based on the ranges established by this poverty map show that in 1991–98, 23 percent of FISE investments went to municipalities with “extreme” poverty, 53 percent to municipalities with “high” poverty, and 24 percent to municipalities with “medium and low” poverty. Both the municipalities with extreme poverty and those with high poverty received a larger share of FISE investments than their share of the population, pointing to a progressive geographic distribution of resources by the standards of the poverty map (table 1). Moreover, the FISE’s allocation of resources to extremely poor municipalities has improved, rising from 11 percent of investments in 1991 to 34 percent of investments in 1998. The FISE has recently updated its poverty map using results of the 1998 LSMS survey and the 1995 census and applying new methodologies that allow the imputation of consumption-based poverty levels.²

The poverty targeting and impact analyses carried out in this study rely on three sources of data: the 1998 LSMS survey, the FISE household survey, and administrative data. The FISE household survey applied the same questionnaire and was fielded at the same time as the 1998 LSMS survey. Both surveys followed the established practices developed in the World Bank LSMS initiative (Grosh and Glewwe 1995). The FISE household survey sampled from households in the area of influence of randomly chosen FISE projects and matched comparison (non-FISE) projects (in health and education only). The area of influence was determined on the basis of service provision norms for schools and health centers and project records on FISE construction for water, sewerage, and latrine projects.

At the sampling stage there was concern that random sampling would not yield sufficient observations of households that actually used the facilities targeted by FISE investments. For this reason, choice-based sampling techniques

1. The poverty map is based on several weighted measurements used to construct a composite poverty score assigned to municipalities based on their basic needs, per capita income, and population size. First, the poverty map is based on three indicators of poverty, each with the following weights: infant malnutrition (40 percent), access to drinking water (40 percent), and the proportion of displaced individuals (20 percent). The results are then weighted to favor the poorest municipalities using a relative poverty indicator (RPI), which measures income levels relative to the cost of a basket of basic goods. Based on the RPI, municipalities are divided into three groups: extreme poverty, high poverty, and medium and lower poverty. Finally, the poverty map score is weighted by the size of municipal populations using estimates based on the 1971 census.

2. For more information on techniques combining census and survey data to estimate poverty rates see Elbers and others (forthcoming).

TABLE 1. Poverty Targeting of FISE Investments Across Municipalities, 1991–98

Municipal poverty ranking	Number of municipalities	Share of population (percent)	Total investment, 1991–98 (US\$ millions) ^a	Average annual per capita investment(US\$)
Extreme	42	18.4	43.6 (22.8)	6.25
High	96	51.6	101.7 (53.2)	5.33
Medium and low	9	30.0	46.1 (24.1)	3.79
Total	147	100.0	191.34 (100.0)	4.98

^aFigures in parentheses are percentage shares of the total.

Source: World Bank (2000).

were applied. Within the randomly chosen set of census segments in the area of influence, all households were classified as either direct beneficiaries or potential but not direct beneficiaries.³ Two samples were drawn, one from the group of direct beneficiaries that were confirmed as users of the social fund investment and one from the group of potential beneficiaries. Weights were constructed to correct for the sampling in the analysis stage (Manski and Lerman 1977). Sample sizes for the FISE survey (which includes households in the area of influence of both FISE projects and non-FISE schools and health posts) are shown in table 2. The sample size for the LSMS survey, from which comparison groups were constructed using propensity score matching methods explained later, was 4,040 households.

The administrative data used in the analysis come from a data file containing the census segments associated with the areas of influence of the universe of FISE health and education projects by type of project. A census segment is included in the database if more than 50 percent of the segment is located within the area of influence of a selected project. This file makes it possible to separate the households in the 1998 LSMS survey into two groups: potential beneficiaries and others.⁴ In addition, the analysis uses data from the poverty map employed by the FISE in targeting its investments. This map contains the estimated poverty head count ratio (share of the population in poverty) for each municipality.

3. For education projects, direct beneficiaries are households that have at least one child in the FISE school. For health projects they are households in which a member has visited the FISE clinic in the past year. For sewerage projects they are households that have a flush toilet connected to the sewer. Water and latrine projects are public access facilities, allowing no distinction between direct and potential beneficiaries.

4. Those living outside the area of influence of a FISE project could decide to benefit from the project. For instance, children living outside the area of influence of a FISE school could enroll in the school. Thus, there is no guarantee that those living outside the area of influence of a FISE project did not benefit from the intervention, potentially biasing the comparison group. This caveat holds for both types of comparison groups.

TABLE 2. Sample Size of FISE Survey
(number of households)

	Treatment or comparison group	Direct beneficiaries	Potential beneficiaries	Total
Education	Treatment	161	79	240
	Comparison	142	99	241
Health	Treatment	165	34	199
	Comparison	164	35	199
Water	Treatment	95	0	95
Sewerage	Treatment	74	30	104
Latrines	Treatment	234	0	234
Total				1,312

Source: World Bank (2000).

II. TARGETING OF FISE INVESTMENTS

The analysis distinguishes between two levels of targeting. First, it explores community-level targeting by examining the characteristics of households in the area of influence of FISE projects—the potential beneficiaries. Second, it investigates household-level targeting by examining the characteristics of households using the FISE investments—the direct beneficiaries. To evaluate the benefit incidence of social fund investments, the analysis applies a conventional benchmark by comparing an implicit transfer with a uniform transfer. The implicit transfer is obtained by assuming that everyone who uses a social fund facility obtains an equal benefit. A uniform transfer assumes an equal transfer to every individual in the population. When the social fund investments reach a larger proportion of the poor than a uniform transfer would, the social fund is considered progressive (on a per capita basis).

Concentration coefficients are used to assess the targeting of FISE investments to the poor. The analogue of Gini coefficients for Lorenz curves, concentration coefficients are derived from concentration curves, which show the cumulative percentage of benefits received by the population ranked according to a welfare measure, in this case per capita consumption. The coefficients range from -1 (all transfers go to the poorest) to 1 (all transfers go to the richest). The concentration coefficient is defined as $1 - 2 \int G(x) dx$, where $G(x)$ is the concentration curve.⁵ A major advantage of using concentration curves is that information on the average probability of benefiting from an intervention is not needed. For any consumption level x , the concentration curve shows the fraction of the population with per capita consumption below x (derived from the LSMS survey) against the fraction of beneficiaries with per capita consumption below x (derived from the FISE beneficiaries survey). The curve can thus be computed using two independent surveys.

5. For information on the concentration curves constructed for this study see World Bank (2000).

The analysis also examines the share of social fund benefits accruing to those below the poverty line and the extreme poverty line used in Nicaragua.⁶ In 1998, 48 percent of the population of Nicaragua lived below the poverty line, and 17 percent below the extreme poverty line. If the share of social fund benefits accruing to these groups is larger than their population share, the investments are progressively targeted to these groups.

The concentration coefficients for FISE investments in education show that they are distributed with a slight propoor bias, although the incidence of benefits is close to neutral for the extreme poor (table 3). This is a common finding in analyses of the benefit incidence of education investments, and it arises mainly from the fact that poor households generally have more children.⁷ When the analysis includes only direct beneficiaries (households with at least one child enrolled in a FISE school) rather than potential beneficiaries, the concentration curve falls slightly higher, indicating that FISE schools have been relatively successful in reaching poor children within the communities where the schools are located.

FISE health interventions reveal a more propoor distribution than the education interventions. This outcome is explained in part by the fact that health posts are typically in rural areas, whereas primary schools are in both rural and urban areas. Whether potential or direct beneficiaries are used in the analysis makes little difference in the targeting results for health interventions, indicating that the likelihood of visiting an FISE facility, conditional on living in an area where one is present, does not depend on income.

The targeting outcomes for water and sanitation investments reveal a great deal of heterogeneity. Latrine investments are the most progressive of all those analyzed in the impact evaluation. Water investments are distributed quite evenly across the population, showing neither a strong prorich nor a strong propoor bias. Sewerage interventions are very poorly targeted, both at the community level (potential beneficiaries) and at the household level (direct beneficiaries).

In considering the poverty targeting results, it should be kept in mind that the nature of projects can affect their potential to reach poor households. Water and sewerage projects need to reach a certain scale to be cost-effective and thus are typically concentrated in more populated areas, which tend to be wealthier. Latrines tend to be used only by the poor, so the success of latrine investments in reaching the poor and the extreme poor reflects the self-targeted nature of this

6. The poverty line is set at \$344, considered to be the level of annual per capita consumption necessary for a person to attain the minimum caloric requirements. The measure takes into account non-food items. The extreme poverty line (also called the food poverty line) is set at \$181, considered to be the level of annual per capita food expenditure necessary for a person to satisfy the minimum daily requirement of 2,226 calories.

7. The benefit incidence of education investments depends in part on the number of children enrolled from a household and the poverty ranking of the household. The choice of welfare measure here—per capita consumption—assumes that there are no economies of scale in household consumption; changing this assumption could lead to reversals in poverty rankings (Lanjouw and Ravallion 1995).

TABLE 3. Targeting of FISE Investments to the Poor and Extreme Poor, 1998

Type of project	Concentration coefficient		Share of benefits reaching extreme poor (percent) ^a		Share of benefits reaching poor (percent) ^b	
	Potential beneficiaries	Direct beneficiaries	Among potential beneficiaries	Among direct beneficiaries	Among potential beneficiaries	Among direct beneficiaries
Education	-0.061	-0.111	18.0	18.1	53.9	59.2
Health	-0.120	-0.115	17.0	12.3	64.1	65.2
Water	-0.004	n.a.	12.3	n.a.	49.9	n.a.
Sewerage	n.a.	0.420 ^c	n.a.	5.1 ^c	n.a.	n.a.
	0.430	0.370 ^d	4.0	8.3 ^d	10.7	8.6
Latrines	-0.301	n.a.	26.9	n.a.	73.3	n.a.

n.a. = Not applicable.

^aThe 1998 LSMS survey observed an extreme poverty rate of 17 percent.

^bIncludes extreme poor. The 1998 LSMS survey observed a poverty rate of 48 percent.

^cBased on broad definition of direct beneficiaries (households with any access to sewerage system).

^dBased on narrow definition of direct beneficiaries (households with flush toilet connected to sewerage system).

Source: Authors' calculations based on 1998 LSMS survey, FISE survey, and FISE administrative data.

type of investment. The targeting outcomes for water and sanitation investments by FISE are consistent with those from other countries (Rawlings and others 2002).

III. IMPACT EVALUATION

The central question posed by the impact evaluation is this: If the FISE had not existed, what would the condition of the beneficiaries have been? The analysis compares this counterfactual condition with the results from the survey of program beneficiaries to estimate the impact of FISE investments in health posts, primary schools, water systems, and sanitation (sewerage systems and latrines) on beneficiaries' access to and use of these basic services as well as their health and education status.

Impact Evaluation Methodology

Because the impact evaluation was designed without the benefit of baseline data, the counterfactual was constructed using a matched comparison technique.⁸ This method defines a comparison group of individuals who did not have the opportunity to benefit from an FISE project. If this group is similar to the treatment

8. For an overview of different methods of impact evaluation see Grossman (1994). An alternative approach is difference in differences, but this was not feasible because of the lack of a preintervention survey. Another alternative is to use instrumental variables. This technique was not applied because there were no good candidates for variables that influence the selection of a community into the program but not the outcome. Such variables typically measure the ability of a community to obtain a project. This information is usually collected through a community questionnaire, which was not included in the 1998 LSMS survey.

group in all relevant preintervention characteristics, a direct postintervention comparison of the comparison and treatment groups provides an estimate of the impact of the FISE intervention. The two groups should be similar in both observable and unobservable characteristics that influence outcomes and selection into the program. Constructing such a comparison group is a nontrivial matter.

A simple comparison of health and education outcomes between areas that benefited from FISE investments and those that did not would not yield a valid estimate of the impact of the investments because of the nonrandom selection process for FISE interventions. These selection issues arise from the allocation process for social fund investments, which takes into account the preferences of both communities and the social fund. Communities take the initiative in applying for a social fund project, including selecting the type of project, such as constructing latrines or rehabilitating a school. Communities' ability to prepare and execute project proposals also determines in part the likelihood that they will receive a project. The preferences of the social fund come into play during the promotion and review of projects. For instance, FISE, using its poverty map, allocates more resources to poorer areas.

Two types of matched comparison methodologies were used to construct a comparison group for estimating the counterfactual. First, for health and education projects only, a FISE comparison group was constructed during the sampling stage of the study, before the FISE survey was implemented. Each FISE facility included in the survey was matched to the nearest non-FISE facility, with the match restricted to facilities of similar size and type.⁹ The FISE survey collected information on households in the area of influence of the FISE facilities as well as households in the area of influence of the non-FISE comparator facilities; this second set of households made up the FISE comparison group.

Second, a propensity score comparison group was constructed separately for each of the interventions from the 1998 LSMS sample using propensity score matching techniques. This score measures the probability that a subject receives an intervention as a function of observable characteristics. Rosenbaum and Rubin (1983) show that if it is valid to match using these characteristics, it is equally valid to match using only the propensity score. This matching method greatly simplifies the problem and allows the inclusion of many variables in the propensity score, thereby reducing the role of unobservables.

One can say little a priori about which comparison group should be preferred for analyzing the impact of FISE health and education investments. Both rely on presumptions about the method that is most suitable for creating a comparable comparison group. The FISE comparison group is based on the notion that the

9. Characteristics used for matching FISE and non-FISE facilities include location (urban or rural) and the poverty category of the municipality. Number of classrooms was also used to match schools and type of facility (according to Ministry of Health norms) to match health posts. Based on these criteria, each FISE facility was matched to the nearest non-FISE facility that did not have an overlapping area of influence.

nearest similar non-FISE health post (or school) and the corresponding households are equivalent to the FISE health post (or school) and corresponding households before the FISE intervention. The propensity score match is valid under the assumption that the variables included in the propensity score functions are sufficient to eliminate the selection bias between the treatment and comparison groups.

Propensity Score Matching Methodology

The variables entering into the propensity score function are chosen with the knowledge that the preferences of both local communities and the FISE determine the final allocation of projects. The variables measure the ability of a community to prepare project proposals, the preferences of the FISE (poverty map data), and, where available, preprogram outcomes (outcome indicators are available only for the water and sanitation interventions). With only a postintervention survey available, the analysis must rely on recall information for preprogram outcomes. All other explanatory variables are valid under the assumption that they have not changed as a result of the FISE intervention. This limitation again emphasizes the need for comparable baseline data when evaluating social programs.¹⁰

Propensity score matching requires one to estimate the probability that an individual lives in the area of influence of a facility receiving an investment. To estimate this function precisely, one needs to know exactly which communities received a FISE investment and which did not. The LSMS survey might appear to be the most appropriate source for this information, since it asks households whether or not they benefited from an FISE investment. But many households do not realize that their community received an FISE investment—and worse, many households whose community did not receive an FISE investment think that it did.¹¹ The analysis therefore relies on FISE administrative data, which provides the census segments associated with the areas of influence of all FISE health and education projects by project type. This file, merged with the LSMS survey results, separates households between those that are in the area of influence of an FISE project and those that are not. This allows an estimate of the propensity score associated with living in the area of influence of an FISE project for individual i :

$$(1) \quad \text{Pr}(\text{potential beneficiary}) = F(X_i\beta).$$

10. The 1993 LSMS survey could not serve as a baseline because it covered different communities than the 1998 LSMS survey, and because of sample size limitations relative to the population of FISE beneficiaries.

11. The FISE survey includes information on whether or not respondents are in a community that had received an FISE investment. The results show, for instance, that only 90 percent of the households classified as direct beneficiaries of an FISE education project claimed that they had benefited from an FISE education investment, whereas 25 percent of those in the FISE comparison group claimed that they had benefited from such an investment.

The function F is estimated using a probit model. X_i are the observed characteristics of individual i . They include community characteristics and FISE targeting instruments.

Modeling the propensity score for a FISE water, sewerage, or latrine investment is easier than constructing the corresponding model for a health or education investment. Almost all the water and sanitation projects were included in the FISE beneficiaries survey, and thus the likelihood that a household included in the LSMS survey is in a community that received one of these projects is negligible. Moreover, the area of influence of water and sanitation projects is geographically defined. The propensity score function is therefore estimated using the combined data of the LSMS survey and the FISE beneficiaries survey, with the assumption that none of the households in the LSMS survey benefited from a water and sanitation project. Sampling weights are used to correct for the choice-based sampling (Manski and Lerman 1977).

Because there is no limit on the number of explanatory variables that can be included in the propensity score function, the analysis uses a fully interacted model for health and education investments. But because the coefficient estimates are difficult to interpret, the article presents the estimates of the probit models in which none of the variables is interacted.

To ensure comparability between the results for the FISE comparison group and those for the propensity score comparison group, the treatment population is always defined as those identified as (potential) FISE beneficiaries in the FISE household survey. The propensity score comparison groups are drawn from the LSMS sample, restricted to the areas in which the FISE has no projects of the type being investigated. The population from which the match is drawn depends on the impact variable used. If the focus is on children's enrollment, for instance, the comparison group is restricted to school-age children. The population is also limited to the geographic region in which the treatment population lives, based on the assumption that households within a region share characteristics that are not fully captured by the regional dummy variable in the model. Limiting the selection of comparison group subjects to those living in the same region as the treatment group increases the likelihood that the two groups will be similar. The geographic restriction did not affect the ability to find a good match for every treatment group.

PREDICTING PARTICIPATION IN FISE PROJECTS. Estimation results for the probability of living in the area of influence of an FISE health or education project as defined in equation 1 show that the geographic variables are highly significant (table 4). This finding reflects the tendency of the FISE to invest in poorer areas, a preference confirmed by the benefit incidence analysis and reconfirmed by the significant positive effect of the poverty head count ratio of the municipality from the poverty map. Results for the access road variables, included as a proxy for the remoteness of the municipality, show that households with worse access roads have a higher chance of living in the area of influence of an FISE project.

TABLE 4. Probit Estimates of Geographic Location of FISE Education and Health Projects

Variable	Education projects		Health projects	
	Coefficient	<i>t</i> -statistic	Coefficient	<i>t</i> -statistic
Managua	0.594	4.72	n.a.	n.a.
Pacific urban	0.757	8.08	-0.545	-11.19
Pacific rural	0.434	5.28	0.571	15.71
Central urban	0.132	1.43	-0.515	-11.06
Atlantic urban	0.193	2.09	-1.332	-21.87
Atlantic rural	-0.287	-2.83	-0.034	-0.81
Log per capita consumption	-0.040	-1.01	-0.049	-2.57
Paved road to house	-0.030	-0.44	-0.133	-3.52
Dirt road to house	0.141	2.04	0.213	6.54
Non-FISE projects from which household benefited	0.040	1.13	0.045	2.35
Non-FISE projects from which household benefited and in which it participated	-0.202	-2.40	0.023	0.60
Total membership of community organizations	-0.047	-1.32	-0.005	-0.31
Distance to school or clinic	-0.152	-5.74	-0.048	-15.11
Head count ratio in municipality based on FISE poverty map	0.014	4.89	0.001	0.84
Gini coefficient in region	1.089	3.71	1.514	9.92
Constant	-0.387	-0.50	-0.774	-2.17
Pseudo R^2	0.052		0.111	

n.a. = Not applicable.

Note: The data in the table are not the estimates for the propensity score function. The propensity score function applies the same explanatory variables but is fully interacted. Dependent variable = 1 if household lives in the area of influence of a health or education project.

Source: Authors' calculations based on matched FISE administrative data on geographic locations of projects and LSMS sample.

The number of non-FISE projects from which a household has benefited and the number of such projects in which a household has participated are included as proxies for a community's ability to develop projects and obtain project financing. The results reveal that the number of non-FISE projects a community has received has no significant effect on the probability of its receiving a FISE education project. By contrast, community participation has a negative effect, possibly because once a community has obtained a non-FISE education project, it is less likely to seek a similar FISE project. The number of non-FISE projects has a positive effect on a community's ability to obtain FISE health projects.

As expected, distance to a FISE facility has a negative effect on the probability of living in its area of influence. Income inequality in the region, as measured by the estimated Gini coefficient, has a positive effect on the probability of obtaining a FISE project.

The R^2 in education is 0.052 and increases to 0.1038 when the fully interacted model is used, whereas the R^2 in health is 0.111 and increases to 0.1632.

A low R^2 does not necessarily mean that the propensity score function is not good. In the extreme case, when the allocation of projects has been de facto random, the R^2 would be zero, but the resulting propensity score comparison group would be perfect.

The estimated propensity score functions for latrine and sewerage projects include a higher-order term for consumption than do those for water projects because of the strong targeting bias found (propoor for latrine projects and prorich for sewerage investments). Interaction terms are not used for water and sanitation projects. Experiments with interaction terms for these models, which have fewer degrees of freedom, found that their use worsens the overlap of the propensity score functions. It was therefore decided to continue with a limited set of descriptive explanatory variables, which yielded a good overlap. (The results, omitted here because of space limitations, are available in Pradhan and Rawlings 2000.)

MATCHING PROCESS. Beneficiaries observed in the FISE sample are matched to similar individuals from the 1998 LSMS survey who did not live in the area of influence of a FISE project. Individuals can be matched only once—that is, without replacement. To test the quality of the propensity score match, propensity scores were plotted for the treatment and comparison groups for each area of investment under evaluation (education, health, water, sewerage, and latrines). Except for a few observations in the health treatment group with high propensity scores, the curvatures of the functions observed for each treatment group come very close to overlapping with those of the comparison group. These results indicate strong similarities between the treatment and comparison groups and a high-quality match.¹²

The Impact of FISE Investments on Beneficiary Households

An unbiased estimate of the average treatment effect of an FISE intervention can be obtained by simply comparing mean outcomes in the comparison and treatment groups. For the treatment group, constructed using propensity score matching, the t -test for equal means has to take account of the uncertainty arising from the fact that the comparison group sample is based on an estimated coefficient vector in the propensity score function. The standard errors for this comparison are calculated using bootstrapping with 400 replications. In each iteration a new comparison group is constructed using a random draw from the estimated distribution of the coefficient vector of the propensity score function. Following the usual bootstrap procedures, a random sample of equal size is drawn from the matched sample with replacement. The observed difference in means in the bootstrapped sample takes account of both the uncertainty arising from the fact that the comparison group sample is based on an estimate and the fact that the treatment group estimate is based on a sample of limited size.

12. The probit estimates for the treatment and comparison groups are compared in Pradhan and Rawlings (2000).

EDUCATION. The average impact of living in the area of influence of a FISE school is estimated for several indicators—enrollment, the education gap (the difference between the ideal educational attainment, given a child’s age, and the highest grade attended), age for grade, repetition, attendance, and age in first grade. Enrollment appears to have increased as a result of the FISE investments (table 5). The net enrollment ratio for the treatment group is almost 10 percentage points higher than that for the propensity score comparison group, though the difference is smaller—4.5 percentage points—and insignificant for the FISE comparison group.¹³ Results for both comparison groups confirm the impact of the FISE in reducing the education gap from around 1.8 years to 1.5. The effect is significant for both comparison groups. No significant effect is found for the share of children in the correct grade for their age. However, the age of first grade children dropped sharply—from 8.6 to 7.9 years—as a result of FISE education investments, a finding confirmed by results for both comparison groups. Nonetheless, absenteeism is high in FISE schools, averaging 6.8 days a month. Although this rate is slightly better than that observed in the FISE comparison group, it is significantly worse than that observed in the propensity score comparison group, rendering the results inconclusive.

Results based on the two comparison groups in education are fairly consistent and are also significant. This suggests that the significant, positive effects of FISE investments in primary education on enrollment, the education gap, and age in first grade are robust.

The effects of FISE education investments are also estimated separately for different consumption quintiles and by gender (table 6). Results based on both comparison groups confirm that FISE education investments have a greater effect on girls’ enrollment than on boys’. They show that the investments have a greater effect in reducing the education gap and increasing the share of children in the correct grade for age for children in poorer quintiles.

HEALTH. The effects of FISE interventions in health are less clear, rendering the results inconclusive. Beneficiary households had a higher contact rate (that is, were more likely to have visited a health post or health center in the past month) than the propensity score comparison group, but there was no significant difference between the treatment group and the FISE comparison group (table 7). Estimation results based on the FISE comparison group indicate an improvement in the contact rate for children with diarrhea, but those based on the propensity score comparison group show no significant effect. Although results based on the FISE comparison group point to an improvement in indicators of acute malnutrition resulting from FISE investments, those based on the propensity score comparison group do not confirm this finding. For most of the other outcome variables the differences between the treatment and comparison groups are in-

13. An evaluation of the Peruvian Social Investment Fund also found a positive effect of social fund investments on school enrollment (Paxson and Schady 2002).

TABLE 5. Impact of FISE Education Investments on School Outcomes

Indicator	Treatment group		FISE comparison group		<i>t</i> -test on means	Propensity score comparison group		<i>t</i> -test on means
	No. observations	Mean	No. observations	Mean	(<i>p</i> -value)	No. observations	Mean	(<i>p</i> -value) ^a
Net primary enrollment ratio (percent)	341	91.7	358	87.2	0.056	341	82.1	0.073
Education gap (years) ^b	338	1.5	357	1.7	0.039	335	1.9	0.0279
Children in correct grade for age (percent)	341	26.0	358	25.5	0.889	341	21.8	0.5094
Days of school missed in past month	302	6.8	313	7.3	0.394	259	1.7	0.000
Age in first grade (years)	76	7.94	85	8.60	0.001	77	8.56	0.0875

^aBased on bootstrapped estimates with 400 replications.

^bDifference between ideal educational attainment, given a child's age, and the highest grade attended.

Source: Authors' calculations based on 1998 LSMS survey, FISE survey, and FISE administrative data.

TABLE 6. Impact of FISE Education Investments by Consumption Quintile and Gender

Indicator and population group	Quintile ^a					Boys	Girls
	1 (poorest)	2	3	4	5 (richest)		
<i>Net primary enrollment ratio (percent)</i>							
Treatment group	82.8	96.1	96.4	94.7	90.2	90.0	93.9
FISE comparison group	85.9	86.9*	97.9	82.0*	84.6	87.1	87.4*
Propensity score comparison group	69.2*	93.3	85.1*	73.9*	89.1	82.4*	81.7*
<i>Education gap (years)^b</i>							
Treatment group	1.8	1.4	1.7	1.3	0.6	1.6	1.3
FISE comparison group	2.2*	2.0*	1.5	1.5	0.7	2.1*	1.3
Propensity score comparison group	2.6*	2.0*	1.8	1.7	0.6	2.1*	1.7*
<i>Children in correct grade for age (percent)</i>							
Treatment group	16.8	23.6	25.4	24.8	55.3	21.8	31.5
FISE comparison group	12.4	19.5	36.7	27.0	48.2	22.9	28.2
Propensity score comparison group	4.5*	9.1*	21.2	43.1	66.4	17.1	27.9
<i>Days of school missed in past month</i>							
Treatment group	6.0	9.2	6.5	6.9	4.5	6.9	6.6
FISE comparison group	8.9*	8.1	7.6	7.4	3.5	8.6*	6.3
Propensity score comparison group	1.6*	2.0*	2.2*	0.6*	1.9	1.7*	1.9*

*Difference between treatment and comparison groups significant at the 5 percent level.

^aBased on the national distribution of per capita consumption as observed in the 1998 LSMS survey.

^bDifference between ideal educational attainment, given a child's age, and the highest grade attended.

Source: Authors' calculations based on 1998 LSMS survey, FISE survey, and FISE administrative data.

significant. Estimation of the average treatment effect by gender and consumption quintile leads to similar inconclusive results.

WATER SYSTEMS. Impact estimates for FISE water projects show that the investments had a significant, positive effect on water supply (table 8). The variables for change in access to infrastructure, constructed using recall information from 1993, before the FISE investments, are equivalent to difference-in-difference estimators. The results show that the share of households with access to piped water increased by about 21 percentage points more in areas where the FISE invested than in areas where it did not. Variables for rates of malnutrition and diarrhea all indicate an improvement in health status, but the results are not significant.

SEWERAGE SYSTEMS. The FISE has had a significant, positive impact on access to sewerage systems in the areas where it has invested (see table 8). The treatment group is defined as direct beneficiaries—households with a flush toilet connected to the sewerage system. The propensity score comparison group is drawn from the eligible population—households not connected to a sewerage system in 1993, based on recall data on access to water and sanitation facilities in that year. Without a FISE intervention, only 8.7 percent of households in the propen-

TABLE 7. Impact of FISE Health Investments on Health Outcomes

Indicator	Treatment group		FISE comparison group			Propensity score comparison group		
	No. observations	Mean (percent)	No. observations	Mean (percent)	<i>p</i> -value for equal means	No. observations	Mean (percent)	<i>p</i> -value for equal means ^a
Contact rate ^b	1,169	10.3	1,196	11.1	0.523	1,169	5.6	0.029
Contact rate for children under age 6 ^b	223	23.4	207	19.4	0.315	223	5.6	0.008
Contact rate for people over age 5 ^b	946	7.2	948	9.6	0.053	946	5.7	0.425
Incidence of diarrhea in past month in children under age 6	220	27.0	207	22.6	0.286	220	16.9	0.153
Contact rate for children with diarrhea ^b	50	43.3	40	18.1	0.009	47	17.0	0.255
Incidence of cough or other respiratory disease in past month	1,169	22.5	1,196	23.5	0.562	1,169	18.9	0.342
Share of women giving birth in past five years who had at least one prenatal checkup in that period	104	76.1	107	69.3	0.271	104	87.4	0.293
Share of institutional births	104	69.0	107	55.0	0.036	104	70.8	0.881
Share of births attended by skilled health staff ^c	104	97.7	107	94.5	0.236	104	94.9	0.324
DPT vaccination coverage ^d	36	86.7	25	94.2	0.320	36	96.3	0.518
Polio vaccination coverage	36	93.6	25	97.3	0.491	36	99.8	0.564
Prevalence of wasting (low weight for height) ^e	164	0.4	144	4.7	0.020	164	1.1	0.739
Prevalence of stunting (low height for age) ^e	164	20.5	144	24.2	0.436	164	17.3	0.717
Prevalence of underweight (low weight for age) ^e	164	10.1	144	19.5	0.021	164	11.4	0.739

^aBased on bootstrapped estimates using 212 replications.

^bContact rate shows the percentage of individuals who visited an outpatient public health care provider in the past month.

^cGynecologist, nurse, nurse assistant, or midwife.

^dDPT is diphtheria, pertussis (whooping cough), and tetanus.

^eModerate malnutrition with z-scores less than -2 for children under age 6.

Source: Authors' calculations based on 1998 LSMS survey, FISE survey, and FISE administrative data.

TABLE 8. Impact of FISE Water and Sanitation Investments on Health and Infrastructure
(percent, except where otherwise specified)

Indicator	Treatment group		Propensity score comparison group		<i>p</i> -value for equal means ^a
	No. observations	Mean	No. observations	Mean	
<i>Water investments</i>					
Incidence of diarrhea in past month among children under age 6	79	18.8	157	25.4	0.399
Prevalence of wasting (low weight for height) ^b	102	3.4	108	3.6	0.946
Prevalence of stunting (low height for age) ^b	102	13.6	108	24.0	0.204
Prevalence of underweight (low weight for age) ^b	102	15.6	108	18.5	0.690
Distance to water source in 1997 (km)	95	0.0090	189	0.075	0.334
Change in distance to water source between 1993 and 1997 (km)	95	-0.1298	189	-0.042	0.157
Share of households with piped water in 1997	95	84.6	189	56.5	0.0000
Change in share of households with piped water between 1993 and 1997 (percentage points)	95	27.3	189	5.9	0.0000
<i>Sewerage investments</i>					
Incidence of diarrhea in past month among children under age 6	23	9.4	45	21.9	0.237
Prevalence of wasting (low weight for height) ^b		0		0	n.a.
Prevalence of stunting (low height for age) ^b	31	12.2	30	16.9	0.683
Prevalence of underweight (low weight for age) ^b	31	16.0	30	6.9	0.414
Share of households with flush toilet in 1997	31	100.0	61	8.7	0.000
Change in share of households with flush toilet between 1993 and 1997 (percentage points)	31	100.0	61	8.7	0.000
<i>Latrine investments</i>					
Incidence of diarrhea in past month among children under age 6	226	29.16	451	24.52	0.365
Prevalence of wasting (low weight for height) ^b	313	5.8	312	4.7	0.694
Prevalence of stunting (low height for age) ^b	313	23.7	312	22.4	0.817
Prevalence of underweight (low weight for age) ^b	313	12.7	312	13.9	0.798
Share of households with no toilet in 1997	224	1.86	447	23.00	0.000
Change in share of households with no toilet between 1993 and 1997 (percentage points)	224	-31.87	447	-13.19	0.000

^aBased on bootstrapped estimates with 200 replications.

^bModerate malnutrition with z-scores less than -2 for children under age 6.

Source: Authors' calculations based on 1998 LSMS survey, FISE survey, and FISE administrative data.

sity score comparison group managed to obtain a flush toilet.¹⁴ None of the health-related impact variables is significant, but the results may reflect small sample sizes.¹⁵

LATRINES. Again using recall data for 1993, the analysis finds that in areas receiving FISE investments in latrines, the share of households with access to sanitation facilities increased by nearly 20 percentage points more than it did in areas without FISE investments. No significant results are found for the impact on diarrhea or malnutrition.

IV. CONCLUSIONS

This article presented estimates of the impact and benefit incidence of the Nicaraguan Emergency Social Investment Fund. Impact estimates were derived using two comparison groups. One was constructed on the basis of geographic proximity and similarities with the facilities (schools and rural health posts) receiving the social fund investments (the FISE comparison group). The other was constructed using propensity score matching techniques and drawing from the household data collected by the 1998 Living Standards Measurement Study survey (the propensity score comparison group).

The benefit incidence analysis indicates that FISE investments in the health and education sectors, which together receive the largest share of FISE financing, have a pro-poor bias. Latrine investments also are strongly biased toward the poor. By contrast, sewerage investments generally benefit the better-off, while water investments are equally distributed, favoring neither the poor nor the rich.

The impact evaluation shows that FISE investments in education have had a positive impact on enrollment and the education gap, although the size and significance of the effect found depends on the comparison group used. As a result of FISE investments, children enroll half a year earlier on average. Enrollment ratios improved more for girls than for boys, and the share of children in the correct grade for their age increased more among the poor than among the better-off.

14. When potential FISE beneficiary households (all those that could have connected to the FISE-financed sewerage system) are matched to similar households, the analysis reveals a 34.4-percentage-point increase in the share of households with a flush toilet from 1993 to 1998 in the treatment group, compared with a 2.5-percentage-point increase in the propensity comparison group. Thus the net increase in access to flush toilets resulting from FISE investments is almost 32 percentage points.

15. When potential FISE beneficiaries are matched to their corresponding propensity comparison group, estimation results show that FISE-financed sewerage investments have a significant impact on the incidence of diarrhea in children under age six. This suggests that sewerage investments may have a community-level effect even in the absence of high rates of connection to the sewerage system. The larger sample size obtained when matching potential beneficiaries (rather than the smaller sample of direct beneficiaries with toilets) also underscores the importance of having sample sizes large enough to estimate specific impacts, especially for a particular population such as children under six.

The results for FISE investments in health are less clear. The strongest result points to a 5-percentage-point increase in the share of households using health clinics. But this effect is found only when the propensity score comparison group is used and is not confirmed by the FISE comparison group. When the FISE comparison group is used, FISE investments are found to have had a significant effect on acute malnutrition, but this effect is not confirmed by the propensity score comparison group. There is similar inconsistency for other indicators of impact. This lack of consistency undermines confidence in the results for FISE investments in health.

Social fund investments in water and sanitation improved the physical infrastructure and the share of households with access to services. They also appear to have had a positive effect on health indicators, but the effects are generally insignificant, possibly as a result of the small samples.

The results of the evaluation were discussed at length in two workshops held in Managua, Nicaragua, with the FISE's management, representatives of its principal multilateral and bilateral donors and representatives of government agencies working closely with the FISE, including the Ministries of Health and Education. The evaluation informed key policy changes. For instance, the FISE suspended investments in sewerage for two years in response to the findings of poor poverty targeting and lack of measurable effects on health.

In addition, the evaluation results helped generate World Bank support for a new pilot project aimed at increasing the development impact of FISE investments for the extreme poor. The pilot project will provide subsidies to households that send their children to school and use health services for basic preventive care. Finally, the results helped inform policy debates within Nicaragua, particularly those relating to the development of the country's Poverty Reduction Strategy, and helped foster an evaluation culture within the country.

REFERENCES

- Alderman, Harold, and Victor Lavy. 1996. "Household Responses to Public Health Services: Cost and Quality Tradeoffs." *World Bank Research Observer* 11(1):3–22.
- Bermudez, Gustavo, Lilia Maria Castro Monge, and Luz Marina Gracias Fonseca. 1999. "Análisis Institucional del FISE." GB Consultants, Managua, Nicaragua.
- Brockerhoff, Martin, and Laurie F. Derosé. 1996. "Child Survival in East Africa: The Impact of Preventive Health Care." *World Development* 24(12):1841–57.
- Dehejia, Rajeev H., and Sadek Wahba. 1998. "Propensity Score Matching Methods for Nonexperimental Causal Studies." NBER Working Paper 6829. National Bureau of Economic Research, Cambridge, Mass.
- Elbers, Chris, Peter Lanjouw, and Jennifer Lanjouw. Forthcoming. "Welfare in Towns and Villages: Micro-Measurement of Poverty and Inequality." *Econometrica*.
- Glewwe, Paul. 1999. *The Economics of School Quality Investments in Developing Countries: An Empirical Study of Ghana*. Studies on the African Economies. New York: St. Martin's.

- Goodman, Margaret, Samuel Morley, Gabriel Siri, and Elaine Zuckerman. 1997. "Social Investment Funds in Latin America: Past Performance and Future Role." Inter-American Development Bank, Evaluation Office and Social Programs and Sustainable Development Department, Washington, D.C.
- Grosh, Margaret E., and Paul Glewwe. 1995. "A Guide to Living Standards Measurement Study Surveys and Their Data Sets." LSMS Working Paper 120. World Bank, Washington, D.C.
- Grossman, Jean. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *World Bank Research Observer* 9(2):159–80.
- Hanushek, Eric A. 1995. "Interpreting Recent Research on Schooling in Developing Countries." *World Bank Research Observer* 10(2):227–46.
- Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65(2):261–94.
- Hotchkiss, D. R. 1998. "The Tradeoff between Price and Quality of Services in the Philippines." *Social Science and Medicine* 46(2):227–42.
- Jorgensen, Steen, and Julie van Domelen. 1999. "Helping the Poor Manage Risk Better: The Role of Social Funds." World Bank, Human Development Network, Washington, D.C.
- Kremer, Michael R. 1995. "Research on Schooling: What We Know and What We Don't. A Comment on Hanushek." *World Bank Research Observer* 10(2):247–54.
- Lanjouw, Peter, and Martin Ravallion. 1995. "Poverty and Household Size." *Economic Journal* 105(433):1415–34.
- Lavy, Victor, John Strauss, Duncan Thomas, and Philippe de Vreyer. 1996. "Quality of Health Care, Survival and Health Outcomes in Ghana." *Journal of Health Economics* 15(3):333–57.
- Lee, Lung-fei, Mark R. Rosenzweig, and Mark M. Pitt. 1997. "The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High-Mortality Populations." *Journal of Econometrics* 77(1):209–35.
- Manski, Charles F., and Steven R. Lerman. 1977. "The Estimation of Choice Probabilities from Choice-Based Samples." *Econometrica* 45(8):1977–88.
- Paxson, Christina, and Norbert Schady. 2002. "The Allocation and Impact of Social Funds: Spending on School Infrastructure in Peru." *World Bank Economic Review* 16(2):xxx–xxx.
- Pradhan, Menno, and Laura B. Rawlings. 2000. "The Impact and Targeting of Nicaragua's Social Investment Fund." World Bank, Latin America and the Caribbean Region, Human Development Department, Washington, D.C.
- Rawlings, Laura B., Lynne Sherburne-Benz, and Julie van Domelen. 2002. "Evaluating Social Fund Performance: A Cross-Country Analysis of Community Investments." World Bank, Social Protection Network, Washington, D.C.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.
- World Bank. 2000. "Nicaragua: Ex-Post Impact Evaluation of the Emergency Social Investment Fund (FISE)." Report 20400-NI. Latin America and the Caribbean Region, Human Development Department, Washington, D.C.

Impact Evaluation of Social Funds

The Allocation and Impact of Social Funds: Spending on School Infrastructure in Peru

Christina Paxson and Norbert R. Schady

Between 1992 and 1998 the Peruvian Social Fund (FONCODES) spent about US\$570 million funding microprojects throughout the country. Many of these projects involved constructing and renovating school facilities. This article uses data from FONCODES, the 1993 population census in Peru, and a 1996 household survey conducted by the Peruvian Statistical Institute to analyze the targeting and impact of FONCODES investments in education. A number of descriptive and econometric techniques are employed, including nonparametric regressions, differences in differences, and instrumental variables estimators. Results show that FONCODES investments in school infrastructure have reached poor districts and poor households within those districts. The investments also appear to have had positive effects on school attendance rates for young children.

Since the creation of the Emergency Social Fund in Bolivia in late 1986, social funds have been established in dozens of countries, often with support from multilateral organizations and international donors. Social funds like Bolivia's were originally put into place to mitigate the social costs of structural adjustment programs. Since then they have been proposed as a safety net for the poorest people; as a means of generating employment and transferring income; as an efficient mechanism for constructing small-scale infrastructure, especially in outlying, traditionally neglected areas; and as a way of building on (or generating) local social capital by involving communities in choosing, preparing, operating, and maintaining projects (Rawlings and others 2002).

This article analyzes the targeting and impact of investments by the Peruvian Social Fund (Fondo Nacional de Compensación y Desarrollo Social, or FONCODES) between 1992 and 1998. Specifically, we look at the investments FONCODES made in education, addressing two questions. First, who benefited

Christina Paxson is Professor of Economics and Public Affairs and Director, Center for Health and Wellbeing at Princeton University. Norbert R. Schady is Senior Economist, Latin America and the Caribbean Region at the World Bank. Their e-mail addresses are cpaxson@princeton.edu and nschady@worldbank.org, respectively. For helpful comments and suggestions, the authors thank François Bourguignon, Angus Deaton, Olivier Deschenes, Rutheanne Deutsch, Esther Duflo, John Gallup, Doug Miller, Martin Ravallion, Julie van Domelen, and Juliana Weissman; participants in a World Bank seminar and in the Northeast Universities Development Conference at Harvard University on October 8-9, 1999; and two anonymous referees. They are also grateful to staff at FONCODES and INFES and to Livia Benavides for providing data. An earlier version of the article was issued as World Bank Policy Research Working Paper 2229.

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

from FONCODES education investments? This is a question of targeting. FONCODES aims to transfer resources, including investments in education, to poor areas and poor households within those areas. The article evaluates the extent to which it was successful in doing so. Second, did FONCODES transfers improve education outcomes? This is a question about how investments in school facilities affected measures of school attendance.

Although the article describes and evaluates a specific program, it adds to an ongoing debate about the relationship between education inputs and outcomes (for a summary see Hanushek 1995 and the response by Kremer 1995). A growing body of literature suggests that expenditures on school facilities have high rates of return in many developing economies (for example, Duflo 2001, Glewwe and Jacoby 1994, Glewwe and others 1995, Hanushek 1995, Hanushek and Harbison 1992). The results here suggest that expenditures on school infrastructure in Peru improved the attendance rate of young children. Because expenditures by FONCODES on education were well targeted toward poor districts and (though less clearly) poor households, improvements in attendance rates were concentrated among the neediest.

I. THE SETTING

Peru made substantial economic progress between 1992 and 1998. After a brief recession following the adoption of stringent stabilization and adjustment measures in 1990, growth was generally strong, inflation low, and poverty reduction sustained (World Bank 1996, 1999). Investments in the social sectors increased dramatically. The Peruvian government attempted to target these social investments to the poor—though with only partial success (World Bank 1999).

FONCODES was created in 1991 with the stated objectives of generating employment, helping to alleviate poverty, and improving access to social services (World Bank 1998). Between 1992 and 1998 FONCODES funded almost 32,000 community-based projects for an aggregate outlay of about 760 million soles.¹ These community-based projects included initiatives in health, education, agriculture, community centers, rural electrification, and water and sanitation. Most of those in education entailed constructing and renovating classrooms (table 1). Before 1995, however, FONCODES also had education projects focusing on constructing and renovating sports facilities and providing textbooks and other educational materials to students. In addition, FONCODES executed a series of centrally designed special projects. Those in education included a school breakfast program and the distribution of uniforms for schoolchildren. Between 1992 and 1996 FONCODES spent about 160 million soles on all special projects, including those in education.

1. All reported expenditures are in 1992 soles, unless otherwise noted. The December 1992 exchange rate was 1.63 soles to the U.S. dollar.

TABLE 1. FONCODES Projects and Project Funding, 1992-98

Year	All projects		Projects to construct and renovate classrooms			Other education projects		
	Number of projects	Funding (m 1992 soles)	Number of projects	Funding (m 1992 soles)	Funding (percent total)	Number of projects	Funding (m 1992 soles)	Funding (percent total)
1992	2,813	102.7	1,185	26.2	25.0	386	6.9	6.7
1993	5,238	144.9	2,327	49.4	34.1	430	8.0	4.0
1994	4,551	110.4	2,380	48.7	44.1	100	1.3	1.2
1995	3,056	79.3	1,037	24.7	31.0	42	0.7	0.9
1996	4,222	83.4	987	15.0	18.0	14	0.3	0.4
1997	5,807	114.8	607	11.0	9.6	9	0.2	0.2
1998	6,088	123.8	636	12.0	9.7	1	0.0	0.0
Total	31,775	759.2	9,160	187.1	24.6	981	13.0	2.0

Note: Includes only expenditures on community-based projects.

Source: FONCODES.

FONCODES has much in common with other social funds in the region. Two features particularly important for this article are the demand-driven and targeted nature of its projects. FONCODES projects are demand-driven in that communities themselves choose a project and prepare a proposal for funding. FONCODES then functions as a financial intermediary: rather than execute projects itself, it approves proposals and releases funds to the *nucleo ejecutor*—a group of community members elected for that purpose.

To target its investments, FONCODES uses a poverty map to allocate resources (for details, see section on targeting). FONCODES staff members also conduct an informal on-site assessment of the poverty of a community requesting a project. Since 1993 the demand for FONCODES projects has far exceeded the program's budget. As a result FONCODES has had a backlog of project proposals and has had to ration its investments. Although decisions about which projects to fund within a district have often been ad hoc, an attempt is made to give preference to projects in communities that the FONCODES evaluators deem to be poorer. However, no attempt is made to target households within a community.

II. THE DATA

The evaluation of the targeting and impact of FONCODES expenditures on school infrastructure uses several sources of data. These include district-level information on the geographic distribution of FONCODES allocations and expenditures, kept by FONCODES, and on district characteristics from a 1993 Population and Housing Census. The analysis also uses household-level information from a household survey conducted by the Peruvian Statistical Institute (Instituto Nacional de Estadística e Informática, or INEI) in 1996 and from two Living Standards Measurement Study (LSMS) surveys, conducted in 1994 and 1997.

District-Level Data

Monthly records on the number of FONCODES projects and amounts spent in each district are available for 1992 through 1998. Expenditures do not include administrative and overhead costs and are available only for the community-based projects, not the special projects. Similar information is available for expenditures by a second school infrastructure program in Peru, Instituto Nacional de Infraestructura Educativa y de Salud (INFES), but only for 1995. Though both INFES and FONCODES are central government programs, an important difference is that INFES has (mainly) built or renovated secondary schools in urban areas, whereas FONCODES has (mainly) renovated primary schools in rural areas. INFES has also spent considerably more on school infrastructure than FONCODES: in 1995 it spent about 350 million soles on school infrastructure, whereas FONCODES spent 25 million soles.

An important district-level variable for the analysis is the FONCODES index, a district-level poverty measure. This index forms the basis of the poverty map that FONCODES has used to allocate resources. Specifically, since 1992 FONCODES

has allocated resources to each of its 24 regional offices in a two-step process. First, FONCODES makes a “referential allocation” to each district (before 1996, to each province) by weighting the population of that district by the FONCODES index.² The allocation to district i is given by:

$$(1) \quad \text{Allocation}_i = \frac{\text{Index}_i * \text{Population}_i}{\sum_{j=1}^n (\text{Index}_j * \text{Population}_j)}$$

Second, it sums these referential allocations over the districts covered by a FONCODES regional office. This determines the budget for each office. Regional offices are instructed to follow the original allocations across districts as closely as possible. Because these instructions require the regional offices to favor poorer districts in the allocation of funds, the FONCODES index provides a useful measure of the priority that projects in any given district should be given.

The FONCODES index is an ad hoc composite of different measures—including access to schooling, electricity, water, sanitation, and adequate housing and measures of illiteracy and chronic malnutrition. (All these are drawn from the Population and Housing Census conducted in 1993, except the rate of chronic malnutrition, which is based on a census of height and age among schoolchildren also conducted in 1993.) Composite indexes invariably involve some arbitrary weighting of indicators. FONCODES standardizes each indicator in its index by dividing it by the lowest value measured, multiplies the rate of chronic malnutrition by seven, and then adds all the indicators.³ For ease of interpretation, FONCODES then standardizes the index by dividing all index values by the lowest value. The resulting index ranges from 1 to 36.38.

Another district-level measure, used in the analysis of targeting, is imputed per capita income, constructed by INEI. In Peru there are no survey-based estimates of income or expenditures at a level more disaggregated than the department: for example, household surveys conducted by INEI, which generally have samples of 15,000–20,000 households, can be used only to compare income across “natural regions” and departments.⁴ INEI has attempted to get around this prob-

2. Provinces and districts correspond to the two levels of local government in Peru. In 1997 there were 194 provinces and 1,812 districts in Peru (Webb and Fernández Baca 1997, p. 112). The median population of a district is about 4,000, but the population size varies considerably: rural districts can have fewer than 500 people, whereas urban districts can have more than 100,000.

3. This procedure had the unintended consequence of giving the greatest weight to the indicators with the greatest variance. Thus although the intended weights were 50 percent for the rate of chronic malnutrition and 7.14 percent for each of the seven other measures, the actual weights in the index turned out to be 15.3 percent for chronic malnutrition, 3.4 percent for illiteracy, 2.2 percent for school attendance, 3.0 percent for overcrowding in homes, 38.3 percent for inadequate roofing on houses, 8.8 percent for access to water, 7.4 percent for access to sewerage, and 21.6 percent for access to electricity (World Bank 1996, p. 7).

4. These natural regions are Lima and the urban and rural areas of the coast, the sierra (highlands), and the selva (jungle).

lem by combining variables common to both the 1993 census and one such survey conducted in 1995 and imputing district-level measures of income and poverty (INEI 1996).⁵ Although crude, the imputed income measure provides a useful measure of district-level welfare.

Schady (2002) shows that in Peru various district-level measures of welfare (including the INEI income measure and the FONCODES index) are highly correlated with one another and do not differ significantly in their ability to separate poor from nonpoor districts. To make the analysis of geographic targeting more easily comparable to the analysis of household targeting, which is based on household per capita income, this article uses this measure of district per capita income rather than the FONCODES index for the targeting results.

Household-Level Data

The main source of household information used is the 1996 INEI survey, a standard multipurpose survey. It has a relatively large sample—more than 18,000 households in 403 districts. The INEI survey collected information on household income, education levels, other household characteristics and benefits from various social programs. Households with at least one member attending public school were asked about recent improvements in school facilities and whether these had been carried out by (separately) FONCODES, INFES, or the local parents' committee (akin to the parent-teacher associations in the United States). These data are used to evaluate the household-level targeting of FONCODES investments in school facilities.

To test the robustness of the results of this evaluation, some of the results are reproduced using two LSMS surveys, conducted in 1994 and 1997. The LSMS surveys are both smaller than the INEI survey—covering about 3,500 households each, in 199 districts (1994) or 228 districts (1997). But they offer an advantage in that a very similar questionnaire was applied in both years. The LSMS data set includes a panel covering just over a quarter of the households in the two samples.

Because the INEI and LSMS surveys were not designed specifically for an evaluation of FONCODES, they have some shortcomings for the analysis. Three are worth noting. First, the surveys did not collect information on the quality of education as measured by, for example, scholastic achievement, pupil-teacher ratios, and the amount of time spent in school. Second, there appears to be a large amount of measurement error in the FONCODES "treatment" variable in the 1996 INEI survey, a point to which we return. Third, in the 1996 INEI sur-

5. Specifically, INEI estimated income in 1995 on the basis of the household survey and then regressed income in every department on its correlates—household composition, education levels, access to basic services (water, sewerage, electricity), ownership of durable goods (radio, TV, refrigerator), and other variables included in both the census and the survey. The coefficients from the 24 department-level regressions were then used to impute average income in every district and the fraction of the population in each district below an income-based poverty line. The methodology applied by INEI for these imputations is similar in spirit to that proposed in Hentschel and others (2000), although there are differences in how it was applied (Schady 2002).

vey, questions about benefits from FONCODES programs were asked only of families with children in school. The survey results therefore cannot be used to determine whether children not in school had access to a FONCODES-improved school.

III. THE TARGETING OF FONCODES INVESTMENTS IN EDUCATION

Targeting education resources is important in Peru because there are large differences in measures of educational attainment across regions and income groups. At any given age, children in the poorest 25 percent of districts lag behind those in the richest 25 percent in years of schooling attained (figure 1). The differences across quartiles increase with age, so that by age 16 there is almost a 2-year difference between children in the poorest and richest districts.

A host of factors probably contribute to differences in the educational attainment of children, including differences in income, ethnicity, employment opportunities, and the education of other household members. Many of these factors cannot be changed through public policies in the short run. But educational attainment is also likely to be a function of the quantity and quality of teachers, learning materials, and classrooms in a community. Poor districts and poor households may therefore need additional resources, including resources spent on school facilities, to catch up with their better-off counterparts.

Geographic Targeting

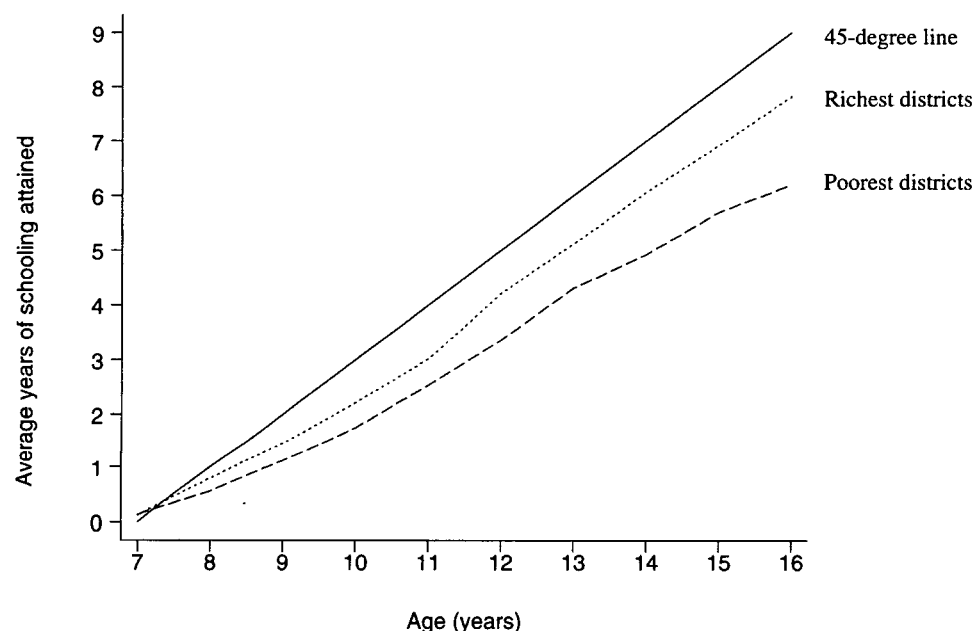
Has FONCODES effectively reached poor districts? Two aspects of the geographic targeting of FONCODES investments in school infrastructure are considered: changes in targeting over time, and district-level expenditures by FONCODES compared with those by INFES.

Regression results show that FONCODES expenditures on school infrastructure were targeted to poorer districts. A regression of per capita FONCODES expenditures on education, summed over 1992–98, on the log of district per capita income indicates that a 10 percent increase in district per capita income is roughly associated with a 1 sol decrease in per capita expenditures. (The regression coefficient on log income is -10.69 , with a standard error of 0.52 .)⁶ Moreover, targeting appears to have improved over time. Nonparametric regressions of per capita FONCODES expenditures on school infrastructure on log per capita income in three typical years—1992, 1995, and 1998—show that districts with lower per capita income clearly received more FONCODES education expenditures, especially in 1995 and 1998 (figure 2).⁷

6. All regression results reported are weighted by district population. Alternatively, per capita expenditures could have been regressed on the (imputed) poverty rate for each district. Ravallion (2000) shows that if there is no targeting within districts, so that poor households within a district are equally likely to receive a transfer whether they live in relatively “rich” or “poor” communities, the coefficient on such a regression can be interpreted as the difference between spending on the poor and spending on the nonpoor.

7. All nonparametric regressions are Fan regressions with a quartic kernel (see Fan 1992).

FIGURE 1. Average Years of Schooling Attained, by Age, for Children in the Poorest and Richest 25 Percent of Districts in Peru, 1996



Source: Authors' calculation based on the 1996 INEI survey.

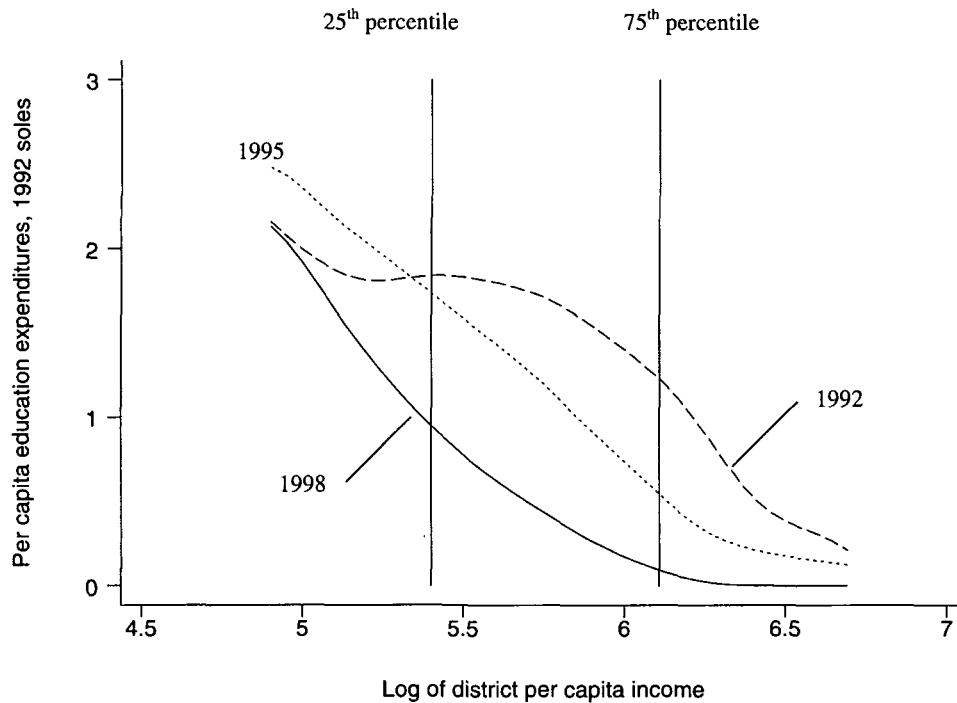
How well do FONCODES targeting outcomes stand up to those of a comparable program? In regressions of per capita expenditures by FONCODES and INFES in 1995 on the log of district per capita income, the coefficient on FONCODES expenditures is significantly negative (-1.60 , with a standard error of 0.18), whereas the coefficient on INFES expenditures is positive though insignificant (1.30 , with a standard error of 0.80). Nonparametric regressions show that per capita expenditures on education infrastructure by INFES in 1995 were much larger but much less well targeted. Per capita expenditures by FONCODES decreased monotonically with district per capita income, and those by INFES were concentrated in the middle of the distribution and were lowest for the districts with the lowest per capita income (figure 3).

Household-Level Targeting

In Peru there is considerable heterogeneity in the distribution of welfare within districts. For example, a simple decomposition of the variance in per capita income in the 1996 INEI survey into inter- and intradistrict components suggests that only 22 percent of the variance is explained by differences across districts. Reaching poor districts is therefore only a weak proxy for reaching poor households.

To examine household-level targeting, the household-level incidence of FONCODES benefits is calculated using information from the 1996 INEI survey

FIGURE 2. Geographic Targeting of FONCODES Education Projects, Various Years

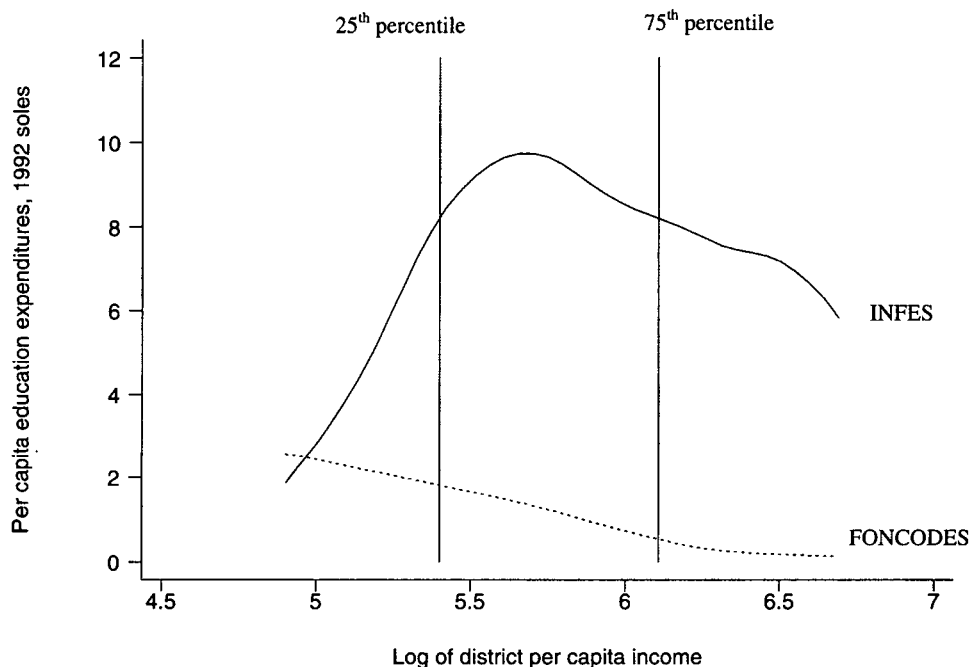


Note: The top and bottom 1 percent of the distribution of log per capita income have been trimmed.
 Source: Authors' calculations based on FONCODES data.

on access to education infrastructure and per capita income. Three separate variables are defined that take the value of one for households that reported having benefited from spending by FONCODES, INFES, and the parents' committees. These variables are regressed on the log of household per capita income. The weighted logit regression results (with the weights given by the expansion factors in the survey) suggest that poorer households are more likely than better-off households to benefit from FONCODES investments: the estimated marginal effect of the log of household per capita income on the probability that the household benefits from FONCODES is -0.010 (with a standard error of 0.001). Although poorer households are also more likely than better-off households to benefit from parents' committees, the marginal effect of income on the probability of benefiting is -0.005 (standard error of 0.002), only half as large as the estimate for FONCODES. Poorer households are less likely than better-off household to benefit from INFES: the marginal effect of income is 0.009 (standard error of 0.001).

Nonparametric regressions are used to capture possible nonlinearities in the relationship between investments and log income. The results confirm that households with lower per capita income are more likely to benefit from FONCODES

FIGURE 3. Geographic Targeting of FONCODES and INFES Education Projects, 1995



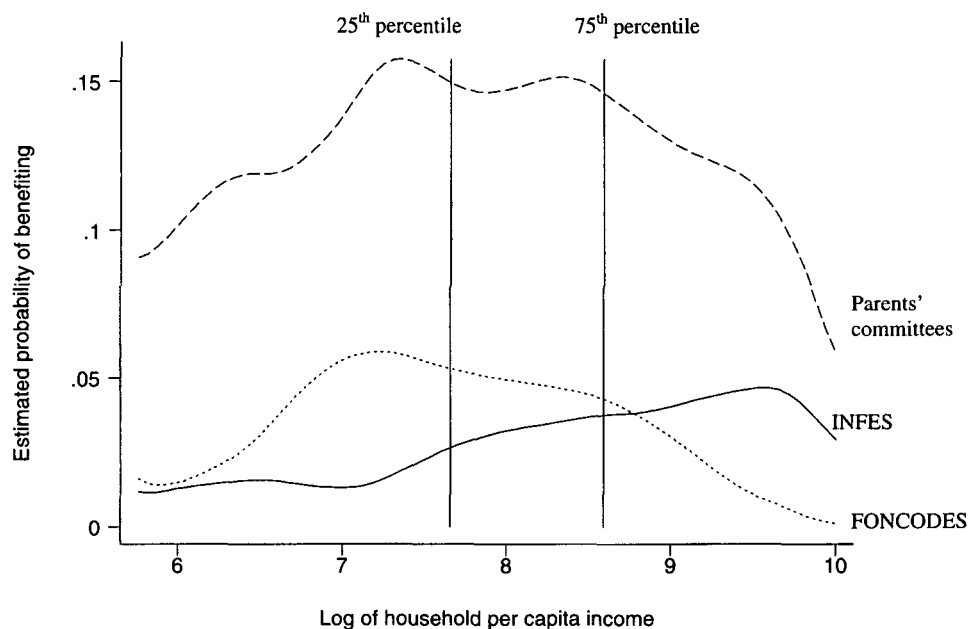
Note: The top and bottom 1 percent of the distribution of log per capita income have been trimmed.
Source: Authors' calculations based on data from INFES and FONCODES.

spending than from INFES spending (figure 4). To some extent this no doubt reflects INFES's emphasis on secondary school infrastructure in urban areas and FONCODES's emphasis on primary school infrastructure in rural areas: in Peru, as in many other countries, the poor are less likely to send their children to secondary school and more likely to live in rural areas.⁸ The nonparametric regressions also show that the FONCODES distribution slopes upward at very low levels of (log) per capita income: the poorest 7 percent of households are less likely than their (slightly) better-off counterparts to benefit from FONCODES investments in education infrastructure.

Measurement error is an important concern for the estimates of household targeting. Rural households in Peru are likely to have little choice of primary school. In the absence of measurement error, one would therefore expect a high

8. This also helps explain why the fraction of households that reported having benefited from INFES (2.6 percent) is smaller than the corresponding fraction for FONCODES (3.6 percent), despite the massive differences in the programs' budgets. In Peru projects to repair primary schools have tended to be small and relatively inexpensive (with low-cost materials and community participation in construction), whereas projects to construct or repair secondary schools are more expensive because they are larger and more elaborate (with higher-end materials and payment of all labor costs for a contractor).

FIGURE 4. Household-Level Targeting of School Infrastructure Expenditures by FONCODES, INFES, and the Parents' Committees, 1996



Note: The top and bottom 1 percent of the distribution of log per capita income have been trimmed.
 Source: Authors' calculations based on the INEI survey.

degree of consistency in the answers given by households within a rural community to questions about the presence of FONCODES-funded education projects. Unfortunately, this is not always the case. Consider households in rural areas that have only children attending primary school. In 4 rural communities in the sample all such households reported that FONCODES had financed improvements to the local school, and in another 107 all such households reported that FONCODES had not financed improvements. But in 46 rural communities different households provided different responses, suggesting that households in the 1996 INEI survey did not always report program benefits accurately. Measurement error of this sort will bias the estimates of program incidence if it is correlated with income so that richer (or poorer) households are more (or less) likely to report that they benefited from FONCODES.

To further explore issues related to measurement error, the analysis tests whether households that did not respond to questions about infrastructure improvements in the 1996 INEI survey differed systematically from responding households. The INEI survey first asked households whether any member attended a public school and, for those answering affirmatively, whether they had "knowledge of any improvement to this public school in the last 12 months." Next, the survey asked about the kind of improvement undertaken and finally about the

agency that financed it. About 15 percent of households with a child in public school did not recall whether there had been a recent improvement, and about 10 percent with knowledge of school improvements did not know who financed them.

The analysis finds that nonresponding households differed in some ways from responders, but the differences tended to be small. For example, the mean education of the household head was 7.1 years for households knowing of school improvements and knowing who financed them, 7.1 for households knowing of school improvements but not knowing who financed them, and 6.57 for households not knowing whether there had been any school improvements. For these same groups of households, the log of household per capita income was 8.04, 7.94, and 8.04.⁹ Although these differences are sometimes significant, they are small and do not invalidate the analytical approach used.

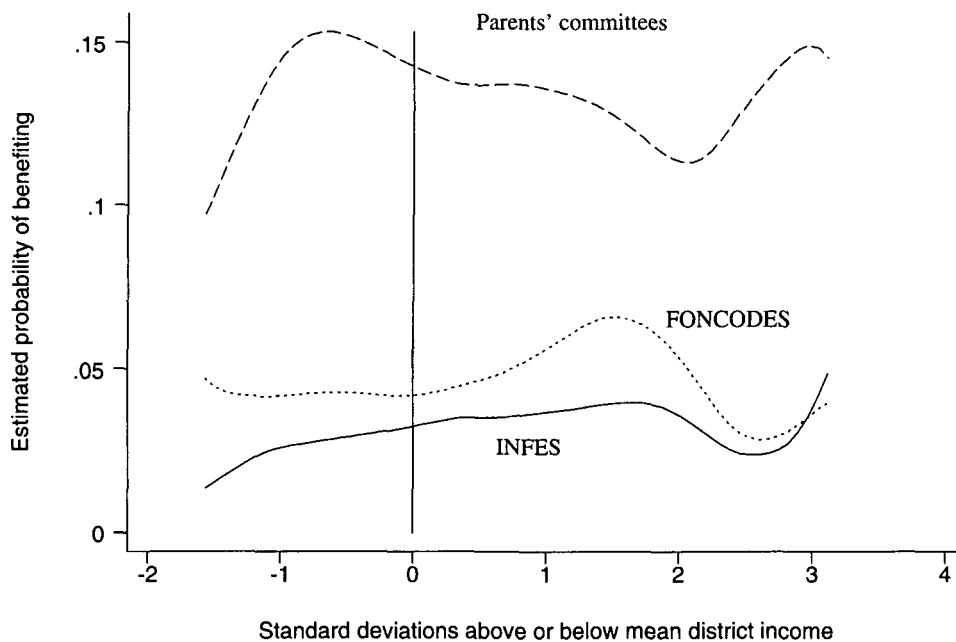
FONCODES has placed a great deal of importance on geographic targeting and less on other forms of targeting, such as means testing (World Bank 1996). A comparison of figures 2 and 4 suggests that it has done better reaching the poorest districts than it has reaching the poorest households. To explore this issue further, the estimated probability of benefiting from school investments by INFES, the parents' committees, and FONCODES is graphed on the number of standard deviations that the income of household i in district j is above or below the mean income in district j when both household and district incomes are calculated using the 1996 INEI survey. (Both the mean and the standard deviations are district-specific.) The nonparametric regression line for FONCODES school infrastructure is humped, peaking at about 1.5 standard deviations above mean district income (figure 5). Within a given district households that are somewhat better off than their counterparts are more likely to benefit from FONCODES investments in school infrastructure. This suggests that there was essentially no (positive) intradistrict targeting of FONCODES resources in 1996.

This finding adds to a debate about the relative importance of central and community-level targeting and about the level at which targeting decisions should be made (for example, Alderman 1998 and Galasso and Ravallion 2000). FONCODES is a central government program that has chosen how to allocate resources across districts from the center. Largely, decisions about which community projects to finance have been left to employees in FONCODES's regional offices—a much more aggregate level than the provinces and districts that form the basis of the poverty map. This targeting scheme has been effective at reaching poor districts but not at reaching the worst-off households within those districts.

Without more information—such as a comparison with a similar small-scale infrastructure program using community-based targeting within districts—it is hard to know whether FONCODES's within-district targeting performance is better or worse than the alternatives. The analysis does suggest, however, that there

9. The authors thank an anonymous referee for this suggestion to analyze differences in the characteristics of responding and nonresponding households.

FIGURE 5. Probability of Benefiting from School Infrastructure Expenditures by Number of Standard Deviations of Household Income Above or Below Mean District Income



Note: The top and bottom 1 percent of the distribution of log per capita income have been trimmed.
 Source: Authors' calculations based on the INEI survey.

are limits to the extent to which central government programs can reach poor households without such targeting mechanisms as indicator targeting or self-targeting through the provision of inferior infrastructure.

IV. THE IMPACT OF FONCODES INVESTMENTS IN EDUCATION

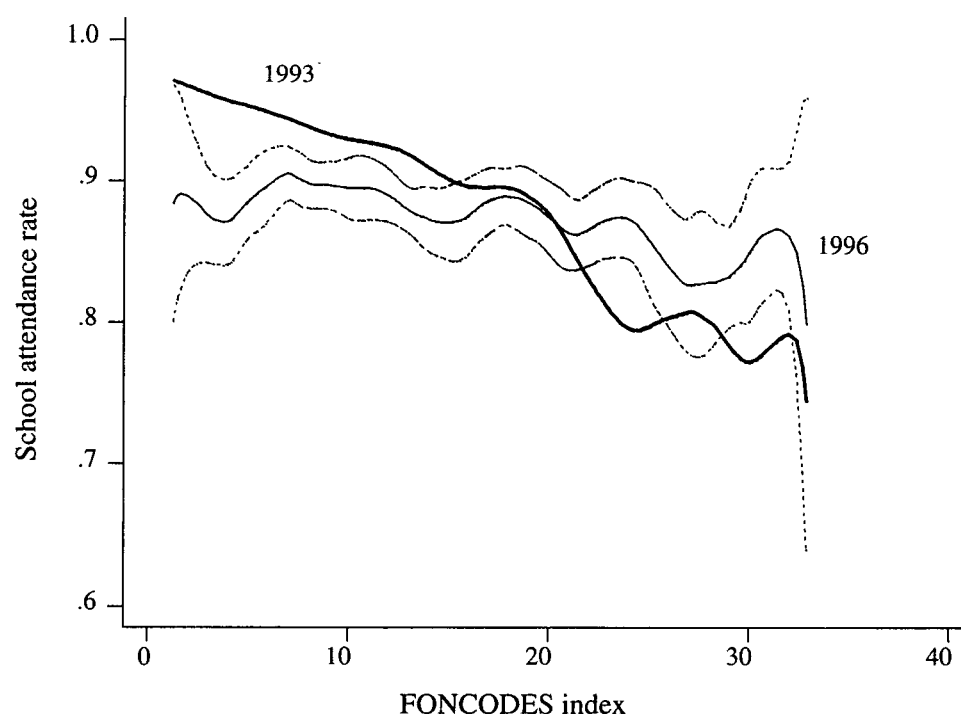
Although it is too early to assess the long-term impact of FONCODES education investments, the program has been in existence long enough to have had short-run effects, such as increasing school attendance rates. In this section district-level data are used to examine the relationship between school attendance rates and FONCODES spending on school infrastructure. Using school attendance data from the 1993 census and the 1996 INEI survey, the analysis shows that there is a positive association between FONCODES education funding and gains in primary education: districts that received the largest per capita allocations of FONCODES funds for education experienced the largest increases in school attendance for children ages 6–11.

The analysis begins by looking at the associations between district-level school attendance rates for children ages 6–11 and the FONCODES index. As noted ear-

lier, the FONCODES index is higher for poorer districts, which were given priority in FONCODES funding decisions. To avoid a mechanical relationship between the FONCODES index and school attendance, the analysis modifies the index to exclude one of its usual elements—the fraction of children ages 6–11 who are not attending school. This is done by replacing the district value of the fraction of children ages 6–11 who are not attending school with the countrywide average. Otherwise the index is identical to that used by FONCODES.

Nonparametric (Fan) regressions of attendance rates in 1993 (based on the census data) and 1996 (based on the INEI survey data) on the modified FONCODES index, using observations on 349 districts, indicate that the relationship between attendance rates and the FONCODES index changed during the period (figure 6). As might be expected, there is an obvious negative relationship between school

FIGURE 6. Nonparametric Regressions of District School Attendance Rate on (Modified) FONCODES Index



Note: Figure is based on regressions that use a bandwidth of three and are weighted by the population of children in the district so that less-populated districts get less weight. The confidence intervals for the 1996 results (shown as dotted lines) were computed using a bootstrap procedure: drawing random samples (with replacement) from the original INEI sample (with the probability of being drawn proportional to the sampling weight), estimating the nonparametric regressions 50 times using the micro-level data, and computing the standard deviation of the estimate at each value of the FONCODES index on the x-axis. The confidence lines show the point estimate at each value of the FONCODES index plus and minus two standard deviations.

Source: Authors' calculations based on 1993 census data and 1996 INEI household survey data.

attendance and the FONCODES index in 1993, so that poorer districts—with higher index values—have lower attendance rates. But this negative relationship is much less pronounced by 1996. In other words, worse-off districts had large gains in school attendance, but better-off districts did not.

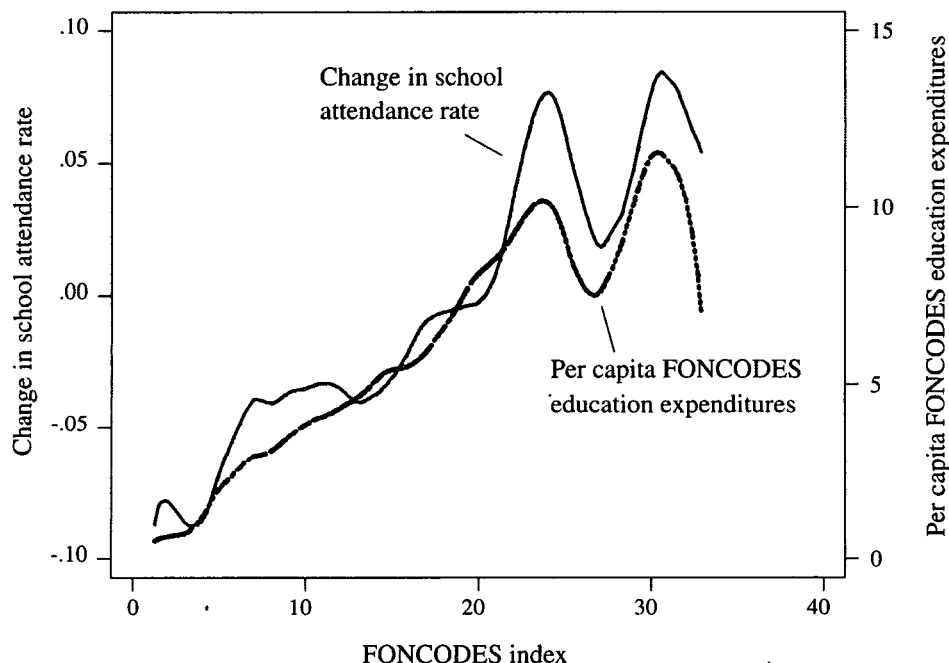
One puzzling feature of figure 6 is that attendance rates appear to have declined for children in well-off districts between 1993 and 1996. But this decline may at least in part reflect the timing of the surveys, coupled with the fact that both surveys explicitly asked about school attendance rather than school enrollment. In Peru the school year runs from April to December. The 1993 census data were collected in June, relatively early in the school year, whereas the 1996 INEI survey data were collected in November. Attrition in attendance over the school year could account for the lower mean attendance rates in 1996.

The LSMS surveys provide some evidence that attrition does affect measures of district-level school attendance rates. The 1994 LSMS survey, conducted between June and August, shows a drop in attendance rates of primary-school-age children of two percentage points (from 96.6 percent to 94.6 percent) between June and July, after which attendance rates appeared to stabilize. The 1997 LSMS survey, conducted between September and November, shows no systematic decline in attendance. The decline from June to July 1994 was concentrated among children in poorer districts with a higher FONCODES index value. For example, for children in districts with a FONCODES index greater than 14 (roughly the median), the attendance rate declined from 96 percent to 92 percent between June and July. Because the 1993 census was conducted in June, part of the high attendance rate in 1993 (shown in figure 6) may reflect the higher attendance early in the school year. Moreover, because attrition after June is more likely for poorer children, the results may understate the gains made by children in poor districts relative to those in rich districts between 1993 and 1996.¹⁰

Were the districts that experienced the largest gains in school attendance also those that received the most FONCODES funding for school infrastructure? Figure 7 graphs both the change in the school attendance rate and the total per capita FONCODES expenditures on school infrastructure in 1992–95 as a function of the FONCODES index. (FONCODES expenditures are summed over 1992–95 rather than 1993–96 because it is assumed that expenditures on school infrastructure cannot affect attendance until the year after they are made.) This figure shows that poorer districts that experienced greater gains in school attendance also received more funding for school improvements. The degree of comovement between attendance gains and school funding is striking. A regression of the predicted value of the attendance gain on the predicted value of school expenditure

10. To double-check the results, the analysis of figure 6 was repeated using the 1994 and 1997 LSMS surveys, including and excluding observations from June. Although the results based on the LSMS surveys are somewhat noisier, because there are fewer districts, they are similar to those based on the census and the INEI survey. Excluding observations from June has almost no effect on the relationship between the gain in attendance and the FONCODES index.

FIGURE 7. Nonparametric Regressions of Change in District School Attendance Rate and Per Capita FONCODES Education Expenditures in 1992–95 on (Modified) FONCODES Index



Source: Authors' calculations based on 1993 census data, the 1996 INEI household survey data, and FONCODES data.

yields a coefficient of 0.0151, which implies that a one-standard-deviation (9.6) increase in per capita FONCODES spending on school infrastructure is associated with a gain in the attendance rate of 14.5 percentage points.

Another interesting feature of figure 7 is that the relationships it shows are nonlinear and nonmonotonic. Districts with FONCODES index values between 22 and 26 had greater attendance gains and higher school infrastructure spending than did poorer districts with index values between 26 and 28. Both measures increased again for even poorer districts with index values greater than 28. These nonlinear patterns are not the result of few observations at very high values of the FONCODES index. About 15 percent of districts (accounting for 14 percent of children) had FONCODES index values between 22 and 26, 6.3 percent of districts (4.7 percent of children) had index values between 26 and 28, and 9.6 percent of districts (8.4 percent of children) had index values greater than 28. Furthermore, these nonlinearities remain even when a wide range of bandwidths is used for the nonparametric regressions. Although these nonlinearities are striking, it is not clear what drives them. The districts with index values between 26 and 28 were allocated more FONCODES funds than the wealthier

districts with slightly lower index values but apparently did not apply or receive approval for greater funding for school infrastructure projects.

Figures 6 and 7 show that there is a positive association between gains in school attendance and FONCODES spending on school infrastructure. A key question is whether these gains in school attendance were caused by FONCODES spending. Two features of the program make causality difficult to ascertain. First, FONCODES is demand-driven, with community groups supplying proposals for specific projects. The unobserved characteristics that prompt community groups to apply for funds may be correlated with the outcomes of interest. For example, a poor district in which people begin to care more about education may have larger increases in school attendance rates and generate more proposals for FONCODES school funding than an equally poor district in which preferences for education do not change. In this example the positive association between gains in school attendance and FONCODES education spending is driven by a third, unobserved factor—changes in district-level preferences about education—and there may be no causal relationship between FONCODES spending and gains in attendance.

Second, FONCODES funding was targeted to poorer districts. There was no explicit randomization of FONCODES funds across districts, and no obvious natural experiment that resulted in different funding levels across similar districts. It is therefore difficult to assess whether the gains in attendance in the poorer districts that received greater FONCODES funding were driven by FONCODES or by unobserved factors correlated with district-level poverty. For example, it is possible that returns to education increased more in poorer districts than in wealthier districts over this period, prompting more parents in poorer districts to send their children to school.

The analysis turns to instrumental variables techniques to deal with these two problems. The first problem—that changes in district-level tastes for education may be correlated with applications for FONCODES funds for school infrastructure—is the more easily handled. The district-level gain in school attendance is regressed on FONCODES spending on school infrastructure, with the infrastructure spending instrumented with the FONCODES index and with a set of variables reflecting the political preferences of the district population, as measured by the fraction voting for Alberto Fujimori in the 1990 and 1993 national elections. Because the FONCODES index was used by regional offices to prioritize allocations, it should be correlated with district-level spending on school infrastructure. Moreover, because the FONCODES index is based on measures of a district's unmet needs in 1993, it is plausibly uncorrelated with unobserved changes in tastes for education between 1993 and 1996. The use of the political variables as instruments is motivated by previous research indicating that districts that moved against Fujimori between 1990 and 1993 were subsequently treated more generously by FONCODES, presumably in an attempt to regain votes (Schady 2000). Under the assumption that the political preferences of districts were not correlated with changes in preferences for education, the political measures are valid instruments.

The second problem—that district-level poverty may be correlated with unobserved factors that affect attendance rates—is more difficult to handle convincingly. The fact that FONCODES spending was targeted to poorer districts—combined with the possibility that there could have been other, unobserved reasons that poorer districts experienced larger attendance gains—suggests that the FONCODES index may not be an appropriate instrument for FONCODES spending on schools. Results are presented in which the FONCODES index is included as an explanatory variable in the second-stage regressions. This specification allows initial district-level poverty to have an independent effect on the subsequent gain in school attendance and relies solely on the political variables as instruments. For this strategy to be valid, the political variables must not have affected other resource flows to districts that had an effect on school attendance. Although no conclusive evidence is available on this point, the analysis examines whether school infrastructure spending by INFES in 1995 is related to the political measures and to attendance gains. These results are discussed.

Table 2 shows ordinary least squares and instrumental variables estimates of regressions of the district-level gain in school attendance on FONCODES spending on school infrastructure. Because the school attendance rate is bounded between 0 and 1, the gain in attendance between 1993 and 1996 is measured as the (approximate) change in the log odds that a child ages 6–11 attends school. Specifically, the gain in school attendance is measured as

$$(2) \quad \text{Gain} = \ln\left(\frac{p96}{1-p96}\right) - \ln\left(\frac{p93}{1-p93}\right) \approx \frac{p96-p93}{p93(1-p93)}$$

where $p96$ and $p93$ are the fractions of children ages 6–11 who attended school in the district in each year. Because the measure of $p96$ is derived from the INEI survey rather than the census, there are some districts in which the observed fraction of 6–11-year-olds attending school equals 1. It is for this reason that the analysis uses the approximation of the change in the log odds, which does not require division by $1 - p96$.¹¹

In panel A of table 2 the assumption is maintained that the FONCODES index affects gains in attendance only through its effect on FONCODES spending. (This assumption is loosened in panel B.) The first column in panel A shows the ordinary least squares estimate from a regression of the gain in attendance on per capita spending on school infrastructure. The point estimate indicates that a 1-sol increase in per capita FONCODES spending on school infrastructure increases the log odds that a child attends school by 0.059. This corresponds to a gain in the attendance rate of about 0.75 percentage point for districts with an initial attendance rate of 85 percent (about 25 percent of children were in districts with at-

11. All regressions shown were also estimated using $p96 - p93$ as the dependent variable. These estimates are qualitatively similar to those using the log odds specification, although they are sometimes slightly less precise.

TABLE 2. Effects of FONCODES Expenditures on District-Level Gain in School Attendance

Variable	(i) OLS	(ii) IV	(iii) OLS	(iv) IV
A				
Per capita FONCODES school infrastructure expenditures	0.059 (0.011)	0.195 (0.027)	0.056 (0.010)	0.152 (0.033)
Per capita FONCODES "other" expenditures			0.014 (0.004)	0.032 (0.016)
χ^2 test of overidentifying restrictions (degrees of freedom in parentheses, <i>p</i> -values in brackets)		7.62 (6) [0.267]		4.16 (5) [0.526]
B				
Per capita FONCODES school infrastructure expenditures	0.022 (0.011)	0.109 (0.069)	0.023 (0.011)	0.145 (0.080)
Per capita FONCODES "other" expenditures			0.008 (0.004)	0.031 (0.019)
FONCODES index	0.068 (0.009)	0.035 (0.027)	0.062 (0.009)	0.004 (0.036)
χ^2 test of overidentifying restrictions (degrees of freedom in parentheses, <i>p</i> -values in brackets)		3.69 (5) [0.595]		3.48 (4) [0.480]

Note: Columns (i) and (iii) are based on ordinary least squares regressions, and columns (ii) and (iv) on instrumental variables regressions. Standard errors in parentheses. The dependent variable is the approximate gain in the log odds that a child ages 6–11 attends school, as given by $Gain = (p96 - p93) / (p93 [1 - p93])$, where $p93$ and $p96$ measure the fractions of children attending school in 1993 and 1996. Each regression is estimated using data from the 349 districts for which the political measures and 1996 school attendance measures were available. The instruments for the instrumental variables regressions in panel A include the FONCODES poverty index, the fraction of pro-Fujimori votes in the province in 1990 and in the district in 1993, and interactions of the two vote measures with the FONCODES poverty index. The instruments in panel B include the vote measures and interactions with the FONCODES index but not the FONCODES index itself (which is included as an explanatory variable in the second-stage regressions). All regressions are weighted by the number of children in the district in 1993.

Source: Authors' calculations.

tendance rates equal to or less than 85 percent in 1993). This effect is large, given that median per capita spending on school infrastructure was 6.06 soles.

The second column shows the instrumental variables estimate of the effect of school infrastructure spending on the attendance gain. The instruments in the first-stage regressions include the FONCODES index, the fraction of the vote received by Fujimori in the province in 1990 and the district in 1993 (in logs), and interactions of the two vote variables with the FONCODES index. (District-level voting data are not available for 1990.) The first-stage regressions indicate that districts with a smaller share of pro-Fujimori votes in 1993, holding the province-level vote in 1990 fixed, received more FONCODES funding between 1993 and 1996. The positive association between the erosion in the pro-Fujimori vote and FONCODES expenditures is stronger for wealthier districts. The instruments in the first-stage regressions are jointly significant ($F[7,341] = 15.12$, $p = 0.0000$),

and the test for the validity of overidentifying restrictions is easily passed. The instrumental variables estimate of the effect of school infrastructure spending on the gain in school attendance is 0.195 and is significantly different from zero. This point estimate implies that a 1-sol increase in per capita FONCODES spending on school infrastructure would result in an increase in the attendance rate from 85 percent to 87.5 percent.

In addition to spending on school infrastructure, FONCODES also funded noneducation spending, and results in the third and fourth columns of panel A examine whether other spending affected school attendance. It is possible that "other" spending could have had a positive effect on school attendance: although most of this spending went to noneducation infrastructure, a small portion went to school supplies and other education inputs. Moreover, the income gains associated with general FONCODES spending could have increased attendance. But other spending is unlikely to have had as large an effect on school attendance as school infrastructure spending. Consistent with expectations, the ordinary least squares results indicate that although other spending was associated with gains in school attendance, the effect was a quarter of the size of the effect of school infrastructure spending. The instrumental variables estimates also indicate that the effect of other expenditure was substantially smaller.

Thus the results in panel A show that FONCODES spending on school infrastructure had a substantial impact on school attendance for children ages 6–11 and a much larger effect than other FONCODES spending. But the instrumental variables estimates do not account for the potential problem of unobserved factors that may have resulted in increased attendance in the poorer districts that received more FONCODES funds. This line of reasoning suggests that the FONCODES index should be made an explanatory variable in the second-stage regressions, and this is done for the results in panel B. Removing the FONCODES index from the list of instruments and including it in the second stage comes at a cost: the political variables that remain in the list of instruments, although related to FONCODES expenditures, are fairly weak instruments, especially for school infrastructure expenditure. The F -statistic for the political variables is $F(4,343) = 2.60$ ($p = 0.036$) in the school infrastructure equation and $F(4,343) = 5.53$ ($p = 0.0003$) in the other expenditure equation. The instrumental variables results in panel B should therefore be treated with caution.

The results provide some evidence that, even controlling for poverty in 1993 (as measured by the FONCODES index), districts that had higher school infrastructure spending had greater gains in school attendance. The effects of FONCODES school infrastructure spending are generally smaller and less precisely estimated than those in panel A. For example, the ordinary least squares estimate of the effect of infrastructure spending is cut in half when the FONCODES index is included. But the instrumental variables estimates in the last column of panel B (which includes both infrastructure and other spending) indicate that the effect of school infrastructure spending is nearly equal to that in the corresponding

column of panel A, though with a much larger standard error. In this specification neither the FONCODES index nor other spending is significant.

The instrumental variables results in panel B are identified through the effects of the political variables on FONCODES spending. It is possible that voting patterns across districts affected flows of other resources—for example, education expenditure by INFES. Unfortunately, no data are available on how INFES spending changed over time, so it is impossible to assess whether districts that moved away from Fujimori between 1990 and 1993 received increases in INFES spending. Data are available on district-level INFES spending in 1995, however, and the models shown in table 2 were reestimated with a measure of per capita INFES spending in that year. In no case was the coefficient on INFES spending significant, and it had minimal effect on the size and significance levels of the other parameter estimates. The instrumental variables models were also reestimated with INFES spending as an additional (although exogenous) regressor, with similar results. INFES expenditure could not be instrumented, because the sets of instruments used in table 2 were not significantly related to INFES spending.

V. CONCLUSION

This article analyzes the targeting and impact of FONCODES projects in the education sector. Nonparametric regressions are used to evaluate the geographic and household incidence of FONCODES investments. The findings show that FONCODES reached poor districts and, to the extent that they lived in those districts, poor households. The targeting of FONCODES projects in education compares favorably with the targeting of a comparable public-sector program. Geographic variation in expenditures and school outcomes is used to analyze the impact of FONCODES spending on school attendance rates. The results show that districts with the highest levels of FONCODES spending on school infrastructure between 1992 and 1995 had the biggest improvements in attendance between 1993 and 1996.

The results in the article are consistent with a causal relationship between spending on school facilities and improvements in attendance, especially of poor children. An earlier version of the article reached similar conclusions through analyses that used household-level data from the 1994 and 1997 LSMS surveys.¹² The results thus add to a growing literature that finds evidence of a positive association between school-based inputs and measures of educational attainment (for example, Angrist and Lavy 1999, Case and Deaton 1999, Glewwe and Jacoby 1994, Krueger 1999). Nonetheless, the analysis in this article was constrained by important limitations in the data. Three areas deserve attention.

First, the usefulness of the instrumental variables results hinges on the validity of the identifying assumptions—in this case that the political variables and

12. These results are available from the authors on request.

(in panel A of table 2) the FONCODES index are determinants of FONCODES spending in a district but are uncorrelated with unobserved factors that affected school attendance. Although these assumptions are plausible, an evaluation based on a randomized treatment and control strategy would have been preferable. Because applications for FONCODES funds have exceeded the amounts available, an opportunity exists for randomization in the allocation of funds (across equally poor districts). The successful use of randomization in the Bolivian Social Fund and in Programa de Educación, Salud y Alimentación (PROGRESA) in Mexico highlights the benefits of randomization for program evaluation. (On the Bolivian Social Fund see Newman and others 2002; on PROGRESA see Behrman and Todd 1999, Schultz 2001.)

Second, as a result of the absence of credible village-level measures of FONCODES funding, all the estimates of the impact of education projects are based on district-level measures of FONCODES expenditures. Further disaggregation could be important, especially in urban districts, which can be very large. Moreover, village-level measures of “treatment” would make it possible to use statistical matching as an estimation strategy.

Third, lack of disaggregated information on such measures as scholastic achievement, pupil–teacher ratios, and the time children spend in school precluded analysis of the impact of FONCODES education projects on school quality. Collecting such data and understanding the mechanisms whereby improvements in school infrastructure in Peru interact with other school-level changes to result in more learning should be priorities in future research.

REFERENCES

- Alderman, Harold. 1998. “Social Assistance in Albania: Decentralization and Targeted Transfers.” LSMS Working Paper 134. World Bank, Washington, D.C.
- Angrist, Joshua, and Victor Lavy. 1999. “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics* 114(2):533–75.
- Behrman, Jere, and Petra E. Todd. 1999. “Randomness in the Experimental Samples of PROGRESA.” International Food Policy Research Institute, Washington, D.C.
- Case, Anne, and Angus Deaton. 1999. “School Inputs and Educational Outcomes in South Africa.” *Quarterly Journal of Economics* 114(3):1047–84.
- Duflo, Esther. 2001. “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment.” *American Economic Review* 91(4):795–813.
- Fan, Jianqing. 1992. “Design-Adaptive Nonparametric Regression.” *Journal of the American Statistical Association* 87(420):998–1004.
- Galasso, Emanuela, and Martin Ravallion. 2000. “Decentralized Targeting of an Anti-Poverty Program.” Policy Research Working Paper 2316. World Bank, Development Research Group, Washington, D.C.
- Glewwe, Paul, and Hanan Jacoby. 1994. “Student Achievement and Schooling Choice in Low-Income Countries.” *Journal of Human Resources* 29(3):843–64.

- Glewwe, Paul, Margaret Grosh, Hanan Jacoby, and Marlaine Lockheed. 1995. "An Eclectic Approach to Estimating the Determinants of Achievement in Jamaican Primary Education." *World Bank Economic Review* 9(2):231–58.
- Hanushek, Eric A. 1995. "Interpreting Recent Research on Schooling in Developing Countries." *World Bank Research Observer* 10(2):227–46.
- Hanushek, Eric A., and Ralph Harbison. 1992. *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. New York: Oxford University Press.
- Hentschel, Jesko, Jean Olson Lanjouw, Peter Lanjouw, and Javier Poggi. 2000. "Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador." *World Bank Economic Review* 14(1):147–65.
- INEI (Instituto Nacional de Estadística e Informática). 1996. *Metodología para determinar el ingreso y la proporción de hogares pobres*. Lima.
- Kremer, Michael R. 1995. "Research on Schooling: What We Know and What We Don't. A Comment on Hanushek." *World Bank Research Observer* 10(2):247–54.
- Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114(2):497–532.
- Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia. 2002. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* 16(2):241–274.
- Ravallion, Martin. 2000. "Monitoring Targeting Performance When Decentralized Allocations to the Poor Are Unobserved." *World Bank Economic Review* 14(2):331–45.
- Rawlings, Laura B., Lynne Sherburne-Benz, and Julie van Domelen. 2002. "Letting Communities Take the Lead: A Cross-Country Evaluation of Social Fund Performance." World Bank, Social Protection Network, Washington, D.C.
- Schady, Norbert R. 2000. "The Political Economy of Expenditures by the Peruvian Social Fund (FONCODES), 1991–1995." *American Political Science Review* 94(2):289–304.
- . 2002. "Picking the Poor: Indicators for Geographic Targeting in Peru." *Review of Income and Wealth* 48(3):417–33.
- Schultz, T. Paul. 2001. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." Economic Growth Center Discussion Paper No. 834. Yale University, New Haven, Conn.
- Webb, Richard, and Gabriela Fernández Baca. 1997. *Perú 1997 en números: anuario estadístico*. Lima: Cuánto.
- World Bank. 1996. "Did the Ministry of the Presidency Reach the Poor in 1995?" Country Operations Division I, Country Department III, Latin America and the Caribbean Region, Washington, D.C.
- . 1998. "Implementation Completion Report: Peru, Social Development and Compensation Fund Project (FONCODES)." Country Management Unit for Bolivia, Ecuador and Peru, Human Development Sector Management Unit, Latin America and the Caribbean Region, Washington, D.C.
- . 1999. *Poverty and Social Developments in Peru, 1994–1997*. World Bank Country Study. Washington, D.C.

Introduce a Friend or Colleague to

OXFORD
UNIVERSITY PRESS

THE WORLD BANK RESEARCH OBSERVER

Simply photocopy this page, fill in the name and address, and Oxford University Press will send a **FREE** sample copy of *The World Bank Research Observer* without obligation!

Name: _____

Address: _____

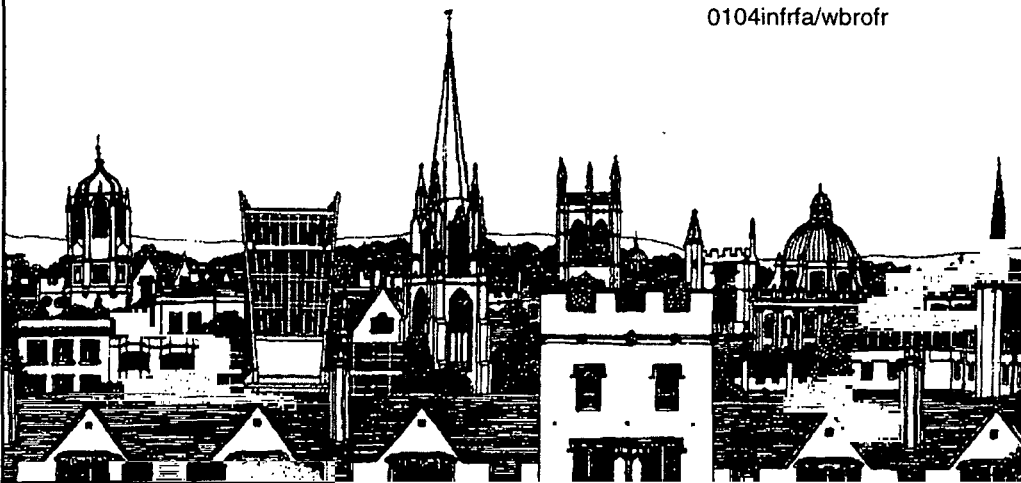
City: _____

State: _____

Zip: _____

Country: _____

0104infrfa/wbrofr



Contact Oxford University Press

In the Americas:

Oxford University Press

2001 Evans Road, Cary NC 27513

Tel: (919) 677-0977 Fax: (919) 677-1714

E-mail: jnlorders@oup-usa.org

Elsewhere:

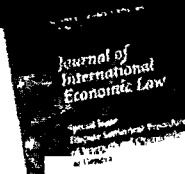
Oxford University Press

Great Clarendon Street, Oxford OX2 6DP, UK

Tel: +44 (0) 1865 267907 Fax: +44 (0) 1865 267485

E-mail: jnl.orders@oup.co.uk

Journals in Economics from OXFORD UNIVERSITY PRESS



Journal of International Economic Law

Dedicated to encouraging thoughtful and scholarly attention to the relation of law to international economic activity.

www.jiel.oupjournals.org



Contemporary Economic Policy

A journal of Western Economic Association International, *CEP* publishes quality research and analysis on policy issues of widespread concern.

www.cep.oupjournals.org



Contributions to Political Economy

A forum for academic discussion of original ideas and arguments drawn from lines of thought associated with the works of classical political economists Marx, Keynes, and Sraffa.

www.cpe.oupjournals.org

Oxford Economics ONLINE!

Subscribers can BROWSE TABLE OF CONTENTS and VIEW SAMPLE ISSUES. Register free of charge for the Table of Contents Alerting Service and an e-mail will be sent every time a new issue appears!

For further details go to: www.economics.oupjournals.org



Oxford Economic Papers

A general economics journal, *OEP* publishes refereed papers in economic theory, development and history, as well as applied economics and econometrics.

www.oep.oupjournals.org



Economic Inquiry

A journal of Western Economic Association International, this peer-reviewed, general economics journal publishes original research from top scholars in the field.

www.ei.oupjournals.org



The Review of Financial Studies

The Review of Financial Studies

Provides significant new research in financial economics, striving to establish a balance between theoretical and empirical studies.

www.rfs.oupjournals.org

Contact Oxford University Press

0104econfa-wbr01a

In the Americas Elsewhere

Oxford University Press

111 Evans Road, Cary, NC 27513

Tel: (919) 677-0977 Fax: (919) 677-1714

E-mail: jnlorders@oup-usa.org

Oxford University Press

Great Clarendon Street, Oxford OX2 6DP UK

Tel: +44 (0) 1865 267907 Fax: +44 (0) 1865 267485

E-mail: jnlorders@oup.co.uk

OXFORD JOURNAL

New to OUP
in 2003!

Socio-Economic Review

<http://www.ser.oupjournals.org>

Editors:

*Alexander Hicks, Emory University
David Marsden, London School of
Economics*

SER is a major new international journal for economists, political scientists, sociologists, and anyone interested in the management and policy sciences. It will publish work on the relationship between society, economy, institutions and markets, moral commitments, and the rational pursuit of self-interest. It seeks to build a bridge among the social sciences to increase their relevance to such issues as the evolution of welfare, the formation of economic policy, the constitution of supranational bodies and organizations, and similar multifaceted issues.

⇒ **FREE e-toc service!**
Stay on top of what *SER* will be publishing, whether you subscribe to it or not. Have the table of contents of each issue e-mailed to you ahead of publication for FREE. Visit www.oupjournals.org/tocmail to register today.

*For further information about *SER*, visit the Web site listed above or contact us at the information below.*

Contact us at:
Journals Dept • E-mail: jnls.cust.serv@oup.co.uk
Tel: +44 1865 267907 • Fax: +44 1865 267835

In the Americas, please contact:
Journals Dept • E-mail: jnlorders@oup-usa.org
Tel: +1 800 852 7323 (USA/Canada)
or +1 919 677 0977 (outside USA/Canada)
Fax: +1 919 677 1714

OXFORD
UNIVERSITY PRESS

Coming in the next issue of

**THE WORLD BANK
ECONOMIC REVIEW**

Volume 16, Number 3, 2002

- **Gender-Differentiated Effects of Social Security Reform:
The Case of Chile**
Alejandra Cox Edwards
- **Does Gender Inequality in Education Reduce Economic
Development? Evidence from Cross-Country Regressions**
Stephan Klasen
- **Gender, Time Use, and Change: Impacts of the Cut Flower
Industry in Ecuador**
Constance Newman
- **The Distributional Impacts of Indonesia's Financial Crisis
on Household Welfare: A "Rapid Response" Methodology**
James Levinsohn and Jed Friedman
- **Density versus Quality in Healthcare Provision: The Use
of Household Data for Budgetary Choices in Ethiopia**
Paul Collier, Stefan Dercon, and John Mackinnon
- **A Firm's-Eye View of Commercial Policy and Fiscal Reforms
in Cameroon**
James Tybout, Bernard Gauthier, Isidro Soloaga



THE WORLD BANK

1818 H Street, NW

Washington, DC 20433, USA

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

E-mail: wber@worldbank.org

ISBN 019-851661-4

