Contents

Preface v
Acknowledgments vii

PART I. HOUSEHOLD BEHAVIOR AND HEALTH

Estimating the Extent of Patient Ignorance of the Health Care Market 3
Mukesh Chawla

Public Transfers and Migrants' Remittances: Evidence from the Recent Armenian Experience 25
Edmundo Murrugarra

PART II. COMMUNITIES AND WELFARE

Better a Hundred Friends Than a Hundred Rubles? Social Networks in Transition—The Kyrgyz Republic 51
Kathleen Kuehnast and Nora Dudwick

An Empirical Investigation of Collective Action Possibilities for Industrial Water Pollution Abatement: Case Study of a Cluster of Small-Scale Industries in India 89
Smita Misra

PART III. LOCAL GOVERNMENTS AND BASIC SERVICES

An Assessment of the Impact of Decentralization on the Quality of Education in Chile 117
Emanuela Di Gropello

Who Benefits from Increased Access to Public Services at the Local Level? A Marginal Benefit Incidence Analysis for Education and Basic Infrastructure 155
Mohamed Ihsan Ajwad and Quentin Wodon
PART IV. FIRMS AND GOVERNMENTS UNDER UNCERTAINTY

Contractual Savings, Capital Markets, and Financing Choices of Firms 179
Gregorio Impavido, Alberto R. Musalem, and Thierry Tressel

Public Expenditures and Risk Reduction 223
Shantayanan Devarajan and Jeffrey S. Hammer
Preface

This volume presents eight outstanding papers from the second World Bank Economists' Forum, held May 3-4, 2001 in Washington, D.C. Launched by then chief economist Joseph Stiglitz in 1999, the Economists' Forum showcases recent research by staff from across the Bank, and especially those in regional operations units. Under the direction of chief economist Nicholas Stern, the Forum 2001 carried on the tradition by including sixteen sessions grouped around the two pillars of the Bank's development strategy, "Enabling Investment, Empowering the Poor." The Forum also had plenary presentations by Mr. Stern, Thomas Schelling (University of Maryland), Michael Mussa (International Monetary Fund), and Paul Collier (World Bank).

The papers published here were chosen from among the 46 papers presented at the Forum, which in turn were selected from more than 100 submitted for consideration. Many of the selections in this volume focus on the key question of "empowerment": how can societies ensure that poor people have the education, health care, social protection, and mechanisms for voice that are necessary for them to participate in economic growth and social development?

We are very grateful to the Bank staff who lent their expertise to this effort—the committee members who helped select papers for the Forum, the session chairs and discussants whose comments improved them, the referees who helped us select from among the papers nominated for inclusion, and the other staff who contributed their expertise. Their names are listed on the next page. We are also very grateful to Susan Graham, our production editor, who shepherded this volume into existence, and to Nick Stern and Ian Goldin for their continued support.

Shantayanan Devarajan
Chief Economist
Human Development Network
World Bank

and

Halsey Rogers
Senior Economist
Office of the Chief Economist and
Development Research Group
World Bank
Acknowledgments

We would like to thank the many World Bank staff who helped make the Forum and this volume possible:

Sadiq Ahmed
Amar Bhattacharya
Carlos Braga
Penelope Brook
Barbara Bruns
Nisangul Ceran
Ariel Dinar
Ishac Diwan
Yahaya Doka
David Dollar
William Easterly
Robert Ebel
Gunnar Eskeland
Deon Filmer
Varun Gauri
Alan Gelb
Cheryl Gray
Charles Griffin
Christian Grootaert
Margaret Grosh
Jose Luis Guasch
Jeffrey Hammer
Trina Haque
Karla Hoff
Robert Holzmann
Gregory Ingram
Roumeen Islam
Emmanuel Jimenez
Steen Lau Jorgensen
Elizabeth King
Daniela Klingebiel
Aart Kraay
Michael Kremer
Barbara Lee
Maureen Lewis
Kathy Lindert
Ashoka Mody
Pradeep Mitra
Govind Nair
John Page
Alexander Preker
Martin Ravallion
Ritva Reinikka
Jo Ritzen
Neil Roger
Marcelo Selowsky
Alfred Thieme
Marilou Uy
Tara Vishwanath
Deborah Wetzel
David Wheeler
Michael Woolcock
Roberto Zagha
Part I

Household Behavior and Health
Estimating the Extent of Patient Ignorance of the Health Care Market

Mukesh Chawla

Abstract

The wide dispersion observed in prices for health services in seemingly competitive markets is not fully explained by physician or consumer characteristics, or by variations in quality of care. For a variety of reasons, including the urgent nature of consumption and the asymmetry of information, patients in the market for health services face high search costs. They balance the prospect of finding a physician willing to accept lower fees against the greater costs involved with the gathering of information and the search for such a physician. At the same time, physicians also balance the prospect of securing a higher fee against losing the patient. In either or both of these cases, stable market equilibria may exist with the same physician charging different fees to different consumers, and with different physicians charging different prices, even if the service provided is fairly homogenous and standard.
This paper presents an estimate of the degree of incomplete consumer and provider information about prices of health care services in a developing country. Following Gaynor and Polachek (1994), "consumer ignorance" is defined as the difference between the observed market price and the lowest price that the provider is willing to accept, and "provider ignorance" as the difference between the highest price that the consumer is willing to pay and the observed market price. Estimates of ignorance are obtained using two-tiered generalized stochastic frontier techniques to separate the dispersion in observed prices into a random two-sided variation attributable to measurement errors, and left- and right-side variations attributable, respectively, to imperfect consumer and provider information.

The results indicate that patients have considerably less than full information about the physician market and, on average, pay substantially higher fees than they would have if they had better information. In particular, patient ignorance is markedly higher for high-severity items like surgery than for routine general practitioner visits. Physicians also lack full information, but not to the same extent as consumers of health care. The fact that physician fees vary significantly in markets that otherwise appear to be competitive has implications for policies suggested to strengthen the role of market forces, consumer empowerment, and government regulation.

The wide dispersion observed in prices for health services in seemingly competitive markets is not fully explained by physician or consumer characteristics, or by variations in quality of care. For a variety of reasons, including the urgent nature of consumption and asymmetry of information, patients in the market for health services face high search costs and balance the prospect of finding a physician willing to accept lower fees against the greater costs involved with the gathering of information and search for such a physician. At the same time, physicians also balance the prospect of securing a higher fee against losing the patient. In either or both of these cases, stable market equilibria may exist with the same physician charging different fees to different consumers, and with different physicians charging different prices, even if the service provided is fairly homogenous and standard.

A large number of studies have documented dispersion in physician fees in the United States and elsewhere (see, for instance, Feldstein 1970; Gaynor 1994; Hsiao 1980; Kleverick and McGuire 1987; McCarthy 1985; Newhouse 1970; Rizzo and Zeckhauser 1992; Sloan 1976). The large dispersion in physician fees has commonly been interpreted as indicative of incomplete market information. Stigler and Kindahl (1973) argue that the consumer search process by itself may
not reduce variance in prices if inflation reduces consumer information about price. Van Hoomissen (1988) also shows that obsolescence of information caused by inflation leads to greater price dispersion. Pratt, Wise, and Zeckhauser (1979) present theoretical models with and without learning and show that an equilibrium may involve variance in prices. Phelps (1992) argues that the relation between dispersion in prices and incomplete consumer information is robust to variations in quality, despite the fact that quality and product differences are also reflected in price. Gaynor and Polachek (1994) find little evidence of one-to-one correspondence between price differences and quality differences, even after controlling for such other factors as physician education, specialty, location, practice style, and clientele. In an examination of physician prices for a standard office visit in Dayton, Ohio, they write, "It is hard to believe that quality alone is responsible for such wide specialty variations."

The fact that physician fees vary significantly in markets that otherwise appear to be competitive has implications for concepts used to study market behavior and policies suggested to strengthen the role of market forces and government regulation. The consequences of incomplete market information and the ineffectiveness of market forces to yield a price indicative of quality and treatment alone are even more serious for equity and access in developing countries, especially in those where the private sector plays a significant role in providing health services. Despite an urgent need to understand reasons for dispersion in physician fees, we are not aware of any study that has attempted to measure the nature and extent of market ignorance in the health sector in developing countries.

Studies that have examined the role of information and prices in the health sector in the United States have typically focused on the effects of advertising on prices (Kwoka 1984; Gaynor and Mullahy 1993), and it was only recently that the first estimates of buyer and seller ignorance in the health market were prepared. Using data on a national sample of 6,353 physicians practicing in five different subspecialties, Gaynor and Polachek (1994) measured incomplete patient and physician information in this market. They found that patient information was significantly less complete than physician information for all types of services, especially office consultations, hospital follow-up visits, blood counts, chest X-rays, and D&Cs.

We believe this paper to be the first attempt to fill this gap in a developing country setting, and provides an estimate of the degree of incomplete consumer and provider information using data on the pric-
ing of health services as obtained from household and provider surveys carried out in 1994 in Egypt. We define "full-information market equilibrium" as one in which a unique price evolves such that it is equal to both the highest price the consumer is willing to pay and the lowest price that the seller is willing to accept. Following Gaynor and Polachek (1994), we measure "consumer ignorance" as the difference between the observed market price and the lowest price that the provider is willing to accept. Similarly, "provider ignorance" is measured as the difference between the highest price that the consumer is willing to pay and the observed market price. Using the two-tiered generalized stochastic frontier technique developed by Polachek and Yoon (1987), we demonstrate that consumer and provider ignorance can be estimated by separating the dispersion in observed prices into a random two-sided variation attributable to measurement errors, and left- and right-side variations attributable to imperfect consumer and provider information. Note that these one-sided variations are equal to zero in a full-information equilibrium.

Our results indicate that patients have considerably less than full information about the physician market and, on average, pay substantially higher fees than they would have if they had better information. In particular, patient ignorance is markedly high for high-severity items like surgery compared to routine general practitioner visits. Physicians also lack full information, but much less so than the consumers of health care across specialties, as well as regions.

Implicit in the interpretation of results is the assumption that the dispersion in physician fees is indicative only of information imperfections in the market. Surely, there are other factors that can result in price differences, such as product quality, credit policies, accompanying services, and personal relations, all of which are unobserved. Unobserved differences in consumers and providers can potentially create an upward bias in the imperfect information measures if these differences cause prices to vary (Polachek and Yoon 1987). We get around this problem by looking at relative measures according to different strata, because providers and consumers are more likely to be homogenous within each stratum than across the whole sample. Accordingly, we evaluate imperfect information separately for general physicians, gynecologists, cardiologists, and surgeons, and for urban and rural populations. Moreover, our unit of analysis is a consultation visit in the private clinic, a service that is more homogenous than other medical services such as surgical interventions. Wide variations in the prices of such services are less likely to be caused by
unobserved variations in quality, particularly within urban and rural sampling frames.

The paper is organized as follows. We present some institutional details of the physician market in Egypt in the first section. The model and methodology are presented next, followed by data sources, then estimation results. A discussion on policy implications in the Egyptian context precedes our concluding remarks.

The Market for Physicians in Egypt

Egypt's health care delivery system is sharply dichotomized, in the sense that the government and public sector provide almost all the inpatient care, whereas private providers dominate in providing ambulatory care (HSPH 1995). Health care in the government and public sector is financed, produced, and delivered by the Ministry of Health and Population (MOHP), the health insurance organization (HIO), the curative care organization (CCO), university hospitals, teaching hospitals, and facilities owned by other government departments and agencies. The government uses general revenues to provide free health services for all citizens through a network of health facilities it owns and manages. The MOHP runs more than 3,700 primary, secondary, and tertiary health care facilities; and more than 95 percent of the population lives within 5 kilometers of a government health facility. In addition, there is a social insurance program that covers employees in the formal sector and school children. In the fiscal year 1995, for instance, Egypt spent LE 7,519 million (3.7 percent of GDP) on health care ($1=LE 3.39), equivalent to LE 127 per capita. Overall public spending accounted for only 44 percent of total health financing, with the balance coming from private sources. Almost 80 percent of public expenditure on health care comes from general tax revenues, 14 percent from social insurance premiums, and the rest from external donor assistance. There is also a large and rapidly expanding private market that functions in an essentially unregulated environment. Health care in the private sector is provided predominantly in clinics and, to a much smaller extent, in private hospitals.

Graduating physicians are guaranteed employment by the government, and the MOHP is the single largest employer of physicians in the country. In 1996, the MOHP employed 39,900 physicians, 40 percent of whom worked in primary health care and preventive health services and the remaining 60 percent in the curative sector. Physicians working for the government are allowed to have a private
practice. Accurate data on the number of physicians currently in private practice are not available because of high rates of physician emigration and lack of routine updating of practice registration. The provider survey carried out by the Data for Decision Making Project (DDM), however, indicates that there are between 34,447 and 48,403 single-physician private practices in the country (HSPH 1997). This translates to 1.8 physicians per 1,000 population, which is the highest availability of physicians among countries in the Middle East and North Africa.

The distribution of private clinics is heavily biased toward urban areas, and 34 percent of all private clinics are located in urban governorates, 29 percent in urban Upper Egypt, and 21 percent in urban Lower Egypt. The remaining 16 percent of the private clinics are located in rural areas, of which a little more than half are in rural Lower Egypt and the rest in rural Upper Egypt. Ninety-two percent of the private clinics are owned and run by male physicians and 8 percent are run by female physicians.

Many physicians in Egypt work in their own private clinics, in addition to holding salaried jobs in other medical facilities and institutions. In a study of physician labor supply in Egypt, Chawla and others (1997) found that physicians holding salaried jobs in MOHP hospitals worked more in the private clinics compared with those who held salaried jobs in other medical facilities and institutions. They found that wage effects and elasticities were small, and physicians responded to increased earnings in private clinics by modestly increasing the number of hours they work in private clinics. When the market was segmented along urban-rural governorates, however, the wage effect and elasticities increased significantly, with the markets with the greatest potential demand showing the largest physician response. Physicians in principally urban governorates of Cairo, Alexandria, Port Said, and Suez showed a relatively weak response to changes in hours worked in the government and public sector jobs as compared with physicians in other governorates. Their findings suggest that changes in government policies are likely to have different results in urban and rural areas, which are distinguished by potential market demand and other institutional factors. Chawla and others (1997) also found a negative relationship between hours of work in two jobs, and concluded that, as physicians work more in their private clinic, they reduce labor supply in their salaried government or public sector jobs. Their study indicated that rural-urban differences and location of a physician's salaried job have an important bearing on the market for
physicians in Egypt, both in relation to their earnings and to their allocation of hours between the two jobs.

**Methodology**

Stochastic frontier models have been used in a wide variety of settings. Originally developed by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977), stochastic frontier models have been used to measure gender discrimination in labor markets (Goldin and Polachek 1987; Robinson and Wunnava 1989), relative inefficiencies in production between solo and group practice physicians (DeFelice and Bradford 1997), earnings in labor markets (Herzog, Hofler, and Schlottmann 1985; Hunt-McCool and Warren 1993; Hofler and Polachek 1985), and production inefficiency (Jondrow and others 1982; Waldman 1984; Greene 1990). Two-tiered stochastic frontier models, proposed by Polachek and Yoon (1987), have been used to estimate incomplete information of workers and firms (Polachek and Yoon 1987, 1996) and incomplete information of patients and physicians in the health market (Gaynor and Polachek 1994).

Essentially, stochastic frontier estimation decomposes the error associated with each observation into two components: the traditional white-noise error, indicative of such errors that may be caused by measurement and omitted variables, and a one-sided error. In practice, this means that the error generated by maximum likelihood estimation techniques is examined for skew which, if present, is then broken up into a normal component and a right- or left-sided skewed component (DeFelice and Bradford 1997). The degree of skew is then used as a measure of such imperfections associated with the dependent variable as may be suggested by theory. Models that break up the skew into a normal two-sided error and both right- and left-sided error terms are also referred to in the literature as “two-tiered models.” Two-tiered frontier models have been developed by Polachek and Yoon (1987) and Gaynor and Polachek (1994), and our model follows their framework closely. Because the derivation is readily available, we will only provide a brief discussion here.

In a full-information market, patients seeking health care of a given quality will generally be able to find the physician supplying that service at least cost to the patient. Similarly, in such a full-information market, income-maximizing physicians will generally be able to assess a consumer’s willingness to pay and charge the maximum fees possible. In a market with informational imperfections, however, the patients
know the distribution of prices in the market, but not the price charged by any particular physician. Therefore, they choose an optimal amount of search, balancing the costs of the search with expected savings from finding a lower price. Similarly, the problem for physicians is that they are not always aware of a consumer’s maximum willingness to pay, and so they compromise by accepting a fee that is lower than what the consumer is prepared to pay for that quality of service. These gaps between actual price paid and physician reservation price on the one hand and maximum willingness to pay by the consumers on the other reflect, respectively, consumer and physician ignorance of the market.

More formally, let the full-information or patient reservation fee be

\[ p^{\text{FIC}} = p^{\text{AC}} + e^C \]  

(1)

where \( p^{\text{FIC}} \) is the consumer reservation fee, \( p^{\text{AC}} \) the actual fee paid by the patient, and \( e^C \) is a nonnegative random error that depicts the amount by which the two differ. Similarly,

\[ p^{\text{FIP}} = p^{\text{AP}} + e^P \]  

(2)

where \( p^{\text{FIP}} \) is the physician reservation fee, \( p^{\text{AP}} \) the actual fee charged by the physician, and \( e^P \) is a nonnegative random error that depicts the amount by which the two differ. In equilibrium, the price paid by the patient is equal to the price charged by the physicians, so that

\[ p^{\text{AC}} = p^{\text{AP}} \]  

(3)

This equilibrium condition (3) can be written as

\[ p^{\text{FIP}} - p^{\text{FIC}} = e^P - e^C \]  

(4)

Equation (4) can also be expressed as

\[ \phi(P, X) = e^P - e^C \]  

(5)

where \( P \) refers to the observed market price, and \( X \) refers to all those exogenous factors that affect consumer’s and physician’s reservation price. Following Polachek and Yoon (1987), we apply Taylor’s approximation to \( \phi(P, X) \) around \( (P_0, X_0) \), and obtain

\[ \phi(P, X) = \phi(P_0, X_0) + (X - X_0) \frac{\partial \phi}{\partial X} + (P - P_0) \frac{\partial \phi}{\partial P} + e^P - e^C + R \]  

(6)
Solving equation (6) for \( P \), we get

\[
P = \left( \frac{\partial \phi}{\partial P} \right)^{-1} \left\{ \left[ \frac{(\partial \phi / \partial P) P_0 + (\partial \phi / \partial X) X_0 - \phi(P_0, X_0)}{+e - e^C - R} \right] \right\}
\]

which can be expressed as

\[
P = \beta' X + u + v + w
\]

or

\[
P = \beta' X + e
\]

where \( \beta \) is the vector \( \left( \frac{\partial \phi}{\partial P} \right)^{-1} \left( \frac{(\partial \phi / \partial P) P_0 + (\partial \phi / \partial X) X_0 - \phi(P_0, X_0)}{\left( \frac{\partial \phi}{\partial X} \right)} \right) \) and \( X \) is the column vector that has two elements \( [1, X_0]^T \). The augmented remainder term of the Taylor's expansion, \( \left( \frac{\partial \phi}{\partial P} \right)^{-1} R \), is captured by the error term \( u \), which is assumed to be two-sided and random, \( v \) is \( \left( \frac{\partial \phi}{\partial P} \right)^{-1} e^P \), and \( w \) is \( -\left( \frac{\partial \phi}{\partial P} \right)^{-1} e^C \). In equation (9), \( e \) is simply the composite error term. In this specification, \( E(v) < 0 \) represents physician ignorance, and \( E(w) > 0 \) represents patient ignorance.

For purposes of tractability and identification, assume that the random error \( u \in [-\infty, \infty] \) is normally distributed with mean zero and variance \( \sigma_u^2 \), whereas \( v \) and \( w \) are distributed over \( [-\infty, 0] \) and \( [0, \infty] \), respectively, and have an exponential distribution, with mean \( \mu_v \) and \( \mu_w \), respectively.

Polachek and Yoon (1987) compute the density of the composite error term, and derive the likelihood function:

\[
\log L = n \log \left( \frac{\theta_u \theta_v \theta_w}{\theta_v + \theta_w} \right) + \left[ \theta_u \theta_v \sum_i e_i + (n / 2) \theta_v^2 \right]
\]

\[
+ \sum_i \left[ \frac{1 - \phi(\theta_u e_i + \theta_v)}{1 - \phi(-\theta_u e_i + \theta_w)} \right] \exp \left[ -\frac{1}{2} (\theta_u e_i + \theta_v - \theta_w)(\theta_v + \theta_w) \right]
\]

where \( \theta_u = 1/\sigma_u, \theta_v = \sigma_u / \mu_v \), and \( \theta_w = \sigma_u / \mu_w \). \( \sigma_u \) is the standard deviation of the normally distributed error term, and \( \mu_v \) and \( \mu_w \) are means of the single-sided error terms.

**Data**

Data for this analysis were obtained from a survey conducted by the DDM in collaboration with the MOHP in Cairo. The survey sampled
802 physicians drawn from 12 of the 28 governorates, separated according to urban and rural governorates. The sampling unit was the physician’s private clinic, and the sample design was based on data collected in the 1986 Institutions Census. Excluding incomplete responses, our final sample had information on 731 physicians. Summary statistics are presented in table 1.

Most physicians in our sample (92 percent) were males, and the average age of physicians in the sample was 44 years. Eighty-nine percent (653 out of 731) physicians worked in a second job, usually hospital based, in addition to their private clinics. Of these, most (54 percent) worked in MOHP hospitals, followed by university teaching hospitals (14 percent), HIO (7 percent), and CCO (1 percent). The remaining physicians work in private and other hospitals.

Physician education was measured by the highest degree earned. Twenty percent of the sample had a Ph.D., whereas 44 percent had a diploma as their highest earned degree. Specialization in a specific area was measured by specialization as reported by the physician, not necessarily areas in which an advanced degree was obtained. Twenty-eight percent of the sample reported general practice as their area of specialization, 16 percent reported gynecology, 17 percent reported surgery, and 11 percent reported pediatrics. The remaining sample included specialists in cardiology, ophthalmology, otolaryngology, dermatology, orthopedics, neurology, and chest diseases.

For purposes of controlling for regional differences, we grouped the 12 governorates into five categories. The principally urban governorates of Cairo, Alexandria, Port Said, and Suez formed one category. The governorates of Lower Egypt (Dhakeliya, Kalubiya, Gharbeya, and Behera) were put into one category, and the governorates of Upper Egypt (Giza, Beni-Suef, Assuit, and Qena) were placed in the third category. The last two categories were further subdivided into “lower-urban” and “lower-rural” and “upper-urban” and “upper-rural,” giving a total of five governorate-categories. The sample included 257 physicians (35 percent) from the principally urban governorates, 216 (30 percent) from upper-urban, 145 (20 percent) from lower-urban, 58 (8 percent) from upper-rural, and 57 (8 percent) from lower-rural governorates.

The average consultation fee charged per patient examination was LE 10.85. There were many variations across specializations and across regions, however. Surgeons reported the highest consultation fees (LE 13.16), followed by gynecologists (LE 9.40), general practitioners (LE 9.3), and pediatricians (LE 7.83). Physicians in the principally urban
TABLE 1. DATA ON PHYSICIANS: SUMMARY STATISTICS OF KEY VARIABLES

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.921</td>
<td>0.270</td>
<td>673</td>
</tr>
<tr>
<td>Age</td>
<td>43.557</td>
<td>9.407</td>
<td>731</td>
</tr>
<tr>
<td>Experience</td>
<td>17.77</td>
<td>9.58</td>
<td>731</td>
</tr>
<tr>
<td>Diploma in medicine</td>
<td>0.439</td>
<td>0.497</td>
<td>321</td>
</tr>
<tr>
<td>Ph.D. in medicine</td>
<td>0.202</td>
<td>0.402</td>
<td>148</td>
</tr>
<tr>
<td>General practitioner</td>
<td>0.286</td>
<td>0.452</td>
<td>209</td>
</tr>
<tr>
<td>Pediatric</td>
<td>0.108</td>
<td>0.310</td>
<td>79</td>
</tr>
<tr>
<td>Gynecologist</td>
<td>0.160</td>
<td>0.367</td>
<td>117</td>
</tr>
<tr>
<td>Surgeon</td>
<td>0.169</td>
<td>0.375</td>
<td>124</td>
</tr>
<tr>
<td>Cardiologist</td>
<td>0.054</td>
<td>0.226</td>
<td>39</td>
</tr>
<tr>
<td>Ministry of Health and Population</td>
<td>0.536</td>
<td>0.499</td>
<td>392</td>
</tr>
<tr>
<td>Curative care organization</td>
<td>0.007</td>
<td>0.083</td>
<td>5</td>
</tr>
<tr>
<td>Health insurance organization</td>
<td>0.069</td>
<td>0.254</td>
<td>50</td>
</tr>
<tr>
<td>University hospital</td>
<td>0.146</td>
<td>0.354</td>
<td>107</td>
</tr>
<tr>
<td>Upper-urban regions</td>
<td>0.296</td>
<td>0.457</td>
<td>216</td>
</tr>
<tr>
<td>Upper-rural regions</td>
<td>0.079</td>
<td>0.270</td>
<td>58</td>
</tr>
<tr>
<td>Lower-urban regions</td>
<td>0.198</td>
<td>0.398</td>
<td>145</td>
</tr>
<tr>
<td>Lower-rural regions</td>
<td>0.079</td>
<td>0.269</td>
<td>58</td>
</tr>
<tr>
<td>Urban governorates</td>
<td>0.351</td>
<td>0.478</td>
<td>257</td>
</tr>
<tr>
<td>Consultation fees</td>
<td>10.854</td>
<td>10.333</td>
<td>731</td>
</tr>
<tr>
<td>Patient/week</td>
<td>22.81</td>
<td>27.14</td>
<td>731</td>
</tr>
</tbody>
</table>

governorates of Cairo, Alexandria, Port Said, and Suez charge the highest consultation fees (LE 15.66), followed by upper-urban governorates (LE 11.42), lower-urban (LE 9.19), lower-rural (LE 6.72), and upper-rural governorates (LE 6.26).

Estimation Results

We first estimate a variant of equation (9), using (log) physician consultation fees for a standard office visit as the dependent variable. The independent variables in this equation represent exogenous factors that are likely to determine physician reservation price and consumer willingness to pay: physician characteristics, market characteristics, and consumer characteristics. Physician characteristics that influence fees are fields of specialty in which the physician practices, as well as advanced degrees, age, experience, and gender. The specialties that we include in this analysis are general practice, gynecology, pediatrics,
cardiology, and surgery. Advanced degrees include a diploma in medicine and a doctorate in a subspecialty. Experience is measured by the number of years worked in the present clinic.

Most physicians in Egypt have multiple jobs, and typically work in a government, public, or private health facility as salaried employees in addition to running their own private practices. The institutional nature of the organization where physicians work in their salaried job has a significant effect on the number of hours the physicians can work in the private practice, the fee they charge, and the number of patients they see (Chawla and others 1997). We capture these market characteristics by including dummies representing place of work in the list of exogenous variables. Physicians in our sample work in MOHP facilities, HIOs, CCOs, and university teaching hospitals.

In an analysis of the factors that influence utilization of health care in Egypt, Nandakumar, Chawla, and Khan (1999) show that income, education, and living in an urban area are the main determinants of health care-seeking behavior, with all three factors having a significant and positive effect. Results of a household survey conducted by the Data for Decision Making Project in Egypt show that almost 75 percent of persons in the lowest income quintile live in rural areas, and that education levels were significantly higher in urban governorates compared with the rural governorates. Accordingly, we use indicators of location as representative of consumer behavior, and include dummy variables for upper-urban, upper-rural, lower-urban, lower-rural, and principally urban governorates in the set of exogenous variables.

The full set of independent variables thus includes age; experience; age squared; experience squared; dummies for subspecialties such as general practice, gynecology, cardiology, and surgery; dummies for regions, that is, principally urban, upper-urban, lower-urban, upper-rural, and lower-rural; a dummy for male; and a dummy variable indicating whether the physician has a Ph. D. degree.

Table 2 reports the ordinary least squares (OLS) and maximum likelihood estimation (MLE) parameter estimates for the physician consultation fee. The results indicate that physician experience is a significant determinant of fees, a finding that is reasonable and consistent with what one would expect. Physicians with a Ph.D. command a higher fee compared to diploma-holders, which is also a reasonable finding. Physician fees are higher in the principally urban governorates and other urban regions compared with those in the upper-rural region. Fees are low for physicians who hold salaried jobs in MOHP facilities compared with those who work on salaried jobs elsewhere.
Table 2. Determinants of Physician Consultation Fees [Dependent Variable: log(fees); N = 731 (all physicians)]

<table>
<thead>
<tr>
<th>Variable</th>
<th>MLE</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.587*</td>
<td>1.5335*</td>
</tr>
<tr>
<td></td>
<td>(0.5793)</td>
<td>(0.5935)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0225</td>
<td>-0.0125</td>
</tr>
<tr>
<td></td>
<td>(0.0323)</td>
<td>(0.0307)</td>
</tr>
<tr>
<td>Age squared</td>
<td>0.0033*</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.0492*</td>
<td>0.0375*</td>
</tr>
<tr>
<td></td>
<td>(0.0120)</td>
<td>(0.0139)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.0031*</td>
<td>-0.0006</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Male</td>
<td>0.0829*</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>(0.0380)</td>
<td>(0.0571)</td>
</tr>
<tr>
<td>Ministry of Health and Population</td>
<td>-0.1237*</td>
<td>0.0046</td>
</tr>
<tr>
<td></td>
<td>(0.0314)</td>
<td>(0.0423)</td>
</tr>
<tr>
<td>Curative care organization</td>
<td>-0.1342</td>
<td>-0.1167</td>
</tr>
<tr>
<td></td>
<td>(0.1145)</td>
<td>(0.2529)</td>
</tr>
<tr>
<td>Health insurance organization</td>
<td>0.1784*</td>
<td>-0.0274</td>
</tr>
<tr>
<td></td>
<td>(0.0892)</td>
<td>(0.0645)</td>
</tr>
<tr>
<td>University</td>
<td>0.1371</td>
<td>0.2178*</td>
</tr>
<tr>
<td></td>
<td>(0.1180)</td>
<td>(0.0919)</td>
</tr>
<tr>
<td>Ph.D.</td>
<td>0.5761*</td>
<td>0.1813*</td>
</tr>
<tr>
<td></td>
<td>(0.1201)</td>
<td>(0.0805)</td>
</tr>
<tr>
<td>General practitioner</td>
<td>-0.1570*</td>
<td>-0.1146*</td>
</tr>
<tr>
<td></td>
<td>(0.0311)</td>
<td>(0.0391)</td>
</tr>
<tr>
<td>Gynecologist</td>
<td>-0.0146</td>
<td>0.0199</td>
</tr>
<tr>
<td></td>
<td>(0.0332)</td>
<td>(0.0469)</td>
</tr>
<tr>
<td>Cardiologist</td>
<td>0.0643</td>
<td>0.0201</td>
</tr>
<tr>
<td></td>
<td>(0.0790)</td>
<td>(0.0852)</td>
</tr>
<tr>
<td>Surgeon</td>
<td>0.1036</td>
<td>0.0197</td>
</tr>
<tr>
<td></td>
<td>(0.0436)</td>
<td>(0.0468)</td>
</tr>
<tr>
<td>Urban governorates</td>
<td>0.2546*</td>
<td>0.2335*</td>
</tr>
<tr>
<td></td>
<td>(0.0548)</td>
<td>(0.0576)</td>
</tr>
<tr>
<td>Upper-urban regions</td>
<td>0.3036*</td>
<td>0.3249*</td>
</tr>
<tr>
<td></td>
<td>(0.0588)</td>
<td>(0.0562)</td>
</tr>
<tr>
<td>Lower-urban regions</td>
<td>0.1513*</td>
<td>0.2342*</td>
</tr>
<tr>
<td></td>
<td>(0.0581)</td>
<td>(0.0587)</td>
</tr>
<tr>
<td>Lower-rural regions</td>
<td>0.4725*</td>
<td>0.0182</td>
</tr>
<tr>
<td></td>
<td>(0.1441)</td>
<td>(0.0684)</td>
</tr>
<tr>
<td>$\theta_u$</td>
<td>7.1653*</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>(0.8018)</td>
<td>n.a.</td>
</tr>
<tr>
<td>$\theta_v$</td>
<td>1.4546*</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>(0.1845)</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

(Table continues on the following page.)
TABLE 2. (CONTINUED)

<table>
<thead>
<tr>
<th>Variable</th>
<th>MLE</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_w$</td>
<td>0.3711*</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>(0.1284)</td>
<td>n.a.</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.1396</td>
<td>n.a.</td>
</tr>
<tr>
<td>$\mu_v$</td>
<td>0.0959</td>
<td>n.a.</td>
</tr>
<tr>
<td>$\mu_w$</td>
<td>0.3760</td>
<td>n.a.</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.2384</td>
<td></td>
</tr>
</tbody>
</table>

* Significant at 0.01 level.
 n.a. Not applicable.

A major advantage in using the maximum likelihood approach is that it allows for decomposition of the error term into average patient and physician ignorance, estimates of which are obtained from the coefficients on $\theta_w$, $\theta_u$, and $\theta_v$. Our results show average patient ignorance to be 0.3760 and average physician ignorance to be 0.0959, both measured in logarithms, indicating that incomplete information among patients is greater than incomplete information among physicians. Similar results are also obtained in Gaynor and Polachek (1994), who find this evidence supporting “common wisdom concerning information asymmetries” (Gaynor and Polachek 1994, page 823).

Note that equation (8) can be written as

$$P = \exp(\beta' X) \cdot \exp(u) \cdot \exp(v) \cdot \exp(w)$$

and because $v$ and $w$ follow the exponential distribution, the expected values of $\exp(v)$ and $\exp(w)$ denote the effect of patient and physician ignorance, respectively, on actual fee. In this case, we note that $E[\exp(v)] = 1/(1 + \mu_v) = 0.9125$ and $E[\exp(w)] = 1/(1 - \mu_w) = 1.6026$. In other words, physicians earn 91.25 percent of what they could have had if they had had full information on consumer willingness to pay, and patients pay 160.26 percent of what they could have paid had they had full information on physician reservation fees. The combined effect of these two inefficiencies is 1.4623, that is, actual fees are 46 percent higher than the full-information competitive fee.

We also estimate imperfections in patient and physician information for different services and for urban and rural locations. These results are reported in table 3. In all cases, patients have less information than physicians, although the range varies from low variation in cardiology to high in surgery. Very similar findings were reported by Gaynor and Polachek (1994), who also found that the largest differ-
TABLE 3. PATIENT AND PHYSICIAN IMPERFECT INFORMATION, BY SERVICE AND LOCATION (PERCENT)

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Patient incomplete information</th>
<th>Physician incomplete information</th>
</tr>
</thead>
<tbody>
<tr>
<td>General practitioners</td>
<td>134.89</td>
<td>87.75</td>
</tr>
<tr>
<td>Gynecologists</td>
<td>136.12</td>
<td>92.56</td>
</tr>
<tr>
<td>Cardiologists</td>
<td>125.37</td>
<td>92.01</td>
</tr>
<tr>
<td>Surgeons</td>
<td>505.54</td>
<td>95.17</td>
</tr>
<tr>
<td>Rural areas</td>
<td>177.76</td>
<td>89.48</td>
</tr>
<tr>
<td>Urban areas</td>
<td>145.09</td>
<td>88.98</td>
</tr>
</tbody>
</table>

ences between patient and physician information measures were in services purchased infrequently and associated with the greatest severity of illness, such as surgery, in our case. We also find much larger differences in relative patient and physician measures in rural areas than in urban, which is reasonable considering consumers in urban areas generally have higher incomes, better education, and greater access to information compared with their rural counterparts. In fact, the combined effect of these imperfections is the least in urban areas and for general practitioners’ and gynecologists’ services, indicating that even though individual fees may vary, the average fee in these markets is close to full-information competitive fees.

Discussion

To understand the implications of consumer ignorance in the market for physician services in the Egyptian context, we start by looking at the data on utilization, expenditure, and choice of providers for those seeking care. The household survey data show that utilization rates are higher in urban areas and among the higher income quintiles (table 4). About half of all visits take place with private providers—21 percent at MOHP facilities, 10 percent at other government facilities, 10 percent at HIO and CCO facilities combined, and 6 percent at mosque clinics. Regardless of income, region, or gender, utilization of the private services is much higher than for any other provider. Private providers account for a majority of visits even in rural areas, where 54 percent of all outpatient visits take place with private providers. The low rate of utilization of MOH facilities is rather surprising, given that MOH finances, owns, and operates an extensive network of outpatient
facilities and 99 percent of the population has a health facility within 5 kilometers of their place of residence. Even with respect to individuals in the lowest income quintile, only 39 percent of all visits occur at MOH facilities.

Overall, patients make 3.47 outpatient visits to formal providers and spend LE 107 annually. The rate of utilization and expenditure is much lower among the poorest quintile, who make 2.21 visits at an average annual expenditure of LE 39.83, compared to 5.1 visits for the richest quintile, at an average of LE 237.34. About 40 percent of all outpatient visits for the poor take place in the private sector, compared to about 60 percent for the richest quintile. Because the public health facilities provide free health care, most of the out-of-pocket expenditure on outpatient health care takes place in the private sector. If 80 percent of all out-of-pocket expenditure takes place in the private sector, the poorest quintile spend almost 8 percent of their annual per capita income on outpatient care alone.

Our estimates of inefficiencies in the competitive market attributable to imperfect information indicate that physician fees are 46 percent higher than full-information competitive fees. Removing market imperfections can, theoretically at least, reduce the burden of expenditure on health for the poor by as much as LE 14.66, or about 3.7 percent of their annual income. Gains of similar magnitude can potentially be realized for other quintiles as well.

A case for improving consumer information in the market for health services in Egypt is made in Yip and Orbeta (1999) as well. In a study of patient response to quality and price changes for outpatient care, they find a greater response among patients to quality changes than to price changes. Within quality itself, they report greater patient sensitivity to process and access aspects of quality that, in their words, "are easy to manipulate and allow providers to easily quality discriminate" (Yip and Orbeta 1999, p. 30). To protect the patients, who are typically not capable of evaluating the technical aspects of quality, they suggest that "it would be advisable for the government in this case to increase information on quality of care to consumers" (Yip and Orbeta 1999, p. 30).

The validity and significance of our results are not confined to Egypt alone. Many other low- and middle-income countries have similar market, physician, and consumer characteristics: private providers play an important role in delivering health care, a significant proportion of health expenditure is out-of-pocket, and there is little or no organized mechanism to promote consumer information. Even though firm estimates are not available, it is very probable that market imper-
Table 4. Utilization of Outpatient Services and Annual Expenditures for Outpatient Care Per Capita

<table>
<thead>
<tr>
<th>Types of providers (percent of total utilization within variable)</th>
<th>Annual rate of utilization (in LE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ministry of Health and Population Government</td>
<td>National Government</td>
</tr>
</tbody>
</table>

### Region

<table>
<thead>
<tr>
<th>Category</th>
<th>Urban</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sample</td>
<td>3.47</td>
<td>0.76</td>
</tr>
<tr>
<td>1.0</td>
<td>0.85</td>
<td>0.96</td>
</tr>
<tr>
<td>2.0</td>
<td>1.74</td>
<td>1.82</td>
</tr>
<tr>
<td>3.0</td>
<td>2.62</td>
<td>2.79</td>
</tr>
<tr>
<td>4.0</td>
<td>3.50</td>
<td>3.67</td>
</tr>
<tr>
<td>5.0</td>
<td>4.38</td>
<td>4.55</td>
</tr>
</tbody>
</table>

### Income quintiles, from lowest to highest

<table>
<thead>
<tr>
<th>Income quintile</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual total expenditure (in LE)</td>
<td>2.21</td>
<td>2.84</td>
<td>3.43</td>
<td>3.76</td>
<td>5.10</td>
</tr>
<tr>
<td>39.83</td>
<td>64.03</td>
<td>84.57</td>
<td>110.83</td>
<td>237.34</td>
<td></td>
</tr>
<tr>
<td>39.4</td>
<td>28.2</td>
<td>22.8</td>
<td>17.6</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>10.8</td>
<td>10.2</td>
<td>13.7</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>7.7</td>
<td>7.2</td>
<td>9.0</td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td>38.9</td>
<td>44.7</td>
<td>49.7</td>
<td>47.4</td>
<td>59.6</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>2.7</td>
<td>3.2</td>
<td>3.4</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>21.20</td>
<td>10.40</td>
<td>8.37</td>
<td>9.70</td>
<td>6.30</td>
<td></td>
</tr>
</tbody>
</table>

### Gender

<table>
<thead>
<tr>
<th>Category</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sample</td>
<td>3.23</td>
<td>3.71</td>
</tr>
<tr>
<td>1.0</td>
<td>0.98</td>
<td>1.14</td>
</tr>
<tr>
<td>2.0</td>
<td>1.24</td>
<td>0.88</td>
</tr>
<tr>
<td>3.0</td>
<td>1.14</td>
<td>0.58</td>
</tr>
<tr>
<td>4.0</td>
<td>1.46</td>
<td>0.52</td>
</tr>
<tr>
<td>5.0</td>
<td>3.5</td>
<td>2.6</td>
</tr>
</tbody>
</table>

### Age

<table>
<thead>
<tr>
<th>Age</th>
<th>0-4</th>
<th>5-15</th>
<th>16-29</th>
<th>30-39</th>
<th>40-49</th>
<th>50-59</th>
<th>60+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sample</td>
<td>4.39</td>
<td>2.34</td>
<td>2.85</td>
<td>4.27</td>
<td>4.80</td>
<td>4.94</td>
<td>4.80</td>
</tr>
<tr>
<td>86.26</td>
<td>38.90</td>
<td>71.19</td>
<td>134.25</td>
<td>206.10</td>
<td>265.80</td>
<td>288.36</td>
<td></td>
</tr>
<tr>
<td>27.0</td>
<td>24.5</td>
<td>23.0</td>
<td>17.6</td>
<td>17.2</td>
<td>14.8</td>
<td>14.6</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>20.2</td>
<td>9.3</td>
<td>8.6</td>
<td>8.8</td>
<td>10.4</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>10.1</td>
<td>5.4</td>
<td>11.4</td>
<td>15.2</td>
<td>10.9</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>59.8</td>
<td>34.2</td>
<td>49.4</td>
<td>50.6</td>
<td>50.7</td>
<td>55.7</td>
<td>63.0</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>3.5</td>
<td>4.0</td>
<td>5.2</td>
<td>1.3</td>
<td>1.7</td>
<td>0.7</td>
<td></td>
</tr>
</tbody>
</table>

Infections in the form of consumer and physician ignorance are endemic in these health systems as well. For these reasons, prescriptions likely to work in Egypt should also be explored in other low- and middle-income countries as well.

**Conclusion**

The study described in this paper follows previous work by Polachek and Yoon (1987) and Gaynor and Polachek (1994), but extends it to a
developing country setting. The emphasis of the paper is to decom- 
pose the dispersion in prices into measures of incomplete patient and 
physician information. This is done by controlling for factors likely to 
affect health care prices, and then separating the random two-sided 
error and the skewed components. Data were obtained from a survey 
in Egypt, a country that has a very high physician-patient ratio. The 
empirical results support reasonable expectations, in that information 
imperfections among patients are found to be significantly greater 
than information imperfections among physicians. The differences in 
the two measures are smaller for low-severity, high-frequency items, 
such as general practice and gynecology consultations, but much 
higher for high-severity, low-frequency items such as surgery. The 
study suggests that, in view of the relative advantage that physicians 
enjoy over the average patient, governments seeking to involve the 
private sector in providing priority health care should explore options, 
such as setting physician fees in the private markets, increasing con-
sumer information, and otherwise ensuring consumer protection.

Although we make a case for improving consumer information in the 
market for physician services, we leave open the discussion of what 
would be the best method. For any instrument to facilitate intelligent 
consumer choice, it must contain accurate information that is critical to 
its users and that can be easily referred to by the users. In the United 
States, for instance, health plan report cards have gained wide popular-
ity, the motivation for compiling such information stemming from the 
belief that access to appropriate data will enable people to choose their 
health plans more effectively. Little is known, however, about whether 
consumers use this information at all in making plan and provider 
choices (Hibbard, Slovic, and Jewett, 1997). Future research on market 
imperfections in developing countries should also address these issues.

References

and Estimation of Stochastic Frontier Production Function Models.” Journal 
Chawla, Mukesh, A.K. Nanda Kumar, Peter Berman, Elaine Fleming, and 
Production between Solo and Group Practice Physicians.” Health Economics 


Public Transfers and Migrants' Remittances: Evidence from the Recent Armenian Experience

Edmundo Murrugarra

Abstract

This paper examines the link between remittances, public transfers, and health care utilization. First, the paper exploits exogenous variation in the coverage and intensity of the public welfare programs caused by policy changes to identify the crowding-out effects of public transfers on private transfers (remittances). In a second stage, the paper uses the instrumented remittances to identify its impact on health care demand. The variation of remittances is explained by the exogenous changes in Social Assistance transfers and by migrants'
characteristics (age, gender, education, and location of residence). It is found that for each Armenian dram (ADM) of Social Assistance transfers, remittances were reduced ADM 0.26–0.32.\(^1\) Although remittances and health care utilization are strongly related, probit-IV estimates indicate that exogenously determined remittances do not have a significant impact on health care use. This suggests that remittances may be responding to health care needs, rather than the reverse. Social Assistance transfers, by contrast, show significant positive effects on health care demand, but these effects cannot be distinguished from fee-waiver instruments in the health sector.

Understanding the role of private transfers among households in developing countries with significant migration patterns is crucial for several reasons. On the household side, although transfers may have a range of motives, remittances play a role in risk management, usually responding to adverse shocks affecting the recipient household. Remittances have risk mitigation and coping dimensions that need to be emphasized. On the public policy side, a better understanding of the interaction between public and private transfers is crucial to assess the expected impact of Social Assistance, such as targeted cash transfers, and social insurance interventions, such as pensions, unemployment insurance, and other benefits. If, for example, the government decides to provide an additional amount in a targeted transfer program, migrants may respond by reducing in the amount of remittances or simply stop sending them at all.

This paper examines the link between remittances, public transfers, and health care utilization. Because remittances and health care utilization are household responses, measuring the impact of remittances on medical care demand is contaminated by unobserved factors affecting both variables. The paper uses an instrumental variable approach to the problem. First, we estimate the effect of targeted public transfer programs on remittances to assess the extent of crowding-out effects. In this stage, the paper exploits exogenous variation in the coverage and intensity of the public welfare programs. In January 1999, Armenia's Social Assistance (SA) programs were reformed from a multibenefit and untargeted system to a single-benefit, poverty-targeted one. The average size of the transfers was increased as well. Using the cross-sectional variation in transfers, as well as variation over time because of the reform, the analysis mimics a difference-in-difference estimator for crowding-out effects. In the second stage, this paper uses the estimates from the

---

1. The exchange rate in December 1998 was ADM 518 per U.S. dollar.
remittances equation to provide an instrumental variable estimate of the impact of remittances on health care demand. The exogenous variation of remittances is explained by the effects of the changes in Social Assistance transfers and migrants' characteristics, such as age, gender, education, and location. Armenia is an interesting case for examining the impact of income sources on health care demand because it has a health care system that relies heavily on out-of-pocket expenditures. It is found that for each Armenian dram (ADM) of social assistance transfers, remittances were reduced ADM 0.25–0.32. Remittances are significantly related to health care use, but once endogeneity is controlled for, remittances do not have a precise impact on utilization rates. This suggests that remittances might be a response to health care needs rather than an incentive to seek health care. The positive effects of Social Assistance transfers on utilization are very strong, but they cannot be distinguished from fee-waiver instruments in the health sector.

The section titled The Armenian Experience details the recent Armenian changes in its Social Assistance system, and the data used in this study. The next section, Methodological Strategy, discusses the basic motivation for remittances and provides the empirical strategy. In the section titled Armenian Data and Results, the results are presented and discussed. The last section concludes the paper.

The Armenian Experience

This paper uses Armenian evidence to exploit three significant features observed in the areas of interest. First, migration and remittances are very extended phenomena in Armenia. On average, about 25 percent of household incomes come from remittances, either from relatives in Armenia or from abroad. About 12 percent of the population is identified as household members living abroad (State Institute of Statistics 1999). Evidence from a recent household survey indicates that the share of household incomes provided by remittances ranges between 20 and 30 percent, depending on the socioeconomic level of the household; poorer households rely more on remittances.

Second, the Armenian Government participates in a very limited way in the funding of the health system. In 1999 government expenditures on health accounted for only 1.4 percent of gross domestic product (GDP) or 5 percent of the government budget. This has caused a drastic deterioration in the quantity and quality of health services, while shifting most of the costs to patients. About 91 percent of those individuals that received inpatient care in Armenia made some informal payment,
much higher incidence than in Moldova (70 percent), Azerbaijan (78 percent), or the Russian Federation (74 percent) (Lewis 2000). As a result, a significant decrease in the number of patients has also been observed. In this context, the role of public and private income transfers is essential to understanding health care utilization decisions.

Finally, and most important, the social assistance system was reformed in 1999, providing a significant increase in the size of social assistance transfers. Before 1999 the Armenian Ministry of Social Protection provided 26 small, uncoordinated categorical benefits in cash. Eligibility for benefits comprised allowances and compensations based on individual and household demographic characteristics that were associated with “social risk groups” (Nahapetian and others 2001). The allocation was at the individual level. The most important groups were orphans (32,000), single mothers (23,000), disabled individuals (74,000), families with four or more children (99,000), and pensioners living alone or not working (58,000), among others. Other groups receiving benefits were identified by a broad range of indicators, such as area of residence or service in specific wars (see Appendix 1 for a full list of beneficiaries). The prereform social assistance system covered a large number of individuals (470,905 as of December 1998) and provided relatively low benefits (between ADM 1,000 and ADM 4,000 per individual).2

In January 1999, the old system was replaced by a poverty-targeted family cash transfer. The new poverty benefit is awarded on a household basis (not individual) and is significantly higher than any other transfer in Armenia (more than ADM 8,000 per recipient family on average). The reformed social assistance system introduced a means-tested targeting mechanism where households are ranked according to a single-index formula that includes individual and household indicators. Among these indicators are those used in the past, such as disabled individuals, but it also includes new household-level indicators that have strong correlates with poverty, such as ownership of a car or characteristics of the dwelling.3

2. The exchange rate in December 1998 was ADM 518 per U.S. dollar.

3. The reform of the social assistance system in Armenia represented a significant change in the distribution of transfers across socioeconomic groups. Despite the difficulties in identifying the poor for targeting purposes (World Bank 1999), the poorest 40 percent of the population increased its share of benefits received from 53 percent to 64 percent because the program was reformed. The concentration coefficient decreased from −0.22 to −0.30, which suggested a more progressive intervention (Sahakyan 2000).
Initially, more than 330,000 families were receiving the benefit (28 percent of the total number of families). Gradually, because of better screening and improved benefit administration, the number of recipient families was reduced to 217,220 by December 1999. The budget allocated to social assistance was increased from ADM 13.4 billion to ADM 21.1 billion. Comparing the benefits between the old and the reformed systems is difficult because administrative data is provided at the individual or household level for each case, as shown in table 1. The 57 percent increase in social assistance budget was not accompanied by a similar increase in benefits per individual (based on administrative data). Evidence from the 1998 and 1999 household survey, however, indicates a corresponding 41 percent increase in average household benefits (for those receiving benefits). It must be noted that the fraction of the population receiving social assistance benefits was almost unaffected, dropping only one percentage point. The reason for this is that the new poverty family benefit used most of the old individual indicators to construct the single-index score in the reformed system.

This exogenous variation caused by the reform in social assistance is the instrument used to estimate the impact of public transfers on remit-

**Table 1. Description of the Social Assistance Reform in Armenia**

<table>
<thead>
<tr>
<th></th>
<th>Prereform (December 1998)</th>
<th>Postreform (December 1999)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of beneficiaries</td>
<td>470,905 individuals(^a)</td>
<td>217,220 households</td>
</tr>
<tr>
<td>Annual budget (billion drams)</td>
<td>13.4</td>
<td>21.1</td>
</tr>
<tr>
<td>Average benefit (drams per month)</td>
<td>2,379 per beneficiary(^b)</td>
<td>8,095 per hhld. (2,300 per indiv.)</td>
</tr>
<tr>
<td>Eligibility mechanism</td>
<td>Individual allocation by social categories</td>
<td>Means-tested at household level</td>
</tr>
<tr>
<td>Average social assistance (SA) per beneficiary household (hhld)</td>
<td>5,463</td>
<td>7,713</td>
</tr>
<tr>
<td>Average SA per household</td>
<td>725</td>
<td>917</td>
</tr>
<tr>
<td>Households receiving SA</td>
<td>13.3%</td>
<td>11.9%</td>
</tr>
</tbody>
</table>

\(^a\) This is an upper bound because one individual may belong to more than one category, and hence be double-counted in the total. See the appendix for a detailed list of beneficiaries before the reform.

\(^b\) The benefits per individual are underestimated because the same individual may receive several benefits simultaneously.

*Source: Nahapetian and others (2001).*
tances. A similar quasi-experimental strategy has been used to examine the impact of the Child Benefits Reform in the United Kingdom on intra-household allocation outcomes, such as expenditures on children (Lundberg, Pollack, and Wales 1997). The reform of Social Assistance in Armenia does not allow the direct examination of intrahousehold allocation decisions because the delivery of the benefits is not strictly targeted to specific household members, such as wives or husbands.

**Methodological Strategy**

The impact of remittances has been examined in the economic literature, with attention being given to their determinants and the effects on household behavior and even on inequality. Remittances could be motivated by purely altruistic motives (improve relatives' well-being), exchange-based by rationales, or by the incentives of a mutually coinsurance contract (Lucas and Stark 1995). Depending on the motive, remittances may have different effects: they may raise household consumption regardless of income (if motivated by altruism); they may increase household investments (under self-interested or exchange motives); or they may appear as responses to adverse shocks (coinsurance contract). These effects on household living conditions may be substantial.

Most of the studies found support for some forms of exchange motives to remit. For example, Cox and Jimenez (1992) and Cox, Eser, and Jimenez (1998), using evidence from Peru, found that the exchange hypothesis was better able to explain the behavior of both the incidence of remittances and the amount of such transfers. In addition, they found that one additional inti (Peruvian currency) from pensions displaced between 0.29 and 0.42 inti from private transfers. Jensen (1997), using evidence from South Africa and implementing an instrumental variable strategy, found that 1 rand (South African currency) of Social Security transfers crowded out around 0.3 rands of private transfers.4

Turning to the impact of remittances on household behavior, previous studies have emphasized the impact among rural households, probably exploiting the rural-to-urban migration experiences from some countries.5 Evidence from rural areas, where credit or insurance

---

4. The uninstrumented estimates showed lower crowding-out effects.
5. Others have emphasized more aggregate effects on income inequality. See, for example, Stark and others (1986) and Adams (1989).
markets may not be working properly, indicates that remittances operate by loosening risk and credit constraints. These impacts are measured through productivity gains in agriculture in China (Rozelle, Taylor, and deBrauw 1999) and increased land assets in Pakistan (Adams 1998), among other measures. Concerning effects on human capital status, little evidence has been examined. Kanaiaupuni and Donato (1999) used evidence from migratory experience between Mexico and the United States to identify significant reductions in later mortality rates. Remittances, however, did not explain all the reduction in mortality rates. The migratory experience brought new lifestyles, health care behavior, information, and resources, and has had a long-term impact on mortality rates.

The empirical strategy to examine the crowding-out effects and the impact on health care utilization are described next.

Remittance equation. Remittances for household i are modeled as a latent variable \( R_i' \) that is linearly dependent on household income components \( I_i \), demographic characteristics of the household \( D_i \), characteristics of the migrant \( M_i \), and other unobserved components \( u_i \):

\[
R_i' = \beta_0 + \beta_1 I_i + \beta_2 D_i + \beta_3 M_i + u_i
\]  

(1)

The variable \( R_i' \), is not fully observed, however. A number of households received zero remittances, even if they reported migrants abroad. This suggests the use of a censored model, particularly one that allows specific characteristics (location of the migrant) to have different effects on the remittance decision and on the amount to be remitted. For example, the location of the migrant may affect her ability to send money home because of the availability of financial transfer mechanisms. Once remitting is feasible, the migrant location may not affect the amount to be sent.\(^6\) The different processes generating the remittance decision and the size of remittances motivate a generalized Tobit model (Heckman 1979). The estimating equation is

\[
R_i = \delta_0 + \delta_1 I_i + \delta_2 D_i + \delta_3 M_i + E(u_i \mid R_i > 0)
\]  

(2)

\(^6\) Except when different locations are unequally affected by economic or natural shocks, such as a drought.
which includes only those households that reported a positive amount of remittances \((R_i > 0)\). Under normality assumptions for \(u_i\), the model is estimated by maximum likelihood. The crowding-out problem is captured by the coefficients in \(\delta_j\), that is, the effect of income from Social Assistance on remittances. The variation in Social Assistance can be decomposed in cross-sectional variation and variation over time attributed to the reform. Denote now an indicator variable \(Reform_i\) that takes the value of one for those households surveyed during the “reformed” system and zero otherwise. Then, the income component \(\delta_1 I_i\) can be expanded as

\[
\delta_1 I_i = \Sigma_k \left( \theta_{1k} I_i^k + \theta_{2k} I_i^k \cdot Reform_i \right)
\]

where \(k\) denotes the income source \((k = \text{pension, Social Assistance, other income})\). Then, the impact of social assistance on remittances before the reform \((Reform_i = 0)\) is defined as \(\partial R_i / \partial I_i^{SA} \mid Reform = 0 = \theta_1^{SA}\), and using the cross-sectional variation after the reform, the impact is \(\partial R_i / \partial I_i^{SA} \mid Reform = 1 = \theta_1^{SA} + \theta_2^{SA}\). The difference between the two estimators provides a difference-in-difference estimator of the impact of social assistance on remittances, which is

\[
\left[ \partial R_i / \partial I_i^{SA} \mid Reform = 1 \right] - \left[ \partial R_i / \partial I_i^{SA} \mid Reform = 0 \right] = \theta_2^{SA}
\]

The estimation of \(\theta_2^{SA}\) will provide a measure of the crowding-out effect between public and private transfers attributed to the effects of the reform.

**Health care utilization equation.** In this paper we examine utilization of health care for those individuals that experienced illness during the previous four weeks. A number of studies have addressed the problem of estimating health care utilization as conditional on being sick (Gertler and van der Gaag 1990, Lavy and Quigley 1993, Dow 1996). Although the effects from these estimations are subject to a number of caveats because of the endogeneity of health status and the consequential selectivity biases, estimates based on conditional estimation strategies can still be interpreted as short-term effects on health care demand (Dow 1996). Moreover, the Armenian evidence demands a short-term interpretation of the results because we are comparing the effects exploiting short-term variation in public transfers.\(^7\)

---

7. Alternatively, if available longitudinal data were available, one could instrument remittances with past remittances or previous changes in remittances.
The decision to seek health care depends on characteristics of the affected individual \((X_{ij})\), such as age and gender; household income from different sources, such as other income, remittances, and Social Assistance \((I_i)\); and the household demographic composition \(D_i\). The unobserved health-seeking variable, \(H_i^*\), is modeled in a linear fashion, but we observe only the outcome \(H_i = 1\) if the person received health care, that is, if \(H_i^*\) is large enough. Otherwise, we observe \(H_i = 0\) if the sick individual did not receive attention. Then the estimating equation follows a dichotomous model, thus:

\[
E[H_i \mid X_{ij}, I_i, D_i] = P[H_i = 1 \mid X_{ij}, I_i, D_i] \\
= P[\alpha_0 + \alpha_1 X_{ij} + \alpha_2 I_i + \alpha_3 D_i + u_i > 0]
\]

which can be estimated imposing distributional assumptions on \(u_i\).

The next section provides estimates for equations (3), (4), and (5) and discusses the results.

**Armenian Data and Results**

The data used in this paper correspond to the Integrated Living Standard Survey carried out between July 1998 and June 1999, where 300 households were interviewed each month across different regions.\(^8\) The survey comprised 3,600 households, 2,180 of them corresponding to urban areas. Income sources and itemized expenditures were collected for both the previous 30 days and, in some cases, for the previous year. In this paper we use monthly information because (a) annual income would include incomes received before and after the reform, not allowing identification of the impact of the reform Social Assistance, and (b) Social Assistance and remittances were detailed only at the monthly level. A separate module on migration reports the characteristics of those members not present because of migration, such as gender, education, and current residence.

Table 2 describes the urban sample before and after the reform. This shows that no significant variation was observed among household characteristics. Household size is about four members, the average household head is around 54 years old, and about 30 percent of household heads are females. About 70 percent of household heads have had some secondary education (general or technical), and more than 20

---

\(^8\) Armenia is divided into 10 different regions, called "marz," and the capital city, Yerevan.
Table 2. Characteristics of Urban Households in Survey
(before and after the reform)

<table>
<thead>
<tr>
<th>Household head</th>
<th>Before</th>
<th></th>
<th>After</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Age</td>
<td>54.9</td>
<td>(13.5)</td>
<td>53.0</td>
<td>(13.5)</td>
</tr>
<tr>
<td>Primary education</td>
<td>0.071</td>
<td>(0.257)</td>
<td>0.048</td>
<td>(0.214)</td>
</tr>
<tr>
<td>Secondary education</td>
<td>0.457</td>
<td>(0.498)</td>
<td>0.454</td>
<td>(0.498)</td>
</tr>
<tr>
<td>Technical education</td>
<td>0.245</td>
<td>(0.430)</td>
<td>0.225</td>
<td>(0.418)</td>
</tr>
<tr>
<td>Higher education</td>
<td>0.218</td>
<td>(0.413)</td>
<td>0.265</td>
<td>(0.441)</td>
</tr>
<tr>
<td>Married/cohabiting</td>
<td>0.661</td>
<td>(0.474)</td>
<td>0.671</td>
<td>(0.470)</td>
</tr>
<tr>
<td>Single</td>
<td>0.039</td>
<td>(0.194)</td>
<td>0.045</td>
<td>(0.208)</td>
</tr>
<tr>
<td>Widow</td>
<td>0.245</td>
<td>(0.431)</td>
<td>0.224</td>
<td>(0.417)</td>
</tr>
<tr>
<td>Divorced</td>
<td>0.055</td>
<td>(0.227)</td>
<td>0.059</td>
<td>(0.236)</td>
</tr>
<tr>
<td>Female</td>
<td>0.311</td>
<td>(0.463)</td>
<td>0.294</td>
<td>(0.456)</td>
</tr>
<tr>
<td>Income components</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other income</td>
<td>29.65</td>
<td>(133.2)</td>
<td>25.57</td>
<td>(43.22)</td>
</tr>
<tr>
<td>Pension</td>
<td>2.752</td>
<td>(11.58)</td>
<td>2.123</td>
<td>(3.757)</td>
</tr>
<tr>
<td>Social assistance</td>
<td>0.839</td>
<td>(2.71)</td>
<td>1.158</td>
<td>(3.633)</td>
</tr>
<tr>
<td>Remittances</td>
<td>10.37</td>
<td>(49.74)</td>
<td>7.594</td>
<td>(26.45)</td>
</tr>
<tr>
<td>Total income</td>
<td>43.61</td>
<td>(141.7)</td>
<td>36.45</td>
<td>(49.45)</td>
</tr>
<tr>
<td>Demographic composition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction 0–5 years old</td>
<td>0.056</td>
<td>(0.117)</td>
<td>0.054</td>
<td>(0.115)</td>
</tr>
<tr>
<td>Fraction 6–12 years old</td>
<td>0.107</td>
<td>(0.162)</td>
<td>0.105</td>
<td>(0.160)</td>
</tr>
<tr>
<td>Fraction 13–18 years old</td>
<td>0.117</td>
<td>(0.176)</td>
<td>0.122</td>
<td>(0.179)</td>
</tr>
<tr>
<td>Fraction 19–25 years old</td>
<td>0.100</td>
<td>(0.166)</td>
<td>0.133</td>
<td>(0.196)</td>
</tr>
<tr>
<td>Fraction 26–45 years old</td>
<td>0.277</td>
<td>(0.216)</td>
<td>0.285</td>
<td>(0.218)</td>
</tr>
<tr>
<td>Fraction 46–64 years old</td>
<td>0.243</td>
<td>(0.291)</td>
<td>0.238</td>
<td>(0.277)</td>
</tr>
<tr>
<td>Fraction 65+ years old</td>
<td>0.161</td>
<td>(0.294)</td>
<td>0.125</td>
<td>(0.260)</td>
</tr>
<tr>
<td>Household size</td>
<td>3.958</td>
<td>(1.899)</td>
<td>4.031</td>
<td>(1.820)</td>
</tr>
<tr>
<td>Fraction of migrants</td>
<td>0.116</td>
<td>(0.399)</td>
<td>0.103</td>
<td>(0.370)</td>
</tr>
<tr>
<td>Sample size (no. of hhlds)</td>
<td>1,100</td>
<td></td>
<td>1,080</td>
<td></td>
</tr>
</tbody>
</table>

percent pursued higher education. More than 20 percent of their income comes from remittances, but despite the small total income reduction between the two samples, remittances decreased 27 percent after the reform. A parallel 38 percent increase is observed in average Social Assistance, but it is not statistically significant.

Two questions need to be examined empirically. First, is this increase in Social Assistance transfers displacing remittances? Second, to what extent are different income components—such as remittances or Social Assistance—affecting health care utilization?
Public and Private Crowding-Out in Transfers

The first question is addressed by estimating the impact of public transfers on remittances. Previous estimates in the literature have exploited cross-sectional variation to identify this effect (Cox and Jimenez 1992; Cox, Eser, and Jimenez 1998; Jensen 1997). Here we provide estimates from cross-sectional variation, as well as those exploiting the variation caused by the reform.

The income components included in the regression are Social Assistance, pension, and other income. Pension and Social Assistance are specified as equation (3), and a dummy for those households receiving each transfer was included to account for different marginal effects around zero transfers. Demographic composition of the household is summarized by the logarithm of household size and the fraction of different age groups. Household head characteristics, such as age, education level, marital status, and gender, are also included. Finally, other identifying instruments are migrant characteristics. These include the number of migrants, gender (female dummy), education (dummies for technical and higher education), location of the migrant (Russia and other countries, with Armenia as the omitted variable). Regional and monthly dummies were included to control for unobserved differences in income patterns across regions and the existing seasonality in many of the income components.9

Before examining the crowding-out effects, let us review other patterns that emerge from the regressions summarized in table 3. The age of the household head indicate that older heads (over 60) are more likely to receive remittances, consistent with models where migrant children remit to their parents. Those more educated households (higher-educated heads) receive more remittances than the less educated ones (primary-educated heads). This evidence has been interpreted under the exchange explanation for remittances: children from more educated parents tend to be more educated as well, so remittances may be a payback at later ages.

The characteristics of migrants offer interesting insights about remittances. The number of migrants increases the amount of transfers for those who receive, suggesting a pure scale effect. Migrants living in Russia are more likely to remit, and to remit more than those living in other areas. This corroborates anecdotal evidence indicating that

---

9. The importance of seasonality factors in different income components and the need to explicitly mention its correction was suggested by a referee.
### TABLE 3. REMITTANCES EQUATION—GENERALIZED TOBIT ESTIMATION

(urban sample only; standard error in parentheses)

<table>
<thead>
<tr>
<th>Variables</th>
<th>$P[R_i &gt; 0]$</th>
<th>$R_i \mid R_i &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>St. Error</td>
</tr>
<tr>
<td>Other income</td>
<td>-0.0054</td>
<td>(0.0016)</td>
</tr>
<tr>
<td>Receive pension?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pension amount</td>
<td>-0.0005</td>
<td>(0.0026)</td>
</tr>
<tr>
<td>Pension reform*</td>
<td>-0.0439</td>
<td>(0.0140)*</td>
</tr>
<tr>
<td>Receive Social Assistance (SA)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA amount</td>
<td>-0.0176</td>
<td>(0.0203)</td>
</tr>
<tr>
<td>SA reform*</td>
<td>0.0102</td>
<td>(0.0239)</td>
</tr>
<tr>
<td>Reform</td>
<td>0.0912</td>
<td>(0.1290)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0641</td>
<td>(0.0172)*</td>
</tr>
<tr>
<td>Age 2</td>
<td>0.0005</td>
<td>(0.0001)*</td>
</tr>
<tr>
<td>Primary education</td>
<td>0.2041</td>
<td>(0.2341)</td>
</tr>
<tr>
<td>Technical education</td>
<td>0.0026</td>
<td>(0.0987)</td>
</tr>
<tr>
<td>Higher education</td>
<td>0.0893</td>
<td>(0.0627)</td>
</tr>
<tr>
<td>Single</td>
<td>-0.1975</td>
<td>(0.2271)</td>
</tr>
<tr>
<td>Widow</td>
<td>0.0154</td>
<td>(0.1802)</td>
</tr>
<tr>
<td>Divorced</td>
<td>0.2006</td>
<td>(0.1408)</td>
</tr>
<tr>
<td>Female</td>
<td>0.0443</td>
<td>(0.0985)</td>
</tr>
<tr>
<td>Fraction 0–5 years old</td>
<td>-0.0290</td>
<td>(0.5183)</td>
</tr>
<tr>
<td>Fraction 6–12 years old</td>
<td>-0.2908</td>
<td>(0.1835)</td>
</tr>
<tr>
<td>Fraction 13–18 years old</td>
<td>0.1763</td>
<td>(0.3532)</td>
</tr>
<tr>
<td>Fraction 19–25 years old</td>
<td>0.0440</td>
<td>(0.1833)</td>
</tr>
<tr>
<td>Fraction 46–64 years old</td>
<td>-0.0938</td>
<td>(0.1249)</td>
</tr>
<tr>
<td>Fraction 65+ years old</td>
<td>-0.2842</td>
<td>(0.1668)*</td>
</tr>
<tr>
<td>Log (household size)</td>
<td>-0.2503</td>
<td>(0.1538)*</td>
</tr>
<tr>
<td>Number of migrants</td>
<td>-0.1924</td>
<td>(0.1248)</td>
</tr>
<tr>
<td>Male</td>
<td>0.1368</td>
<td>(0.1904)</td>
</tr>
<tr>
<td>Age</td>
<td>0.0076</td>
<td>(0.0141)</td>
</tr>
<tr>
<td>Technical education</td>
<td>0.7326</td>
<td>(0.2881)*</td>
</tr>
<tr>
<td>Higher education</td>
<td>0.0643</td>
<td>(0.2402)</td>
</tr>
<tr>
<td>Russia</td>
<td>0.6066</td>
<td>(0.1935)*</td>
</tr>
<tr>
<td>Other places</td>
<td>-0.2684</td>
<td>(0.1355)*</td>
</tr>
<tr>
<td>$p$</td>
<td>0.1576</td>
<td>(0.1105)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.1133</td>
<td>(0.0394)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.1755</td>
<td>(0.1223)</td>
</tr>
</tbody>
</table>

Sample size 1280 400

Log-likelihood -1560.75

* Indicates that it is significant at the 10% level.

Note: Indicators and variables for regions (marz) and months were included.
Russia and other members of the Community of Independent States (CIS) are the major migratory objective for income purposes.

Now we examine the effects of different income components. The estimates from table 3 are transformed into estimated impacts of different public transfers (pension and Social Assistance) on remittances because the log-linear specification does not allow a direct estimate of the effect ($\theta_2^{SA}$). Table 4 shows the effect of an additional dram from public transfers on remittances for different subsamples.

The first horizontal panel describes the effects on the average household, so they do not differ across columns. The effects, however, can be estimated at the average transfers before and after the reform and as the difference between the two (equation 4). The crowding-out effect from each cross-section represents a decline of between 0.69 and 0.74 in remittances per dram received in transfers. The effect from pensions is lower, only about 0.33 to 0.39 displacement of remittances per dram in pensions. These effects, however, measured as differences between the two samples, suggests minor effects on the average household (about 0.05 displacement). The low displacement found is consistent with relatively low coverage of these transfers: about 12 percent receive Social Assistance and 35 percent receive pensions.

An alternative strategy is to measure the effects on those households that receive these transfers. These results are displayed in the second panel, where the impact on the treated are shown. The first two columns show the effects on those households that receive Social Assistance. Before the reform, the displacement effect of Social Assistance (SA) was only 0.14 per dram, and increased to 0.46 after the reform. Our difference-in-difference estimator provides a crowding-out effect of 0.32, significantly different from zero.

The impact of pensions on this subsample is much smaller and not significant, consistent with an unaltered pension system. If the impact of pensions is measured only for those households that receive pensions, the effects are larger (column 4). Displacement of remittances are 0.32 and 0.50 in each period, accounting for a difference-in-difference displacement estimate of 0.18, which is not significantly different from zero. Finally, if these effects are examined within those households that receive both transfers (columns 5 and 6), the difference-in-difference estimator is -0.26 for Social Assistance (significantly different from zero) compared with -0.17 for pensions (not significant).

In summary, if cross-sectional variation of the data is used to estimate crowding-out effects, we would be getting relatively larger effects in most cases. When exploiting the exogenous change caused
TABLE 4. ESTIMATED IMPACT OF PUBLIC TRANSFERS ON REMITTANCES
(estimates evaluated at the mean of different subsamples)

<table>
<thead>
<tr>
<th>Sample Transfer</th>
<th>Receive social assistance</th>
<th>Receive both transfers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SA</td>
<td>Pension</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before</td>
<td>-0.6885</td>
<td>-0.3263</td>
</tr>
<tr>
<td>After</td>
<td>-0.7427</td>
<td>-0.3923</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.0542</td>
<td>-0.0660</td>
</tr>
<tr>
<td>Treated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before</td>
<td>-0.1419</td>
<td>-0.3256</td>
</tr>
<tr>
<td>After</td>
<td>-0.4645</td>
<td>-0.4085</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.3227</td>
<td>-0.0830</td>
</tr>
<tr>
<td></td>
<td>(0.0774)</td>
<td>(0.3017)</td>
</tr>
</tbody>
</table>

Note: These estimates present combinations of $\theta_i^{SA}$ and $\theta_i^{P}$ at different means of SA and P, respectively.

by the reform, a precise displacement effect between 0.26 and 0.32 is obtained. The Social Assistance difference-in-difference effects are precisely estimated and larger than those of pensions. This is partly explained because the pension system (eligibility, benefits) was not changed between 1998 and 1999. In addition, the different size of the effects of Social Assistance and pensions may be explained by the differences in benefit eligibility and the underlying decisionmaking on different income components. Although only households with elderly members received pensions, Social Assistance benefits had a broader eligibility, which may represent different preferences and needs from the corresponding households. The first column (Receive Social Assistance) should provide a partial answer to this problem: the difference-in-difference estimates of the effects of pensions are still lower than that of Social Assistance, indicating that for the same households, the effects are different. An alternative explanation is based on the different decisionmaker attributed to each income component. For instance, the elderly receiving pensions may have an important word to say about the destination or use of such pension incomes. Social Assistance benefits, however, were based on specific household member characteristics, but their distribution was not tied to those members (for example, benefits were provided if a single mother was pres-
ent, but not given to the corresponding individual). This difference may also explain the difference in the estimated effects of different income sources.  

**Impact on Health Care Demand**

The estimation of equation (5) for those individuals that are sick required additional controls. Given that the survey was carried out during 12 months, weather conditions may have altered the probability of being sick and the severity of the sickness, depending on the month.  

Because detailed information on weather indicators by regions (marz) by month was lacking, we included controls for months (shown in table 5) and marz (not shown).

Table 5 shows the estimates for equation (5) when remittances are taken as exogenous (column 1) and when remittances are treated as endogenous and properly instrumented using the results described above (column 2). The results indicate that, conditional on income, higher educated individuals are more likely to seek health care. Education of individuals certainly affects the desire to seek health care, but also affects the ability to identify an illness. This is precisely the endogeneity problem mentioned in the discussion on conditional health demand. However, we should interpret these results as the direct impact of education on health care utilization. Otherwise, increased education may also affect individuals’ health care behavior and eventually reduce their probability of being sick. The fraction of children aged 0-5—conditional on household size—has negative effects on health care utilization, which suggests that children may represent a competing demand for resources.

---

10. The discussion about the lack of fungibility of income was motivated by an anonymous referee.

11. The Armenian epidemiological profile indicates that about 50 percent of the first diagnoses are classified as respiratory diseases, excluding pulmonary cancer (Ministry of Health 1900).
### Table 5. Health Care Utilization

<table>
<thead>
<tr>
<th></th>
<th>Probit (1)</th>
<th>Probit-IV(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>s.e.</td>
</tr>
<tr>
<td>Remittances</td>
<td>0.0005</td>
<td>(0.0003)*</td>
</tr>
<tr>
<td>Other income</td>
<td>0.0009</td>
<td>(0.0004)*</td>
</tr>
<tr>
<td>Social assistance</td>
<td>0.0080</td>
<td>(0.0027)*</td>
</tr>
<tr>
<td>Age</td>
<td>0.0001</td>
<td>(0.0011)</td>
</tr>
<tr>
<td>Technical education</td>
<td>0.0322</td>
<td>(0.0252)</td>
</tr>
<tr>
<td>Higher education</td>
<td>0.1056</td>
<td>(0.0219)*</td>
</tr>
<tr>
<td>Fraction 0-5 years old</td>
<td>-0.4059</td>
<td>(0.1890)*</td>
</tr>
<tr>
<td>Fraction 6-18 years old</td>
<td>-0.0815</td>
<td>(0.0843)</td>
</tr>
<tr>
<td>Fraction 46-64 years old</td>
<td>-0.0930</td>
<td>(0.0601)</td>
</tr>
<tr>
<td>Fraction 65+ years old</td>
<td>-0.0419</td>
<td>(0.1056)</td>
</tr>
<tr>
<td>Log (household size)</td>
<td>-0.0201</td>
<td>(0.0235)</td>
</tr>
<tr>
<td>August</td>
<td>-0.0674</td>
<td>(0.0398)*</td>
</tr>
<tr>
<td>September</td>
<td>-0.0506</td>
<td>(0.0400)</td>
</tr>
<tr>
<td>October</td>
<td>-0.0409</td>
<td>(0.0494)</td>
</tr>
<tr>
<td>November</td>
<td>0.1523</td>
<td>(0.0371)*</td>
</tr>
<tr>
<td>December</td>
<td>0.0250</td>
<td>(0.0552)</td>
</tr>
<tr>
<td>January</td>
<td>0.0858</td>
<td>(0.0446)*</td>
</tr>
<tr>
<td>February</td>
<td>0.1373</td>
<td>(0.0523)*</td>
</tr>
<tr>
<td>March</td>
<td>0.1078</td>
<td>(0.0617)*</td>
</tr>
<tr>
<td>April</td>
<td>0.1431</td>
<td>(0.0823)*</td>
</tr>
<tr>
<td>May</td>
<td>0.0289</td>
<td>(0.0443)</td>
</tr>
<tr>
<td>June</td>
<td>-0.0136</td>
<td>(0.0417)</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.05190</td>
<td>0.0503</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1,147</td>
<td>1,147</td>
</tr>
<tr>
<td>Wald $\chi^2(9)$</td>
<td>89.13</td>
<td>87.94</td>
</tr>
<tr>
<td>Prob &gt; $\chi^2$</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-726.666</td>
<td>-727.91</td>
</tr>
</tbody>
</table>

* Indicates significant at the 10% level.

Note: Probit estimates for urban areas. The column Coeff. shows marginal effects $dP/dX$. Standard errors in parentheses. The sample comprises all individuals living in urban areas aged 30 or more.

The seasonal pattern is as expected: the probability of seeking health care is higher during the winter months (January–March), which suggests that the severity of illnesses is worse during those months.\(^{12}\)

The results of examining the effects of remittances on health care utilization are described in table 6. Assuming that remittances are

\(^{12}\) Data on monthly temperatures by marz will be provided soon, allowing for direct control of the seasonal pattern in health status and health care demand.
Table 6. Income Effects on Health Care Utilizations

Table 6 presents the percentage point changes in health care utilization per ADM 1,000 for remittances, other income, and Social Assistance, using Probit and Probit-IV models. The results are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Probit</th>
<th>Probit-IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remittances</td>
<td>0.05</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(4.53)</td>
</tr>
<tr>
<td>Other income</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Social Assistance</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.28)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses.

Source: Table 5.

Exogenous, ADM 1,000 in remittances (equivalent to $2) increases utilization rates in a very small, but significant, increase. Once remittances are instrumented, the impact becomes insignificant. This suggests that if remittances are explained by household characteristics (not including health-related variables) and migrant characteristics, an (exogenous) increase in remittances may not necessarily be accompanied by increases in health care utilization. The stronger connection observed in column (1) suggests that remittances may be responding to health care needs, thereby generating the usual endogeneity problem. These results must be interpreted with care because its power hinges on the identification restriction imposed in the paper. Variables, such as migrant characteristics, may also play a role in the health care utilization equation. One valid reason for such inclusion could be the effect of migration on lifestyle and health care behavior, as found in other cases (Kanaiaupuni and others 1999). The argument for suggesting these as proper estimates relies on the short-term interpretation of the estimates, which avoids the long-term effects of migration on health care and lifestyle.13

The impact of other income and Social Assistance is larger and precisely estimated. An additional ADM 1,000 in other income will marginally increase utilization (about 0.09 percentage points). The same ADM 1,000 increase in Social Assistance transfer will represent almost a 1 percentage point increase in utilization rates among those sick. Comparing those households that receive Social Assistance benefits (about ADM 8,000) with similar ones not receiving the benefits, the

---

13. This caveat was emphasized by a referee, which pointed to the potential weak identification of the model.
beneficiaries would have utilization rates that are 6 points higher that the comparison household.

The differential impact of different income sources on different uses has been examined with respect to consumption. Using Russian data, Richter (2000) found that the propensity to consume is higher from regular than from transitory income, and higher from pensions than from child benefits. The Armenian evidence shows a significantly higher impact of Social Assistance rather than that of other income sources. To the extent that income can be spent on either consumption or investments (in human or physical capital), the results from the Armenian exercise support the notion that poverty family benefits are directed toward investments in human capital rather than direct consumption.

The estimates above are subject to a number of caveats. First, the Armenian health sector also provides a basic package free of charge to selected households. The selection of the beneficiaries in the health sector overlaps with some categories described in Social Assistance. Then the large income effect attributed to Social Assistance may be in part explained by the fact that some social assistance beneficiaries may also be eligible for free health services (which will have a price effect). Distinguishing income effects from price effects on health care utilization is crucial for the design of policies to cover the poor. The estimates described above cannot separate them. Second, there might be household characteristics that affect remittance and migration simultaneously. Members from specific households might be “more likely to migrate” and to remit because of unobserved household characteristics. For instance, relatives or friends who migrated a long time before could be affecting both migrants’ characteristics (location) and remittances simultaneously, in cases where migrants went to their relatives and friends abroad. In this context, migrants’ characteristics are correlated with unobserved components. Ideally, a longitudinal data set would have enabled us to control directly for those unobserved household components. In this paper, we have addressed the problem by using old migrants’ history to explain the short-term variations in remittances. Third, the instruments for remittances (particularly, migrants’ characteristics) may not be valid instruments in the health care utilization equation. This would be the case if migratory experiences affected health status and health care utilization beyond the effects that operate through remittances. This could be the case when migratory experiences also affect the lifestyle and health-related behavior of Armenian households, as was found by Kanaiaupuni and others (1999) among the families of Mexican migrants. Finally, the
effects found in this paper could vary by type of health provider. Although the effects of remittances on aggregate health care are imprecise, those effects could be significant for specific types of interventions, most likely those demanding large expenditures on hospitalization or specialized treatments. These caveats pose additional questions to be addressed with more detailed data to understand the links between public interventions, private risk-coping responses, and human development outcomes.

Conclusions

This paper examined the link between remittances, Social Assistance transfers, and health care utilization. The finding that Social Assistance government transfers lead to significant displacement of private remittances underscores the fungibility of money at the household level. The design of Social Assistance transfers should take into account the potential displacement of private transfers.

A second finding is that improving remittance mechanisms (such as reducing the financial costs of sending money) may not have an immediate impact on health care utilization, but may simply loosen the financial constraints for those already seeking health care. In other words, remittances represent an alternative safety net, or risk-coping mechanism. Use of this mechanism has some costs, both fixed (travel and migration costs) and variable (financial cost of remitting). However, the impact of additional remittances caused by incentives to remit is likely to provide beneficial effects through income effects on health status, lifestyle, and investments in other assets (such as durable goods or education). Although the impact of remittances on health care utilization was not found to be significant, it might be the case that remittances play a more important role in seeking specific types of health care that are more expensive. This analysis will be pursued in the research agenda. Similarly, the analysis should be extended to examine the effects on other human capital investments (for example, education).

Understanding the links between private and public transfers, and between those and other public interventions, such as health and education, will help the design of policies that exploit the externalities across public interventions.
### Appendix. Social Assistance Benefits in December 1998

<table>
<thead>
<tr>
<th>No.</th>
<th>Categories of citizens receiving allowances and compensation</th>
<th>Number of recipients</th>
<th>Amount per month (drams)</th>
<th>Total (thousand drams per month)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disabled</td>
<td>7,000</td>
<td>4,000</td>
<td>28,000</td>
</tr>
<tr>
<td>2</td>
<td>Single mothers</td>
<td>23,000</td>
<td>2,000</td>
<td>46,000</td>
</tr>
<tr>
<td>3</td>
<td>Orphans (unilateral/bilateral)</td>
<td>32,000</td>
<td>2,000-3,000</td>
<td>70,000</td>
</tr>
<tr>
<td>4</td>
<td>Recipients of alimony</td>
<td>10,600</td>
<td>2,000</td>
<td>21,200</td>
</tr>
<tr>
<td>5</td>
<td>Enlisted men (privates, corporals, sergeants)</td>
<td>850</td>
<td>2,000</td>
<td>1,700</td>
</tr>
<tr>
<td>6</td>
<td>1st and 2nd degree disabled</td>
<td>15,900</td>
<td>2,000</td>
<td>36,800</td>
</tr>
<tr>
<td>7</td>
<td>Refugees residing in temporary dwelling</td>
<td>4,000</td>
<td>2,000</td>
<td>8,000</td>
</tr>
<tr>
<td>8</td>
<td>Families residing in earthquake area container housing</td>
<td>41,000</td>
<td>2,000</td>
<td>82,000</td>
</tr>
<tr>
<td>9</td>
<td>Families with 4 or more minors</td>
<td>99,000</td>
<td>2,000</td>
<td>198,000</td>
</tr>
<tr>
<td>10</td>
<td>Families with 3 or more minors residing in Gumry, Spitak, and Vanadzor</td>
<td>28,000</td>
<td>2,000</td>
<td>56,000</td>
</tr>
<tr>
<td>11</td>
<td>Single mothers and persons with children under 2 yrs</td>
<td>16,700</td>
<td>1,800</td>
<td>30,060</td>
</tr>
<tr>
<td>12</td>
<td>On partially paid vacation</td>
<td>12,400</td>
<td>1,800</td>
<td>22,320</td>
</tr>
<tr>
<td>13</td>
<td>With a status of unemployed</td>
<td>8,000</td>
<td>1,800</td>
<td>14,400</td>
</tr>
<tr>
<td></td>
<td>Compensation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1st and 2nd degree disabled (except World War II)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Military personnel disabled during World War II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Servicemen (or other equal categories), who became disabled defending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Armenia during military service or after retirement because of injuries,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>disorders, and diseases, as well as the servicemen’s family (spouses,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>children, parents) or other equal categories who died in the line of duty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Relatives of military personnel killed in action.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Military heroes and other distinctions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Forced resettlers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Citizens disabled and deceased because of the Chernobyl disaster,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>and their relatives.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Relatives of citizens in Nagorno-Karabakh</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Families consisting of not-working lonely pensioners or only of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>not-working pensioners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Invalids since childhood</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Personal pensioners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Guides for 1st degree blind invalids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Deaf-mute invalids (over 16 yrs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>470,905</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Nahapetian and others (2001).
References


Part II

Communities and Welfare
Better a Hundred Friends
Than a Hundred Rubles?
Social Networks in Transition—
The Kyrgyz Republic

Kathleen Kuehnast and Nora Dudwick

Abstract

The social networks of poor and nonpoor households in the post-Soviet Kyrgyz Republic have polarized and separated, in a process that parallels the sharp socio-economic stratification that has taken place since national independence in 1991. Not only have the networks separated, each has changed in character. The non-poor, especially those in urban communities, are moving away from relationships based on ascriptive relationships to more “modern,” interest-based networks.
which they successfully exploit to access an expanding array of resources. By contrast, the shrinking networks of the poor have reduced their access to decent health care, good education, and timely social assistance, services that are increasingly mediated by personal “connections.” Given that person-centered social networks still predominate in Kyrgyz society, the deteriorating networks of the poor should be of serious concern to policymakers. Their deterioration signals that an escalating process of social exclusion is now under way.

We have found it difficult to comprehend the politics of survival in economies that are dominated by non-market forces and that reward blat, stability, conformity, and material equality rather than work, risk, creativity, and personal achievements. Because we live in consumer-oriented societies where virtually all goods and services are available to those who have the money to pay for them, we have brought too many Western economic, social, and psychological assumptions to our analyses of Communist systems.

Fleron and Hoffman (1993), p. 174

Ten years ago, the question of whether a hundred friends are better than a hundred rubles in postcommunist Kyrgyz Republic would have been largely rhetorical. In keeping with the sense of this proverb, answers would most likely have confirmed the superior importance of “connections” over cash. Today, however, answers to this question are no longer so predictable. For along with Kyrgyz society as a whole, the scope and function of social networks have been undergoing a dramatic transformation. Because social networks of the poor and nonpoor were moving along different and contrasting trajectories, it became clear that they could best be studied in relation to each other. This study, then, seeks to expand the ways in which issues of the poor have traditionally been studied. It argues that the poor cannot be studied in isolation from the nonpoor, nor can solutions to poverty be devised for them alone. Rather, the success of poverty alleviation in countries in transition depends in equal measure on understanding the emerging nonpoor and their relations to and attitudes toward the new poor.

This study provides a unique vantage point from which to consider these relationships on a continuum between poor and nonpoor, as the Kyrgyz Republic moves on the path from a centralized, planned economy to a market economy. In most developing countries of the world, extreme poverty has been a fact of life for generations. By contrast,
widespread and severe poverty was new to all but the oldest of our respondents. This study thus focuses on a moment in history when rapid impoverishment has polarized the social networks of the poor and nonpoor, in order to capture the dynamics of how the poor both disengage from and are isolated by and from the nonpoor.

The study of social networks in postsocialist countries is an important tool for bridging the policy gap between macro-level economic strategies and micro-level interventions. These networks provide an essential framework for understanding how informal institutions interact with formal institutions in the postsocialist Kyrgyz Republic. The role of social networks in a society and economy in transition has important implications for institutional reform at every level. Informal networks are not only “safety nets”; they are also institutions that can undermine or sabotage apparently well-designed programs intended to target the poor or marginalized. Qualitative poverty studies conducted in the countries of the former Soviet Union (FSU) have found, for example, that the very poorest lack “insider connections” to formal institutions and are therefore most likely to be excluded from formal assistance.¹ In her study on social networks in Cairo, Diane Singerman (1995, p. 133) found that “networks are the political lifeline of the community, allowing individuals and groups to cooperate with other members of the community to achieve individual and collective goals.” A better understanding of the complex relationships between local networks and formal state and international institutions can also yield important insights into derailed reform projects and patterns of corruption (see Stark and Kemeny 1997).

Recent poverty assessments in Central Asia have not fully examined the ways in which access to information and goods depends on social networks (see Dudwick, Gomart, and Marc forthcoming). Understanding how social networks can enhance or restrict peoples’ access to limited resources is particularly important in view of growing economic stratification and the increase in structural poverty throughout the region. Despite the introduction of market principles and the gradual depersonalization of economic relations in the Kyrgyz

¹. These include studies by Kuehnast (Kyrgyz Republic); Dudwick (Armenia and Georgia); De Soto and Dudwick (Moldova); Gomart (Armenia); Wanner and Dudwick (Ukraine); and the Institute for Sociology and Philosophy, Riga, with Dudwick (Latvia), in Dudwick, Gomart, and Marc (forthcoming).
Republic, social networks in the transition period remain as important to survival and social mobility as they were during the Soviet-era "shortage" economy.²

Social relations in Kyrgyz society are based upon person-centered social networks. Thus, in the Kyrgyz Republic, as well as elsewhere in Central Asia, gift-giving and other forms of reciprocity are essential to social life, especially for cultivating, maintaining, and expanding networks important for security and social mobility. During a time when both the state and the market are unreliable, the gift exchange networks still provide social support, personal financing, and mutual assistance in Kyrgyz society. Life-cycle celebrations and rituals that serve as the venue of relationship building usually invigorate such networks. From a study in rural India, Vijayendra Rao (2001) noted that exchange networks, in addition to their central role in helping the poor to cope during difficult times, serve the nonpoor as arenas for status-enhancing competitions. Likewise, in rural communities in the Kyrgyz Republic where banks and other services are unavailable, social networks fueled by these traditions are more valuable than goods or money. Although these networks operate without guidebooks or formal regulations, they can be considered institutions in that they pattern recurrent transactions and exact social consequences for failure to honor agreements.³

Networks vary in composition and form, from horizontal or flat networks that link equals or near-equals to "vertical" networks—including patron-client relationships—that hierarchically link people with unequal power and access to resources. Previous studies of informal social networks in Central Asia found that interhousehold transfers were an important safety mechanism for the poor.⁴ Today, a new reality is emerging. The poor are being excluded or are withdrawing from those social networks that once offered important support. In response to this trend, the nonpoor indicated in interviews that they are less likely to sustain their relation-

---

² In contrast to some of its Central Asian neighbors, the Kyrgyz Republic adopted an aggressive strategy of market reform, including widespread privatization, the introduction of a new currency in 1993, and other macroeconomic reforms unique to the region.

³ For further discussion of informal social networks in postsocialist states, see Rose (1999, p. 3).

ships with poorer relatives because these relationships are financially draining. This attitude toward the obligation to support extended family members is a major shift in the previous family-centered informal welfare system of the Kyrgyz. Consequently, although the networks of the poor are shrinking and becoming more homogeneous, networks of the nonpoor are expanding and diversifying. These changes parallel the growing chasm between the networks of the poor and of the nonpoor, bridged, if at all, by patron-client relationships. Given that person-centered social networks still predominate in Kyrgyz society, the deteriorating networks of the poor should be of serious concern to policymakers. Their deterioration signals that an escalating process of social exclusion is now under way. Yet it is not enough to understand the networks of the poor. A thorough analysis of the networks of the nonpoor is also critical for understanding how the entire society operates through these informal systems, how formal institutions are brought into the web of personal networks, and how uneven the playing field has become in the new “market economies” of the FSU.

The focus of this study on social networks also places it within the purview of social capital research. Discussions of social capital theory distinguish two major approaches to this phenomenon. The first, rooted in the concepts developed by the French sociologist, Pierre Bourdieu, and the U.S. sociologist James Coleman, considers individuals and small groups the unit of analysis, and focuses on the ways in which they manipulate social relationships to gain benefits. The other major approach, of greater interest to development specialists, emerged from the work of the political scientist Robert Putnam, who investigated social capital as an attribute of communities and, in some cases, of nations (1993). Putnam and others argue that social capital arises through a dense associational life that produces norms of generalized trust and reciprocity within a community. The level or “stock” of social capital partially determines why some communities are more able than others to mobilize to pursue shared objectives. These two applications of the social capital concept are quite distinct, and as

5. For further discussion of social stratification in Kyrgyz society, see Mikhalev and Heinrich (1999).
Alejandro Portes (Portes and Landolt 2000, p. 535) suggests, can be contradictory. In the case where individuals or small groups use their connections to bend regulations and gain access to public resources, for example, "individual social capital in such instances consists precisely in the ability to undermine collective social capital, defined as 'civic spirit.'" Portes also points to a confusion between social capital as cause and effect, when in fact high levels of community solidarity might accompany economic growth because both have been shaped by an external factor. We consider this study more in line with social capital theories that consider social capital as a dependent rather than independent variable, and that consider norms and values as separate, albeit related, issues. Interestingly, the only other study of social capital in post-Soviet society undertaken through the World Bank's Social Capital Initiative, Richard Rose's study of social capital networks in Russia, likewise uses this more restrictive definition of social capital to examine networks in post-socialist Russia (Rose 1998). Finally, this study particularly stresses the context-dependent nature of social capital. As Foley and Edwards have argued, the way in which networks are embedded in broader socioeconomic contexts and can link individuals to resources determines whether or not a social network has social capital. In their terms, then, "social capital = resources + access" (see Foley and Edwards 1999). We also support those theorists who argue that norms of trust and reciprocity are more usefully considered separately from social capital, so as not to confuse their cause-and-effect relationships with social networks.

Although similar processes are under way in other post-socialist countries, we chose to pilot this study in the Kyrgyz Republic for several reasons. First, the country has become drastically impoverished since independence. As of 1997, more than half the population lived below the poverty line, and the gap between rich and poor (indicated by a Gini coefficient of 4.7) was second only to that of Russia among the post-socialist countries (World Bank 1999). In addition, informal kinship and neighborhood-based social networks have long played an important role in Kyrgyz society, both during the Soviet period (1917–91) and in pre-Soviet times, when tribal and clan loyalties were based on mutual webs of obligation and protection. Although the Kyrgyz have had a shorter history with Islam than many Central Asian groups, it is nevertheless important to add that they have also been influenced by the emphasis Islam places on the importance of family solidarity and mutual assistance (see Coudouel, McAuley, and Micklewright 1997, p. 202). Finally, the Kyrgyz Republic was one of 23
participating countries in the World Bank's recent "Voices of the Poor" study, and the Kyrgyz case study was managed in the field by one of the authors of this study (Kathleen Kuehnast). By returning to the same sites with the same local interviewers, the authors were able to build upon the interviewers' experience, the relationships they had already established in the field, and the rich qualitative data they had already collected in the poorest oblasts (regions) and Bishkek.

For this study, "networks" are defined as a web of relationships through which goods, services, money, and information are traded, and through which mutual obligation and gift-giving activities directly enhance social status. It is assumed in this study that personalized systems of exchange are based on different motives and values than those of anonymous markets. Interviews were designed to elicit information from respondents on the characteristics of their networks, the kinds of transactions that predominated within each network, and finally, the changes in the structure, size, and importance of their networks during the last 10 years. Although the limited number of sites and respondents does not allow us to generalize the findings for Kyrgyz society as a whole, similar findings from other recent studies in the Kyrgyz Republic suggest that they do indeed represent a countrywide phenomenon. (See Mikhalev and Heinrich (1999) and Rumer (1996).)

The categories "poor" and "nonpoor" used in this study largely refer to how study participants in the poorest three regions (oblasts) of the country identified themselves to interviewers.

The study was primarily designed to develop a more detailed and nuanced understanding of poverty in the Kyrgyz Republic—particularly in rural regions—rather than to define these terms with precision. In general, poor respondents in the study had few assets, participated

<table>
<thead>
<tr>
<th>Oblast</th>
<th>Poor (percentage)</th>
<th>Extreme poor (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bishkek, capital city</td>
<td>6.0</td>
<td>0.8</td>
</tr>
<tr>
<td>Issyk-Kul oblast</td>
<td>64.5</td>
<td>23.8</td>
</tr>
<tr>
<td>Jalal-Abad oblast</td>
<td>73.0</td>
<td>30.3</td>
</tr>
<tr>
<td>Naryn oblast</td>
<td>90.5</td>
<td>58.7</td>
</tr>
<tr>
<td>Osh oblast</td>
<td>65.7</td>
<td>10.1</td>
</tr>
<tr>
<td>Talas oblast</td>
<td>67.0</td>
<td>23.0</td>
</tr>
<tr>
<td>Chui oblast</td>
<td>26.6</td>
<td>3.5</td>
</tr>
</tbody>
</table>

*Source: World Bank (1999).*
in increasingly flat or “horizontal” social networks, and had little or no cash. Nonpoor respondents had sufficient material and monetary resources to overcome financial setbacks, participated in extensive and diverse networks, and either had cash or were able to convert resources into cash easily.

This study was conducted over a 6-week period between April and June 1999. Three local research teams conducted 21 focus groups (involving 210 respondents) and 105 interviews in seven urban, semi-urban, and rural communities of Naryn, Talas, and Jalal-Abad oblasts (regions), plus the capital city of Bishkek, with a purposively selected sample of poor and nonpoor respondents (table 1). To select sites, we identified the three poorest oblasts (Talas, Jalal-Abad and Naryn) on the basis of 1999 World Bank poverty update (World Bank 1999). Local leaders and interviewers then identified two of the poorer villages or towns in each oblast. Focus groups were held at the chosen sites with poor, nonpoor (identified by discussions with local leaders at each site), and “special” groups such as local minorities, or rural migrants in the city, and in-depth interviews of two to three hours were conducted with seven poor and seven nonpoor, with a mix of age and sex in each group. Individual respondents were selected in part by their willingness to be interviewed, their ability to articulate the social issues and, in some cases, on the basis of previous participation in the “Voices of the Poor” study. A high frequency of common themes emerged in the interviews, as did unique situational differences in individual social networks. The respondents identified as “poor” or “nonpoor” prior to the interview, working with the interviewer, filled in detailed matrices documenting the kind and frequency of transactions in which they regularly engaged.

8. The research team consisted of 12 interviewers who had been trained in methods of Participatory Rapid Appraisal (PRA), as well as in-depth interview and focus group techniques. All the interviewers had participated in the previous World Bank study “Voices of the Poor.” The interviewers originated in each of the oblasts in which they did their research, which assured greater comprehension of local conditions and regional problems.

9. The village sites were Urmural and Beisheke (Talas Oblast); At-Bashy and Ak-Kiya (Naryn Oblast); and Kok Yangak and Achy (Jalal-Abad Oblast).

10. A man and a woman were each chosen from the following age categories: under 30, between 30 and 50, and between 50 and 65. One respondent over 65, of either sex, was also interviewed.
In each oblast, one person kept a 6-week diary of his or her transactions. The interviewers recruited women to keep the diaries, because the women were more engaged in the day-to-day transactions that link network participants. As Cynthia Werner (2000) noted from her own fieldwork in Central Asia, women tend to be “more active in maintenance of social networks by serving guests, exchanging gifts, and helping others prepare food for guests...[whereas] men are more active in the manipulation of social networks, as they are the ones who typically ‘call in favors.’”

Qualitative research is very labor-intensive; the time required to review and analyze the 105 interviews and 21 focus group reports was considerable. Common themes, as well as points of divergence, were noted and then analyzed only by returning repeatedly to the original interviews and reports. The end result is a detailed view of how social networks function among some Kyrgyz today.

Background: Social Networks during the Soviet Period

*Better a hundred friends than a hundred rubles.*

Russian proverb popular during the Soviet era

During the socialist period, webs of personal relationships were the principal “currency” in society. Although basic goods and services were heavily subsidized and widely affordable, informal social networks were the most important mechanisms for getting things done, obtaining access to “deficit” goods and services, acquiring accurate information about events and opportunities, circumventing regulation and, in combination with bribes, gaining access to elite education, high-quality health care, and positions of power. This network-based economy of reciprocal favors, referred to in Russian as sviazy (connections), was an important feature of the centralized socialist economy that helped people to compensate for failures of the state.

---

11. Although this proverb was known throughout the FSU, it dates back to the Russian Empire.

Although the networks of ordinary people and the elite largely functioned independently of one another, the relatively egalitarian conditions of Soviet society enabled most people to establish far-reaching networks. Most people perceived their predicaments as similar and were not ashamed to ask favors or borrow money from one another, because guaranteed employment and stable incomes made it likely they could return the debt or favor in the future. In the Soviet shortage economy, who one was and whom one could access were far more important than the money one had saved. Thus, status and power depended less on income than on the extent to which one’s informal networks included people with blat (pull or influence). Such individuals were typically close to sources of political, social, and economic power and were capable of pulling the levers of power within an institution to fulfill a request. As Larissa Lomnitz (1988) concluded from a comparative study of Mexico, Chile, and the Republic of Georgia, the more a social system is “bureaucratically formalized, regulated, planned, and yet unable to fully satisfy social requirements, the more it tends to create informal mechanisms.” Social networks in the Kyrgyz Republic represent one such mechanism.

Such informal networks were not only a response to the inadequacies of formal institutions in the FSU. In Central Asia, these networks emerged from traditional kinship ties that proved to be exceptionally strong. Prior to Sovietization, tribal and clan relations constituted the basis of economic and political collective well-being in Central Asia. No individual could survive without the protective mantle of tightly woven networks of extended relatives who lived across the once nomadic territory. Customary laws that regulated marriage and elaborate rituals of gift-giving centered around life-cycle celebrations provided safety, security, and social status in the pre-Soviet world of the Kyrgyz. This pattern of expansive and influential kinship networks persisted despite attempts of the Soviet regime to weaken them. One typical Soviet prohibition forbade family gatherings of more than 100 people—in a society where hundreds of relatives had traditionally gathered for weddings or funerals.

At the same time, the elaborate system of Soviet collective farms often grouped extended families and clan groups together, thereby reinforcing kinship networks by ensuring that their members lived and worked in the same location. The informal networks that became a Soviet way of life integrated easily with Central Asian practices of gift exchange and Islamic concepts of charity, both of which reinforced mutual support among kinship groups, friends, neighbors, and col-
leagues. Consequently, sorting out the various strands of Soviet networks, Central Asian social obligations, and the practices of an emerging market economy was one of the more challenging tasks of the following analysis.

Findings

Qualitative poverty studies carried out since 1993 in countries of the FSU reveal that the informal social networks of the poor have deteriorated. The purpose of this study, carried out in the Kyrgyz Republic in 1999, was to investigate the impact of socioeconomic change on the characteristics and functions of the social networks of poor and nonpoor households in rural and urban communities. A better understanding of the role of informal networks in Kyrgyz society, it was thought, should help development specialists devise more effective ways to reach out to the poor and excluded, while ensuring that the benefits of development were not simply captured by those with more effective and far-reaching connections. The findings reveal that the social networks of the poor and nonpoor have polarized and separated, paralleling the sharp socioeconomic stratification that has taken place since independence. Poverty and the increased penetration of market relations have significantly altered family- and clan-based networks and, to a lesser extent, networks based on work, friendship, and neighborhood. The disintegration of kin-based networks was striking in cash-starved and isolated rural regions, where the poor could no longer afford to participate in essential gift exchanges or life-cycle celebrations, nor maintain contact with relatives and acquaintances living in other villages or towns.

Not only had networks of the poor and nonpoor begun to separate, they had each changed in character. The nonpoor in urban communities and, to a lesser extent, in rural communities, were moving away from networks based on ascriptive relationships to more “modern,” interest-based networks through which they successfully exploited access to resources (for example, “insider” information, credit, and preferential treatment by government offices). By contrast, the shrinking networks of the poor reduced their access to decent health care, good education, and timely social assistance, services that are increasingly mediated by personal connections. Indigenous systems of self-help, including rotating savings clubs and mutual aid obligations, were moving out of reach of the very poor, who were unable to afford even modest cash contributions. Catastrophic events were even forc-
ing some poor into patron-client relationships and other varied forms of exploitation. These findings have important implications for community-based approaches aimed at empowering the poor and expanding their economic opportunities. Since person-centered networks in the Kyrgyz Republic remain important for regulating access to important resources, interventions should be designed to ensure that the poor, who are increasingly excluded from informal networks and unable to penetrate the expanding sector of nongovernmental organizations (NGOs), are directly represented and specifically targeted. Given the continuing practical role of social networks as informal safety nets, greater attention should also be paid to investing in rural infrastructure, so that deteriorating transportation and communications services do not further isolate poor communities.

The key findings of the study, which illustrate the impact of poverty on the form and function of informal social networks in the post-Soviet Kyrgyz Republic, are summarized below. Given the case study approach and small sample numbers, the findings can be considered propositions to be further investigated and tested, rather than definitive conclusions.

**Finding 1: Social networks continue to be an integral part of everyday life in post-socialist Kyrgyz society.**

> At present, it is more useful to have a wide network than one hundred rubles, because if you have connections in all structures, and acquaintances in different departments and institutions, you can easily solve any problem.

Focus Group with the nonpoor, At-Bashy village

The gradual encroachment of market relations, the curtailment of state support, and the drastic decline in living standards for the majority of the population in the Kyrgyz Republic have intensified people’s reliance on personal networks for support. As they did during the Soviet-era shortage economy, people continue to engage extensively in informal exchanges and barter within networks made up of family members, colleagues or classmates, and neighbors. Even the practice of exchanging favors within a certain circle of friends or acquaintances in order to reach someone with blat has continued to operate. During the Soviet period, people used blat to obtain deficit goods and services that money alone could not buy. Although most goods and services can be acquired today with money, people often resort to blat to help
augment their incomes (for example, by circumventing official procedures to obtain valuable productive assets or lucrative employment).

Even though the rural population is often associated with social networks of long-term duration, the economic problems of transition have left such networks highly vulnerable. It is an unfortunate paradox that at a time when these networks have become ever more critical for survival, poverty has weakened kinship ties and made it more difficult for the poor to maintain critical support networks. In particular, rural poor respondents identified privatization of collective farms as the pivotal event that changed their daily lives and dramatically altered their social networks. Where once the collective farm was the nerve center of the rural economy, providing employment as well as a critical social safety net for households, privatization pushed many households into subsistence agriculture and severed the continuity of relationships developed over decades of collectivization. The “new” poor emerged significantly during the mid-1990s as privatization spread throughout the rural regions.13

For the nonpoor, networks are important not only for maintaining their social standing, but also for ensuring their future security and prosperity, particularly in the absence of institutional stability. Thus, personal connections to people with official or unofficial power and access to important information have become more essential for finding employment, obtaining loans, establishing enterprises, gaining admission to elite educational institutions, or simply avoiding harassment from officials. In many cases, these networks include influential figures or government officials who share information for financial gain. Such exchanges among the nonpoor are reciprocal, equal, and timely.

The importance of having channels for obtaining information, particularly in information-hungry rural areas, can hardly be overstated. Most periodicals are distributed in Bishkek, whereas in outlying areas, the high cost of paper, transportation problems, and lack of funds means that even those who can afford to subscribe may not receive newspapers for weeks at a time. For those who own them, television and radio are the most important sources of information, but since many remote areas lack reliable electricity, people learn even about government decisions and presidential decrees through word of mouth—and often very late (British Broadcasting Corporation 2001).

13. For an extended discussion of the impact of privatization on collective farm communities, see Humphrey (2000) and Roy (2000).
Finding 2: The size of networks and frequency of social encounters have significantly decreased among the poor. As a result, the rural poor find themselves increasingly isolated: economically, geographically and socially. Simultaneously, the nonpoor are increasingly reluctant to provide support to poor relatives.

The rich have relationships with the rich, their equals, and the poor, but the poor have relationships only with the poor. They don’t maintain relationships with the rich because they don’t have enough money to give them expensive presents or to repay them properly for something. So they avoid those networks because they cannot enter them. If you have enough money, you have greater opportunity to maintain relationships with your relatives and acquaintances.

Focus Group with the poor, At-Bashy

Important formal and informal networks of the poor that formerly centered on the workplace and were reinforced by work relationships have disintegrated as privatization and restructuring of industrial and agricultural enterprises have scattered former colleagues. Urban and rural neighborhoods have altered as impoverished households sold apartments or land to former Communist elites, and new entrepreneurs quickly mastered the rules of the new economy. Isolated villages, deteriorating communications infrastructure, and decreased access to affordable transportation have limited the ability of the rural poor to participate in the nascent market economy.

Even when social networks of the poor remain dense, they tend to be relatively flat, linking together those with the fewest resources and least potential to assist one another. At best, they help the poor avoid further impoverishment. The poor have difficulty maintaining networks with the nonpoor because they are unable to afford an acceptable level of traditional gifts. Although some of the poor have deliberately withdrawn from relationships to save face, others have been excluded by newly rich relatives, whose behavior they consider cruel, insensitive, and a shameful violation of kinship obligations. The nonpoor, on the other hand, characterize the growing distance between poor and nonpoor as an inevitable part of a market economy, in which it is necessarily “every person for himself or herself.” Their strategic deployment of social networks to improve their economic and social status is replacing their traditional obligation to support poor relatives financially.
Despite long-standing kinship ties, a network of close relatives in the Kyrgyz Republic today may number as few as 10 to 15 people. Respondents consistently ranked this circle of relatives as their most important network, partly because they considered it “more appropriate” to deal with relatives than nonrelatives. At the same time, they stressed that ties between previously close kin have weakened during the last decade because people are hesitant to rely on relatives for assistance. The economic crisis has even caused rifts in traditionally important sibling relationships, most noticeably among the poor. It has become more difficult to visit relatives because transportation is no longer subsidized, and the poor cannot afford bus tickets that have increased threefold in price. In addition, poorly maintained roads now prevent buses and trucks from traveling to many rural areas in winter. Thus, visits between relatives, typically accompanied by exchanges of gifts, farm produce, and other items at weddings, funerals, or birthday celebrations, take place with less frequency.

As the nonpoor increasingly distance themselves from poor relatives, the latter criticize their lack of support and disdain for traditional kinship obligations. As a result, tensions have increased among extended families. Respondents noted that family relations are best maintained when a family member whose authority is recognized by all relatives actively works to maintain good communication. If such a person moves away or becomes unable to communicate with the extended family, relations deteriorate and contacts diminish, further isolating poorer family members.

Particularly in isolated regions, neighbors often play a more central role than do relatives in the day-to-day lives of the poor, a fact captured by the Kyrgyz saying, “Buy a neighbor, not a house.” Both the urban and rural poor rank neighbors as second only to kin in importance. Neighbors lend each other small sums of money, food, and other basic necessities on a daily basis. They also exchange services and assist each other at weddings or funerals. In rural areas, groups of neighbors sometimes join to purchase diesel fuel and seeds, rent a tractor or combine harvester, irrigate their fields, or locate a market or mill. For these reasons, a good neighbor is valued more than a distant relative.

Yet even neighbors socialize less than in the past, when it was customary to meet several times a day. Now such encounters may take place once a week or even once a month, and then only when they happen to meet on the street or at the bazaar. In the past, when people received a visit from relatives, they also invited neighbors, a practice that created large, overlapping networks of relatives, friends, and
neighbors. Decreased social visits among relatives and less casual socializing among neighbors have drastically reduced opportunities to expand networks in this manner and, consequently, have diminished mutual support.

Finding 3: Money has become central to maintaining informal social networks, making it more difficult for the poor to remain part of them. Although the poor use what little cash they have for survival, the nonpoor use cash as a tool for mobility.

To maintain one's position in the network, one needs to have money, to be wealthy. Those who have no money try at least not to lose the connections that they have, especially connections with relatives. Friends, in most cases, would not think much of you unless you have money and a prestigious job.

Focus Group with the poor, Achy

If you have money, you can resolve any problem. The main thing is to find the right person who can resolve the problem and provide the appropriate amount for a bribe.

Bulul-ezhe, pensioner, Kenesh

In contrast to the situation during the Soviet era, money has become a key mechanism for establishing and mobilizing networks. During the Soviet period, when separate spheres of exchange operated on the basis of different currencies (for example, money, deficit goods, information, favors, and so forth), money was by no means the principal currency. Most people received regular cash salaries that covered basic needs, but relied on extensive informal networks based on mutual obligations to obtain many difficult-to-find consumer items. In most transactions, obtaining access to something was more difficult than paying for it and the amount of money involved was usually nominal. Even for expensive purchases or large bribes, the exchange of money was carefully brokered by trusted intermediaries.

Today, with consumer items readily available for cash, but priced at world market levels, money has assumed greater practical, as well as symbolic, value. Much of the focus within families is now on the need to make money. This is especially true for the poor, because even state pensions are often paid in kind with flour or oil. Money has also become essential in the exchange of gifts, either as the means for purchasing an expensive gift or as the gift itself. In nongift exchanges
involving services, favors, or information, money has also become an important part of the transaction. The emphasis on cash-based exchange has also affected how people perceive relationships. In the past, favors or services were often provided in the context of long-term relationships in which the giver trusted the recipient to return the favor in the future. Trust has since diminished, and has become more short-term. Thus, most people prefer to receive their payment immediately, and in cash.

Although this practice further excludes the poor, it has directly aided the nonpoor, who can more easily deploy financial resources to bypass traditional or well-established networks. Indeed, even among kin, the transactions of the nonpoor increasingly involve money, because they have managed to dissociate wealth from its negative Soviet connotations and no longer think “having money” suggests illegal or immoral activities. Even friendship has become contingent on wealth. Asked if a hundred friends are still better than a hundred rubles, nonpoor respondents generally observed that few problems could be resolved without personal connections, but that important personal connections can no longer be established without money. As a schoolteacher from At Bashy explained, “Many people nowadays can’t participate in networks because they don’t have enough money for it, so they only associate with those who are as poor as they are, because then neither party is obliged to the other and their relations are free of these problems.” As for kin, the nonpoor regularly review and assess the financial implications of maintaining relationships with poor relatives who expect their frequent help.

Limited resources have also taken a toll on friendships among the poor, because gifts—and, therefore, money—are required to sustain them. Friendship is now seen as a luxury and not a necessity. In response to the question of whether a hundred friends are still better than a hundred rubles, a poor respondent replied that no one could afford a hundred friends anymore. He reminisced about pretransition life, when friends frequently gathered to celebrate birthdays and other holidays, attend the cinema and theater, or hike in summer, without ever thinking about how much they spent. With unemployment and poverty, such gatherings have become infrequent, and life for the poor, as they describe it, has become dismal and lonely.

Finding 4: Because the poor find it increasingly difficult to participate in ceremonial events, they are becoming gradually excluded from kinship and other important networks. By contrast, the nonpoor are host-
ing ever-more-lavish social events as a way of diversifying their networks and expanding their access to a vast array of resources.

Sometimes we cannot go to funerals of our close relatives, for such trips require much money. We postpone the trip, reassuring ourselves that we will go to the 40-day commemoration. But we cannot do that either, because besides the money for the trip, we need money for sevet and kiyit [special gifts]. All this requires money that we don’t have. This is why the trip gets postponed to the one-year commemoration. If you do not go to your relative’s one-year commemoration, your relatives will be offended and most likely will not keep ties with you.

Focus Group with the poor, Achy

People in the Kyrgyz Republic, as elsewhere in Central Asia, depend on person-centered informal networks that are reaffirmed through the rich ceremonial and social life that characterizes these societies. Life-cycle celebrations and rituals, toi in Kyrgyz—connected with birth, marriage, and death—are pivotal encounters that help people cultivate, maintain, and expand networks through the reciprocal exchange of gifts and other material and nonmaterial items, including information, favors, and advice. Although gift exchange constitutes a significant portion of an ordinary household’s annual expenditures, people strive to maintain this tradition because they know they must give in order to receive. As the Kyrgyz proverb says, Kattashpasa jakyn tuugan jat bolot (if you don’t stay in touch with your family, they will become strangers to you one day).

Elaborate gift exchange is not only pivotal to the maintenance of social networks; it is essential to Kyrgyz social identity. Through the activities involved in gift giving, families gain social recognition as responsible members of their kinship group, neighborhood, or community. Within the Kyrgyz extended family, gift transactions are val-

14. Toi—A celebration that takes place for such events as births, circumcisions, marriages, anniversaries, or housewarmings. The most important tois are: sunnot toi—circumcision, ulonuu toi—marriage of a son, kyz toi—marriage of a daughter, iu toi—house warming, beshik toi—presentation by the parents of a new mother to her and her family upon the birth of the first child.
15. For a detailed investigation of similar rituals and networks in rural China, see Yan (1996).
ued as an indication of upstanding moral behavior, even though achieving this level of morality may entail considerable economic deprivation or indebtedness. The understanding that family honor depends on appropriate participation in obligatory gift exchange pressures poor families to borrow beyond their capacity, a practice some people attribute to the negative Kyrgyz trait of sokur namys (blind pride). (See Rao (2001) for useful comparative data.)

Ironically, the poorer rural population appears to celebrate many more events, ceremonies, and traditions than do their more prosperous urban counterparts. In a region where there are more unpredictable calamities such as drought, extreme cold, and other hardships, these celebrations provide not only a respite from such difficulties, but also an important venue for reaffirming old ties and creating new ones as a form of life insurance. Consequently, rural participants in the study noted that the profit they earn during the year through hard manual labor is mostly spent in the autumn on these celebrations. Even though these traditions and practices are considered burdensome at times, the participants agreed that they are essential to social relations, because such relations are the primary conduit for finding jobs, locating food and fuel at low prices, and gathering important information on everything from changing governmental policies to future marriage partners for their children. Rao argues that in addition to their “direct utility,” such social networks are an essential element in poverty alleviation strategies. Thus, such celebrations and rituals observed within social networks provide the public arena in which families are scrutinized and tested, where reputations are made, broken, or enhanced. According to Rao (2001, p. 89), life-cycle events become theaters where public reputations are maintained, and stadiums where people compete in games of status competition. “Because these structures provide rules for what is considered appropriate behavior, they determine the criteria by which people are judged.”

It is at such life-cycle events that obligatory gift exchange is transacted. As Caroline Humphrey and Stephen Hugh-Jones have noted, gift exchange underwrites social relations and is concerned with social reproduction (Humphrey and Hugh-Jones 1992, p. 7). Similarly, gift giving in Kyrgyz society operates according to specific rules and norms that vary according to the type of network. Giving a gift creates indebtedness on the part of the recipient, who is obliged to repay the giver at some future date with a gift of equal or greater value. Failure to do so puts the reputation of the indebted individual or household at risk. Indeed, most households carefully record gift exchanges in a spe-
cial notebook, which they consult when they receive or issue an invitation, to remind them what gift they should give or expect.

In Central Asia, weddings and funerals are two life-cycle rituals pivotal to maintaining social status and preserving social networks. Because marriage has the primary purpose of linking kinship groups rather than individuals, some respondents judged the wedding to be the most important community celebration. Although Soviet authorities outlawed traditional practices of *sep beruu* or dowry (payments by the bride's family) and *kalym* (payment made by the groom's family to the bride's family), these traditions never completely disappeared and are now reasserting themselves.16 From pre-Soviet times, kalym, usually paid in the form of cattle, displayed the wealth, influence, and prestige of the groom's family. Many rural families still give kalym in the form of money, sheep, or horses and such items as fabric, blankets, and clothing to important members of the bride’s extended family. Collection of the dowry begins with the birth of a girl and is given to the groom’s family at the time of engagement.

Since 1991, the value of gifts given for kalym and dowry, as well as the actual cost of weddings, has significantly increased. Respondents estimated that payment for gifts exchanged at the first meeting of future parents-in-law required 10,000 to 15,000 soms (at the 1999 exchange rate of 48 soms to the U.S. dollar, some $208-312), whereas an average wedding cost between 50,000 and 250,000 soms (some $1,040-5,200). Today the “start-up costs” of marriage, including elaborate preparations for the wedding and recruitment of neighbors to host out-of-town guests, mean that rural families have fewer opportunities to establish close ties with people living in better-serviced urban areas, worsening the geographic isolation of their villages. In one case, rural respondents forced their son to renounce his chosen bride from Bishkek because they could not afford the required visits to her family in the city. This case illustrates a change from the Soviet period, when

16. *Sep beruu*—Kyrgyz, a bride’s dowry may consist of furniture, refrigerator, washing machine, a television set, blankets, kitchenware, and a clothing chest.

17. *Kalym* usually consists of cash, *kiit* and *keshik*. *Kiit* refers to the clothing traditionally presented to all close relatives and to relatives of high social status who are close to the bride’s parents. *Keshik* is boiled mutton or foal meat given when visiting a daughter or the son-in-law’s parents. It must be fat, high-quality meat.
young Kyrgyz men often wed urban women to expand their family’s network into urban areas, a strategy that opened up an array of educational and employment opportunities for the groom’s entire family.

In customary Kyrgyz practice, funerals are also socially important, both to display respect for the deceased and to demonstrate the worth of his or her life. There are strong social expectations that a proper funeral ceremony will be organized when someone dies. The threat of social exclusion for failure to do so pressures poor households to take on large debts. The funeral itself is followed by further commemorations, such as one that takes place after 40 days, when relatives, friends, and colleagues meet at the home of the deceased. Many items and large amounts of money are exchanged at funerals, and serious conflicts may result if the goods presented are less in value than those received on a previous occasion. Because they are unable to pay for the trip or the required gifts, some of the poor have stopped attending funerals, 40-day commemorations, and other death-related events, despite their knowledge that failure to attend may provoke offended relatives to sever ties.

Thus, a major concern of the poor is the high cost of hosting or attending such celebrations or rituals, and bringing the obligatory gifts. Consequently, the poor are increasingly withdrawing from participation. For poorer households, celebrations once attended by hundreds of relatives and neighbors now include only the very closest relatives. As a result, poor households have fewer and fewer opportunities to maintain relationships, especially with relatives living in other communities. At the same time, the nonpoor increasingly refrain from inviting poorer relatives to events; in part, they wish to spare them the burden of purchasing gifts or the disgrace of failing to do so. Some nonpoor respondents candidly explained that maintaining relations with poor relatives is no longer beneficial to them. In fact, the nonpoor have increased their expenditures for such social and ceremonial events, which they see as useful opportunities for creating important alliances and strategically displaying their wealth and position. Elaborate funerals may cost more than $10,000 and involve 1,500 guests, all of whom will be accommodated with assistance from neighbors and extended family members.

Although the nonpoor, particularly the wealthier among them, have actively escalated the size and scope of ceremonial exchange, many joined poorer respondents in condemning lavish events as wasteful luxuries, and contrasted such excesses with the thriftier and more “rational” behavior of other ethnic groups. During interviews, poor and nonpoor respondents alike approvingly recalled how Soviet
authorities had once reprimanded Communist Party leaders for organizing expensive commemorations. They commented that the pressure to organize and participate in extravagant ceremonies was damaging to households. Although many respondents felt that local elders should use their moral authority to encourage less costly funeral commemorations, they observed that some elders actively promote lavish expenditures in the name of tradition. The nonpoor, especially those who are less wealthy, are ambivalent about the custom. They feel pressured to compete, fearful that they will lose face if they fail to live up to traditional expectations. Interestingly, respondents from very different socioeconomic backgrounds called for government authorities and the mass media to publicly oppose this form of ritual competition. This common perspective shared by two segments of society that are rapidly diverging in income and opportunity may well reflect the legacy of egalitarian values absorbed during 70 years of Soviet rule.

**Finding 5:** Indigenous forms of cooperation, such as rotating savings clubs and mutual aid obligations, still operate. The requirement for cash contributions is making them inaccessible to the poorest, but they are useful mechanisms of advancement for the nonpoor.

The poor are ashamed to go to a special event held by their relatives because they are unable to contribute 100 soms to razha, and as a result they gradually drop out of the family network. In some instances, relatives promise to make their contributions later. This might work for one or two events; however, when they systematically fail to contribute money, they are "simply forgotten" to be invited for the next event. That is how someone is dropped from the family networks.

Focus Group with the poor, At Bashi

Mutual aid, referred to in Kyrgyz as razha, or yntymak in some regions, is rendered through small monetary exchanges that exist in most communities, whether poor or nonpoor, rural or urban, Kyrgyz or non-Kyrgyz. Of course, this practice is by no means unique to the Kyrgyz Republic or Central Asia, but occurs throughout the world. Razha generally involves the practice of collecting small amounts (30-500 soms—$0.63-10.00), from members of a given social network on the occasion of a wedding or funeral. Most people participate in multiple razha networks of kin, neighbors, colleagues, and friends. People are automatically a part of kin-based razha networks from birth.
and from a very young age learn to be responsible to relatives and assume formal razha obligations after marriage. Respondents recalled that during the Soviet period, when most people had enough money to make such contributions, razha was regularly practiced among relatives and neighbors. The required contributions remained modest, however, because Soviet authorities, as noted above, punished attempts to hold large-scale private festivities.

This informal institution remains essential for the poor, because it is the only way they can hope to pay for a wedding or funeral. A normal contribution consists of 50 soms, which in large kinship networks of 100 or so people can cover the cost of the horse that should be butchered at the ceremony. But today, many family ties are weakening because not every family member can contribute even this small amount. And because most people participate in multiple razha networks, they have multiple obligations. Repeated failure to contribute means exclusion from the network, and exclusion means that an individual’s household will not benefit in the future from razha contributions. The nonpoor still participate in razha exchanges, but more to maintain face than because they need this modest support.

Rotating savings associations, referred to by the Kyrgyz term sherine, or occasionally, in Russian as chërnaya kassa (literally, black cash register or till), are also found in the Kyrgyz Republic. These informal associations consist of people who make regular cash contributions to a fund that is given in whole or in part to each contributor in turn. A worldwide phenomenon, rotating savings and credit associations (ROSCAs) are popular among poorer (but not the poorest) segments of the population.18 During the Soviet period, ROSCAs were widespread among middle-income people, who usually participated in these associations at their workplace. In the Kyrgyz Republic, sherine has become particularly popular among the nonpoor.19 Amounts of up to 250,000 soms collected on a single occasion are used toward the purchase of cars or expensive personal items, or to make investments. Such informal institutions can respond quickly to members’ needs. By providing

---

18. See, for example, Low (1995). Deniz Kandiyoti (1998) notes the practice of chërnaya kassa in Uzbekistan, where the practice functions as a rotating savings club. See also Kandiyoti (1999).

19. Ardener (1995, p. 1) suggests that Taiwan is also a good example of a transition economy in which rotating savings clubs appeal to the emerging elite.
a reliable way to quickly raise large sums of money, they compensate for ineffective or nonexistent banking systems. There is a remarkably low rate of default on what are in effect loans, because participants are intensely concerned to avoid both social disgrace and exclusion from this useful exchange network.

Rotating savings clubs are more than just a way of raising money. They also provide the occasion for enjoyable social functions that provide people an opportunity to exchange information and professional advice. Half a dozen or more friends may take turns hosting each other, with or without families, using a portion of the cash contributions to prepare a lavish meal. Some sherine networks are exclusively male or female. A nonpoor female respondent, for example, described how she met with female friends each month to share a meal and discuss issues of personal interest. The women tended to use the money collected to purchase items for their household or expensive jewelry or clothing for themselves. People who are relatively poorer but have reliable incomes may participate in more modest sherine networks, to which they contribute only 100 soms, gathering over a meal to share news as well as to sing and dance. In such networks, however, recipients prefer to use the money to provision their household with several months' worth of staples such as flour, rice, and oil.

**Finding 6: The relative importance of blat has increased for people without cash and decreased for people with cash. Access to public institutions and employment has declined for the poor, because such access is increasingly mediated by influential social networks. By contrast, the nonpoor are using enlarged and diversified networks to expand their access.**

*Those who have no connections will never be treated fairly. My son had a traffic accident. He was just sitting in a car parked by the side of the road, and another car, with a son of a high government official at the wheel, ran into it. First, the man admitted that it was his fault and even promised that he would pay for the repairs, but then he sued my son instead. Powerful connections let the man win the case, and my son was imprisoned.*

Middle-aged woman, Focus Group Discussion, At Bashy

The importance of social networks for regulating access to public institutions and services is hardly new to post-Soviet society. Ironically, this importance has increased in the post-Soviet era.
Respondents were unanimous in asserting that blat had become essential for finding work, being admitted to a competitive university department or resolving a traffic dispute. Although blat often depends on bribery, it is nevertheless important to distinguish between these two modes of interaction. A key difference is that bribery is illegal, whereas legal codes do not refer to blat (for example, in terms of conflict of interest or nepotism). Rather, in local terms, bribery "implies a conflict of interest where one is to be 'compensated' for doing something one would not do otherwise, while blat is a form of cooperation and mutual support with a long-term perspective, implying trust rather than compensation for risk."20

The success of blat depends on effective and supportive social networks, whereas bribery may or may not have to be supported by personal networks. During the Soviet period, bribery depended on blat, since bribes had to pass through trusted personal connections to the ultimate recipient. Today, it has become easier to rely primarily on bribery as the most expedient way of getting things done in the new economy, because the practice now has fewer legal and social repercussions. Although bribery allows people to circumvent networks because middlemen are no longer so essential to transactions, insider connections (sviaz) remain important, since it is often through such connections that one learns who can or should be bribed, what constitutes reasonable payment, and how to time the payment. Bribery requires specific techniques, depending on the organization involved (for example, a university, a tax or customs department, or a hospital). Such "technical" knowledge is local and specific, and depends on information provided through personal relationships. Even the nonpoor who move from rural to urban areas must obtain access to local social networks to identify which powerful individuals they should bribe to achieve their specific objectives.

Nonpoor respondents described how people used blat and bribery to gain important official positions. Despite the importance of blat during the Soviet period, many respondents argued that bright and talented people had more opportunities at that time to achieve positions

20. As a nonmonetary use of influence, blat was not new to Soviet Russia. As Alena Ledeneva (1998, p. 12) points out in her book, the term "blat" was derived from the Polish blat, which means someone who provides an umbrella, a cover. Prerevolutionary dictionaries imply that blat had connotations of criminal activity, but of the lesser order, such as petty thievery.
of importance without blat. Today, they assert, blat is essential for obtaining government positions and surviving in the new market environment. The nonpoor, for example, use blat to solve problems with tax inspectors, to deal with customs officers when they conduct commerce across borders, to favorably resolve a law suit, to expedite a loan, or to evade military service. Respondents’ perceptions concerning the increased importance of blat are well worth noting. It is well documented from a host of studies that blat remains an active component of transactions in post-socialist societies.

Nonpoor respondents were reluctant to detail their own experiences with bribery, although they claimed that the practice was flourishing as never before (a perception possibly influenced by the fact that during the Soviet period, bribes were transacted covertly through personal networks). In general, they noted that people have become more open about bribery. They feel freer to demand bribes or to directly inquire how much they should pay for a specific favor.

The system of education in the Kyrgyz Republic exemplifies the continuing importance of bribery. In the Soviet times, bribes were frequently used to assure admittance to a school or university, but the amount required was usually manageable. Large sums are now needed to enroll a child in university and to find them employment after graduation. The practice is equally widespread in other public institutions. To register for child benefits, for example, one must pay 50 soms for the registration form and 17 soms for the application form. Officials openly keep the benefits for the first two months, a practice respondents are willing to endure as long as they eventually receive some money. Likewise, when a postal worker delivers a pension, he or she generally keeps 5 or 10 soms as the “delivery fee.”

Finding 7: The poor are becoming increasingly indebted and forced into patron-client relations with the nonpoor.

If you owe money to a wealthier person and cannot return the money on time, the wealthy person will say, “You have to work for that money.” He then gives an amount of work that is usually more than equivalent to the debt. For example, you owe 100 soms and your wealthier neighbor makes you build a fence around his house. Certainly, this work costs much more than 100 soms, but you do not have a choice and so you do the job.

Focus Group with the poor, Archa-Beshik, migrant community in Bishkek
The phenomenon of indebtedness in Kyrgyz society has serious consequences that range from shame to ruptured social relations, ostracism, and what some respondents referred to as “enslavement.” For their daily needs, the poor generally borrow small sums of money (15–20 soms) from each other, usually agreeing beforehand when the money is to be repaid. Exchanges of money or goods must be equal and, in contrast to the Soviet past, people now keep exact accounts of what they borrow or lend. Failure to return the loan seriously strains relationships between neighbors, friends, and acquaintances, and within networks. A participant in the focus group of Kok Yangak described the change in neighborly relationships as follows, “In the past, we were not counting how many presents we gave. Today, our relations are measured by kilos. If you take a kilo of flour, half a bottle of oil, or half a kilo of sugar, you must return the same amount. Otherwise, you may get into trouble, or lose the trust of your neighbor. The next time, he will politely refuse to give you anything because last time you were dishonest.”

To participate in more elaborate exchange networks or to attend ceremonial and social gatherings, however, the poor are often forced to borrow greater amounts, sometimes becoming so indebted that no one in their networks will lend further to them. One resident of At Bashi said, “I constantly ask my close relatives for support without giving them anything in return. I believe that if it goes on in this fashion, I will lose my network of relatives because they do not have enough money to support me like this.” Economic and social pressures have pushed some of the poor into patron-client relationships with the nonpoor. These relationships are one way in which the latter exploit their networks and kinship norms, because they are able to depend on the cheap or free labor of poorer relatives.

The poor feel uncomfortable asking the nonpoor for assistance. Such transactions are unequal from the start: the poor understand that when they do borrow, failure to repay in full means that their creditor may call back the debt in the form of labor worth much more than the original loan. Nevertheless, personal tragedies, unforeseen economic shocks, and social obligations create predicaments in which the poor become indebted to their wealthier neighbors, former friends, and even relatives. A pattern described in some villages is as follows: the poor exhaust their food stocks over the winter, borrow money from wealthier neighbors, and then repay this debt the following summer by cultivating the neighbor’s land. The pattern of indebtedness intensifies when public celebrations or a funeral impose social obligations.
on the poor, whose further borrowing limits their chances of escaping 
an increasingly vicious cycle of obligation and debt. Although non-
poor respondents referred to the help they rendered poor relatives and 
neighbors, the poor preferred to describe these “helping relationships” 
as a modern form of slavery in which the nonpoor exploited them to 
further their own economic advancement.

**Finding 8:** There is increasing differentiation in the form and function 
of social networks of the poor and the nonpoor. The polarization of 
these networks reflects the increasing socioeconomic stratification of 
the population.

> If you have nothing to offer a friend, you will just avoid their 
> company altogether. Nonpoor people have different interests, and 
> it's easier to make friends with those who have the same problems 
> and understand one.

**Focus Group with the nonpoor, At Bashy**

> The poor talk a lot. They keep discussing my money and the way 
> I make it. They have no idea how hard I work to make money. 
> They get it all wrong and believe that I make a lot more than I 
> actually do, so I try to avoid their company.

**Nonpoor female shopkeeper, Kok Jangak**

Kyrgyz society, relatively egalitarian during the Soviet period, has 
become strikingly unequal. This inequality is reflected in the increasing 
dissimilarity of informal social networks of the poor and nonpoor 
and the separation of these networks from each other. The chasm 
between the poor and nonpoor is also widening in relation to their 
social values. Many of the poor describe those who have money as 
“thieves, crooks, or cheats.” As a poor woman from Urmara expressed 
it, “It is very difficult to gain wealth by honest work. Usually, people 
made their fortune by dishonest means.”

Likewise, criticisms of the poor are made by the nonpoor, who call the former “lazy” and accuse 
them of “wanting to use their wealthier relatives as conduits to jobs or 
opportunities” instead of working hard.

---

21. This attitude of the poor toward the nonpoor was repeatedly docu-
mented in focus group discussions in a previous World Bank study, *Kyrgyz 
Republic—Consultations with the Poor* (World Bank, 2000).
Networks of the poor have become flat, linking people with similar incomes and assets. They have also shrunk in size and geographic reach, comprising increasingly smaller groups of people who live near each other. Links with people from higher-income groups have increasingly taken on the character of patronage relations. Even kinship networks, which once functioned as highly secure, dependable social safety nets that linked urban and rural relatives, have ceased to provide long-term security.

Networks of the nonpoor, on the other hand, have become more extensive geographically and more dense socially, reflecting the importance of networks for social and economic mobility. To maximize their utility, the nonpoor have attempted to reshape their networks, discarding some and cultivating others, thereby creating "modern" relationships of more practical value to their new economic circumstances. The separation of networks sometimes takes on visible form. One respondent described a recent funeral in which participants divided themselves into separate groups based on the quality of their clothing. After the funeral, wealthy and poor guests entered the house of the deceased in two separate groups.

Not only have networks taken on different characteristics according to income level, they are also influenced by location. Because about 80 percent of the poor live in rural areas, networks of the rural poor are most affected. A majority of the rural poor are ethnic Kyrgyz whose traditional social networks were based extensively on elaborate gift-giving exchanges, a tradition now rendered much more difficult by poverty. Because collective farms, nonfarm enterprises, and schools once played a key role in bringing rural people together and cementing social networks, their closure has created additional impediments for social networks. With the demise of the collective enterprises, most rural Kyrgyz now survive on labor-intensive subsistence agriculture, which allows few opportunities for casual or formal socializing. Finally, roads are no longer maintained, spare parts are rarely available to repair buses or trucks, and what few phone lines once existed in rural communities have been largely destroyed by nonferrous metal "pirates" who strip copper from telephone wires to trade in China.

22. In many ways, these findings reinforce Mark Granovetter's observations that economic transactions are embedded in social relations and that economists often underestimate the significance of personal relationships and networks of relations in reform economies. See Granovetter (1985).
Although the nonpoor in rural areas lack the wealth of their urban counterparts, they still act as gatekeepers in their communities, regulating access to goods, services, and information. Yet the reach of their networks is also limited by some of the same obstacles that confront the poor.

In major urban areas, recent migrants from rural areas and refugees from Tajikistan are among the poorest segments of the population. Both these groups of poor try to reestablish local networks with others from the same place of origin. These reestablished networks are both assets and liabilities. On the one hand, by joining people of similar origin and situations, they provide a buffer and some degree of assistance. Yet they also hinder the poor from extending their networks beyond their small groups and forming new relationships with others who might provide greater access to urban resources.

The urban nonpoor are in the most advantageous position. They have created a multiplicity of networks that reach into rural areas of the Kyrgyz Republic, as well as abroad, and include relationships with relatives, friends, schoolmates, colleagues, and neighbors of varying income levels. With easy access to a variety of resources, they maintain networks with rural relatives and friends, enjoying the respect and authority their continued attention brings. In urban areas, they have extensive access to a wealth of goods, private and public services, and most importantly, information about business, investment, and employment opportunities in the Kyrgyz Republic and abroad.

Conclusions and Policy Implications

Since the beginning of the move toward a market economy, the multiple and overlapping informal networks that once linked relatives, neighbors, colleagues, and friends from different backgrounds, professions, and geographic regions have become increasingly polarized. In traditional Kyrgyz culture, elaborate gift exchange and a rich ceremonial life once structured social identity, status, and morality, creating supportive links among hundreds of people. The collective organization of Soviet life, much of which centered on workplace relationships, further enhanced the salience of social networks. Today, the transformation of the economy has dramatically transformed and polarized social relationships among the poor and nonpoor.

Impoverishment in the post-Soviet context has had a doubly negative impact on the poor, first, by reducing their ability to maintain support networks, and second, by increasing the need to rely on networks
to maintain access to services. In some cases, debt and dependency have pushed the poor into patron-client relationships, sometimes under the guise of “helping relationships” with wealthier relatives or neighbors. The nonpoor, by contrast, have actively reshaped their networks. Marketization and new forms of competition have robbed traditional and morally sanctioned relationships of their value, while creating important incentives for the nonpoor to expand and diversify interest-based relationships as a means of enhancing social mobility. Reluctant to maintain financially draining relationships with poor relatives, the nonpoor consciously espouse values that diminish the importance of ascriptive identities, and strategically deploy long-established as well as recently formed networks with people who have equal or greater resources.

The dynamics of post-socialism and the changing availability of resources have also affected the social capital of individuals and groups. Because the maintenance of most social networks requires resources, the new poor have found themselves with diminished social capital simply because they have few resources. For example, previously well-connected individuals whose networks were embedded in a sector now in decline, such as collective agriculture, are likely to find that their social capital has completely eroded. Although they may have maintained their networks, the types of resources to which these networks now provide access have lost their usefulness and therefore their value. Thus, not only has the size of the networks of previously well-established people shrunk, the relationships that have survived now primarily link people who have few resources. Moreover, generalized trust, often considered a component of social capital, can hardly be said to exist in the studied communities. Such trust as does exist is highly context bound, and can be characterized as the confidence that participants have in a transaction that their counterparts will honor their part of the contract. Yet in the current situation, even such transaction-based trust has diminished. People tend to prefer short-term exchanges, preferably in cash.

The diminished social capital of the poor raises important policy considerations. As Rao (2001) argues, the infusion of market-driven values and mechanisms is eroding links between the social networks of the poor and the nonpoor. Although these relationships were previously based on traditional systems of status, the nonpoor now find the emphasis on the shared values of reciprocity and assistance counterproductive to their own interests, especially in the unstable new economic environment. For the nonpoor, attending the celebrations or
funerals of poor relatives no longer enhances their social capital. Because social pressure within the extended family to adhere to traditional familial obligations has also diminished, the nonpoor are even less likely to provide an informal safety net for struggling relatives. In fact, as Rao demonstrates and as our Kyrgyz respondents noted, the nonpoor’s economic standing may increase their need to demonstrate their upward mobility to their peers, thus explaining why lavish displays of wealth at life-cycle celebrations are on the increase for the nonpoor and why the nonpoor do not feel obligated to invite or attend to the needs of their poorer relatives.

These findings argue for both the continued importance of supporting formal institutions that serve the poor and for assessing ways in which development interventions can directly reach the poor. Supporting formal institutions is important for providing viable alternatives to patronage relations that force the poor to rely on the dense and resource-rich networks of the nonpoor. Increased support of formal institutions with a stronger emphasis on transparency and complete and timely information could help compensate for the inability of the poor to muster powerful connections to access services. Yet development specialists must also carefully consider avenues for bridging economic differences between the poor and nonpoor rather than further exacerbating these differences, for example, by introducing market reforms too abruptly. In the Kyrgyz Republic, the closure of Soviet-era collective farms and state enterprises has caused particular hardships for the poor, for whom these institutions served as a hub of important social relationships, as well as provider of social services. Thus, the study reinforces Mark Granovetter’s observations that economic transactions are embedded in social relations, and reminds us not to underestimate the significance of personal relationships and networks of relations in transition economies (see Granovetter (1985).

Although the World Bank policies increasingly stress the importance of ensuring inclusion, empowerment, and security for the poor, the capacity to address poverty is weakened by an exclusive focus on the poverty side of the equation. This study argues that the complete story, with attention to the relationships between the poor and nonpoor, must be told. There is little doubt that it is more difficult to “study up,” yet grasping how the nonpoor use social networks in

---

23. This phrase refers to advice by the anthropologist Laura Nader on the importance of expanding the profession’s traditional focus on the poor and vulnerable to include the rich and powerful (see Nader 1972).
their daily lives is essential for understanding how prevailing norms and beliefs about the poor operate in a given society. This understanding is also essential for developing policies that create incentives for the nonpoor to act in ways that enhance inclusion rather than increase the exclusion of the poor.

Greater attention to affordable and reliable rural infrastructure, from roads to communications, could also assist the poor in better maintaining their social networks, which still play a role in their everyday and long-term survival. One of the more constructive ways to assist these networks is by maintaining both rural roads and roadways that are on the outskirts of cities, where many of the urban poor live. Roads allow the poor not only to access support networks, but also make it possible to access employment, markets, schools, and medical care. Telephones give people moral support, as well as the means to exchange useful information. With the support of formal institutions and infrastructure, opportunities for sustainable income generation would become more feasible.

Community-based programs that assist the poor must also recognize the indigenous support systems in the Kyrgyz Republic as viable mechanisms for their programs. Many such support systems (for example, razha and sherine) have been in place for generations and are already familiar to the community. Community-based projects that use these fundamental building blocks of Kyrgyz society could leverage established social relationships to achieve wider inclusion of the very poor. Finally, development interventions that stress direct outreach to the poor must be carefully designed with knowledge of the powerful gatekeeping role of local elites, especially in rural regions. It should be recalled that local NGOs are predominately staffed by such elites. If project interventions do not carefully take into account their complex role, resources may well end up in the pockets of the gatekeepers and not in the hands of poorer community members.

In summary, increased formal institutional support, improved infrastructure, sustainable employment opportunities, and well-

24. For a discussion of how the Grameen Bank originated within the context of indigenous rotating savings clubs in Bangladesh, see Ardener (1995), p. 3.

25. See discussion by Narayan (1999) on the relationship between social exclusion and social capital, which brings out the importance of power differentials and the potentially exclusionary nature of social capital.
designed community programs that reach out directly to the poor could help level an economic playing field that is growing ever more uneven in the post-socialist Kyrgyz Republic. In addition, exploring the interrelationships between the poor and nonpoor, especially their social networks, is a first step toward developing new ways to bridge the growing gulf between these socioeconomic groups. Engaging the nonpoor directly in poverty alleviation efforts and finding new incentives for them to maintain or create linkages with the poor should be part of the social development agenda.
References


Fleron, Frederic J., Jr., and Erik P. Hoffmann. 1993. “Post-Communist Studies and Political Science: Peaceful Coexistence, Détente, and Entente.” In Frederic J. Fleron, Jr., and Erik P. Hoffmann, eds., Post-Communist Studies


An Empirical Investigation of Collective Action Possibilities for Industrial Water Pollution Abatement: Case Study of a Cluster of Small-Scale Industries in India

Smita Misra

Abstract

A case study of the Nandesari Industrial Estate in Gujarat, India, demonstrates the roles played by different agents in industrial water pollution abatement: affected parties, polluters, nongovernmental organizations, regulators, and the court. The study empirically estimates the "benefits" and "costs" of water pollution abatement for a cluster of 250 small-scale industries at Nandesari, and uses these estimates for a social cost-benefit analysis. Benefits are estimated using the contingent valuation method, with a "willingness to accept" format for the rural village areas, and a "willingness to pay" format for the urban area of the city of Vadodara. The study considers costs of command and control, market-based solutions, and the option of common effluent treatment as alternatives. It discusses how joint abatement
at a common effluent treatment plant (CETP) by the 250 industries makes it possible to meet the State Pollution Control Board norms, which was not possible for the industries acting individually over the last 20 years. Finally, a detailed social cost-benefit analysis has been undertaken to estimate the net present social benefits (NPSBs) with and without the CETP. The cost-benefit analysis shows that "collective action"—joint treatment with CETP institutional arrangement—is economically and socially preferable to other approaches for water pollution abatement. Moreover, this conclusion is robust to the inclusion of shadow prices for investment, foreign exchange, and labor, and it also holds when equity considerations are introduced into the calculations.

Nandesari Industrial Estate (NIE) is located in the Indian state of Gujarat, on the banks of the Mini River and its tributary the Mahi River, 20 kilometers north of the city of Vadodara. The first major chemical industry came to this area in 1960. The Mini and Mahi Rivers and their easy accessibility served as excellent disposal agents, and attracted other industries to the area. Currently, the NIE has 250 small-scale industries that produce organic and inorganic chemical compounds, pharmaceuticals, and drugs.

The extent of pollution of the Mahi River and related fish loss was first reported in October 1968. The inhabitants of nearby villages reported to the local government authorities about contamination of their village water tanks and wells, and the related death of fish and cattle, especially around the confluence of Mini and Mahi. These kinds of complaints became a regular feature in this area. The continuous dumping of effluent wastes of diverse kinds into the Mahi and Mini Rivers made their waters inhospitable for aquatic life and unsuitable for human consumption. Nearby villages increasingly suffered from groundwater contamination problems. As a consequence of these reports on the increased pollution loads of the rivers, the Government of Gujarat appointed a technical committee to review the matter. The committee recommended the construction of a 56 kilometer-long effluent channel, the effluent channel project (ECP), to divert the industrial wastewaters from Nandesari to Jambusar, for discharge into the estuary at the Gulf of Cambay. The industries were required to treat the effluents in their own treatment plant before disposing them into the collection wells at Dhanora, from which they were conveyed into the channel. Over the years, this channel has been used by the farmers in surrounding areas as a free source of irrigation water.

A study by Sharma (1995) shows high metal concentrations of
nickel, lead, and zinc at upstream river points. Also, analysis of groundwater from wells located 50–200 meters from the effluent channel shows high levels of total solids, total dissolved solids, and chemical oxygen demand, as well as chlorides, sulfates, nitrates, and metals. In addition, fruits, vegetables, and cereal grains grown in the channel areas have a much higher metal content than those grown in other areas. There has also been rapid erosion in the quality of estuarine flora and fauna at the Gulf of Cambay. With increasing loads of pollutants in the effluent channel, uncontrolled pilferage of channel water for irrigation, and continuous disposal of untreated effluents into the Mini and Mahisagar Rivers, the underground water is undesirable for human consumption and soil unfit for human subsistence.

There are records of numerous complaints made during the 1970s and 1980s by local people about the quality of water in the surrounding areas. For example, various consumer groups filed complaints in the local courts about the surface and groundwater quality, newspaper reports carried stories of water pollution in the area, and the State Pollution Control Board (SPCB) filed about 200 cases against defaulting industries. Remedial action, however, was not taken by the government until Mr. Padiwal, an environmentalist and lawyer, filed a Public Interest Litigation (PIL) case against the Nandesari Industries, leading to closure of industries in 1995.¹ This court order forced the Nandesari Industries Association to undertake measures for water pollution abatement.

This study analyzes the costs and benefits related to water pollution abatement, and explores the possibility of joint treatment and collective action by industries as an alternate institutional arrangement for addressing water pollution abatement problems in an industrial estate.

Costs and Benefits Related to Water Pollution Abatement at Nandesari

Major stakeholders in the Nandesari political economy for water pollution abatement are polluters, affected parties (water users and others), government organizations, and nongovernmental organizations

¹ Public Interest Litigation was introduced in India to provide redress where a legal wrong is caused to a determinate class of persons, who for reasons of poverty, helplessness, or social or economic disadvantages are unable to approach the court for relief.
The costs and benefits to these different agents with and without water pollution abatement can be identified on the basis of the "scenarios" listed in tables 1 and 2.

Various Government of India acts during the 1970s and 1980s, including the Water (Prevention and Control of Pollution) Act of 1974, the Water (Prevention and Control of Pollution) Cess Act of 1977, and the Environment (Protection) Act of 1986, provided the correct impetus for assessing the nature and impact of environmental problems. As a result of these acts, the Gujarat Industrial Development Corporation (GIDC) constructed the above-mentioned effluent channel to convey the waste water from Nandesari to the Gulf of Cambay in the mid-1980s. Also, given the financial, technical, and space problems with small-scale industries, the GIDC constructed a common effluent treatment plant (CETP) in 1984. Meanwhile, the State Pollution Control Board (SPCB) came into existence, and standards were laid down for

### TABLE 1. COSTS AND BENEFITS WITHOUT ABATEMENT

<table>
<thead>
<tr>
<th>Agents</th>
<th>Costs</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polluters</td>
<td>1. Charges paid to pollution control boards for not meeting standards or threat of closure.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Bribes paid to local authorities and the regulator.</td>
<td>Savings in costs for not meeting standards.</td>
</tr>
<tr>
<td>Affected party</td>
<td>Damages caused by water pollution:</td>
<td>1. Employment and income generation from industries</td>
</tr>
<tr>
<td></td>
<td>1. User:</td>
<td>2. Infrastructural development benefits due to proximity of industries.</td>
</tr>
<tr>
<td></td>
<td>(a) Waterborne diseases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) Costs incurred to treat water for drinking purposes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) Losses to farmers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(d) Degradation of soil fertility and increases in toxicity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Nonuser:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a) Degradation of water aquifer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) Degradation of the Mahi and Mini Rivers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) Degradation at Gulf of Cambay</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2. Costs and Benefits with Abatement

<table>
<thead>
<tr>
<th>Agents</th>
<th>Costs</th>
<th>Benefits</th>
</tr>
</thead>
</table>
| **Polluters:**
| 1. Without joint treatment or collective action of polluters | Costs of unilateral treatment by polluters | n.a. |
| 2. With joint treatment or collective action by polluters | Costs of primary (within industry) and secondary treatment (CETP) to meet standards | Savings in cost from economies of scale |
| **Affected party** | |
| 1. Transaction costs of collective action, including costs of organizing a club or consumer forum | 1. Effluent control costs ensure User benefits: (a) Savings due to waterborne diseases avoided (b) Savings in costs of supply of drinking water (c) Degradation of soil fertility reversed (d) Loss in fish productivity reversed |
| 2. Costs of legal action | |
| **Government** | 1. Costs incurred (catalytic role) for ensuring joint treatment or collective action: (a) Incentives in the form of financial support (b) Providing technical know-how | Savings in costs of enforcement and policing |
| 2. Costs incurred for continuous legal threats to defaulters | n.a. |
| **NGOs** | 1. Transaction costs incurred in filing legal cases against defaulters. | n.a. |
| 2. Costs incurred on education/awareness raising for preservation of water quality | |
| 3. Costs incurred in providing technical information to polluters | |

n.a. Not applicable.
discharge of effluents into the rivers, the effluent channel, and the Gulf of Cambay. At the same time, innumerable court cases were filed by consumer groups and local people against the Nandesari Industries. NGOs, such as the Society for Clean Environment (SOCLEEN), played a very important role in disseminating information about water pollution to the local people and the nearby city of Vadodara. Finally, the court ordered closure of the industries in 1995. Under increasing pressure from the court, the Nandesari Industrial Association (NIA) was forced to consider seriously water pollution abatement activities.

The Emergence of an Alternative Institutional Arrangement at Nandesari

Although GIDC set up the CETP in 1984, it could not function regularly for various technical and financial reasons, and was unable to meet the SPCB standards. The main problem was the nature of the heterogeneous mix of effluents from various industries, which could not meet the CETP influent norms. Over the years, and especially after SPCB filed several cases against the defaulting industries, the industries tried various ways of meeting the CETP influent water norms. It became increasingly clear that each industry has to treat its wastewater in a primary treatment plant in order to meet the CETP influent requirements and enable the CETP to function on a regular basis. Meanwhile, the NIA isolated 26 highly acidic industries for setting up a common primary treatment plant at a cost of Rs 15 million ($0.4 million). These 26 industries—producing mainly Vinylsulphone, H-acid, reactive dyes, J-acids, sulfuric acid, S-sulpho anthranic acid—had been individually unable to meet the CETP influent norms. Gradually, the NIA was realizing the benefits from joint treatment.

Economies of scale, characteristic of water pollution reduction, were a strong incentive for industries to seriously consider the CETP arrangement instead of unilateral action (Misra 1998). To take care of the heterogeneous mix of industries and the related effluents, a need arose to establish a special type of institutional arrangement: primary treatment at each industry, common treatment by a subcoalition of industries (as in Nandesari in 1995), and a further common treatment by the grand coalition of industries. The NIA bought the CETP from GIDC in 1995 for Rs 30 million ($0.8 million) and took up the opera-

---

2. All equivalent dollar values are on the basis of 1995-96 exchange rates.
tions and maintenance of the CETP to ensure a smooth operation. The NIA members financed this purchase from their own funds on an equity basis. The capital cost of the CETP plant was shared proportionately among the industries on the basis of their consumption of water, at the rate of Rs 2.50 ($0.07) per kiloliter of water consumed. In 1996 the NIA spent an additional Rs 15 million ($0.4 million) to update the technology of the plant, shared proportionately among the industries on the basis of water consumed. By 1998 the NIA had an established institutional arrangement with primary treatment at individual industry level and a common secondary treatment at the CETP level.

Table 3 shows the range of effluent standards of the Nandesari Industries—industry level and after common treatment for 26 industries, the CETP influent norms, achievable effluent standards, and the ECP and SPCB standards in 1995. As can be seen in table 3 (columns iii and iv), the common primary treatment plant for the 26 industries could not meet the influent norms required by the CETP. Neither could the individual industries (columns ii and iv). Also, given the design of the CETP, it could neither meet the ECP standards nor the SPCB standards. This led to various court cases filed against these 26 industries and against the NIA as well. A court order directed the 26 industries to shut down in 1995. Under pressure from the court, the NIA took over the CETP, as well as the responsibility of ensuring that the final effluent discharged by the NIA will meet the ECP standards. This in turn made the NIA responsible for verifying that each industry undertakes the treatment at the industry level for meeting the technical requirements of the CETP. The NIA also had to solve the cost-sharing arrangements for the CETP. This led to the emergence of rules and conditions as well as self-monitoring schemes for all industries at Nandesari. The perceived benefits motivated the industries to act jointly for undertaking water pollution abatement (see figure 1).

**Estimation of Benefits and Costs**

A carefully designed and administered contingent valuation survey (Misra 1997) adapted to the requirements of the local situation has been used for estimating benefits (damages avoided) from water pollution abatement for the Nandesari area, comprising six affected villages surrounding Nandesari Industrial Estate and the city of Vadodara.

The first pretesting round used the willingness to pay (WTP) format for both urban and rural areas, as is the conventional practice.
TABLE 3. WATER QUALITY CHARACTERISTICS AND STANDARDS\textsuperscript{a} AT NANDESARI, OCTOBER 1995 (BEFORE IMPROVEMENT)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Industry level (actual range)</th>
<th>Common primary treatment for 26 industries (actual, after treatment)</th>
<th>CETP for all industries\textsuperscript{b}</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
<td>Before treatment (required influent norms)</td>
<td>After treatment (achievable effluent standards)</td>
<td>ECP effluent standards</td>
</tr>
<tr>
<td>BOD</td>
<td>2,700-6,000</td>
<td>2,000-3,000</td>
<td>1,750-2,100</td>
<td>123-220</td>
<td>100</td>
</tr>
<tr>
<td>COD</td>
<td>10,000-21,000</td>
<td>8,000-9,000</td>
<td>2,800-4,500</td>
<td>240-450</td>
<td>250</td>
</tr>
<tr>
<td>SS</td>
<td>3,700-11,000</td>
<td>600-800</td>
<td>400-750</td>
<td>112-210</td>
<td>100</td>
</tr>
<tr>
<td>PH</td>
<td>0.25-0.75</td>
<td>6.5-8.5</td>
<td>6.5-8.5</td>
<td>7-7.5</td>
<td>6.5-8.5</td>
</tr>
</tbody>
</table>

\textsuperscript{a} All values are in milligrams per liter, except for pH.

\textsuperscript{b} The CETP effluent standards refer to the parameter values the CETP could achieve in 1995, provided that its influent norms were met, and it was functioning on a regular basis. If the influent norms were not met, the CETP was unable to function on a regular basis.

However, the WTP format failed to work in the rural areas. There were 90 percent protest bids by the poor rural villagers for payments to clean up water pollution in the area, which according to them was the responsibility of the polluting industries. But the villagers were willing to estimate their losses as a result of water pollution and willing to accept a compensation for damages. Hence a willingness to accept (WTA) format was tried, and it worked. Because a loss had already occurred, the WTA was the natural and appropriate measure for assessing damages and welfare losses for the respondents of the rural areas. The Nandesari experience supports the view that an understanding of the “reference position” and “entitlements and property rights” are important factors in determining an appropriate measure for assessing a welfare change (Knetsch 1994).

The contingent valuation survey was undertaken for two separate areas: the urban area of Vadodara city, and the rural area comprising six villages surrounding Nandesari. These areas have been identified
as the affected areas because of pollution from the Mahi and Mini Rivers. The kinds of damages suffered in the two areas are very different. In the urban area, the damages relate to contamination of drinking water and consumption of toxic fruits and vegetables, but because their impacts are not immediate, they are not perceived to be severe. On the other hand, in the rural areas, there are significant damages relating to losses in livelihoods (losses in crop production) and health, which are very noticeable and severe. Thus, two separate questionnaires were designed that related to the respective damages. Special cards illustrated the kinds of user and nonuser damages the respondents suffered. A special card explained the nonuser values related to the Mini and Mahi Rivers, showing the water quality in 2015 with and without water pollution abatement. This particular method of elicitation was chosen to minimize various biases, including strategic biases, hypothetical bias, starting point bias, and scenario misspecification bias.

The urban willingness to pay (WTP) user survey covering 386 households gave an estimate of Rs 74 (US $2.09) per capita per year and urban WTP nonuser survey covering 366 households gave an estimate of Rs 57 (US $1.61) per capita, per year. The total willingness to pay for user and nonuser values estimated for urban Vadodara are Rs 126 million (US $3.6 million) and Rs 97 million (US $2.7 million) per year, respectively. The results show that WTP for user values by urban households significantly depends on per capita earnings, age, size of family, awareness of water pollution-related problems, ideal solution for polluters treating their effluents, expenditures incurred for purifying water (for example, filters), membership in conservation groups, and expenditures on treatment costs caused by damage to health. The WTP for nonuser values by urban households significantly depends on per capita earnings, age, whether or not they have higher education (above school level), the size of the family, the responsibility they attach to Nandesari for water pollution problems, awareness levels,

3. Economic values are classified as user and nonuser values. User values refer to the benefits enjoyed by direct users of water resources, including water quality for drinking and irrigation. Nonuser values refer to “existence,” “bequest,” and “option” values. Existence values recognize that welfare of individuals could increase simply with the knowledge that a resource exists and is preserved. Bequest values recognize that the resource should be preserved for future generations. Option values take into consideration the guarantee that the resource will be available for any future use.
ideal solution of Nandesari adopting pollution abatement strategies, and the quality of water in the surrounding aquifers.

A rural survey was conducted in six villages, with a total of 7,890 households. The sample covering 405 households gave a per capita, per-year WTA estimate of Rs 2,709 ($76.5). The total WTA user values in rural area surveyed was Rs 107 million ($3.0 million). The results show that the WTA damages by the rural population significantly depend upon the per capita earnings, size of the family, education levels higher than class nine, employment and earnings from industries, time losses in collecting water, state of surrounding environment, cost of damages to health, and economic losses due to decline in crop productivity. The rural population did not attach significant nonuser values to the quality of water.

Specific information was collected from the urban respondents about their costs for defensive expenditures on filters or aquaguards for purifying drinking water. Similarly, detailed information was collected from the rural respondents about their actual crop productivity losses and treatment costs for diseases related to water pollution. This information has been used to check for validity of estimated WTP and WTA. The results show that on an average, both WTP and WTA are lower bounds to the actual losses suffered by the respondents, thus validating the use of WTP format for the urban survey and WTA format for the rural survey (Misra 1997).

Costs of abatement have been compared under command-and-control regime, market-based solutions, and the alternate institutional arrangement with CETP technology (Misra 1998), based on the current SPCB standard of 250 milligrams per liter of chemical oxygen demand (COD). The estimates of abatement cost functions enable us to deter-

---

4 The command-and-control regime refers to regulatory instruments that include standards specifying ceilings on emissions, as set by the Central Pollution Control Board in India. Hence, the cost calculations for the command-and-control instruments are reflective of the existing situation in India (if each industry has to independently abate to meet the SPCB standards). Market-based solutions (instruments) provide economic incentives (price or quantity based) for industrial units to abate. Since market-based instruments are currently not operative in India, the numbers in this regard are illustrative. The alternate CETP technology refers to the actual institutional arrangements working in Nandesari (independent treatment by industry plus a joint treatment at CETP). Theoretically, it can be shown that the costs of compliance are generally higher when command-and-control are used than when economic incentives (such as taxes or marketable permits) are used (Baumol and Oates 1988).
mine which institutional arrangement will efficiently internalize the externalities for the cluster of small-scale industries at Nandesari and to verify the validity of the actual solution chosen by the NIA. The cost of water pollution abatement under a command-and-control regime (that is, without a CETP) is estimated at Rs 424 million ($12 million). The cost of abatement under the least-cost, market-based solution is estimated to be Rs 355 million ($10 million). As an alternative, abatement costs have also been estimated under a third institutional arrangement with CETP technology, which is actually operating at Nandesari. This alternative includes a two-step institutional set-up, based on primary treatment at the industry level and a joint abatement at the CETP. To estimate the full cost of abatement (that is, primary + CETP), the annualized cost of CETP is added to the annual primary treatment cost. For meeting the SPCB requirement of 250 milligrams per liter of COD, the total cost of abatement (that is, primary + CETP) is Rs 122 million ($3.4 million).

Figure 2 shows the relationship between the three institutional arrangements, CC (command-and-control), MS (market-based solution), and CETP technology (that is, primary with CETP) under current SPCB requirements of 250 milligrams per liter and a more relaxed requirement of 500 milligrams per liter of COD (for comparison purposes). The currently operating CETP institutional arrangement at Nandesari turns out to be the most economical arrangement. The figure also shows that the total user and nonuser benefits estimated for six villages and the city of Vadodara are Rs 330 million ($9.3 million) (Misra 1997). Thus large potential net benefits can be generated using a CETP for water pollution abatement at Nandesari Industrial Estate. By contrast, if command-and-control methods are used, the cost of abatement exceeds the benefits. This analysis thus confirms that collective action and a joint abatement is the ideal solution for internalizing the water pollution externalities of a cluster of small-scale industries.

5. The Nandesari case shows that the CETP technology is cost-effective even when compared with the market-based solution. This is because of economies of scale being reaped by the joint abatement under the CETP technology, which is not possible with independent end-of-pipe treatment under market-based solutions (taxes / marketable permits). The case of market instruments promoting joint abatement is not considered here.
FIGURE 2. COSTS OF WATER POLLUTION ABATEMENT UNDER ALTERNATE INSTITUTIONAL ARRANGEMENTS

Existing Cost-Sharing Arrangements at Nandesari

The CETP institutional arrangement set up by a cluster of small-scale industries enables the member industries to take advantage of scale economies in wastewater treatment and thus save on the costs the industries will have to bear if they have to individually meet the SPCB standards. The distribution of these savings in costs among the industries depends on the cost sharing arrangements of the CETP. Hence, it is interesting to examine the existing cost-sharing arrangements at Nandesari and to see how they evolved historically.

Table 4 presents the charges to industry for water treatment during 1995–98. The charges during 1995–96 and 1996–97 were not efficient in sustaining the alternate institutional arrangement with CETP technology. This was the reason that the CETP could not function for continuous periods, which resulted in innumerable court cases filed against the defaulting industries. The cost-sharing arrangements in 1997–98 were on the basis of pollution load of polluters and hence were more efficient. The industries that required a minimal CETP treatment had to pay Rs 5 ($0.14) per kiloliter of wastewater discharged toward establishment, electricity, maintenance, and agency charges to the association for operating the CETP. The rationale for
TABLE 4. COST SHARING ARRANGEMENTS AT NANDESARI

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
<td>(iii)</td>
</tr>
<tr>
<td>Charge per kiloliter (Rs)</td>
<td>2.70</td>
<td>3.30</td>
<td>5</td>
</tr>
</tbody>
</table>

(i) Industry requiring minimal CETP treatment.
(ii) Inorganic chemical manufacturing unit.
(iii) Organic dyes and intermediates manufacturing unit.

Source: Nandesari Industries Association.

This change was that even though their COD level was within norms, their wastewater was mixed with the treated water from the CETP before it was finally discharged to the effluent channel, and the NIA is fully responsible for overall effluent characteristics of the discharged wastewater. The inorganic chemical manufacturing units were charged at the rate of Rs 7 ($0.20) per kiloliter of wastewater treated because their COD effluent load was low. The organic dyes and intermediate manufacturing units were charged at the rate of Rs 10 ($0.28) per kiloliter of wastewater treated because their COD concentration was very high. The NIA arrived at these rates after several months of chemically analyzing effluent samples of these units. This policy of price discrimination on the basis of pollution load made the CETP institutional arrangement sustainable and provided incentives to the industries to abate jointly, so that each polluter enjoyed a cost advantage according to his pollution load. However, those industries that were paying Rs 5 ($0.14) per kiloliter of wastewater discharged continued to be at a disadvantage, perhaps because of their weak bargaining power. The question of fairness and bargaining strength would finally determine the mutually accepted prices.

The problem of making credible commitments is resolved through the process of mutual monitoring. Given court orders for closure of defaulting industries, it is in the interest of each industry to abate under the CETP arrangements and benefit from economies of scale in water pollution abatement. Further, the CETP requires a particular concentration of the wastewater that it can treat. The NIA has taken the responsibility of ensuring that each industry meets the criteria for CETP treatment. This is a necessary condition for assuring cost advantage to all. The association monitors the effluents of the industries and takes action against the defaulting industries.

Three scenarios have been considered to estimate the costs and benefits for water pollution abatement practices at Nandesari Industrial Estate:6

1. Costs and benefits of water pollution abatement with the joint treatment at the common effluent treatment plant (CETP) by industries to comply with standards, namely, collective action by polluters.
2. Costs and benefits of water pollution abatement with independent treatment at industry level to comply with standards, namely, without CETP and without the collective action of polluters.
3. Damages when there is neither joint treatment nor independent treatment at the industry level, and standards are not met.

Benefits from Abatement

The user and nonuser benefits from abatement practices at Nandesari have been evaluated using the contingent valuation technique (outlined in the section Estimation of Benefits and Costs).

Costs of Abatement

The water pollution abatement costs for realization of SPCB standards for Nandesari Industrial Estate can be apportioned to (a) industries, (b) the SPCB and ECP, (c) court, (d) nongovernmental organizations, and (e) Padiwal’s Public Interest Litigation Case. The costs to SPCB, ECP, and the court represent costs to the regulator. The costs to NGOs and toward Padiwal’s Public Interest Litigation Case represent the costs to the affected parties.7

6. The first two scenarios assume collective action by the affected parties, and standards are realized. The third scenario assumes neither collective action by the polluters nor any collective action by the affected parties, and standards are not realized. The damages in the third scenario are equivalent to benefits not being realized; hence no further estimations are made for this scenario.

7. See Misra (1998) for details of costs accruing to various agents.
Industries. Detailed information on capital cost and operation and maintenance costs for water pollution abatement was collected in a survey conducted in April 1997 (details in Misra 1998). The capital cost details provide information about the domestic materials, skilled labor, and unskilled labor used in the construction of the effluent treatment plants at the industries. The operation and maintenance cost details provide information about energy, materials, skilled labor, and unskilled labor used for operation and maintenance of the effluent treatment plants at the industries. The capital cost and operation and maintenance cost details were also collected for the CETP.

Government—SPCB and ECP. A series of discussions were held with the SPCB and ECP (Vadodara) officers in January 1997 and April 1997, and data was collected from the balance sheet and annual reports of the SPCB and the ECP. Ten percent of the capital costs and operation and maintenance costs incurred by the SPCB and ECP office at Vadodara has been apportioned for Nandesari, on the basis of their time spent on Nandesari as a percentage of total time spent for the area. The operation and maintenance cost is based on the number of trips they make to Nandesari for monitoring and policing the effluents, the laboratory expenses for analyzing samples, and expenses of court cases filed against defaulters.

If Nandesari Industries did not operate the CETP and each industry abated independently to comply with standards, the cost for SPCB would increase because monitoring and so forth would now have to be done for individual industries. On the basis of the expenditures incurred by SPCB, the costs to SPCB for Nandesari would increase to about 30 percent in this case. The cost to ECP will not increase even if the industries abate independently because they are monitoring the water effluents discharged from the entire industrial estate and not from each industry.

Court. Information was gathered from the District Collector’s Office for the expenses incurred because of court cases filed against the Nandesari Industries. This has been estimated on the basis of an average number of hearings per year, including salaries, wages, and rental and energy charges, for each hearing.

Nongovernmental organizations. The NGO SOCLEEN is currently involved with pollution abatement activities at Nandesari. Detailed discussions were held with Prof. Modi (Professor, Maharaja Sayaji
University, Vadodara, and member of the managing committee of SOCLEEN, and data was collected from the balance sheet and annual reports of SOCLEEN. The function of SOCLEEN is to bring about environmental awareness in and around Vadodara. SOCLEEN is also monitoring and policing the effluents of Nandesari per court orders. Ten percent of the annual operation and maintenance cost of SOCLEEN could be attributed to Nandesari activities. If industries at Nandesari have to abate independently, the costs to SOCLEEN will double.

Public Interest Litigation Case. Mr. Padiwal filed a public interest litigation case against Nandesari Industries Association in March 1995, bearing all the court costs himself. Information about the costs of filing the case was obtained through detailed discussions with Mr. Padiwal at Ahmedabad in January 1997. The lawyer devotes one day a week of his time on the Nandesari case. The actual expenses for court fees, petition memos, photographs, and stationery were obtained. In addition, the value of the lawyer's time, rental value of his chamber, and electricity charges were imputed. If industries abate independently to meet SPCB standards, the cost to Mr. Padiwal would increase by about 30 percent. Mr. Padiwal estimated this on the basis of increases in costs for individual cases filed against each industry.

Estimation of “Social” Benefits and Costs

Based on the information outlined in the section Identification of Benefits and Costs to the Agents: Affected Parties, Industries, and the Government, an attempt is made in this section to estimate the social benefits of collective action for water pollution abatement in Nandesari. The approach of Dasgupta, Marglin, and Sen (1972) has been used for cost-benefit analysis, and corrections are attempted in these flows for shadow prices of investment, unskilled labor, foreign exchange, and income distributional preferences of the government. The social rate of discount for the economic appraisal of public investment projects in India is recommended at 12 percent (Murty and others 1992), and this rate is adopted in this study. Also, an attempt is made to estimate the benefits with respect to alternate rates of discount in the range of 10–20 percent. The life of CETP and other investments has been taken as 25 years. Table 5 provides estimates of the net present benefit (NPB) at market prices from water pollution abatement practices in alternative scenarios (with CETP and without CETP, that is, independent treatment). There are considerable benefits with water
pollution abatement, because the internal rate of return (IRR) with CETP is 95 percent and without CETP (independent treatment) is 30 percent (see Misra 1999).

Estimates of net present social benefits (NPSB) of water pollution abatement practices after making the corrections for the shadow price of foreign exchange and unskilled labor are presented in table 6. There are various methods to estimate shadow exchange rate, including revealed preferences methods and equilibrium exchange methods. Murty and others (1992) have given estimates of shadow exchange rate for India, using some of these methods. Relying on these estimates, the social premium of 15 percent is used in this study. The ratio of shadow price of unskilled labor to the project wage rate is taken as 0.40.8

Table 7 presents the estimates of net present social benefits of water pollution abatement practices after making corrections for the shadow prices of foreign exchange, unskilled labor, and the price of invest-

---

**Table 5. Net Present Benefits, 1995-96 Prices (Rs Millions)**

<table>
<thead>
<tr>
<th>Rate of discount</th>
<th>With CETP</th>
<th>Without CETP (independent treatment)</th>
<th>Benefits due to CETP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1791</td>
<td>1219</td>
<td>572</td>
</tr>
<tr>
<td>0.12</td>
<td>1488</td>
<td>938</td>
<td>550</td>
</tr>
<tr>
<td>0.20</td>
<td>796</td>
<td>309</td>
<td>487</td>
</tr>
</tbody>
</table>

---


<table>
<thead>
<tr>
<th>Rate of discount</th>
<th>With CETP</th>
<th>Without CETP (independent treatment)</th>
<th>Benefits due to CETP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1696</td>
<td>1048</td>
<td>648</td>
</tr>
<tr>
<td>0.12</td>
<td>1404</td>
<td>781</td>
<td>623</td>
</tr>
<tr>
<td>0.20</td>
<td>738</td>
<td>186</td>
<td>552</td>
</tr>
</tbody>
</table>

---

8. For the State of Gujarat, Rs 30 ($0.85) per day is the wage for sowing, weeding, or harvesting, as reported in table 5.1, page 467, of Ministry of Agriculture (1996). The project-specific wages for unskilled labor in Nandesari Industrial Area are about Rs 70 ($1.98) per day. Hence, the shadow wage for unskilled labor is $0.40 for this study.
ment. Because the actual level of savings and investment are less than what government determines as an optimal level of savings, a rupee of savings or investment at margin is socially more valuable than a rupee of consumption. The shadow price of investment in the Indian economy is taken as Rs 1.80 (recommended in Murty and others (1992)). A premium of 80 percent implies that the social cost of a rupee of investment in a project or the social benefit of a rupee of savings from a project at market prices is Rs 1.80 ($0.05). This depends on the social rate of discount, rate of savings and the rate of return on investment. The social rate of discount is discussed above. The rate of savings is taken as 0.3 for urban residents, 0.05 for rural residents, 0.4 for industry owners, zero for unskilled labor, and 0.24 for government.

An attempt is made to assess the impact of pollution abatement practices on income distribution in the economy. For this purpose, the beneficiaries of the project can be identified as urban residents, rural residents, unskilled labor, government, and industries. Some recent studies (Murty and others 1992) provide the estimates of inequality aversion parameter, "e," for India in the range of 1.75-2.00. The Economic Survey, Government of India (Ministry of Finance (1996-97)), provides an estimate of per capita net national product as Rs 9321.4 ($263) at 1995-96 prices. Using Atkinson’s measure, the income distribution weights attributable to different beneficiaries identified for this study are given in Misra (1999).

Taking the national per capita income as the numeraire and e = 1.75, a rupee of income accruing to industry owners has the least social value, equivalent to Rs 0.08 ($0.002), whereas a rupee of income accruing to rural workers has the highest social value, equivalent to Rs 4.0 ($0.11). The social valuation of benefits to industry owners is very low, because their income is about eight times the national per capita income. Table 8 shows net present social benefits of water pollution

<table>
<thead>
<tr>
<th>Rate of discount</th>
<th>With CETP</th>
<th>Without CETP (independent treatment)</th>
<th>Benefits due to CETP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1981</td>
<td>942</td>
<td>1038</td>
</tr>
<tr>
<td>0.12</td>
<td>1525</td>
<td>697</td>
<td>828</td>
</tr>
<tr>
<td>0.14</td>
<td>1227</td>
<td>521</td>
<td>706</td>
</tr>
<tr>
<td>0.20</td>
<td>731</td>
<td>208</td>
<td>523</td>
</tr>
</tbody>
</table>
TABLE 8. NET PRESENT SOCIAL BENEFITS FROM POLLUTION ABATEMENT PRACTICES: WITH EQUITY CONSIDERATIONS (RATE OF DISCOUNT = 0.12, e = 1.75; Rs million at 1995–96 prices)

<table>
<thead>
<tr>
<th>Beneficiary groups</th>
<th>With CETP</th>
<th>Without CETP (independent treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>156.0</td>
<td>156.0</td>
</tr>
<tr>
<td>Rural</td>
<td>2,994.0</td>
<td>2,994.0</td>
</tr>
<tr>
<td>Unskilled labor</td>
<td>14.5</td>
<td>22.3</td>
</tr>
<tr>
<td>Government</td>
<td>-35.3</td>
<td>-36.4</td>
</tr>
<tr>
<td>Industries</td>
<td>-70.0</td>
<td>-120.0</td>
</tr>
<tr>
<td>Total</td>
<td>3,059.2</td>
<td>3,015.9</td>
</tr>
</tbody>
</table>

abatement practices at Nandesari after taking into account the distributional effects. The table shows that although the net present social benefits to rural and urban areas remain the same when standards are met (with or without CETP), the benefits to unskilled labor increase, costs to the government remain more or less at the same level and the costs to industries increase very significantly with independent treatment (without CETP). Overall, these results show that the net present social benefits from the CETP at Nandesari, including shadow prices of foreign exchange, unskilled labor, investment, and income distributional weights, are significantly large.

**Reasons for Industries to Cooperate**

The central question is: Why will the polluters cooperate, and why will they not defect? This could be analyzed in terms of: economies of scale in water pollution abatement; mutual expectations; institutional framework; and sustainability considerations.

**Economies of Scale**

There are economies of scale in water pollution abatement (Misra 1998) that can be reaped with the help of the CETP technology. This technology can be used only with the collective action of polluters. Hence, under court pressure from the Public Interest Litigation Case, the individual industry seeks the CETP arrangement to ensure minimum costs. The economies of scale act as externalities, providing incentives for an optimum investment in the abatement technology with cooperation by all industries.
Mutual Expectations

Given the economies of scale and the institutional arrangement (primary treatment at individual industry and a secondary treatment at CETP), each industry’s propensity to cooperate toward a CETP arrangement depends on the expectations of the behavior of other industries. If the industry does not expect mutual cooperation, there will be a tendency to defect. The standards set by the SPCB, however, as well as cases filed against polluting industries, provide the incentives for mutual cooperation.

Institutional Arrangement

The institutional arrangement should ensure a fair sharing of costs and benefits from water pollution abatement for each industry. The savings in costs from using the CETP arrangement accrue to all industries, depending on their pollution loads. The total gains from the CETP arrangement are shared on a mutually agreeable “fair” basis. The “size” of the industrial group and the “effluent heterogeneity” problem is also solved with the institutional arrangements of a “primary” and a “common secondary” treatment plant. Having an association to enforce effluent control at the industry level, according to the nature of the effluent, resolves the problem of noncooperation among the heterogeneous industrial groups. The Nandesari case empirically verifies the “Coase in politics” solution given by Becker (1983). Credible threats by the Pollution Control Board, as well as retaliatory action by the affected parties, enforce compliance by the polluters.

Sustainability

The state regularly establishes rules through the State Pollution Control Board and other agencies. Monitoring and enforcement of these rules will strengthen and increase the efficiency of joint compliance by the polluters. A “monitoring committee” that includes public and industrial representatives and NGOs has been set up by the court for monitoring effluents of the industries. Appropriate linkages are thus developed for the sustainability of the institutional arrangement and preservation of “water quality.” A complementarity of interests of the local community, the government, and the industries can be seen clearly in the preservation of “water quality” in Nandesari industrial area.
Collective Action at Nandesari—A Response to Court Action

Historically, Prisoner’s Dilemma (Dawes 1973), Tragedy of the Commons (Hardin 1968), and Logic of Collective Action (Olson 1965) are all portrayed as unsuccessful collective action models. Their emphasis is on difficulties of voluntary collective action, based on moral commitments, habits, individual benefits, and the free-rider problem. However, Coase (1960) and Becker (1983) have successfully argued that voluntary collective negotiation and competition of political interest groups help in correcting market failure. Although a Coase solution (voluntary collective action) may be one end of a spectrum with self-enforcing behavior of the individuals, contractual arrangements with complete state regulations could be the other extreme. The latter framework depends on the extent of distortions in the market and the failure of Coasean assumptions. In between would lie a whole string of arrangements to influence the outcome. This study considers one such arrangement, namely, collective action in response to the judicial pronouncement. In the Nandesari context, collective action cannot be strictly defined in the sense of a fully voluntary negotiation leading to efficient abatement; rather, it is a response to a judicial pronouncement on complaints filed by the affected parties and the Public Interest Litigation case.

Given the judicial pronouncement, two outcomes are possible:

1. Culprits abate voluntarily.
2. Administration and policing are required.

In this study, we define (1) as the (limited) collective action case.

Conclusions

The Nandesari case study accomplished the following:

- illustrated an institutional alternative for controlling industrial water pollution;
- investigated the role of various agents: polluters, affected parties, NGOs, regulators, and the court in the political economy of water pollution abatement;
- examined the forces determining the demand and supply of water pollution abatement;
- quantified the benefits and costs from industrial water pollution abatement;
illustrated efficiency and equity gains from water pollution abatement with the help of a detailed social cost-benefit analysis.

Collective action can be seen as an alternative institutional arrangement for bringing about water pollution abatement. The various actors in this game are polluters, affected parties, and regulators. Communities, with the help of social organizations, NGOs, and the court, find ways of enforcing environmental laws. They influence the implementation and tightness of enforcement, which formal regulators have not been able to accomplish.

The role of the regulator is no longer confined to a coercive role of enacting, monitoring, and enforcing standards. The regulator plays a catalytic role of building environmental information and infrastructure. This helps to raise awareness and encourages voluntary non-governmental organizations to address the local problems. It regularly monitors and disseminates information on the ambient quality of local receiving bodies and rivers, provides technical advice on abatement alternatives, and transfers pollution abatement experience from other locations. The regulator thus "levels the playing field" for the communities, strengthening their environmental awareness and bargaining power for effective negotiations with the local industries.

The role of the polluters depends on the bargaining strength they enjoy in the local area. It also depends on how much environmental reputation matters for them and how this affects their expected costs and revenues, as determined by their customers, suppliers, stakeholders, export orientation, or multinational ownership. For reputationally sensitive industries, public certification of good or bad performance may translate into large expected gains or losses over time.

The study analyses collective action in field settings, and identifies various problems, such as physical and institutional settings, that are likely to determine the course of collective action, the agents involved, the strategies they will adopt, the costs of these actions, the outcomes that can be achieved, how actions are linked to outcomes, what information can be available, how much control collective groups can exercise, and what payoffs can be assigned to particular combinations of actions and outcomes. With this kind of rich empirical information, one can capture the essence of collective action problems and provide solutions in the right direction. The effectiveness of this model, however, will be different in different field situations, depending on the levels of environmental awareness, education, income, and the commitment of the concerned agents.
References


Sharma, A. H. 1995. “Environmental Impact Assessment, along the Effluent Channel from Baroda to Jambusar and the Confluence with Mahi Estuary at the Gulf of Cambay with Special Reference to Heavy Metal.” Ph.D. Thesis, Vadodara University, Gujarat; India.

Part III

Local Governments and Basic Services
Abstract

Chile decentralized its primary and secondary education to the municipal level at the very beginning of the 1980s. The reform, which involved all the country’s municipalities, was also extended to the school level at the beginning of the 1990s. A whole strand of literature argues that the transfer of responsibility in the delivery of education services from the central government to subnational levels of government and, even more, to the schools, should make it possible to deliver a service of higher quality. According to this view, what makes the higher quality possible is a better match between supply and demand and the increased accountability of the service providers to the local community.

Does the Chilean case confirm this view? This paper aims at making an original contribution to the debate on the decentralization of education, trying to assess the impact of different forms of decentralization on the quality of edu-

Emanuela Di Gropello (edigropello@worldbank.org) is a Human Development Economist in the Latin America and the Caribbean vice presidency of the World Bank.

The findings, interpretation, and conclusions are the author’s own and should not be attributed to the World Bank, its Executive Board of Directors, or any of its member countries.

cation in Chile. Within the framework of an “extended” education production function, it tests the impact on educational achievement of several decentralization measures, which were provided by an extensive ad hoc survey implemented to complement the existing information. The analysis covers 50 municipalities and is restricted to the period 1992–96/97. The paper carries out both a cross-section (CS) and a value-added (VA) analysis, in order to exploit the variation in the indicators across both space and time, and thereby to get estimates that are as reliable as possible. Among the most significant results of the analysis is that pedagogical and curricular decentralization at the school level and the level of school involvement in local financing decisions both have a significant positive impact on educational achievement. Some econometric evidence also shows that municipal training expenditure and wage incentives have a significant positive impact on educational achievement. In contrast, the impact of some other measures of local administrative autonomy and local financial decentralization is found to be unexpectedly negative. This second set of results suggests that the impact of decentralization might not be as clear-cut as expected and that both the form of decentralization (institutional level involved, functional area decentralized) and the surrounding institutional and socioeconomic environment have an important influence on the results.

“Decentralization,” following the definition given by Rondinelli and Nellis (1986), refers to “the transfer of responsibility for planning, management, and the raising and allocation of resources from the Central Government and its agencies to field units of government agencies, subordinate units or levels of government, semiautonomous public authorities or corporations, areawide, regional or functional authorities or nongovernmental private and voluntary organizations.”

A growing literature addresses the issue of the effect of decentralization on the social efficiency, technical efficiency, and quality of the delivery, and argues that this effect should be positive. The basic assumption of all this literature is that subnational units have better access than the central government to information on local preferences, needs, and conditions, and that as a result, they will make decisions that are more responsive to these local aspects. This will increase the social efficiency of delivery (through a better fit with local preferences) and the technical efficiency, and quality of delivery (through the innovative and creative approaches adopted to fulfill needs and cha-

---

1. Covering different functional areas (financing, administrative, and planning, pedagogical, or curricular areas) and institutional levels (the municipality and the school).
acteristics, as well as the higher level of external accountability produced by the closer link between providers and users).

This same type of reasoning applies to the more specific case of the education sector. Focusing just on the quality issue, the transfer of responsibility in the delivery of the service from the central government to the subnational units, and even to the school level, should in theory make it possible to deliver a service of higher quality. This quality would be achieved through a better match with local needs and characteristics and the increased accountability of the service providers to the local community. In both cases, the positive impact on the quality of education would be enhanced by high levels of participation in the decisionmaking process of the users (teachers, parents, students), because this would increase external accountability (through local monitoring and control) and the fit with local needs and characteristics (through the direct expression of users’ needs). A whole strand of literature specific to education argues that a decentralized and participatory decisionmaking process at the subnational and, above all, the school level, has good potential for improving student performance (taken as a proxy of educational quality) through mechanisms that are basically the above-mentioned ones.

However, there has been little, even if increasing, systematic and rigorous evidence on the impact of decentralization on the quality of service delivery. Recent studies on the impact of decentralized management on the quality of education include the studies of King and Ozler (1998) on school autonomy reform in Nicaragua, Jimenez and Sawada (1999) on El Salvador’s EDUCO’s schools, Ross and others (1998) on the decentralization of decisionmaking to schools in Memphis, Filmer and Eskeland (2002) on autonomy and participation in Argentinian schools, Wossmann (2000) on the cross-country relation between educational institutions and student performance, Paes de Barros and Mendonca (1998) on the determinants of educational achievement in Brazil, and Jimenez and Pagueo (1996) on the impact of local financial decentralization on public schools in the Philippines.

Most studies find a positive and significant relationship between decentralization at the school level and educational achievement. More specifically, in several studies, a strong positive relationship with learning is found for variables measuring autonomous decisionmaking in teacher management, which would lead to more informed staffing decisions, increased monitoring of teacher activities, and increased accountability on the part of teachers (this is the case in King and Ozler (1998), Jimenez and Sawada (1999), Wossmann (2000), and Ross and others.
In the studies, this is followed by a strong impact on learning of variables measuring autonomous decisionmaking in pedagogical processes (see Wossmann 2000 and Ross and others 1998), which would lead to teaching practices more suitable to the local school community's characteristics. By contrast, the evidence is more ambiguous on the impact of autonomy in school decisionmaking on financial issues, with Wossman (2000) showing some negative effects (caused by the opportunistic behavior of schools in the context of generally insufficient accountability frameworks) and Paes de Barros and Mendonca (1998) some positive ones (caused by schools' access to better information about resource allocation than the central level and higher-community oversight at the school level). The importance of community oversight is also highlighted by the findings of several studies that a positive relationship exists between educational achievement and measures of community involvement in the school decisionmaking process (see Jimenez and Sawada (1999) and Paes de Barros and Mendonca (1998)).

By contrast, very little evidence exists on the impact of decentralization to subnational units (local or intermediate governments) on educational achievement. Only Wossmann (2000), Jimenez and Paqueo (1996), and, to a minor extent, King and Ozler (1998) provide some evidence on the decentralization at this level, and they focus on financial decentralization. The first two studies find that a larger share of funds provided by the local (Jimenez and Paqueo 1996) and intermediate levels (Wossmann 2000) has a positive impact on the efficiency and quality of the educational process at the school level. The Nicaragua study finds that the success or failure of increased local school financing is crucially dependent on the degree of impoverishment of surrounding communities, and finds no clear positive effect on educational achievement. Related to this, I should also point out that most studies highlight the importance of the surrounding political, institutional, and socioeconomic environment in the success of the reforms—it is this environment that ensures that the right accountability and governance framework is finally in place.

Within the context of this literature, the paper aims at providing an original contribution to the debate on the impact of decentralization reforms on the quality of education by providing some further evidence on the following:

---

2. This positive impact is thought to result from the better knowledge of subnational units on the needs of the local communities and schools, as well as their higher levels of accountability to local communities, induced by the impact of their policies on the financial contributions of the community's people.
• the effect on educational achievement of the decentralization of decisionmaking to different levels—specifically, the impact of decentralization to the municipal and school levels;
• the impact on educational achievement of decentralized decisionmaking in different spheres—specifically, the impact of decentralization in the financial, administrative, and pedagogical or curricular areas;
• the combined effect of decentralized decisionmaking at different levels and in different areas; and
• the surrounding institutional and economic or financial environment that is conducive to a positive impact of decentralization.

The paper does this by analyzing the Chilean experience with decentralization. Chile decentralized its primary and secondary education by placing responsibility at the municipal level at the very beginning of the 1980s. The reform, which involved all the country's municipalities, was also extended to the schools at the beginning of the 1990s for some specific dimensions. The existence of this “double” level decentralization, combined with the comprehensive nature of the decentralization process and the time frame of the reform, make this experience particularly valuable to assess.

The Model

This section introduces a model of educational achievement and discusses how it can be applied to the Chilean case; later sections develop the model further and estimate it using Chilean data.

The Education Production Function Framework

The “economics of education” literature provided me with the main methodological framework for the analysis through the “education production function” methodology, according to which a measure of educational achievement (the outcome of the process, taken as proxy for the quality of education) is related to a series of inputs that determine it. This sort of function has been widely used in the literature. A standard cross-section (CS) education production function is the following:

---

where $EA_i = \alpha + \beta X_i = u_i$ (1)

However, to discriminate among alternative interpretations and draw more reliable conclusions on the causal nature of the relationship between two variables, a commonly used model of educational achievement using past data is the so-called value-added (VA) model. This model attempts to explain the change in educational achievement over a period, instead of the level of that variable.

A standard VA education production function is the following:

$$EA_t - EA_{t-1} = (\alpha_0 - \alpha_0^*) + \alpha_t X_{it} + (\beta_t \beta_{t-1}) G_i + u_{it}$$ (2)

Such a model has several advantages. First, it makes it possible to concentrate only on the present value of the inputs. In principle, by the cumulative nature of education, educational achievement depends on both present and past educational inputs. In fact, CS analyses omit to measure a set of variables that are likely to be significant in the explanation of the outcome, which reduces the explanatory power of the model and generates another reason for correlation with the error term, insofar as past inputs are correlated with present input values. Second, insofar as time-invariant unobservable characteristics are assumed to be additive-fixed constants with a constant impact in time, they drop from the above formulation. This basic model, however, is simplistic because it assumes that the impact of past inputs on present outcome is the same as the impact of present inputs on present out-
come. In other words, it assumes a constant impact of inputs over time, which is unrealistic. To allow for a decreasing impact over time, the model should be rewritten as follows:4

$$E_{A_{it}} = (\alpha_0 - \lambda \alpha_{i',r}) + \lambda E_{A_{it-1}} + \alpha_1 X_{it} + (\beta_1 - \lambda \beta_{i-1}) G_i + u_{it}$$

Taking an education production function as a starting point, I have integrated it with elements taken from the public finance literature (including aspects related to local financial decentralization, to the local match between demand and supply, and to the involvement of local community) and from the sociology of education literature (aspects of internal school organization and management). Consideration of these elements, which are not included in standard education production functions, focused more on traditional educational inputs, makes it possible to test the effect on educational achievement of decentralization measures. More precisely, the function of education production was estimated including a set of “autonomy” variables, which describe the extent of decentralization per municipality (see below for more explanation).5 Why include these decentralization indicators in the education production function? Are they not working through the other inputs? The reason is that all the schools’ characteristics that, according to theory, might have a direct

4. In this model, the relation between the educational achievement in both periods depends on a coefficient $\lambda$, which, if smaller than 1, can be thought of as the decay rate of the impact. The coefficient $\lambda$ can also have other interpretations, however. Allowing for differential growth in achievement, depending on the initial score, it can capture the fact that it might be more difficult to obtain further gains in achievement as achievement grows.

5. This “extended” education production function approach, which goes beyond the estimation of the inputs that (according to standard theory) should have a direct impact on educational achievement, is now increasingly adopted in the economics of education literature. One of the first attempts was made within the World Bank with the analysis of the determinants of education achievement in Jamaica published by Glewwe and others (1995). The analysis takes an “eclectic” approach by integrating the production function framework with the concerns of sociologists regarding school organization and management. More recent examples that integrate education production functions with institutional factors, including measures of organizational autonomy and community participation in the education sector, can be found in the above-mentioned papers of Jimenez and Sawada (1999), Filmer and Eskeland (2002), Wossmann (2000), and Paes de Barros and Mendonca (1998).
impact on educational outcomes are difficult to measure in practice, which leads generally to incomplete analysis of the determinants of educational achievement and omitted variable biases. For instance, in the regressions below, the variables “teacher motivation” and “interactive or vertical teaching practices” might not have been measured. Under these conditions, some “autonomy” indicators, which capture these and other unmeasured dimensions, might easily be significant in the regression. They complement and even outperform the more traditional school indicators in the explanation of educational achievement.

Application to the Chilean Case

In the Chilean case, a national reform was implemented contemporaneously in all the country’s municipalities (meaning that there is a “control group” of municipalities that did not decentralize or that decentralized at a later stage) without any reliable baseline data collected before the decentralization reform. As a result, the impact of decentralization needs to be captured through the variability across municipalities in the “actual” levels of autonomy associated with the different municipal behaviors in relation to the given decentralization framework. This is what the “autonomy” variables included in the “extended” education production function attempt to capture.

Because the “potential” levels of autonomy are the same across all the municipalities, I could not measure just a municipality’s “capability” of taking certain decisions; rather, I had to measure in some way the “actual” use of this capability by the municipalities. This entailed conceptual difficulties in measuring autonomy, because this concept is more about capability than action, and the use of it could indeed lead to a variety of decisions that cannot be identified beforehand. I have, however, implicitly assumed that “proactive” behaviors, within the potential for autonomous behavior specified by the existing legislation, indicate a more “intensive use” of potential autonomy than “passive” behaviors that lead municipalities to rely purely on centralized rules (even if it is their actual choice to do that). Examples of such proactive behaviors include directing an important amount of local

---

6. Summarized by P. Glewwe (see Glewwe and Grosh 2000) into three main categories: material inputs, teacher characteristics, and pedagogical techniques.
resources to education, introducing local administrative initiatives, and promoting school autonomy.

Additionally, independent of the choice issue, higher levels of local funds spent on education, higher levels of local flexibility in personnel management, and higher levels of school autonomy all constitute measures of initiatives taken locally that capture the impact of some form of decentralized actions on educational outcomes. Beyond this, they are themselves a source of higher autonomy (for instance, it is clear that a higher proportion of local funds increases the potential for local initiatives). This is why I define the indicators presented in the next section (Selected Decentralization Indicators in the Context of the Chilean Decentralization Process) as “decentralization” and “autonomy” indicators. Ultimately, in the weakest possible interpretation, I am at least capturing the impact of innovations induced by the decentralization process and therefore assessing, if not the impact (that is, the effect relative to the appropriate counterfactual), at least the direct consequences or effects of the Chilean municipalization reform.

When it comes to the estimation of the models, two main problems in the Chilean case are that no reliable data on educational achievement and other important socioeconomic and school dimensions are available before the very end of the 1980s or early 1990s, and that it is not possible to construct many of the decentralization measures on the basis of the existing information. This led me to limit the analysis to the period 1992–96/97 and to carry out an ad hoc survey to complement the existing data with decentralization-related information. To measure the impact of decentralization on educational achievement, in this first stage of the analysis I chose the municipality as the unit of measure and regressed an aggregated measure of educational achievement on a series of inputs, also aggregated at the local level, including the decentralization indicators. I took advantage of the variability of these indicators across municipalities and across time to assess their effect on educational outcomes.

The main database was therefore produced by an ad hoc survey that was carried out from July 1997 to December 1997. The survey consisted of five different questionnaires that were submitted to the local municipal authorities and to school directors, which were completed through individual interviews. The questionnaires covered aspects of autonomy and efficiency of delivery, as well as aspects of involvement of the local and school community in the decisionmaking process. In all, 50 municipalities (out of the country's 335 municipalities) and 96 schools within these same municipalities were covered. The popula-
tion of municipalities was stratified by geographic location and size of municipalities to make it possible to select a sample that would be representative of these dimensions. The main reason for stratifying the population by these two dimensions was that geographic location—defined as location in the northern, central, or southern areas—and municipal size—measured in terms of population—were supposed to be an important source of variability in municipal behavior in Chile and, consequently, they were expected to produce some variability in the sample. This database was then complemented by some secondary data sources on several dimensions (for example, teachers' characteristics, enrollment figures, socioeconomic characteristics, and educational achievement data) to constitute the complete data set to be used as a basis for all the regressions.

The CS model was run over 1996 data; whereas in the VA model, I concentrated on the time span between 1996 and 1992. The absence of longitudinal follow-up of the individual students made it impossible to remove, or control for, the time-invariant unobservables aggregated over that level in the VA; but the model, as it was applied, constituted at least a panel at the municipal level. Another difference with the standard VA approach was that instead of analyzing the determinants of the change in test scores from one minor grade to a higher grade (that is, of the educational achievement gain of a further year of education, properly defined as value-added), I focused on the change in the same grade, implying that the terminology VA would not be strictly appropriate in this case. The two models finally estimated are as follows:

\[ EA_i = a + bLSEV_i + cLSC_i + e_i \]  \hspace{1cm} (4)

and

\[ EA_{it} = a + bEA_{it-4} + c_iLSEV_{it-3} + d_iLSC_{it-3} + e_{it} \]  \hspace{1cm} (5)

where \( EA_i(\text{EA}_{it}) = \text{vector of educational achievement variables (proxied by test scores at the end of the fourth year) of the} \ i\text{th municipality in 1996} \)

\[ 7. \text{This 4-year span corresponds to a period full of innovations in educational policy in Chile. Four years should also be a time span sufficient to notice a preliminary impact of local and central innovations on educational achievement.} \]
\[ EA_{it-4} = \text{vector of educational achievement variables (proxied by test scores at the end of the fourth year) of the } i \text{th municipality in 1992} \]

\[ LSEV_i (LSEV_{i-3}) = \text{vector of local socioeconomic variables of the } i \text{th municipality in 1996(1993)} \]

\[ LSC_i (LSC_{i-3}) = \text{vector of local school characteristics of the } i \text{th municipality in 1996(1993), which can in turn be disaggregated into the following groups of variables (including the decentralization indicators): SMI (school material inputs); STC (school teacher characteristics) and SPP (school pedagogical practices) } \]

The \( b \) coefficient in the VA can have several interpretations, more or less close to the interpretations given for the function estimated at the individual case (see above). I would expect \( b \) to be smaller than 1 to express the rate of decay of the impact. In this specific case, being interested in the period (1992–96), I plugged the 1992 test scores in the equation and considered \( b \) the decreasing rate of impact of the inputs of 1992 (and before) on the 1996 test scores. As far as the other regressors of the VA equation were concerned, inclusion of the 1992 test scores in the equation made it safe to concentrate only on the inputs of the period (1993–96). In theory, all 4 years should have been plugged in, but for collinearity reasons, this seemed unfeasible. Period averages might have been an alternative solution, but averages might be difficult to interpret because they are subject to simultaneity bias. Eventually, I decided to plug in the 1993 values, when possible and appropriate, to avoid any difficulty in interpretation. As long as time-invariant characteristics can be controlled for through the inclusion of past achievement, the risk of correlation with the error term caused by omitted heterogeneity would be reduced, too.

**Selected Decentralization Indicators in the Context of the Chilean Decentralization Process**

The Chilean education decentralization mode can be seen as a principal-agent type of model where the principal (the Ministry of Education) transferred the responsibility for the provision of the education services to the agents (the municipalities) at the beginning of the

---

8. A more detailed treatment and explanation of the selected decentralization indicators can be found in Di Gropello (2001).
1980s, but retained some important tasks in the supervision, financing, and planning areas. From 1992 onwards, the reform took two main directions. First, pedagogical and curricular decentralization was promoted directly from the center to the school level bypassing the municipalities. Second, a partial "recentralization" of human resources management occurred. This was followed by a set of political, financial, and institutional measures that encouraged more autonomous decisionmaking at the municipal level.

For the purpose of this paper, delivery of education was divided into three main functional sectors (financing, administrative, and planning-curricular-pedagogical sectors). Also decentralization indicators for each of these three sectors were worked out, when applicable, at both the municipal and school levels.

**Decentralization in the Financing Area**

A first indicator of decentralization measures the extent of decentralization in the education sector from the financial side. In Chile municipalities are free to complement the funds of the center (given as per-student subsidies and with a predetermined use) with their own funds, over which they have total autonomy. This means that a measure of decentralization in the financing area is provided by the following indicator:

\[
\text{FINANCIAL AUTONOMY} = \frac{F_{\text{mun}}}{F_{\text{tot}}}
\]

where \(F_{\text{mun}}\) = the municipality's own funds directed to the education sector, and \(F_{\text{tot}}\) = total funds directed to the education sector.

9. This indicator, however, is not completely appropriate to capture the effective degree of financing autonomy in the education sector in Chile, because in many cases, the central education subsidy does not even cover all the personnel costs, which makes it necessary for the municipality to close the gap. The municipal funds used for that purpose cannot be considered autonomous, because they are, in a way, predetermined by the minimum working needs of the services. However, the attempts that were made to construct alternative financial indicators were not very successful, and I decided to stick to the indicator constructed above. Keeping FINANCIAL_AUTONOMY does at least have the advantage of explaining the main determinants of local financing, with all its consequences, and it casts some light on the decentralization process.
Higher ratios of local-to-total funds in the context of similar amounts of central funds per student imply a higher local mobilization effort, which will lead to additional resources given to education. It should also lead to higher levels of local accountability of the municipalities and to services that are more responsive to local needs (above all, if the increased use of local funds is accompanied by the increased participation of the beneficiaries in the local decisionmaking process). All this should result in higher educational achievement.

In the same area, where school autonomy measured by the proportion of funds directly raised at or decentralized to the school level is minimal, a form of "restricted" autonomy at the school level is provided by the school involvement in the local decisionmaking process, including decisions on fund-raising and, above all, resource allocation at the local level. According to the school-based management literature, direct school involvement in local decisions on expenditure allocation is expected to make it more responsive to the needs of parents and students and, additionally, to lead to a higher level of school accountability to the community and accountability of local administrators to the schools. The index of school involvement is constructed by aggregating three different indicators: the number of school participation mechanisms at the local level, the number of meetings of these participation mechanisms, and the contrasting perceptions of the education department directors and the schools' directors on the degree of involvement of the school in the decisionmaking process. The resulting composite index was named INVOLVEMENT_INDEX.  

Decentralization in the Administrative Area

The Chilean decentralization reform was largely an administrative one, with the responsibility for the direct administration of the schools being completely transferred to the municipalities. However, an important administrative aspect—the administration of human resources—went through different stages of decentralization over time. Autonomy over decisionmaking in labor policy issues is a crucial aspect. Beyond the transfer of the responsibility for the direct

10. Constructed as a simple average of the standardized version of each single indicator.
11. The decentralization of 1981 was accompanied by an abrupt change in the employment conditions of teachers. Their status changed from one of civil
administration of the education services to the new agents, an effective
decentralization policy should, at least partly, decentralize the adminis-
tration of human resources to the new providers if gains in flexibility
are to be achieved.

In the administrative area, I have constructed indicators that
attempt to measure the extent of local initiatives and the actual level of
flexibility in the management of human resources by municipality,
given the existing restrictive framework. Within the current legisla-
tion, municipalities have some flexibility in adopting the following
measures:

- introducing local wage incentives, complementing existing
  ones set centrally (Art 42 of Law 19.070 and Law 19.410);
- setting up local training programs and/or applying training
  incentives, through grants, special agreements with universi-
ties, and so forth (Art 12 of Law 19.070 and 19.410); and

servant subject to public-sector legislation to one of municipal staff subject to
the private-sector labor code. This reform was aimed at increasing the level of
decisionmaking autonomy of the municipalities in all the main aspects of
labor: hiring, firing, and determining wages and career prospects. This reform
was a far-reaching one, but discontent among teachers eventually led to the
introduction and application from 1991 of the Law 19.070, known as the
Estatuto Docente (Teacher Statute), aimed at reintroducing some labor stability
and a new wage structure. The Estatuto meant a partial recentralization of
labor policy insofar as it decreased the autonomy of municipalities in fixing
wages and, above all, in making changes in the size of the teaching staff. The
situation did not change until the end of 1995, when a new law (Law 19.410)
was finally approved after a long period of negotiations. This law provided for
an increase in the amount of the total subsidy transferred to the municipalities
and made it possible to dismiss teachers where numbers became excessive
because of the cancellation of courses or the merger of schools (Art. 22 and
52-i). After further negotiations between different governmental actors and the
Teacher College, however, it was decided to suspend this provision of the law.
The new law also facilitated teacher reallocations and early retirements, as
well as voluntary withdrawals. The last two were fostered by the launching of
a special program whereby teachers were entitled to receive an indemnity cor-
related with their years of service in the municipal sector, and the Ministry of
Education committed itself to sharing the indemnity cost with the municipal-
ity. Most municipalities adhered to this program, which was renewed to the
end of 1997.
• hiring fixed-term teachers (as opposed to permanent ones, hired through a competitive exam) subject to private law and not to public sector law, up to a limit of 20 percent of the total teacher-hours (Art 26 of Law 19.070 and Law 19.410).

As far as training is concerned, I have stuck to an expenditure indicator reporting the amount of funds spent on training over total working capital expenditures (referred to as TRAINING), which includes all types of expenditures on training (for example, on grants, local courses, and special agreements). This is probably not the best indicator, but it is the only one that could be constructed, given the data. This indicator makes it at least possible to determine whether decentralization led to extra-training and whether this extra-training had a positive impact on educational achievement. Insofar as training has a potentially important impact on educational achievement through its impact on teachers' skills, I might hope the policy would be widely applied and its effectiveness enhanced by local design (programs might be designed to respond to the particular skills requirements of the teachers in that area).

Turning to wages, I have adopted an indicator that measures the average wage incentive rate (in terms of the base salary) that was applied, referred to as WAGE_INCENTIVE. I might expect wage incentives to have an important impact on teacher motivation in the Chilean case because of the low wages (especially low for young teachers), which would justify their extensive application. Above all, I might expect, as in the training case, their effectiveness would be enhanced by the application of local criteria.

Probably the most important labor decentralization measure that can be applied by the municipalities under the strict Estatuto Docente is the mix between fixed-term and permanent contracts. Three main reasons explain why having a fair proportion of fixed-term teachers in the Chilean context might be desirable. First, introducing more flexibility in the administration of the teaching staff, because these teachers are not subject to the public sector law, should make it possible to have a staff responding better to local needs. In other words, it makes it easier to adapt the composition of the teaching staff to local needs. Second, municipalities with more fixed-term contracts should be able

---

12. Even if, ultimately, the ideal proportion depends on the local needs and conditions faced by each local area.
to adjust the number of teachers and hours to the number of students more easily. Third, the possibility of easy dismissal makes the fixed-term teachers more accountable to parents and students than permanent teachers, with possibly a positive impact on their work commitment. The indicator of fixed-term teachers that was used, referred to as FIXED_TERM, was constructed as the proportion of fixed-term teachers over the total teachers hired.

*Decentralization in the Pedagogical, Curricular, and Planning Area*

Up to 1992, the Chilean decentralization was above all an administrative one with little planning, curricular, and pedagogical autonomy transferred to the regional, municipal, or school level. The contents of what had to be taught and the decision over the main instructional inputs, as well as the organization of the schooling time and the educational objectives, were dictated by the center, with little flexibility left to the other levels. From 1992 onwards, the decentralization process deepened along the planning, curricular, and pedagogical dimensions, which involved the schools above all, and which initially bypassed the municipalities. In 1992, the so-called *Proyectos de Mejoramiento Educativo* (PMEs), or Educational Improvement Projects, were launched, with the objective of fostering curricular and, above

---

13. Much depends, however, on the motivations that lead municipalities to incorporate fixed-term teachers into their staff. Less good reasons to do so exist (for example, search for cheaper workforce, and unwillingness to face the longer and more complicated process of filling permanent positions). Another limitation is related to the fact that, by law, fixed-term teachers cannot be contracted for more than one or two years in a row and should preferably be appointed to teach technical, innovative, optional subjects or to replace permanent teachers. This would seem to restrict the utility of this measure of flexibility. In practice, however, the condition about the type of functions to be assigned to it is interpreted so broadly that fixed-term teachers can be used in all types of subjects, even if for relatively short periods.

14. The Ministry of Education set the national curriculum, as well as the main objectives for primary and secondary education in 1980 (with Decree 4002/80) and 1981 (with Decree 300/81), respectively.

15. In 1995, however, the process was partially reversed with Law 19.410; this law made it compulsory for all municipalities to prepare a yearly local education plan (called PADEM) and encouraged municipalities to adopt autonomous goals, developing their sense of initiative.
all, pedagogical innovation at the school level. The central idea of the program is to finance projects (such as PMEs), at the school level, which lead to an improvement in the quality and equity of education. Typically, these projects should consist of implementing innovations in pedagogical practices applied to one or several subjects (above all, Spanish) which, in some cases, can also be extended to the syllabus of these same subjects, and which can involve the use of new infrastructure (for example, laboratories and libraries) and teaching equipment and materials (for example, audiovisual devices and new textbooks). Their main peculiarity is that they have to be designed and implemented by the schools' teaching staff in an autonomous way and, as such, are a clear example of educational decentralization.¹⁶

Inducing the schools to compete for the financing of a PME is an indirect way of encouraging a better fit between supply and demand for the municipalities, as schools would be applying pedagogical practices (and other innovations) designed by their own teaching staff on the basis of their pupils' needs and characteristics, in contrast to traditional centrally suggested practices. This improved fit might rest on the adoption of interactive and participatory pedagogical practices, an increase of learning time, or other individual practices. Municipalities with more schools (or students) covered by a PME can be said to be more pedagogically decentralized than municipalities with a smaller school (or student) coverage. The indicator of pedagogical and curricular autonomy based on the implementation of these projects that I adopted was based on the implementation of these projects, and consisted of the proportion of students of municipal schools covered by a PME (referred to as PEDAGOGICAL_AUTONOMY). This indicator made it possible to determine whether municipalities with higher coverage perform better than those with lower coverage, on the basis of the comparative performance of "decentralized" schools (that is, schools with PMEs) versus centralized ones (that is, without PMEs).

In the same area, and beyond the above-mentioned indicator, a measure of "restricted" school autonomy (see above) was provided by the school involvement in the local decisionmaking process, including the selection of the main local education objectives and instruments to

¹⁶ A yearly selection process carried out by the Ministry of Education determines the projects that will receive financing according to the availability of funds, the quality of the project, and the level of deprivation of the school.
reach them. The index of school involvement was constructed by aggregating the same three indicators used above to create the variable INVOLVEMENT_INDEX: the number of school participation mechanisms at the local level, the number of meetings, and the contrasting perceptions of the education department directors and schools' directors on the degree of involvement of the school in the decisionmaking process. The resulting composite index was named INVOLVEMENT_INDEX2.17

**Estimation Results**

Table 1 provides a synthesis of the main variables that were constructed for the estimation of the equations (only the variables included in the ordinary least squares (OLS) estimations presented in table 2 are included), and indicates units of measure and data sources. In table 2, the empirical estimation of the CS (equation (4)) and VA (equation (5)) equations are shown, by OLS. The specification of the CS included in the table comprises a few insignificant variables that are included just for comparison with the VA specification. Because the residuals appeared to be normally distributed, but slightly heteroscedastic, I replaced the traditional standard errors with White's heteroscedastic consistent standard errors in the estimation.18 19 Equation (5) was estimated using two different specifications: a parsimonious one and a slightly more general one for comparison with the CS results. They are shown in columns (2) and (3). The two VA models were found to have residuals normally distributed and homoscedastic. A semilog specification was also attempted. Since this did not produce any significant change in the results, I stuck to the initial specification. The calculations of table 2 were obtained by including only the decentralization indexes, which turned out to be significant in the explana-

---

17. Constructed as a simple average of the standardized version of each single indicator.
18. The joint skewness and kurtosis test for normality yielded a $\chi^2$-statistic of 0.26 insignificant at whatever level of significance (Pr($\chi^2$) = 0.87), meaning that the normality assumption was accepted.
19. I performed the Cook and Weisberg test for heteroscedasticity. The test yielded a $\chi^2$-statistic of 2.98 significant at 10 percent, meaning that I could not accept the hypothesis of constant variance of the residuals at that level of significance.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable definition</th>
<th>Construction and unit of measure</th>
<th>Time span</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCORES96</td>
<td>Fourth year SIMCE test score(^a) (only municipal schools) in Spanish and mathematics (simple average of the two)</td>
<td>Weighted average of school scores constructed as the percentage of correct answers on total valid answers. Min:0; Max:100</td>
<td>1996</td>
<td>Ministry of Education</td>
</tr>
<tr>
<td>SCORES92</td>
<td></td>
<td></td>
<td>1992</td>
<td>(SIMCE department)</td>
</tr>
<tr>
<td>PRIV_SHARE</td>
<td>Private subsidized sector market share (primary education)</td>
<td>Student enrollment in the private subsidized sector as a proportion of total enrollment (%)</td>
<td>1996</td>
<td>Ministry of Education</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1993</td>
<td>(Enrollment Data Base)</td>
</tr>
<tr>
<td>DEPRIVATION</td>
<td>Index of school deprivation (only municipal schools)</td>
<td>Weighted average of schools' indexes constructed as a weighted average of years of mothers' schooling and anthropometric measures of first-year students(^b) Min: 0 = less deprived; Max: 100 = more deprived</td>
<td>1996</td>
<td>Ministry of education</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1993</td>
<td>(&quot;JUNAEB&quot; department)</td>
</tr>
<tr>
<td>MSIZE, MSIZE</td>
<td>Binary dummy capturing the size of the municipality</td>
<td>MSIZE: 1 = large; 0 = other MSIZE: 1 = small; 0 = other(^c)</td>
<td>Fixed from</td>
<td>Chilean Institute of National Statistics (INE)</td>
</tr>
<tr>
<td>MGEOLS, MGEOLN</td>
<td>Binary dummy capturing the geographic location of the municipality</td>
<td>MGEOLS: 1 = south; 0 = other MGEOLN: 1 = north; 0 = other(^d)</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>Pupil-to-teacher ratio per municipality (primary education)</td>
<td>Proportion of pupils per teacher</td>
<td>1996</td>
<td>Ministry of Education</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1994</td>
<td>(Teacher Census) and ad hoc survey</td>
</tr>
</tbody>
</table>

(Table continues on the following page.)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable definition</th>
<th>Construction and unit of measure</th>
<th>Time span</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIXED_TERM</td>
<td>Fixed-term municipal teaching staff (idem)</td>
<td>Fixed-term teachers as a proportion of total teachers (%)</td>
<td>1996, 1993</td>
<td>Ad hoc survey and Ministry of Education</td>
</tr>
<tr>
<td>TRAINING</td>
<td>Local training expenditure ratio (municipal sector, all levels)</td>
<td>Training expenditures over total working capital expenditures (%)</td>
<td>Aver. (92-96), Aver. (93-95)</td>
<td>“SUBDERE”e</td>
</tr>
<tr>
<td>WAGE_INCENTIVE</td>
<td>Local wage incentive (municipal schools)</td>
<td>Average wage incentive rate as a proportion of the base salary (%)</td>
<td>1996, 1995</td>
<td>Ad hoc survey</td>
</tr>
<tr>
<td>SCHOOL_HOURS</td>
<td>Dummy variable capturing the number of learning hours (primary education-municipal schools)</td>
<td>1 = 28 hours, or less, a week; 0 = other</td>
<td>1996</td>
<td>Ministry of Education (Teacher Census)</td>
</tr>
<tr>
<td>INFRASTRUCT</td>
<td>State of school infrastructure (municipal schools) per municipality</td>
<td>Schools in good state as a proportion of total schools (%)</td>
<td>1996, 1992</td>
<td>Ad hoc survey</td>
</tr>
<tr>
<td>P900COV</td>
<td>Schools covered by the P900 program (municipal schools)</td>
<td>Municipal schools covered as a proportion of total schools (%)</td>
<td>1995</td>
<td>Ministry of Education</td>
</tr>
<tr>
<td>TOT_EXP</td>
<td>Total education expenditure per pupil (municipal schools, all levels)</td>
<td>Education expenditures of the municipal sector divided by the total students in the sector. Chilean pesos per pupil at 1990 prices</td>
<td>1996, 1993</td>
<td>Ad hoc survey and “SUBDERE”e</td>
</tr>
<tr>
<td>PEDAGOGICAL AUTONOMY</td>
<td>Proportion of students covered by a project of pedagogical and curricular innovation (PME) (municipal schools)</td>
<td>Percentage of primary education students covered by a PME (%)</td>
<td>1995</td>
<td>Ad hoc survey and Ministry of Education</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------</td>
<td>-------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>INVOLVEMENT INDEX</td>
<td>Composite index on school involvement in local financing decisions</td>
<td>See subsection, Decentralization in the Financing Area (composite index. Max = 100; Min = 1)</td>
<td>1996</td>
<td>Ad hoc survey</td>
</tr>
<tr>
<td>MATER_WAIT</td>
<td>School waiting time for getting teaching materials from the municipality</td>
<td>Average number of days</td>
<td>1996</td>
<td>Ad hoc survey</td>
</tr>
</tbody>
</table>

a. The SIMCE test is a standardized test administered to all the country's schools, with the exception of small multigrade schools located in the rural areas. It covers two main subjects (Spanish and mathematics) and applies, on alternate years, to the fourth and eighth grades of primary education.

b. More precisely, the composite index was constructed as a standardized weighted average of the percentage of mothers with eight and fewer schooling years (with weight = 3.03), the percentage of students with dental cavities (with weight = 1.03), the percentage of students with the ratio of height to age below 1 s.d. from the reference measure established by the Chilean Institute of Health (with weight = 0.46), the percentage of students with the ratio of weight to age below 1 s.d. from the reference measure established by the Chilean Institute of Health (with weight = 0.41), and the percentage from the rural sector (with weight = 1.03). The index was constructed according to a similar methodology since 1993, which means that the index is comparable over the period 1993-96, but not before.

c. Large = equal or more than 100,000 inhabitants; small = equal to or fewer than 20,000 inhabitants.

d. South = regions VII and VIII; north = regions I, II, III, IV.

e. Regional Development Sub-Secretary: institution depending on the Ministry of Domestic Affairs, which coordinates the activity of the country's municipalities. Among other things, it collects and revises the municipal balance sheets. All expenditure data came either directly from the ad hoc survey or from the municipal balance sheets of the education sector gathered by the sub-secretary.
## Table 2. OLS Estimation (Dependent Variable: SCORES96)

<table>
<thead>
<tr>
<th>Variables</th>
<th>CS model (robust OLS)</th>
<th>VA model (OLS)</th>
<th>VA model (OLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (t-ratios)</td>
<td>Coefficient (t-ratios)</td>
<td>Coefficient (t-ratio)</td>
</tr>
<tr>
<td>SCORES92</td>
<td>-</td>
<td>0.47</td>
<td>0.295</td>
</tr>
<tr>
<td>PRIV_SHARE</td>
<td>-0.114</td>
<td>-0.071</td>
<td>-0.094</td>
</tr>
<tr>
<td>DEPRIVATION</td>
<td>-0.073</td>
<td>-</td>
<td>-0.043</td>
</tr>
<tr>
<td>MSIZEL</td>
<td>0.456</td>
<td>1.99</td>
<td>1.632</td>
</tr>
<tr>
<td>MGEOLS</td>
<td>0.681</td>
<td>2.22</td>
<td>1.849</td>
</tr>
<tr>
<td>INFRASTRUCT</td>
<td>0.061</td>
<td>0.025</td>
<td>0.038</td>
</tr>
<tr>
<td>P900COV</td>
<td>-0.060</td>
<td>-</td>
<td>-0.027</td>
</tr>
<tr>
<td>MATER_WAIT</td>
<td>-0.014</td>
<td>-0.048</td>
<td>-0.038</td>
</tr>
<tr>
<td>FIXED_TERM</td>
<td>-0.177</td>
<td>-0.091</td>
<td>-0.082</td>
</tr>
<tr>
<td>TRAINING</td>
<td>0.886</td>
<td>0.60</td>
<td>0.869</td>
</tr>
<tr>
<td>PT</td>
<td>-0.183</td>
<td>-0.37</td>
<td>-0.39</td>
</tr>
<tr>
<td>WAGE_INCENTIVE</td>
<td>0.058</td>
<td>0.064</td>
<td>0.056</td>
</tr>
<tr>
<td>PEDAGOGICAL</td>
<td>0.027</td>
<td>0.052</td>
<td>0.042</td>
</tr>
<tr>
<td>AUTONOMY</td>
<td>0.027</td>
<td>0.052</td>
<td>0.042</td>
</tr>
<tr>
<td>SCHOOL_HOURS</td>
<td>-3.50</td>
<td>-</td>
<td>-1.68</td>
</tr>
<tr>
<td>TOT_EXP</td>
<td>-0.046</td>
<td>-0.093</td>
<td>-0.089</td>
</tr>
<tr>
<td>INVOLVEMENT</td>
<td>0.0530</td>
<td>0.0366</td>
<td>0.0476</td>
</tr>
<tr>
<td>R2</td>
<td>0.74</td>
<td>0.74</td>
<td>0.76</td>
</tr>
<tr>
<td>Adj-R2</td>
<td>-</td>
<td>0.64</td>
<td>0.65</td>
</tr>
</tbody>
</table>

- Not applicable.

a. All 1996 values, except 1995 for P900COV and PEDAGOGICAL_AUTONOMY, and average of 1992–96 for TRAINING.

b. All 1993 values except: 1992 for INFRASTRUCT, MATER_WAIT, and INVOLVEMENT_INDEX; 1995 for WAGE_INCENTIVE, PEDAGOGICAL_AUTONOMY, and P900COV (corrected for nonrenewed projects before 92: P900COV952); average of 1993–95 for TRAINING; and 1996 for SCHOOL_HOURS.

* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.
tion of educational achievement, in contrast to previous, more general CS estimations that included all decentralization indicators, as well as some other variables, with the purpose of explaining how the different indicators work (see Di Gropello 2001).20

Several main points can be made in interpretation of these results. First, on the VA estimation, even in the parsimonious model, it is quite clear that the coefficient on SCORES92 is smaller than 1, which justifies the decision to introduce past achievement among the explanatory variables.21

Second, the comparison between the CS and VA estimations shows that—even if the inputs before 1993 seem to explain present achievement, as indicated by the fact that the coefficient on the past SIMCE score introduced in the VA is significantly positive—the coefficients of the other inputs included in all VA specifications maintain, in general, a strong significant explanatory power, which confirms the CS results. This suggests that the determinants of educational achievement changed somewhat in the period under analysis compared with the previous periods; otherwise, there would likely have been a larger drop in the coefficients and level of significance of the regressors. I was, in fact, expecting such a change, considering all the innovations that occurred in the period under analysis.

Third, in general, the inclusion of past achievement leads to a slight decrease in the size and level of significance of the coefficients of most of the inputs. The magnitude of the decrease is quite heterogeneous across the variables, and it depends, ultimately, on the intensity and sign of the correlation between the included variables and past educational achievement. This in turn depends on the intensity of the correlation between present and past inputs, on the evolution over time of the association between the included variable and the outcome, and on

20. In fact, a composite index constructed on the basis of the index of school involvement in the financing area and school involvement in the planning area might have been included instead of the mere index in the financing area; this would have produced a slightly higher coefficient and the same t-ratio. The very small difference led me to stick to a simpler indicator of easier interpretation. We should keep in mind, however, that the extent of school involvement in the planning area has effects very similar (even if slightly weaker) to the one in the financing area.

21. The hypotheses that the coefficient on SCORES92 = 1 was rejected at every level of significance (F(1,36)=47).
the possible correlation of both past achievement and the included variable with a common unobservable variable.\footnote{In some cases, SCORES92 might proxy just itself (and not some unobservable correlated characteristic) if the variables have been “allocated” according to the past quality of education (see, for instance, the variable \textsc{P900COV}).} As far as individual heterogeneity is concerned, however, some clarification should be made. As noted above, one of the advantages of VA models is that, under certain assumptions, they should be able to remove time-invariant unobservables from the error term. These assumptions are, however, quite restrictive.\footnote{Referring to the VA version used here (see equation (3)), this means assuming either of two possibilities. One is that the impact of unobservables is fixed over time (that is, assuming the time stationarity of the coefficient of the linear projection of educational achievement on the individual effect) and that the unobservable drops out because of the collinearity with the included past achievement. A second possibility is that, even if the impact of unobservables changes over time, the effect of the unobservable on achievement in different periods follows a geometric pattern similar to the effect of the time-variant independent regressors that are included (this second example is illustrated by Boardman and Murname (1979)).} Ultimately, I want the unobservable to affect the level, but not the change in the outcome. Apart from this last very specific case, this will generally be true if the unobservable is correlated, with the same sign and intensity, with both the present and past level of achievement. Only then, by introducing past achievement, can I effectively control for the unobserved variable and break the correlation between the explanatory variable included and the error term. This condition might seem plausible for some unobservables, but not for others.\footnote{In this case, considering that I compared educational achievement at the same level over two points in time, it might be very plausible that some unobservables, such as teacher motivation, affect both educational achievement functions in a similarly strong way (assuming that the impact of teacher motivation on educational achievement differs across levels but not years), and it might even be possible to find unobservables whose effect fades in time (for example, some initial advantage of some municipalities in managing local issues because of the different quality of previous administrators). Other unobservables, however, might have had a changing pattern and effect over time, which would have made it unlikely to control for them through the inclusion of past achievement in the regression.} In the following discussion, we have to keep in mind that all unobservables are not necessarily controlled for through the inclu-
sion of past achievement and that the results I got must still be interpreted with caution. Concentrating more specifically on the decentralization indicators, we see the following.

**Local Financial Autonomy in Education**

The indicator of local financial autonomy was found to be negatively but insignificantly related to educational achievement in previous CS analysis. The main reason for this is that, as was detected through auxiliary regressions, FINANCIAL_AUTONOMY works completely through other included regressors. In particular, it works through the expenditure per student (higher in municipalities with higher shares of local funds), the proportion of fixed-term teachers, and the use of wage incentives (both higher as well in municipalities with higher shares of local funds). Now, as we will see, both expenditure and fixed-term teachers are negatively related to educational achievement, which explains the negative sign on the financial autonomy variable, and which, in the Chilean context, is more indicative of policy constraints (financial and institutional) and scarce municipal capacity than of autonomy and local innovation.

**School Involvement in Local Decisionmaking**

The variable measuring the intensity of schools' decisionmaking in the financing decisions at the municipal level, INVOLVEMENT_INDEX, is positively related to educational achievement in all estimations. Its level of significance and the size of its coefficient are slightly lower in the VA. The positive correlation between the involvement index and past achievement, which leads to a decrease in the coefficient of the former variable, might be explained by the correlation of SCORES92 with an unobservable municipal characteristic also correlated with the index (for instance, skills and sense of leadership of local schools' directors), but is more likely to be caused by the strong correlation between past and present participation within the framework of a relatively stable relation between participation and educational achievement. In quantitative terms, an increase of one standard deviation of

---

26. In fact, a survey made by Rounds (1994a) in Chile shows that the directors with more sense of leadership tend to be more involved in municipal issues.
the 1996 involvement index (or 1.1 standard deviation of the 1992 index) leads to an increase of between 0.7 and 1.1 points in test scores, equivalent to, respectively, 18 percent and 27 percent of one standard deviation. These results indicate that INVOLVEMENT_INDEX works only to a small extent through the included regressors. It is likely to capture, in fact, the impact that school involvement has on the type and quality of the specific investment programs designed and implemented by the local education administrators.

Local Training and Wage Incentives

Both TRAINING and WAGE_INCENTIVE are positively related to educational achievement. Their coefficients and level of significance are stable across all specifications. This is explained by the fact that local training expenditures and wage incentive practices really developed only from 1993 onwards (with the Estatuto Docente and the newly elected mayors in 1992), explaining the low or very low correlation with the inputs included in SCORES92. In quantitative terms, an increase of one standard deviation (or 1.2 percentage points) in TRAINING leads to an increase between 1.1 and 0.7 points in test scores, equivalent to, respectively, 27 percent and 17 percent of one standard deviation. An increase of one standard deviation (or 15 percentage points) in WAGE_INCENTIVE leads to an increase of between 1.5 and 0.8 points in test scores, equivalent to, respectively, 37 percent and 20 percent of one standard deviation.

The sign on the TRAINING coefficient is the expected one because this ratio measures the priority attributed by each municipality to training expenditures and, supposedly, a higher priority should be related to a higher level of educational achievement through a better-trained teaching staff.\textsuperscript{27} Evidence is scarce on in-service training, but it points, in general, to a positive impact of the variable on educational achievement.\textsuperscript{28} The results seem to suggest that the variable really proxies for the impact of trained teachers on educational achievement and not for the impact of some municipal characteristic, such as a

\textsuperscript{27} Worked out over the period 1992–96 in the CS to correct for data volatility (1993–95 in the VA).

\textsuperscript{28} Fuller and Clarke (1994) found evidence of significant positive impact of in-service training on educational achievement in 8 of the 13 studies reviewed that included that variable.
sense of innovation generated by some specific cultural feature, on outcome. Otherwise, I would have expected a decrease in the coefficient following the inclusion of SCORES92 (assumed to be correlated with this sense of innovation).

The sign on the wage incentive variable, too, is the expected one, because I would expect wage incentives to have a positive impact on educational achievement through its positive impact on teacher motivation. The previous evidence on the positive impact of wage incentives is weak. Teacher salaries are found to be significantly positively related to educational achievement in only 9 of the 60 studies, that include the wage variable, surveyed by Hanushek (1986), and in 4 of the 11 studies, that include that variable, surveyed by Fuller and Clarke (1994). The fact that local wage incentives are mainly allocated among teachers according to a merit criterion, encouraging them to improve the level of their performances, might explain the positive impact of wage incentives in the Chilean case; the very low initial salaries may also play a role.

Local Proportion of Fixed-Term Teachers

In all estimations, FIXED_TERM is statistically negatively related to educational achievement, even if in the VA estimation, its coefficient and level of significance drop slightly. Quantitatively, an increase of 10 percentage points in the proportion of fixed-term teachers (equivalent to one standard deviation in 1993 and to 1.1 of one standard deviation in 1996) would lead to a decrease in test scores of between 1.8 points and 0.82 points, equivalent, respectively, to 45 percent and 20 percent of one standard deviation. The negative impact of the variable might seem surprising, but it actually makes sense. The fixed-term variable works in part through some excluded identified regressors and in part through unobservable variables, with both effects leading to a negative relationship with educational achievement. Previous CS estimations indicated that the proportion of fixed-term variables is negatively correlated with teachers' education and teachers' years of experience, both variables that are positively correlated

---

29. Considering the strong correlation between FIXED_TERM93 and FIXED_TERM92, included in SCORES92, the reduction is not surprising. Some changing impact of FIXED_TERM on educational achievement could have occurred to explain the persistent significant negative sign.
with educational achievement. As long as FIXED_TERM is associated with low-experience teachers who are, additionally, less educated than the average, we should not be surprised that it has a negative impact on educational achievement.

Given this, I should point out that FIXED_TERM remained significant in the original general CS even with the inclusion of the variables measuring teacher's education and experience, which means that it must also work on educational achievement through some unobserved variable. It is very plausible that teacher motivation and sense of duty might be an important unobservable captured by the fixed-term variable. Hiring teachers on a fixed-term basis might help to solve the chronic problems of inadequate resources by making the teaching staff more mobile, but the employment status of these teachers might also reduce their commitment to their task and involvement with the children, knowing that they will not stick with the same class for more than 1 or 2 years. In fact, hiring teachers on a fixed-term basis instead of on a permanent one might have negative consequences on educational achievement through two channels. The first is increased turnover, which makes the pupil-teacher relationship by definition more volatile; the second is the impact of this turnover on teachers, who will be less prone to commit completely to their educative task.

**School Pedagogical Decentralization**

PEDAGOGICAL_AUTONOMY, the variable measuring the school coverage of PMEs, is very close to the 10 percent level of significance in explaining educational quality in the CS, but it shows a level of significance of 5 percent in the VA (with a coefficient almost doubling). The increased impact in the VA is caused by the fact that, even if the

---

30. We can see this effect when we add these two variables to, and exclude FIXED_TERM from, a more general version of the CS (the high collinearity between the three variables led me to include only FIXED_TERM, more significant than the two other variables, in the parsimonious model).

31. No negative correlation was found between PT and FIXED_TERM, which indicated that fixed-term teachers typically are not extra teachers hired for introducing new subjects, but rather are hired instead of permanent teachers. The positive relationship between these two variables indicates, in fact, that fixed-term teachers have been used as an adjusting device.
PME program was not a targeted one, low-quality schools and municipalities were more motivated to participate in it (pushed, moreover, by the Ministry of Education to do so), explaining the negative correlation between SCORES92 and PEDAGOGICAL_AUTONOMY, which produces a bias in the CS results. In other words, once we control for the fact that the pedagogical autonomy program was not allocated randomly, but rather favored slightly the areas with lower educational quality (that is, lower past educational inputs), the variable becomes significant at 5 percent in explaining educational achievement. In quantitative terms, in the VA, an increase of one standard deviation in the pedagogical autonomy variable (or 22 percentage points) leads to an increase between 0.9 and 1.1 points in test scores, equivalent, respectively, to 23 percent and 28 percent of one standard deviation. These results suggest that schools' own initiatives in the design and implementation of pedagogical practices and, to a lesser extent, curricular innovations, have a positive impact on educational achievement. As shown in previous CS analysis, the pedagogical autonomy variable mainly captures unmeasured pedagogical and curricular practices, designed according to local needs and involving, in many cases, high levels of interaction between teachers and students, which has a positive impact on educational achievement.

Finally, for clarity, I should add a word on some of the control variables.

---

32. Corr (PEDAGOGICAL_AUTONOMY, SCORES90) = -0.12, Corr (PEDAGOGICAL_AUTONOMY, SCORES92) = -0.27.
33. See Pitt, Rosenzweig, and Gibbons (1993) and Rosenzweig and Wolpin (1986) for a discussion and formal treatment of the issue of nonrandom allocation of health and education programs.
34. Little evidence exists on the impact of curricular and pedagogical practices in general on educational achievement, but some evidence is available on the impact of active pedagogical practices. Arancibia (1996) found that active and innovative pedagogical techniques have a positive effect on educational achievement in five of the six studies surveyed. Fuller and Clarke (1994) found that an active, complex pedagogy has a positive impact on educational achievement in three of the eight studies surveyed. Finally, Wolff, Schiefelbein, and Valenzuela (1994) found that active pedagogical strategies have a positive impact on educational achievement in six of the nine studies reviewed.
School Deprivation and P900 Program School Coverage

The index of school deprivation (DEPRIVATION) provides a general measure of the socioeconomic status of the students attending public sector schools. As indicated in table 1, DEPRIVATION is a composite index that combines the schooling of the students' mothers with some health status measures of these students and the rural-urban location of the school, which gives a higher weight to the first variable. This deprivation index should have a strong direct and indirect impact on the average SIMCE score. The direct impact works through the strong proved relation between students' learning environments and educational achievement. The indirect one works through the relation between family background and school characteristics, insofar as the organization and quality of the local education sector is generally expected to reflect the surrounding socioeconomic level of families. As expected, the index has a negative significant sign in the CS, indicating that the higher the level of deprivation of municipal schools, the lower the SIMCE score. The impact of this variable (which, according to individual studies of educational achievement, would be expected to be the stronger predictor of test scores) is, however, weaker than expected. Part of the explanation for this smaller-than-expected impact might be found in the aggregation bias involved in testing the impact of individual and school-level data at a more aggregated level. Problems of omitted variable bias tend to increase along with the level of aggregation because,

35. As a proxy for the learning environment of the children taking the SIMCE test in 1996, DEPRIVATION93 (constructed on the basis of information covering first-year students in 1993) would be more appropriate than DEPRIVATION96, which would capture exactly the socioeconomic status of that cohort of students. However, I am also interested in measuring the impact of the current socioeconomic status of families on educational achievement through its possible impact on the current quality of the education sector. As it turned out, replacing the 1996 index with the 1993 index did not produce any significant difference in any of the coefficients.

36. Attempts to plug indicators of rural population ratios, poverty ratios, and human development indexes per municipality gave much less significant results.

37. As shown by most studies estimating education production functions at the individual level. See Arancibia (1996) for an extensive survey showing, among other factors, the impact of the characteristics of children and families on educational achievement.
typically, individual, school, and district or state variables tend to be cruder in aggregate studies than in micro ones. This leads to the exclusion of some important variables that produce a bias in the coefficients of the variables included. In this case, it might well be the case that socio-economic status (SES) has not been completely accounted for (meaning that the deprivation index used here is a poor measure of SES) and that the included regressors pick up some of the effect of SES.

Another explanation for this smaller-than-expected impact, however, relates to the inclusion of the variable P900COV in the regression and its impact on the deprivation index. P900COV measures the amount of schools of the municipality covered by the P900 program, a program that was introduced at the beginning of the 1990s to provide schools with school equipment (specifically libraries) and instructional materials. Because one of the main purposes of the program was to provide the most deprived and low-quality schools with these facilities, the CS setting makes it impossible to disentangle the impact of deprived and low-quality schools from the impact of the availability of facilities on educational outcomes. P900COV becomes another measure of deprivation and, above all, low-quality schools (schools with very low past SIMCE scores), explaining its negative sign and the collinearity with the deprivation index, which leads to a decrease in the coefficient of this latter variable. Overall, a one standard deviation increase in both DEPRIVATION and P900COV leads to a combined reduction of test scores of 2.5 points, equivalent to more than 60 percent of one standard deviation. This would also imply that, through these two variables, school socioeconomic status should, eventually, be adequately captured, making it unlikely that the decentralization indicators would pick up the effect of SES.

In the VA, both variables lose significance. In both cases, this is not surprising. Because of the slow evolution in time of this index, DEPRIVATION93, used in the VA, proxies quite well for DEPRIVATION89 (which measures the level of deprivation of the student cohort passing the SIMCE in 1992 and, thus, captured by SCORES92), becoming, consequently, insignificant with the inclusion of SCORES92. Once we control for the negative correlation between the past achievement score and P900COV through the inclusion of SCORES92, P900COV becomes insignificant, and loses its function of being a proxy for deprived and low-quality schools.\(^{38}\)

\(^{38}\) In particular, \(\text{Corr (P900COV, SCORES90)} = -0.42\) and \(\text{Corr (P900COV, SCORES92)} = -0.48.\)
Proportion of Subsidized Private Sector Enrollment

The negative significant impact of PRIV_SHARE on educational achievement in all estimations indicates that municipalities with lower proportions of subsidized private sector enrollment over total enrollment outperform municipalities with higher ratios. The robustness of the results to the VA estimation (with 1993 data) indicates that no reverse causality is to be feared, but that this variable is capturing the impact of poor socioeconomic background and motivation in the public sector schools on educational quality.\textsuperscript{39} In other words, at the same level of school deprivation (which is negatively correlated with PRIV_SHARE because municipalities with higher private market shares are less poor, on average, than municipalities with lower shares), municipalities with higher private sector market shares would have a less favorable composition in terms of socioeconomic background and motivation of their public sector students than municipalities with a lower relative share.\textsuperscript{40}

Total Expenditure in Education per Student

The significant negative impact of the expenditure variable (TOT_EXP) in all estimations\textsuperscript{41}, including the VA model with 1993 data, indicates that reverse causation has to be ruled out. The likely

\textsuperscript{39} The larger the proportion of subsidized private sector schools, the easier it is for the sector to attract students from the middle and even lower-middle classes which, otherwise, would have stuck to the municipal schools. Along the same line of reasoning, private subsidized schools would also attract the most motivated students (or at least the ones with the most motivated parents) of the lower-middle and lower classes which, again, would otherwise have stuck to the public sector schools.

\textsuperscript{40} This type of explanation was also put forward by McEwan and Carnoy (1999) in their analysis of the impact of competition on public school quality in Chile. In their school level analysis at the national level, they found the private enrollment share to have a slight negative impact on SIMCE scores in their first-difference regression, and mentioned that part of the explanation might reside in the large-scale sorting of students across public and private schools that occurred in Chile.

\textsuperscript{41} This finding of a negative effect is uncommon, although a majority of studies has not found positive significant effects of expenditure per pupil on educational achievement, but rather insignificant positive or negative effects.
explanation for this unexpected result is that from the early 1990s, higher levels of total expenditure were increasingly used to pay seniority benefits of the aging staff. As long as they are associated with teachers with low levels of motivation because they are waiting to retire, such seniority benefits will be negatively related to educational achievement. In other words, in my specifications, total expenditure would be capturing the impact of teacher motivation on educational achievement, which, because it is time-variant during the period under analysis, is not picked up by SCORES92.

Conclusions

The impact of decentralization on educational achievement is not clear-cut. Contrary to expectations, financial autonomy and some measures of labor autonomy turn out to have a negative impact on educational achievement, whereas some other measures of labor autonomy, as well as pedagogical and curricular autonomy and school involvement in local financing issues, turn out to have a positive impact. However, the negative impact of the first set of measures seems to be mainly caused by a combination of factors that should not be too difficult to modify, which suggests that the negative results are not irreversible. Additionally, all estimates, positive or negative, require caution in interpretation for at least three main reasons: the small sample size, the lack or limited reliability of past data, and, in some cases, the aggregation bias involved in testing the impact of individual and school-level data at

Hanushek (1986), for instance, found the expenditure per pupil variable to be significantly positively related to educational achievement in only 13 of the 65 studies surveyed (with 16 cases significant or insignificantly negatively related).

42. As was shown in a detailed CS analysis of the determinants of education expenditure in Chile. See Di Gropello (2001).

43. The Estatuto Docente of 1991 led to an improvement in the employment status of teachers and, apart from getting greater job stability, teachers were entitled to higher financial compensations for leaving. Most municipalities could not pay these compensations, so teachers preferred to stay than leave. The wide success of the early retirement program in 1996 made it clear that many senior teachers were in fact waiting for the appropriate compensation to retire.

44. Getting a positive impact of these variables might, however, require local preconditions (skills, information, political will) that are not necessarily present and whose importance could not unfortunately be effectively tested here.
a more aggregated level. Given this, the following summarizes what seem to be some of the main conclusions of the paper.

- The econometric analysis provides some evidence that pedagogical and curricular decentralization at the school level, measured by the proportion of students covered by projects of pedagogical and curricular innovation, has a significant positive impact on educational achievement, once the not entirely random allocation rules of the decentralization projects are controlled for. This evidence is reinforced by the significant positive impact on educational achievement of the level of school involvement in local financing decisions, which can be seen as a “restricted” measure of school autonomy. If we assume that school involvement is largely associated with substantial local initiatives driven by schools, we reach the conclusion that decentralizing initiative to the schools in major areas, including the pedagogical and curricular ones, seems to increase for educational achievement.

- Some econometric evidence was also provided that municipal training expenditure and wage incentives, that is, measures of locally decentralized staff management, have a significant positive impact on educational achievement.

- By contrast, the paper also provides some econometric evidence that another measure of decentralized staff management—that is, the use of fixed-term teachers—generally has a significant negative impact on educational achievement. The analysis has also shown, however, that the negative impact seems to be related to the combination of three main factors: the choice, usually, of inexperienced and uneducated fixed-term teachers; the use of fixed-term teachers as an adjusting device; and the existence of rules that produce a negative impact on teacher motivation, making it impossible to hire a fixed-term teacher for more than 1 or 2 years. These last two factors (and, as a consequence, also the first one) might be changed through modification of the legislation aimed at introducing elements of competition in an otherwise restricted labor market, through extending employment opportunities for successful fixed-term teachers, and through making it possible to fire permanent teachers for a specific set of reasons.\(^4\) Only then can fixed-term

\(^4\) As we have seen, the modifications in the Estatuto Docente are at least going in the direction of facilitating firings.
teachers become a real alternative and the impact of different ratios of centrally and locally ruled teachers be assessed thoroughly, leading to recommendations for labor market reforms.

- There is, as well, some econometric evidence that the level of local financial autonomy is negatively related to educational achievement through its positive relation with total expenditure in education and the proportion of fixed-term teachers. Again, the negative impact of FINANCIAL_AUTONOMY can be explained. In Chile, financial decentralization has mainly been a response to the insolvency problems caused by the extremely high personnel costs that municipalities had to face from the beginning of the 1990s, instead of being dictated by some autonomous decision to mobilize local funds for delivering better education services. As a consequence, the high levels of expenditure associated with local financial decentralization are almost entirely associated with personnel costs that, as seen above, have a significant negative impact on educational achievement. Additionally, within the restrictive legislative framework in place, fixed-term teachers were increasingly used as a flexibility device by municipalities facing high personnel costs, which explained the positive correlation between the proportion of fixed-term teachers and financial decentralization and also the negative impact of the proportion of fixed-term teachers on educational achievement. Ultimately, more financially decentralized municipalities end up being associated with a mix of demotivated senior teachers and young, inexperienced, and similarly demotivated (but for other reasons) teachers which provide the main explanation for the negative relation between financial decentralization and educational achievement.

The main conclusion that can be extracted from this and the previous paragraph is that financial decentralization and partial measures of labor decentralization cannot be really effective if they are promoted within the framework of very rigid employment legislation. The introduction of ad hoc measures of labor autonomy (local wage incentives, training expenditures, and fixed-term contracts) is positive, but a more global and integrated approach to administrative and, particularly, labor practices seems to be needed to ensure a level of flexibility that is conducive to a successful decentralization. Only then will it be possible to use local funds for truly local purposes, including locally designed investment programs, local wage incentives and training expenditures, and local pedagogical and curricular innovations.
References


Who Benefits from Increased Access to Public Services at the Local Level?
A Marginal Benefit Incidence Analysis for Education and Basic Infrastructure

Mohamed Ihsan Ajwad and Quentin Wodon

Abstract

Do poor people benefit more or less than the nonpoor from an expansion in access to public services? And do those benefits depend on the existing level of access? Answering these questions is essential to strategies for empowering (or "investing in") poor people, but the lack of panel data or repeated cross-sectional data in poor countries has often made it impossible. This paper proposes a methodology for answering these questions using data from only a
single cross-section survey. We argue that the methodology may be useful for monitoring the allocation of public expenditures in a context of decentralization, and we demonstrate this by applying it to local-level data from Bolivia and Paraguay. The results indicate that the marginal benefit incidence is higher (or at least not systematically lower) for the poor than for the nonpoor in education, but this is not the case for many basic infrastructure services. More generally, the poor seem to gain access only once the nonpoor already have high levels of access. This suggests that pro-poor policies must be implemented if the poor are to reap the benefits of gains in access faster.

Latin America has made substantial movement toward decentralization during the 1990s (Burki, Perry, and Dillinger 1999). As a result, expenditures for education, health, and access to basic infrastructure services tend to be managed more and more at the local level. Argentina and Brazil were among the first countries to decentralize, and other countries have followed suit. The best-known recent example is probably Mexico (Giugale and Webb 2000), although smaller countries, such as Bolivia and Paraguay, have also adopted decentralization laws.

Two of the main arguments in favor of decentralization are related to the ideals of efficiency and empowerment. From an efficiency point of view, it is often argued that local authorities have better information than central governments for deciding what types of programs and policies to implement, and how to target these interventions so that the poor benefit from them. From an empowerment perspective, it is also argued that providing resources and delegating decisions at the local level is good in itself because it lets local communities decide what they want and how to achieve their goals. When mechanisms are designed to channel more resources to poorer municipalities, decentralization has the potential to empower the poor.1

Although the flow of financial resources to local authorities has increased considerably in Latin America over the last decade, good accountability mechanisms by which the allocation of the funds at the local level may be monitored are still missing. In Mexico, for example, allocations to states and municipalities for new basic social infrastructure are now based on a formula that takes into account unmet basic needs. The formula has dramatically increased funding for the poorest

---

1. For a discussion of empowerment in the context of poverty reduction, see World Bank (2002).
states. One remaining challenge, however, is to design appropriate institutional management and control mechanisms to ensure that the funds are well spent. Many local governments lack the expertise and personnel to manage the funds, and few resources have been made available to help them hire new staff, train existing staff, or modernize their administration. Another potential danger lies in the risk of a political use of the funds at the local level, especially in states and municipalities where control mechanisms by civil society are weak.

Another important issue with the trend toward decentralization in Latin America is whether the funds allocated to local authorities benefit the poor, which would "empower" them. To measure who benefits from an increase in access to public services made feasible by the financial transfers to local authorities, it was necessary to conduct a marginal benefit incidence analysis. While traditional benefit incidence analysis provides information on who the current beneficiaries of access to public services are, marginal benefit incidence analysis focuses on the beneficiaries of improvements in access. In principle, to measure the distribution of gains in access, panel data—or at least repeated cross-sectional data—are necessary. In many countries, however, such data are not available, or are not comparable over time. The question, then, is whether marginal benefit incidence can be measured with a single cross-section of data. Following work by Lanjouw and Ravallion (1999), this paper argues that it is indeed feasible to measure marginal benefit incidence with a single cross-section of data. A key difference between this paper and previous work is that within the context of decentralization, we focus on marginal benefit incidence at the local, rather than at the national, level. Another difference is that we analyze marginal benefit incidence in a broader social welfare framework that takes into account relative deprivation, whereby individuals and households assess their level of well-being not only in absolute terms, but also by comparing themselves to others, the "others" being defined here as their geographic neighbors.

Our empirical work is based on household survey data from Bolivia and Paraguay, two countries that made substantial efforts toward decentralization in the 1990s. The administrative structure of Bolivia consists of 9 departments and 311 municipalities. Decentralization has

---

2. For the brief review of the decentralization process in the two countries that follows, we are indebted to Diego Zavaleta for Bolivia and Estanislao Gacitua-Mario for Paraguay.
been promoted in this country through three main laws. First, in 1994, the Popular Participation Law doubled the share of national income channeled to local authorities, and modified the allocation mechanism from a formula based on local tax generation to a distribution according to population. The law also transferred to local authorities the management of the health and educational infrastructure, as well as that of local roads and sanitation systems. Second, the Administrative Decentralization Law adopted in 1995 redefined the departmental level by merging existing public organizations into prefectures. The law also transferred public investment responsibilities and resources to the departments, and it created coordination mechanisms with local (that is, municipal) authorities. Third and last, the National Dialogue Law adopted in July 2001 completed the transfer of the management of current expenditures for education and health to the municipalities. As in Mexico, the law also established a resource allocation criterion whereby municipalities with high rates of poverty receive a larger share of the debt relief transfers provided by the international community to the country as part of its participation in the Highly Indebted and Poor Countries (HIPC) initiative.

In Paraguay, departments and municipalities have also acquired important responsibilities and autonomy. Paraguay is composed of 16 departments, plus the capital city of Asunción, and 220 municipalities. According to the 1992 constitution, departments and municipalities have political, administrative, and financial autonomy. The departmental government consists of a governor and a departmental council (junta departamental) elected by popular vote to serve 5-year terms. The municipal government consists of a mayor (intendente) and a municipal council (junta municipal). The functions of the departments include (a) the coordination with the municipal governments of the delivery of public services, such as water, electricity, and others, that by their characteristics involve more than one municipality; (b) the preparation with the junta departamental of departmental development plans with a budget; and (c) the coordination with the central government of the provision of health and education services. Municipal governments are responsible for urban development and zoning, public education, health, water, sanitation, and social services, as well as the maintenance of municipal roads and public infrastructure.

Because Bolivia and Paraguay have both made important strides in the decentralization process, they represent interesting case studies for analyzing the marginal benefit incidence analysis of public services at the local level. It is important, however, to stress several of the limita-
tions of this paper. The main limitation is that we do not claim that the analysis provided here constitutes a thorough evaluation of the local allocation mechanisms observed in the two countries. A more detailed analysis would have to be undertaken to perform such an evaluation, especially given that there may be a disconnect between the responsibilities granted in principle to local authorities and the reality.³

A second limitation of the paper is that we focus on the measurement of the marginal benefit incidence at the local level rather than on the determinants of the local allocation of resources. As noted by Ajwad and Wodon (2001), a sizable literature explains the allocation of public services across and within jurisdictions. Tiebout (1956) has argued that if the residents of different areas value public services at different levels, varying levels of public provision should be allocated across areas, with voters sorting themselves into areas where the level of public goods and services maximize their utility (for more recent work along these lines, see Brueckner (2000); Hoxby (2000); Behrman and Craig (1987)). An unequal allocation of services between or even within areas (say, by municipality within a department) may also result from assigning weights to different groups in the objective function of local governments (for example, Ravallion and Wodon 2000, Ajwad 1999, Shoup 1989). Another strand of research argues that if the cost of providing public services varies from one area to another, this may also lead to different levels of provision across and within areas (for example, Hoxby 1999; Ajwad and Wodon 2001). This unequal allocation may be observed even if voters are homogenous in their preferences and governments weigh welfare gains equally across regions.

Finally, a cautionary note should be struck about the difference between locally based and nationally based marginal benefit incidence analysis. In general, one cannot assume that the results of a locally based analysis apply at the national level and vice versa. Assume, for example, that the unit of analysis at the local level is the department, such that households are ranked in various income groups (say, quin-

³ In Paraguay, for example, the decentralization process has been hindered by a lack of financial resources, a lack of professional staff, and a lack of clear organic laws. As mentioned earlier, departmental governments should in principle get substantial resources from the central government. In reality however, even though departmental funding has increased, the central government continues to control most of the resources, and transfers at the local level do not necessarily take needs into account.
tiles) within their department. This method of ranking has its benefits in the context of the evaluation of local allocation patterns. It must be noted, however, that although the poorest household in the richest department may be richer than the richest household in the poorest department, they will be treated in the same way in a locally based analysis, which may not be appropriate for an assessment of marginal benefit incidence at the national level. On the other hand, in a decentralized environment, or in a cross-country study, we believe that a local ranking is more appropriate.

An important result is that marginal benefit incidence at the local level appears to be strongly pro-poor only when the level of access (the benefit incidence) is very high. In primary education for example, where access rates are high, the poor do benefit much more than the nonpoor from increases in access. By contrast, for telephones, where access rates remain low, the nonpoor benefit from the bulk of the gains in access. Thus, a threshold effect exists (as pointed out by an anonymous referee), whereby the poor gain in access only once the nonpoor already have fairly high levels of access. This does not imply that local authorities favor the nonpoor: As discussed in Ajwad and Wodon (2001), the observation that, in general, gains in access to education are more pro-poor than gains in access to basic infrastructure is consistent with a policy by local authorities to maximize local access rates (that is, a policy that specifically targets neither the poor nor the non-poor). The results, however, do suggest that active pro-poor policies may be needed if the poor are to reap the benefits of increases in access earlier in the process of expanding access.

The paper is structured as follows. The first section presents a simple social welfare framework in which to consider marginal benefit incidence analysis. Together with a technical appendix, the next section presents the methodology used to estimate the marginal benefit incidence of public services. This is followed by the results for Bolivia and Paraguay. The paper concludes with a summary of the findings.

**Analytical Framework**

In this section, we provide a simple analytical framework for analyzing the inequality in the distribution of access to basic services and the impact on inequality of the distribution of new access. The objective

---

is to provide summary statistics to identify the current beneficiaries of access, and the beneficiaries of an increase in access. To use the tools developed for traditional welfare analysis, the simplest way to proceed is to assume that we know the value of access to a service, and that this value has been incorporated into the income or consumption aggregate of the household. In other words, because access to primary education for children or a connection to the electricity grid has a certain value for a household, this value is considered an income source. We also assume that access means usage (because it is usage that typically generates value), such that take-up of the service among those who have access does not need to be considered. Finally, we do not discuss the fees that users may have to pay for access. The bottom line of all these assumptions is that we limit our analysis to the distributional characteristics of who has access now and who gains access at the margin when access rates are improved.

If we denote by \( \bar{y} \) the mean income (per capita or per equivalent adult) in the population and by \( F(y) \) the normalized rank of a household (weighted by the household’s size and expansion factor) in the distribution of income (this rank takes a value of zero for the poorest household and one for the richest), the Gini coefficient of inequality, denoted by \( G_y \), is defined as

\[
G_y = \frac{2 \text{cov}[y, F(y)]}{\bar{y}}
\]  

(1)

When combined with mean income, the Gini coefficient can be used to derive the following social welfare function:

\[
W = \bar{y}(1 - G_y)
\]  

(2)

In this function, a higher mean income leads to a higher level of social welfare. Higher inequality lowers social welfare. Sen (1976) and Yitzhaki (1982) provide different rationales for the use of this welfare function. In the case of Yitzhaki, the rationale relies on relative deprivation theory, whereby people assess their welfare in part by comparing themselves with others, which seems appropriate in a decentralization context if the peer comparison group is geographically defined.\(^5\)

---

5. For a derivation of the connection between relative deprivation and the Gini coefficient, see Chakravarty (1990) and Yitzhaki (1982). Ebert and Moyes (2000) offer an axiomatic characterization.
The benefits from access to a service are denoted by $x^A$. For simplicity, we assume that the level of the benefits, denoted by $B$, is the same for all those who have access. That is, if $A$ is a dichotomous variable that denotes access, following Siaens and Wodon (2002), we have the following:

$$\begin{cases} x^A = B & \text{if } A = 1 \\ x^A = 0 & \text{if } A = 0 \end{cases}$$

The Gini income elasticity (GIE hereafter) of the benefits from access to the service is then

$$\eta^A = \frac{\text{cov}[x^A, F(y)]}{\text{cov}[y, F(y)]} \frac{\bar{y}}{\bar{x}^A}$$

where $\bar{x}^A$ is the mean benefit from access computed across the population as a whole, including those who do not have access (that is, if the share of the population with access is denoted by $p$, $\bar{x}^A = B \times p$). When considering a new project, only the additional access provided by the project should be taken into account in the evaluation of the project’s impact on the distribution of income. Yet equation (4) is useful to assess the project’s distributional implications when new access is distributed in the same way as current access. Using a result from Yitzhaki (1999), it can be shown that if those gaining new access to the service have the same position in the distribution of income as those who currently have access, increasing access at the margin by multiplying the share of households with access by $1 + \Delta$, with $\Delta$ small, will generate a gain in social welfare equal to

$$dW = \left(\bar{x}^A \Delta \left(1 - \eta^A G_y \right) \right)$$

Of course, new access need not be distributed in the same way as current access. Imagine, for example, that new access to the service is distributed randomly among the households without access. In this case,
case, the GIE for the benefits of new access would be equal to the following:\(^8\)

\[ \eta^{NA} = \frac{\text{cov}[x^{NA}, F(y)]}{\text{cov}[y, F(y)]} \frac{\bar{y}}{\bar{x}^{NA}} \]  

(6)

where

\[
\begin{align*}
x^{NA} &= 0 \text{ if } A = 1 \\
x^{NA} &= B \text{ if } A = 0
\end{align*}
\]  

(7)

To find the impact on social welfare of the distribution of new access specified by (7), it suffices to replace \( \eta^A \) by \( \eta^{NA} \) in equation (5). Note also that if \( p \) is the population share with access, we have the following:

\[ \eta^A * p + \eta^{MA} * (1 - p) = 0 \]  

(8)

Although the distribution of new access could follow the pattern of current access, or of the current lack of access, it could also follow any other pattern. If we denote by \( x^{MA} \) (where MA stands for marginal access) the benefits from the actual new pattern of access, the GIE that we are interested in, is

\[ \eta^{MA} = \frac{\text{cov}[x^{MA}, F(y)]}{\text{cov}[y, F(y)]} \frac{\bar{y}}{\bar{x}^{MA}} \]  

(9)

**Marginal Benefit Incidence Analysis with a Single Cross-Section of Data**

With a single cross-section of data, estimating \( \eta^A \) and \( \eta^{NA} \) is easy. Information on marginal benefit incidence, however, is needed to estimate \( \eta^{MA} \). This typically requires panel data, or at least repeated cross-sections to look at the distribution of changes in access over time. Unfortunately, panel data or repeated cross-sections are often not available in developing countries. Even when repeated cross-sections are available, they are often not comparable. This section discusses how to estimate the marginal benefit incidence of new access with a single cross-section of data.

---

8. In equations (6) and (7), the value of B is not actually part of the income aggregate of those who do not have access, and it remains included in the income aggregate of those who have access through the variable \( x^A \). For computing the GIE at the margin, the expression is nevertheless appropriate.
Two papers—Ajwad and Wodon (2001) and Lanjouw and Ravallion (1999)—have proposed methodologies that use a single cross-section of data to identify the distribution of increases, at the margin, in access rates to public services or in outlays for social programs. Both studies used the variation in access rates across regions in a country to capture the expected evolution of access over time, assuming that the distribution of new access in lagging regions will follow the pattern observed in regions where access rates are higher.

At the conceptual level, the approaches used by Ajwad and Wodon (2001) and Lanjouw and Ravallion (1999) differ in the method used for ranking individuals, municipalities, or any other entities that are the basic units of observations. Lanjouw and Ravallion classify individuals as poor or rich according to their rank in the national distribution of income. Ajwad and Wodon classify individuals according to their rank in the local (that is, departmental) distribution of income, rather than at the national level. Under a decentralized system of government, a local ranking may be more appropriate. The social welfare framework presented above also stresses relative deprivation, which leads to a local ranking if the peer groups, according to which individuals assess their welfare, are geographically defined. For an assessment of the national impact of policies, however, a national ranking is probably more suitable.

At the empirical level, two differences exist between the approach of Ajwad and Wodon (2001) and that of Lanjouw and Ravallion (1999). The first difference lies in the manner in which the endogeneity bias in the estimation of the marginal benefit incidence analysis is dealt with. The technique used in both papers consists of regressing the access rate in a given quintile against the mean access rate. The mean access rate, however, includes information from the access rates in each quintile. To purge the mean from this endogeneity, Ajwad and Wodon use the leave-out mean as their right-hand side variable. That is, the access rate in any given quintile is regressed against the average of the access rates across all quintiles, except for the quintile for which the regression is performed. Lanjouw and Ravallion, on the other hand, use an instrumental technique, whereby the actual mean is instrumented by the leave-out mean. The second difference is that Ajwad and Wodon constrain the estimates of the marginal benefit incidence analysis to sum to one, and show that without such a constraint, the estimates will be biased downward.9

9. The estimates reported in Lanjouw and Ravallion (1999) are lower than one on average, but it would be easy to apply a similar constraint for their estimation.
This paper uses the method proposed by Ajwad and Wodon (2001). The method is outlined in some detail in the appendix. One last methodological issue must be dealt with before presenting the results. The method for estimating marginal benefit incidence provides information at the quintile level, not at the household level. This is the level of aggregation that must be used to compute the GIE for the distribution of improvements in access. It is well known that using group data implies a downward bias in estimates of inequality because the within-group component of the inequality measure is ignored. Wodon and Yitzhaki (2002c), however, show that using aggregate data for the estimation of the GIE rather than the Gini itself need not necessarily lead to a large bias. In this paper, since we estimate the GIE for marginal increases in access using quintile data, we also estimate with quintile data the GIE for the current distribution of access, and for an increase in access that would be randomly distributed among those who do not currently have access.

Empirical Results

The data employed, for both Bolivia and Paraguay, are nationally representative households surveys. In Bolivia, for education, we use the 1997 *Encuesta Nacional de Empleo*. For access to basic infrastructure, we use the 1999 *Encuesta Continua de Hogares—Condiciones de Vida*. In Paraguay, we use the 1999 *Encuesta Permanente de Hogares*. In each country, the household-level observations are divided into five income intervals, or quintiles, with the ranking being local (the quintiles are defined within departments). As mentioned earlier, Bolivia has 9 departments, and Paraguay has 16. The question we are trying to answer is whether, at the local level, poorer households benefit more or less than other households from an increase in access to a number of public goods or services.

Table 1 presents basic statistics on access. The variables can be divided into two clusters, namely, enrollment in various education cycles and access to basic infrastructure services. In the education cluster, the preschool, primary school, and secondary school net enrollment rates are defined as the number of children of the appropriate age enrolled at each level of schooling divided by the number of stu-

10. We use the 1997 Bolivian survey for the education indicators because in the 1999 survey, due to the formulation of the questionnaire, the measures of school enrollment for the children are affected by holidays.
students who fall into the appropriate age category. In the basic infrastructure cluster, access rates of electricity, pipe water, sewerage, and telephone are computed by dividing the number of households with access by the total number of households.

In Bolivia, the average enrollment rates are 89.7 percent and 48.7 percent for primary schools and secondary schools, respectively. Preschool enrollment appears to be very low, at 6.1 percent, but this may be because the questionnaire asks about enrollment only among children of at least five years of age. In Paraguay, the average enrollment rates are 22.0 percent, 94.8 percent, and 38.7 percent for preschools, primary schools, and secondary schools, respectively. In Bolivia, 71 percent of all households are connected to the electricity grid, 67 percent have access to pipe water, 40 percent have sewerage access, and about a quarter of all households have a telephone. The proportions for Paraguay are similar with access to electricity, water, sewerage, and telephone at 88 percent, 40 percent, 69 percent, and 18 percent, respectively. Table 1 also indicates that access rates vary widely by income quintile. As expected, a strong positive correlation exists between the levels of access to public services and per capita

<table>
<thead>
<tr>
<th>Income quintile</th>
<th>Preschools</th>
<th>Primary</th>
<th>Secondary</th>
<th>Electricity</th>
<th>Water</th>
<th>Sewerage</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest</td>
<td>0.048</td>
<td>0.852</td>
<td>0.241</td>
<td>0.372</td>
<td>0.382</td>
<td>0.136</td>
<td>0.036</td>
</tr>
<tr>
<td>Q2</td>
<td>0.058</td>
<td>0.888</td>
<td>0.425</td>
<td>0.643</td>
<td>0.585</td>
<td>0.246</td>
<td>0.087</td>
</tr>
<tr>
<td>Q3</td>
<td>0.066</td>
<td>0.907</td>
<td>0.520</td>
<td>0.808</td>
<td>0.743</td>
<td>0.400</td>
<td>0.191</td>
</tr>
<tr>
<td>Q4</td>
<td>0.054</td>
<td>0.923</td>
<td>0.580</td>
<td>0.904</td>
<td>0.827</td>
<td>0.590</td>
<td>0.358</td>
</tr>
<tr>
<td>Richest</td>
<td>0.090</td>
<td>0.947</td>
<td>0.686</td>
<td>0.974</td>
<td>0.933</td>
<td>0.801</td>
<td>0.708</td>
</tr>
<tr>
<td>Mean</td>
<td>0.061</td>
<td>0.897</td>
<td>0.487</td>
<td>0.711</td>
<td>0.668</td>
<td>0.403</td>
<td>0.246</td>
</tr>
</tbody>
</table>

income. Enrollment rates in preschools, primary schools, and secondary schools increase with household income. The same is observed for access to electricity, pipe water, sewerage, and telephones.

The data in table 1 provide measures of mean benefit incidence (current access rates), but they do not inform us about the distribution of marginal gains in access when overall access rates are increased. To obtain marginal benefit incidence indicators, we proceeded as explained in the appendix. The marginal benefit incidence indicators provided in table 2 have been normalized, such that a value of one means that the households in a given income quintile benefit as much as the average household from an increase in access. If the marginal benefit incidence is below (or above) one, it means that the households in that income quintile benefit less (or more) from an increase in access than the average household. For example, the households in the first quintile in Paraguay benefit less than the average household from increases in access (or usage) for preschools, water, and telephone; more than the average household for access to primary education; and about as much as the average household for increases in access to secondary education and sewerage. Importantly, even when

<table>
<thead>
<tr>
<th>Income quintile</th>
<th>Preschools</th>
<th>Primary School</th>
<th>Secondary School</th>
<th>Electricity</th>
<th>Water</th>
<th>Sewerage</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest</td>
<td>1.144</td>
<td>1.816</td>
<td>1.327</td>
<td>1.228</td>
<td>1.037</td>
<td>0.801</td>
<td>0.665</td>
</tr>
<tr>
<td>Q2</td>
<td>1.287</td>
<td>0.613</td>
<td>1.361</td>
<td>1.414</td>
<td>1.482</td>
<td>0.716</td>
<td>0.234</td>
</tr>
<tr>
<td>Q3</td>
<td>1.216</td>
<td>1.180</td>
<td>1.744</td>
<td>1.215</td>
<td>1.312</td>
<td>1.359</td>
<td>1.444</td>
</tr>
<tr>
<td>Q4</td>
<td>0.897</td>
<td>0.897</td>
<td>0.581</td>
<td>0.645</td>
<td>0.794</td>
<td>1.348</td>
<td>1.851</td>
</tr>
<tr>
<td>Richest</td>
<td>0.457</td>
<td>0.495</td>
<td>-0.014</td>
<td>0.497</td>
<td>0.374</td>
<td>0.776</td>
<td>0.807</td>
</tr>
<tr>
<td>Mean</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income quintile</th>
<th>Preschools</th>
<th>Primary School</th>
<th>Secondary School</th>
<th>Electricity</th>
<th>Water</th>
<th>Sewerage</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest</td>
<td>0.785</td>
<td>2.019</td>
<td>0.955</td>
<td>1.218</td>
<td>0.697</td>
<td>0.996</td>
<td>0.368</td>
</tr>
<tr>
<td>Q2</td>
<td>0.875</td>
<td>0.558</td>
<td>1.314</td>
<td>1.437</td>
<td>1.056</td>
<td>1.314</td>
<td>0.760</td>
</tr>
<tr>
<td>Q3</td>
<td>1.169</td>
<td>0.494</td>
<td>1.208</td>
<td>1.074</td>
<td>1.174</td>
<td>1.120</td>
<td>1.125</td>
</tr>
<tr>
<td>Q4</td>
<td>0.894</td>
<td>1.164</td>
<td>0.746</td>
<td>0.744</td>
<td>1.084</td>
<td>0.963</td>
<td>1.318</td>
</tr>
<tr>
<td>Richest</td>
<td>1.277</td>
<td>0.764</td>
<td>0.776</td>
<td>0.527</td>
<td>0.989</td>
<td>0.608</td>
<td>1.428</td>
</tr>
<tr>
<td>Mean</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

the marginal benefit incidence suggests that the poor benefit less than
the nonpoor from gains in access, the poor still benefit more at the
margin than they do currently. (See figure 1, which presents graphs of
most of the results presented in tables 1 and 2. All estimates in figure
1 are normalized, that is, divided by the mean access or increase in
access.)

In most cases, the marginal benefit incidence analysis gives similar
results for Bolivia and Paraguay. Improvements in access to primary
school are the most pro-poor, simply because most other groups of
households already have access. Improvements in access to telephones
are the least pro-poor, because in this sector, even those in the highest
quintiles still lack universal access. Electricity and secondary schooling
tend to be pro-poor at the margin, whereas the distribution of the gains
in access for water and sewerage are more evenly distributed.

To summarize the quintile data provided in tables 1 and 2, we pres-
ent GIEs in table 3. As discussed earlier, the GIE for access captures
the current distribution of access. The GIE for lack of access represents
how redistributive a marginal increase in access would be if it were
distributed randomly among the households that do not currently
have access. Because those with access tend to be less poor than those
without access, the GIE for the lack of access is smaller (that is, more
redistributive at the margin) than the GIE for the current pattern of
access. The GIEs for the marginal benefit incidence are our estimates
for the distribution at the margin of the gains in access. These GIEs are
based on the marginal benefit incidence estimates presented in table
2. In most cases, the GIE for the marginal benefit incidence is within
the interval provided by the GIE for the current pattern of access and
the GIE for the lack of access. This is not very surprising, given that
the richer among those who do not have access have a higher proba-
bility of getting access once access rates are improved. In Bolivia,
however, for the three education indicators, the GIE for the marginal
benefit incidence is slightly more pro-poor than if the gains in access
were randomly distributed among those who currently do not have
access.

Finally, figure 2 presents a scatter plot with the GIEs for all the ser-
ices and for the two countries as a function of the mean access rate. A
second order polynomial is fitted through the scatter plot to suggest
the relationship. Services with low access rates have higher GIEs than
services with low access rates. In other words, the higher the mean
benefit incidence of the public service, the more pro-poor will be the
distribution at the margin of an increase in access. For instance, pri-
FIGURE 1. NORMALIZED BENEFIT AND MARGINAL BENEFIT INCIDENCE FOR VARIOUS SERVICES

Bolivia

Preschool Net Enrollment

Income interval

Richest

3

Poorest

0

0.2 0.4 0.6 0.8 1 1.2 1.4 1.6

Paraguay

Income interval

Richest

3

Poorest

0

0.2 0.4 0.6 0.8 1 1.2 1.4

Primary School Net Enrollment

Income interval

Richest

3

Poorest

0

0.5 1 1.5 2

Secondary School Net Enrollment

Income interval

Richest

2

Poorest

0

0.5 1 1.5 2

Access to Electricity

Income interval

Richest

3

Poorest

0

0.2 0.4 0.6 0.8 1 1.2 1.4 1.6

Access to Sewerage

Income interval

Richest

3

Poorest

0

0.2 0.4 0.6 0.8 1 1.2 1.4 1.6

Access to Telephones

Income interval

Richest

2

Poorest

0

0.5 1 1.5 2.5 3
TABLE 3. GINI INCOME ELASTICITIES FOR THE MARGINAL BENEFIT INCIDENCE

<table>
<thead>
<tr>
<th></th>
<th>Bolivia</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With access</td>
<td>Without access</td>
<td>Marginal benefit</td>
</tr>
<tr>
<td>Preschool</td>
<td>0.200</td>
<td>-0.013</td>
<td>-0.268</td>
</tr>
<tr>
<td>Primary</td>
<td>0.038</td>
<td>-0.335</td>
<td>-0.358</td>
</tr>
<tr>
<td>Secondary</td>
<td>0.326</td>
<td>-0.309</td>
<td>-0.526</td>
</tr>
<tr>
<td>Electricity</td>
<td>0.289</td>
<td>-0.712</td>
<td>-0.313</td>
</tr>
<tr>
<td>Water</td>
<td>0.282</td>
<td>-0.568</td>
<td>-0.283</td>
</tr>
<tr>
<td>Sewerage</td>
<td>0.583</td>
<td>-0.393</td>
<td>0.082</td>
</tr>
<tr>
<td>Telephone</td>
<td>0.921</td>
<td>-0.300</td>
<td>0.267</td>
</tr>
</tbody>
</table>

|                | Paraguay         |            |            |
|                | With access      | Without access | Marginal benefit |
| Preschool      | 0.175            | -0.049     | 0.155      |
| Primary        | 0.031            | -0.557     | -0.294     |
| Secondary      | 0.334            | -0.210     | -0.143     |
| Electricity    | 0.061            | -0.456     | -0.265     |
| Water          | 0.402            | -0.276     | 0.078      |
| Sewerage       | 0.204            | -0.450     | -0.144     |
| Telephone      | 0.699            | -0.157     | 0.343      |


FIGURE 2. GINI INCOME ELASTICITY FOR THE DISTRIBUTION OF GAINS IN ACCESS AND ACCESS LEVELS
primary schools in Bolivia and Paraguay have enrollment rates of 95 percent and 90 percent, respectively, according to the surveys, and GIEs are -0.294 and -0.386. The negative relation between the benefit incidence and the GIEs suggests that on average, the very poor start to benefit from public services only once the services are widely available to the nonpoor.

Conclusion

Within the context of decentralization in Latin America, the allocation of investments at the local level is an important decision for policymakers. Although funding for municipalities and departments has been increasing over time, good monitoring systems to assess how these funds are spent are lacking. The risk of capture by the better-off of the funds allocated to the social sectors and to the provision of basic infrastructure services may well be larger at the local level than at the national level. This is why it is important to provide good methodologies for measuring the distribution of the benefits from public expenditures at the local level.

This paper has proposed one such methodology. When it is applied at the departmental level (as we did in the empirical work), the methodology provides estimates of how, on average across all departments, increases in access to basic services are distributed within departments. To obtain measures of marginal incidence at a lower administrative level, the methodology could be applied by ranking households within their municipality instead of their department. In any case, the main empirical result of the paper is that the poor, and especially the very poor, appear to benefit from an increase in access to public services only once the nonpoor are already well served. In primary education, for example, the poor benefit more than the nonpoor from gains in access, because coverage is already high. In basic infrastructure services, however, the nonpoor continue to reap a large part of the gains in access.

If the objective is to reach the very poor, the results may inform priority sectors of investments, even though considerations other than marginal benefit incidence estimates should, of course, be reviewed before making sectoral policy choices. The results need not indicate that local governments favor the nonpoor, but they do suggest the need for pro-poor policies to accelerate the speed at which the poor benefit from the expansion of public social services.
Appendix: Estimation Procedure for the Marginal Benefit Incidence Analysis

Following Ajwad and Wodon (2001), consider a country with $i = 1, \ldots, N$ departments, and a number of households within each department. The households are ranked by per capita income and assigned to one of $q = 1, \ldots, Q$ income intervals. The ranking is done locally, which means that the intervals are defined within departments. We denote by $x_{ij}^{q}$ the benefit incidence of a program or service in household $j$ belonging to interval $q$ and living in department $i$. This benefit incidence reflects the share of the population with access to the public program or service. The mean benefit incidence in interval $q$ for households in department $i$ is denoted by $X_i^q$, and the overall department mean is denoted by $X_i$. If $J_i^q$ is the number of households in interval $q$ for department $i$, the two means are respectively equal to the following:

$$X_i^q = \frac{\sum_{j=1}^{J_i^q} x_{ij}^{q}}{J_i^q} \quad \text{(A.1)}$$

$$X_i = \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J_i^q} x_{ij}^{q}}{\sum_{q=1}^{Q} J_i^q} \quad \text{(A.2)}$$

To estimate the marginal benefit incidence, that is, who gains from an expansion in the program or service, we use the geographic variation in access both between households and between departments as a source of information for understanding the diffusion process that generates access. This is done by regressing the incidence in each of the intervals in the departments against the departmental means, using $Q$ regressions:

$$X_i^q = \alpha^q + \beta^q \left( \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J_i^q} x_{ij}^{q} - \sum_{j=1}^{J_i^q} x_{ij}^{q}}{\sum_{q=1}^{Q} J_i^q - J_i^q} \right) + \epsilon_i^q \quad \text{for } q = 1, \ldots, Q \quad \text{(A.3)}$$

To avoid endogeneity, the right-hand side variable is computed at the departmental level as the mean on all the households, except for those belonging to interval $q$. Pooling all observations from the various intervals together, we estimate one regression:
In equation (A.4), the intercepts and slopes are allowed to differ for each interval, but there is an implicit restriction. It must be that across the various intervals, the average marginal increase in access from a unitary increase in mean access is one. It can be shown that the restriction is as follows:

\[
\sum_{q=1}^{Q} \frac{\beta_q}{Q - 1 + \beta_q} = 1 \tag{A.5}
\]

Writing \( \beta_Q \), the parameter for interval \( Q \) in relation to the other parameters, yields the following:

\[
\beta_q = \frac{(Q - 1) \left( 1 - \sum_{q=1}^{Q-1} \frac{\beta_q}{Q - 1 + \beta_q} \right)}{\sum_{q=1}^{Q-1} \frac{\beta_q}{Q - 1 + \beta_q}} \tag{A.6}
\]

To take into account the restriction (A.6), (A.4) is estimated with nonlinear least squares. It can also be shown that a change in benefit incidence for the households belonging to quintile \( q \) in response to an increase in the aggregate incidence is as follows:

\[
\frac{\partial X^q}{\partial X_i} = \frac{Q \beta_q}{Q - 1 + \beta_q} \quad \text{for } q = 1, \ldots, Q \tag{A.7}
\]

The right-hand side values in (A.7) are the estimates of marginal benefit incidence. A value larger (or smaller) than one implies that the corresponding group of households benefits more (or less) than the average from an expansion in public programs and services.
References


Part IV

Firms and Governments under Uncertainty
Contractual Savings, Capital Markets, and Financing Choices of Firms

Gregorio Impavido, Alberto R. Musalem, and Thierry Tressel

Abstract

This paper analyzes the relationship between the development and asset allocation of contractual savings and firms’ capital structures. The authors develop a simple model of firms’ leverage and debt maturity decisions. They illustrate the mechanisms through which contractual savings development may affect corporate financing patterns. In the empirical section, they show that the development and asset allocation of contractual savings have an independent impact on the financing choices of firms. Different channels for this impacts are identified. In market-based economies, an increase in the proportion of shares in the portfolio of contractual savings is associated with a decline in firms’ leverage. In bank-based economies, instead, an increase in the size of contractual savings is associated with an increase in leverage and debt maturity in the corporate sector.

The past two decades have witnessed a parallel explosion of equity
markets and institutional investors, especially pension funds. In many stock markets, capitalization and liquidity have soared while institutional investors have become crucial actors in the capital markets not only in developed, Anglo-Saxon economies, but also in a handful of emerging economies (for instance, Chile and South Africa). Demographic evolution—mainly in the Organisation for Economic Co-operation and Development (OECD) countries, but also in emerging economies—is bound to increase pressure on countries to reform their pension systems, and to choose effective investment regulations and policies for the newly created institutions. Pension reforms, designed to ensure a sufficient living standard after retirement, generate a stable source of long-term domestic savings. Recent studies argue that this will foster the development and deepening of capital markets.¹ Ultimately, the array of funding possibilities for domestic firms will be enriched, in particular the access to long-term capital. In the recent context of currency and financial crisis associated with asset-liability mismatch in the balance sheets of firms (and banks), and excess reliance on (foreign currency denominated) short-term debt, it is becoming urgent to evaluate whether the presence of domestic institutional investors tends to reduce firms’ and other economic agents’ vulnerability to interest rate variations and other shocks.² In a more general context, Caprio and Demirgüç-Kunt (1997) show that the lack of long-term finance in emerging economies is not totally explained by firms’ characteristics. The institutional environment and macroeconomic factors significantly affect the supply of long-term finance. This paper attempts to assess both empirically and theoretically the impact of contractual savings development on the financing decisions of firms in a sample of developed and emerging economies.³

The primary objective of a pension reform is to provide sufficient and affordable benefits for old age that can be sustained in the long term. Financial deepening alone should not motivate pension reform. Moreover, history teaches that contractual savings institutions are neither sufficient nor necessary for capital market development.⁴ Still, the issue is the speed of financial development. Whether financial deepen-

---

2. See, for instance, Rodrik and Velasco (1999) and Aghion, Bacchetta and Banerjee (2000) for a theoretical model of monetary policies in such a context.
3. Contractual savings institutions include pension funds and life insurance companies.
ing takes two decades or two generations has very different implications for development strategies. Recent studies (see Catalan, Impavido, and Musalem (2000) and Impavido and Musalem (2000)) suggest, for instance, that the rapid growth of capital markets during the past 15–20 years is partly explained by the development of contractual savings institutions. Pension funds and life insurance companies are becoming essential characteristics of modern financial systems, and as such may significantly modify the corporate sector’s financing choices. Moreover, in the present context of financial instability, developing countries may find it worthwhile to develop a domestic source of long-term financing.

In this paper we address the following questions. First, as contractual savings institutions develop, is there a sizeable impact on the leverage and debt maturity of firms? Second, can such an impact be accounted for by the characteristics of firms in each country? Third, does this effect remain significant once we control for the activity of the banking sector, the size and activity of the stock market in each country, and unobserved fixed characteristics? Fourth, can we disentangle the potential channels through which contractual savings institutions affect firms’ capital structures? Finally, what do our results imply for the resilience of domestic financial systems in the highly volatile environment of the international financial architecture?

The rest of the paper is organized as follows. The first section presents a brief literature survey. The next section, A Simple Model of the Financing Choices of Firms, sketches a model of the financing choices of firms and provides a benchmark for discussing the interaction between informational issues and corporate capital structures in the context of contractual savings development. The third section, Data and Empirical Strategy, introduces the data and discusses the variables used. The next section reports cross-country empirical results. Finally,

5. There are, of course, other central players on capital markets, such as mutual funds, hedge funds, investment companies, or simply nonlife insurance companies. We do believe, however, that pension funds and life insurance companies are particular because of the long-term structure of their liabilities (see Impavido and Musalem (2000), who underline the different impacts of contractual savings and nonlife insurance companies on capital markets).

6. Walker and Lefort (2002) argue that equity investments by fully privately managed pension systems have reduced price volatility in Argentina, Chile, and Peru.
the conclusion is devoted to answering the questions just posed in a
discussion of findings and policy implications.

Survey of the Literature

There exists a rich literature that explores the effect of the institutional
environment on firm financing choices in specific countries and across
countries. First, the legal approach, led by La Porta and others (1998),
shows how legal traditions and the rights of specific creditors and
minority shareholders shape the access to external finance and the cor-
porate ownership structures around the world. Second, Rajan and
Zingales (1995) and Demirgüç-Kunt and Maksimovic (1996b) docu-
ment cross-country regularities in the correlation between corporate
financial structures and various firms' characteristics. Demirgüç-Kunt
and Maksimovic (1996a) explore the impact of stock market develop-
ment on firms' leverage, and Demirgüç-Kunt and Maksimovic (1999)
extend this analysis by looking more closely at the institutional and
legal determinants of capital structure. They find that how much the
firm can grow by relying on external finance does depend on the legal
disentangle the financial, legal, and technological factors that deter-
mine access of firms to external finance. Third, others highlight the
impact of particular institutional arrangements on firms' external
financing possibilities (see, for instance, Hoshi, Kashyap, and
Scharfstein (1991)). Fourth, firms' characteristics will affect the financ-
ing choices: for instance, firms try to match the maturity of their assets
and liabilities. (See Caprio and Demirgüç-Kunt (1997) for a discus-

sion.) Moreover, informational asymmetries affect the choice of secu-
ritv when seeking external finance, and restrict the feasibility set (see,
among others, Barclay and Smith (1995), Stohs and Mauer (1996),
Myers and Majluf (1984), Rajan (1992), Petersen and Rajan (1995),
Diamond (1991), Jensen and Meckling (1976), Myers (1977)). Overall,
the existing literature confirms that the institutional environment,
together with the real characteristics of firms, determines the capital
structures of firms.

---

7. See Beck and others (2000) for a synthetic approach, at three different lev-

eels: firms, industries, and countries.

8. See the survey by Harris and Raviv (1990) and Stulz (2000).
This paper comes as a necessary complement to earlier works stressing the impact of financial and legal institutions on firms' financing patterns in a cross-country perspective. In a world of asymmetric information, in which there are conflicts of interest between external investors and those who manage and control the productive assets, the financial institutions and the legal environment will shape the capital structures of firms, which will lead to systematic differences across countries. To the extent that contractual savings institutions may modify the information set available to all investors, push for compliance with transparency rules and legal rights, or simply modify the relative supply of different securities, one should indeed expect to observe significant cross-country differences associated with contractual savings' characteristics.

A Simple Model of the Financing Choices of Firms

In this section, we briefly sketch the main features and conclusions of models developed in other papers. This model emphasizes informational issues and refinancing risks. More specifically, we describe a simple framework in which firms choose the debt maturity and are also able to issue equity. We discuss the potential benefits associated with the development of stock markets and the nature of investors within a framework in which banks may be subject to term transformation risks. In particular, the model suggests that the development of contractual savings institutions will affect firms' financing choices if: it leads to an increase in the supply of long-term debt; it reduces equity

9. See Demirgüç-Kunt and Maksimovic's papers.
13. More specifically, we focus on adverse selection issues and the role of private information in the credit relationship. The literature has highlighted many considerations also relevant for the debt maturity decision that we won't tackle here: underinvestment (Myers 1977), short-termism (Von Thadden 1995), ex post moral hazard (Rajan 1992 and Petersen and Rajan 1995), among others.
rationing; and it fosters information disclosure and better corporate governance mechanisms on the stock market. More generally, the model predicts that the equilibrium capital structures of firms will be a function of (a) their characteristics (including maturity of assets, profitability, risk, and asymmetry of information); (b) the efficiency of the financial system (for instance, in generating—ex ante and interim—private and public information); and (c) the supply of funds to capital markets that are affected by the nature of investors.

The Corporate Sector

There is a continuum of firms differing with respect to their initial equity $E_R < 1$. Each firm has access to a project that requires an investment $I = 1$. The investment can be spread between date 0 and date 1, under the following constraints:

1) The present value of the two investments $I_0$ and $I_1$, respectively at dates 0 and 1, is equal to the total investment: $I_0 + I_1/R = 1$, where $R - 1$ is the safe interest rate (the return on government bonds).

2) The initial investment $I_0$ must be strictly positive:

$$I_0 \geq \gamma > 0$$  \hspace{1cm} (1)

If the initial investment is realized and is not liquidated at date 1, the project will yield cash flows for all dates $t > 1$. There are two types of firms in the economy, however. Good firms yield strictly positive cash flows $\Pi$ at each date $t > 1$, and can be liquidated at date 1 for $lI_0$, where $l < 1$. Bad firms yield no cash flows. They are worth nothing if the project is terminated at date 1. Firms' types are private information and cannot be credibly signaled to outsiders (creditors and new

---


15. As becomes clear in the analysis of short-term debt, firms that are good risks choose to minimize the first-period short-term debt to reduce the cross-subsidization of bad risks. If no constraint is imposed, they would choose not to borrow short-term at date 0. The constraint we impose here can be endogenized as in Petersen and Rajan (1995) by adding a moral hazard imperfection at date 1.

16. $\Pi$ is assumed to be greater than $R^2$ so that in the perfect information case, all good firms have access to long-term debt.
The uncertainty of the project is measured by $\lambda$, the prior probability assigned by external providers of funds (banks or investors) that the firm is good at date 0.

The firm is run by a manager (who may be the controlling shareholder) who maximizes the expected discounted value of dividends paid to initial shareholders. As dividends will be the same in all periods, for $t > 2$, this is equivalent to

$$\max E_0(D_{t_2} + B)$$

(2)

where $D_{t_2}$ is the date 2 dividend received by the initial shareholders, $B$ the discounted value at date 2 of dividends for $t > 2$, and $E_0$ the expectation operator at date 0.\textsuperscript{17,18}

The firm undertakes the project, however, if and only if it yields a greater cash flow than simply investing in government bonds:

$$\max E_0(D_{t_2} + B) \geq R^2 \cdot E_R.$$  

(3)

We assume that a firm cannot have a mix of short-term and long-term debt. Therefore, external financing ($I - E_R$) possibilities are as follows:

- long-term debt only (the debt is issued at date 0 and repaid at date 2),
- short-term debt only (in this case the debt must be rolled over at date 1),
- a combination of short-term debt and external equity, and
- a combination of long-term debt and external equity.

These financing possibilities are briefly described in the following paragraphs.

**Long-Term Debt**

Banks are perfectly competitive. They gather savings from households and invest them in loans to either the public sector (government

\textsuperscript{17} More precisely, $D_{t_2} = P - R'D$, when no shares are issued, where $D$ is the face value of the debt and $R'$ the gross repayment per dollar borrowed.

\textsuperscript{18} Because dividends paid for dates $t > 2$ are the same in this model, $B$ is equal to $\sum_{j=1}^{t_2} \frac{D_{t_2}}{R^j} = \frac{D_{t_2}}{R - 1}$.
bonds) or to the private sector (corporate bonds). The structure of the economy is similar to Diamond's model with banking and limited access to the stock market (see Diamond 1997). In particular, households are subject to liquidity needs at date 1. As in Diamond and Dybvig (1983), this feature may lead to runs (see the short-term debt section below), and firms may not be refinanced to complete the project. In the case of long-term debt, banks cannot force firms into liquidation when they face sudden withdrawals (and I₀ = 1).

For the sake of simplicity, long-term debt is repaid once and for all at date 2. Ex ante competition among banks implies that banks make zero expected profits on loans: the expected rate of return on loans must be equal to the safe interest rate per period. As a proportion, 1 − λ of loans are never repaid, and banks charge a two-period gross return equal to 1 − λ per unit of capital borrowed.

In this imperfect information world, some firms won't get access to long-term debt (LTD). Indeed, banks refuse to lend whenever the maximum expected return on the loan (LTD = 1 − ER) is less than the return on government bonds:

\[ \Pi < \frac{R^2}{\lambda} \cdot LTD \]  

Hence, if \[ E_r < E_1 = 1 - \frac{\lambda \Pi}{R^2} \]

the rationing region \[ [0, E_1] \] becomes larger if the profitability of good firms falls, if the cost of capital \( R \) increases, or if informational frictions increase (\( \lambda \) increases).

**Refinancing Risk and Financial Institutions**

**Short-Term Debt.** The firm may be able to obtain a short-term loan from a bank when long-term debt is not accessible. The existing relationship between the bank and the borrower allows the former to obtain private information about the quality of the project; by lending at short horizons, the bank can decide not to refinance the project if it obtains bad information on the firm (see Sharpe (1990), Petersen and Rajan (1995),

---

19. In the general case, long-term debt is repaid in \( n \) periods between date \( k \) and date \( k + n \), with \( k, n \) finite, and \( k > 1 \). It is straightforward to show that the qualitative results are not modified by this simplification. See Tressel (2001) for a detailed justification.

20. The interest rate \( r \) on bank loans is given by \( R^2/\lambda = (1 + r)^2 \).
and Stulz (2000) for a survey). In parallel, the bank obtains an informational advantage with respect to other potential lenders, because the latter has less precise information on the quality of the borrower at the interim date (more on this later). Therefore, the initial lender can compensate initial for losses on bad projects by charging a higher interest rate on second period loans. This process makes short-term debt more feasible in an uncertain environment.\footnote{21}

The information game is modeled in the following way: by lending to a firm in the first period, a bank is able to refine the information on the quality of the project. More specifically, we assume that the bank can get two possible signals at date 1: signal down and signal up. Signal down reveals with probability 1 that the firm is of a bad type. Signal up can be received by the bank for both types of firms. However, this signal reduces the uncertainty on the type of the borrower. The probability of being good and given signal up is $\lambda_U$. Other banks also receive a signal; however, their information is less precise: the probability that the firm is good and given signal up is $\lambda_E$, with $\lambda_U > \lambda_E > \lambda$. This difference between private (measured by $\lambda_U$) and public information (measured by $\lambda_E$) creates a captive market for each bank at date 1, composed by the firms that it financed at date 0, and receiving signal up. Other banks will charge a gross rate of return equal to $R/\lambda_E$. The incumbent bank, however, is willing to lend as long as the rate of return on the loan is greater or equal to $R/\lambda_U$, which is strictly less than $R/\lambda_E$. Therefore, that bank will be able to make positive profits on the firms it already finances by proposing a loan with a gross rate of return $R/(\lambda_E - \epsilon)$ (with $\epsilon \to 0$). Finally, ex ante competition among banks (that is, at date 0) implies that expected profits of banks must be zero: their positive profits between date 1 and 2 compensate the losses made between date 0 and 1. This means that more firms are funded at date 0: the threshold value $E_2$ under which the firm is rationed with short-term debt is lower than $E_1$.\footnote{22}

The drawback of short-term debt is that if the bank refuses to roll over the debt, the firm will be forced into liquidation.\footnote{23} We model the refinancing risk in the next section.

\footnote{21. The market power of the incumbent bank, however, may create distortions ex post; see, for instance, Rajan (1992) and Sharpe (1990).}

\footnote{22. More precisely, this is the case only if the refinancing risk described below is not too large.}

\footnote{23. Long-term finance, contrary to short-term debt contracts, may also reduce "short-termism" in the behavior of managers; see Von Thadden (1995).}
Bank Runs and Refinancing Risks. The initial lender will not refinance the project if it receives bad news on the borrower's type (signal down). Inefficient liquidation, however, may also occur, depending on the stability of the banking system, if, for instance, depositors and banks' lenders in general have no confidence in the ability of banks to serve sudden withdrawals. Diamond and Dybvig (1983) have rationalized the possibility of self-fulfilling runs when banks serve investors sequentially by drawing on a small liquidity reserve: it is rational for each individual creditor to join a run, because by doing so, it secures a chance to be at the beginning of the queue and get his money back.\textsuperscript{24} Next, the question of equilibrium selection is addressed in the following way: we adopt the convention that investors coordinate on one equilibrium or the other depending on the realization of a sunspot variable: runs occur with a probability \( \mu \).

The liquidation value of the initial investment \( I_0 \) is \( l \cdot I_0 \) with \( l < 1 \). Banks' assets are firms' liabilities, however. The ability of banks to get repaid in full (by refusing to roll over the debt, which forces the firm into bankruptcy, unless it can find other lenders) depends on the liabilities of the firm. Let us assume that the firm will not find other external funds if the initial lender refuses to roll over the debt (for instance, because of contagion, runs may occur on the whole banking system). Therefore, the firm is forced to liquidate its assets to repay the debt; hence, it goes bankrupt.\textsuperscript{25} The issue here is that bankruptcy occurs because of the mismatch between liabilities and assets in the corporate sector; that is, because short-term debt is used to finance long-term, illiquid, productive investment. If the firm has a low debt-to-equity ratio, it will be able to repay the debt fully because equity acts as a cushion. On the contrary, if the firm is highly indebted, early liquidation implies that the bank cannot get the full value of the debt. Hence, the bank may not be able to obtain enough liquidity if depositors (or generally all banks' creditors) decide to run (or not renew their loans to the bank).

The argument goes like this. Each bank borrows from many investors, and lends to many firms. We assume that investors observe the average capital structure of these firms and hence know the value of

\textsuperscript{24} For the recent literature on the fragility of the banking system in the context of the recent financial crisis and capital flows, see Chang and Velasco (1999), and Rodrik and Velasco (1999), among others.

\textsuperscript{25} Here we formally assume that each bank lends to a homogenous group of firms. This is obviously not realistic. This assumption is purely technical and does not affect the general argument.
bank assets in case of early liquidation. If the value of the short-term debt is less than the liquidation value of the firm, investors know that they will be fully repaid in case of run. Therefore, they have no reason to run in the first place (or to expect other investors to do so). On the contrary, if the date 1 value of the debt is more than the liquidation value of the firm, runs are possible. They will occur with a probability $\mu$.

Formally, runs are possible (and occur with probability $\mu$) if and only if:

$$R_0 \cdot STD_0 > 1 \cdot I_0$$

where $R_0 - 1$ is the rate of interest on the short-term debt $STD_0$.

It can be shown that, for intermediate values of $\mu$, two types of firms will finance their projects with $STD$ rather than $LTD$: (a) firms that have limited internal liquidity, that cannot borrow long-term and that face a positive probability of runs; and (b) firms that have important reserves relative to their borrowing needs and that do not face a refinancing risk. For intermediate values of liquidity reserves, the firm chooses long-term debt.

**Equity Markets.** Firms may also increase their capital by issuing equity on the stock market. We neglect underwriting costs and assume that new shares are sold to dispersed investors so that the initial shareholder keeps all the control rights (see, for instance, Pagano and Röell (1998) and Shleifer and Wolfenson (2000) for a similar assumption). The equity contract for new (minority) shareholders is the following:

- Each minority shareholder $i$ invests $E_i$ in the project at date 0.
- Each minority shareholder $i$ receives a proportion $\alpha_i$ of all future cash flows, net of debt repayment, where $\alpha_i = E_i / E$ ($E$ is total equity).

Investors in the stock market, however, do not observe the type of a firm at date 0: they only know that a proportion $\lambda$ of firms are good. The participation constraint for an investor is therefore

---

26. Here we formally assume that each bank lends to only one firm. This assumption is purely technical. In a more general context, the occurrence of runs should depend on the average capital structure of firms financed by a bank. See Tressel (2001) for a discussion.

27. In this case, $STD$ is cheaper than $LTD$ because of its informational advantage.
\[ \lambda \alpha E_0 (D\alpha_2 + B) \geq R^2 - E_i \]

This condition assumes that (a) investors are risk-neutral (this assumption may depend on preferences and the ability of investors to hold a diversified portfolio); (b) there are no transaction costs on the stock market; (c) there are no liquidity premiums on stocks; and (d) the controlling shareholder of a firm cannot expropriate minority shareholders (in that situation, the effectiveness of investor protection can be crucial; see the theoretical analysis of Shleifer and Wolfenson (2000)).

By relaxing one or several of these assumptions, we can formally derive simple equity rationing rules (that depend on investors’ characteristics and the regulatory environment) affecting each firm willing to raise capital on the stock market.

The objective function of the controlling shareholder is now

\[ \text{Max } E_0 (1 - \alpha)(D\alpha_2 + B) \]

where \( \alpha = \sum_i \alpha_i \).

The initial shareholder may decide to issue equity: to be able to undertake the project (eliminate the rationing situation); either to be able to borrow long-term; or to be able to borrow short-term with no refinancing risk. The cost of raising equity on the stock market is that profits have to be shared with new shareholders.

For intermediate values of the informational advantage of short-term debt (measured by \(1/k - 1/X_1\)) and the refinancing risk (measured by \(m\)), the model predicts the following:

- Firms with few initial reserves will issue shares so that they can borrow with long-term debt (that is, they issue \(E_X = E_1 - E_R\)).

---

28. More precisely, the argument here is that the optimal financing strategy for some projects implies a mix of debt and equity finance, and that debt finance only may not be possible for projects either highly uncertain and/or with few cash flows in the short and medium term. For instance, Eurotunnel had its debt swapped into equity when it became clear that debt repayments were not sustainable. A substantial dilution of property rights followed—at the expense of minority shareholders.

29. We are assuming here that the issuing price of shares is \(p = 1\): the controlling shareholder is not able to extract any of the additional return that investors get by buying shares instead of buying bonds.
• Firms with larger initial reserves issue shares so that they can borrow with short-term debt and no refinancing risk (they issue $E_X = \hat{E} - E_R$).

We can introduce in this setting the possibility for the controlling shareholder to divert part of the cash flows after debt repayment. This can be done by assuming that the controlling shareholder is able to hide (and consume) a nonverifiable proportion $k$ of the dividends and claim that the present value of total dividends is only $(1 - k)(\text{Div}_2 + B)$. Now, a minority shareholder with a stake $\alpha$ in the firm will only receive $\lambda \alpha (1 - k)(\text{Div}_2 + B)$. The participation constraint of investors may imply the possibility of equity rationing for firms that are otherwise able to borrow from a bank. The value of $k$ depends on the legal environment (for example, transparency rules and protection of minority shareholders).

**Predictions of the Model**

This model predicts that firms' capital structures are a function of firms', the banking sector's, and capital markets' characteristics, and more specifically that the supply of long-term capital, the quality of (private and public) information, and corporate governance mechanisms interact in shaping firms' capital structures. As discussed by Impavido, Musalem, and Tressel (2001), the development and investment behavior of contractual savings institutions may affect these factors and therefore have a significant impact on firms' financing choices.

Moreover, the model provides predictions on (a) the characteristics of firms that benefit from an increase in long-term credit or in equity; and (b) the impact of information and corporate governance mechanisms on the debt maturity structure. Testing these predictions in detail is beyond the scope of the paper. Still, they support the argument that contractual savings institutions may affect firms' capital structures through different channels. The model suggests that the characteristics of the banking system and the stock market are crucial for this matter, and leads naturally to the distinction between bank-based and market-based financial systems (Demirgüç-Kunt and Levine 1999) as a first-order approximation.

The predictions are the following:

1) **Consider first an exogenous increase in the supply of long-term credit.** Start, for instance, from a situation in which long-
term debt contracts are not proposed by banks. This may be the consequence of a lack of long-term liabilities in the banking system, making the probability of sudden withdrawal high, so that banks are reluctant to perform their term transformation activity. It is reasonable to assume that an exogenous increase in the maturity of banks' liabilities may make banks willing to propose long-term contracts. The model predicts the following:

A) The firms that will benefit more from long-term loans are
   a) with less initial asymmetric information (λ),
b) with less liquid investments (that is, lower value of I).
c) with intermediate values of internal liquid reserves (that is, firms with \( E_1 < E_R < \hat{E} \)).
d) with no access to STD without refinancing risk (that is, \( E_R < \hat{E} \)); the firms that will benefit more from LTD are the more profitable ones (\( \Pi \) larger).
e) with higher up-front investments (\( \gamma \)).

B) More firms will benefit from an increased supply of long-term debt when
   a) the banking system is more subject to runs (\( \mu \) large) because STD becomes more costly to firms.
   b) the informational advantage \((1/\lambda - 1/\hat{\lambda}\) of STD is lower, because the benefit of STD is reduced.
   c) the market power of banks in the second period is lower (measured by \( \lambda_I/\lambda_E \)): in this situation, banks charge a higher interest rate in the first period; hence, more firms become subject to the refinancing risk.

2) Consider a reduction in equity rationing. This may happen because investors require a lower risk premium, or liquidity premium, to buy shares, transaction costs are reduced, and there is less scope for minority shareholders' expropriation.

A) The average impact on the debt maturity is not clear-cut. In particular, it depends on the initial internal reserves of the firm. If firms are initially relatively well capitalized, the debt

---

30. Although the model does not integrate why banks are or are not willing to offer long-term loans, it allows discussion of the impact on firms' financing choices when LTD contracts are proposed.

31. In each case, all parameters, except the one considered, are fixed.

32. In this case, we extend the model to consider that we have several "sectors," each being characterized by a given value of the parameter \( \mu \).
maturity will decrease; on the contrary, if firms have limited initial capital, the debt maturity will increase.

B) The firms that will benefit more from an easier access to the stock market are those
   a) with low or intermediate internal reserves.
   b) with lower liquidation value (l).
   c) with higher up-front investments (γ).

C) More firms will benefit from an easier access to the stock market when the banking system is more subject to runs (μ large).

3) Information disclosure and corporate governance. We discuss here the impact of exogenous modifications in the informational parameters on firms' financing choices.
   A) If the quality of ex ante public information increases (λ increases), more firms have access to long-term debt (E₁ is lower).
   B) If the quality of interim public information increases relative to the initial public information (λ江东/λ increases), the second period interest rate on STD decreases. Hence, to maintain profitability, banks increase the first period interest rate on STD. This, in turn, increases the risk of early liquidation, which makes STD less attractive relative to LTD.
   C) If transparency increases at the interim date, relative to the private information of banks (λ江东/λ江东 increases), the second period market power of banks decreases, which forces them to increase the first-period interest rate on STD to maintain their expected profitability. Again, the refinancing risk increases, which makes STD less attractive than LTD.
   D) If λ江东, λ江东 and λ江东 increase in the same proportion (both public information and private information), short-term debt becomes relatively less attractive than long-term debt (the reduction in the cost of debt by choosing STD instead of LTD is lower).

Data and Empirical Strategy

A detailed description of the data used in the empirical analysis can be found in Impavido, Musalem, and Tressel (2001). The empirical study

33. This may be interpreted in relation to shareholder activism: shareholder activism increases the transparency on the stock market, and simultaneously increases the efficiency of bank monitoring.
aims at assessing the impact of the development of contractual savings institutions on firms’ capital structures. The two fundamental characteristics that we analyze are: the choice between debt and equity, and the maturity structure of debt. We focus on pooled ordinary least squares (OLS) estimates, robust to heteroscedasticity, and panel estimates (fixed effects).

The dependent variables we consider are: total debt over equity (TDTE), defined as the ratio of long-term plus short-term debt over the book value of equity; long-term debt over the book value of equity (LTDTE); short-term debt over the book value of equity (STDTE); and long-term debt over total debt as a measure of debt maturity (LTDTD).

These variables are self-explanatory. Notice, however, that we choose to use the book value of equity rather than the market value. Although the market value of equity may be a better measure of the “true” value of the firm’s net worth than its book value, using the market value may introduce a spurious correlation between these dependent variables and the contractual savings variables simply because contractual savings investments (for instance, in shares) are evaluated at their market value.\textsuperscript{34} We will return to this issue later.

We use three sets of explanatory variables defined in table 1: firms’ characteristics; macroeconomic factors; and financial system characteristics to obtain the empirical specification:

\[
\text{Capital structure} = F(\text{firms’ characteristics}; \text{macroeconomic factors}; \\
\text{financial system characteristics})
\]

\textit{Firm-Specific Characteristics}

Firm-specific considerations are important in determining corporate financing patterns. The asymmetries of information and risk aspects to which firms are exposed will in general vary from firm to firm. Therefore, the macroeconomic and institutional environment may only partly explain the observed capital structures in different countries. For instance, the apparent lack of long-term finance in developing countries when compared with developed countries may simply

\textsuperscript{34} This is less likely to be the case in highly volatile and illiquid stock markets. Moreover, the market value may deviate from the fundamental if a bubble develops.
be the consequence of cross-country differences at the corporate level rather than institutional factors.\footnote{As shown by Demirg"uc-Kunt and Maksimovic (1999), however, the institutional environment (that is, the development of the financial and legal systems) does affect firms' financing decisions after controlling for cross-country differences in the averaged firms' characteristics.}

We define the following firms' specific control variables (see table 1).

First, in accord with Myers's theory of underinvestment (1977), Barclay and Smith (1995) have shown that firms with more growth options in their investment opportunity sets have less long-term debt in their capital structure. The reason is that stockholders have incentives to reject profitable investments when they have to share their benefits with debt holders. Myers argues that, for a given indebtedness, this incentive problem can be mitigated by shortening the maturity of debt.\footnote{Moreover, Fama (1978) shows that shortening the maturity of debt remains beneficial when stockholders can recapitalize the firm because the price at which they may repurchase the debt will reflect more the value of the investment for short-term debt than long-term debt.} We control for this by including as an explanatory variable the market-to-book ratio (a proxy for Tobin's Q) defined as the ratio of the sum of the market value of equity plus the book value of total debt over the book value of assets (that is, the sum of the book value of equity plus the book value of debt). We expect that, if the market to book ratio is a good proxy for growth opportunities, we will observe a negative correlation between the long-term debt to total debt ratio and this variable.

Second, theories of lending under asymmetric information show that the debt capacity of a firm depends on the availability of collateral. We use the proportion of net fixed assets in total assets as an indicator. Moreover, Stohs and Mauer (1996) have shown that firms in the United States match the maturity of assets and liabilities (as suggested by Hart and Moore (1994), but it is also the case if firms try to limit the risks of illiquidity). Therefore, the maturity of debt may also be positively correlated with this variable.

Third, as argued by Demirg"uc-Kunt and Maksimovic (1999), a high ratio of net sales to total assets may signal a need for short-term financing. To the extent that high sales (relative to total assets) imply high short-term assets (relative to total assets), maturity matching will also lead to a high short-term indebtedness. Thus, the ratio of net sales to total assets is also used as an explanatory variable.
### Table 1. Definition of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Firms' Characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Leverage (TDTE)</td>
<td>Total debt over book value of equity</td>
</tr>
<tr>
<td>Leverage (STDTE)</td>
<td>Short-term debt over book value of equity</td>
</tr>
<tr>
<td>Leverage (Ldte)</td>
<td>Long-term debt over book value of equity</td>
</tr>
<tr>
<td>Debt maturity (LTDTD)</td>
<td>Long-term debt over total debt</td>
</tr>
<tr>
<td>Debt maturity (STDTD)</td>
<td>Short-term debt over total debt</td>
</tr>
<tr>
<td>Growth opportunities (Tobin's Q)</td>
<td>(Market value of equity + total debt)/ (Book value of equity + total debt)</td>
</tr>
<tr>
<td>Net fixed assets (%)</td>
<td>Net fixed assets/total assets</td>
</tr>
<tr>
<td>Net sales (%)</td>
<td>Net sales/total assets</td>
</tr>
<tr>
<td>Size</td>
<td>Ln (net sales) (constant US $)</td>
</tr>
<tr>
<td>Profitability</td>
<td>[1 + (EBIT/total assets)]/[1 + CPI inflation] − 1</td>
</tr>
<tr>
<td>Volatility of earnings</td>
<td>St. dev. (EBIT)/abs [mean (EBIT)]</td>
</tr>
<tr>
<td><strong>Macroeconomic Factors</strong></td>
<td></td>
</tr>
<tr>
<td>Cost of equity</td>
<td>(1 + g)/(P/E) where g is the average rate of growth of earnings over the period and P/E is the closing P/E</td>
</tr>
<tr>
<td>Inflation</td>
<td>Consumer Price Index rate of growth</td>
</tr>
<tr>
<td>Real interest rate</td>
<td>Lending interest rate adjusted for inflation (World Development Indicators)</td>
</tr>
<tr>
<td>Volatility of inflation</td>
<td>St. dev. (inflation)/abs [mean (inflation)]</td>
</tr>
<tr>
<td>Log(GDP/cap)</td>
<td>Ln (GDP/capita) (constant US $)</td>
</tr>
<tr>
<td><strong>Financial System Development</strong></td>
<td></td>
</tr>
<tr>
<td>Credit to private sector (ec2)</td>
<td>Credit to private sector by financial intermediaries (% GDP)</td>
</tr>
<tr>
<td>Stock market capitalization (ec12)</td>
<td>Stock market capitalization (% GDP)</td>
</tr>
<tr>
<td>Stock market liquidity (ec19)</td>
<td>Value traded (% GDP)</td>
</tr>
<tr>
<td>Turnover ratio (TOR)</td>
<td>Value traded (% Capitalization)</td>
</tr>
<tr>
<td><strong>Contractual Savings Institutions</strong></td>
<td></td>
</tr>
<tr>
<td>CS development (% GDP)</td>
<td>Pension funds + life insurance* total (Pension funds + life insurance*)</td>
</tr>
<tr>
<td>(csfaGDP)</td>
<td>Financial assets (% GDP)</td>
</tr>
<tr>
<td>CS development (% Stock Mkt)</td>
<td>Pension funds + life insurance* total (Stock market capitalization)</td>
</tr>
<tr>
<td>(csfamkt)</td>
<td>Financial assets (% stock market capitalization + total outstanding debt on domestic debt market)</td>
</tr>
<tr>
<td>CS portfolio allocation</td>
<td>Shares % financial assets</td>
</tr>
<tr>
<td>(cssfa)</td>
<td>(Pension funds + life insurance*)</td>
</tr>
<tr>
<td>CS shares (% CAP)</td>
<td>Pension funds + life insurance* shares (%)</td>
</tr>
<tr>
<td>(cssshCAP)</td>
<td>Stock market capitalization</td>
</tr>
<tr>
<td>CS shares (% GDP)</td>
<td>Pension funds + life insurance* shares (%)</td>
</tr>
<tr>
<td>(csssh GDP)</td>
<td>(% GDP)</td>
</tr>
<tr>
<td><strong>Dummy Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Book reserve system</td>
<td>= 1 for Germany, Austria, Italy, and South Korea; 0 otherwise.</td>
</tr>
<tr>
<td>Centrally managed</td>
<td>= 1 for Singapore and Malaysia; 0 otherwise.</td>
</tr>
</tbody>
</table>

* Life and nonlife insurance for Argentina and Mexico.
Fourth, the size of the firm may be an important determinant of the firm indebtedness. A positive correlation between leverage and size is expected if the size is a proxy for the public information and the reputation of the firm. A similar correlation is expected with the debt maturity. Barclay and Smith (1995) find that large firms have more long-term debt in their capital structure.

Fifth, several studies in the past (Rajan and Zingales (1995) for developed economies and Demirgüç-Kunt and Maksimovic (1996b) for emerging countries) have found a negative correlation between profitability and leverage. Although this correlation is not clearly explained, we also use a profitability measure in our regressions (defined as earnings before taxes and interest expenses over total assets, deflated for inflation).

Finally, risk considerations seem to be important determinants of corporate financing decisions (Graham and Harvey 2001). Our risk control variable at the firm level is defined as the ratio of the standard deviation of earnings and the average of earnings over the period (in absolute value).

Macroeconomic Factors

Various macroeconomic factors may affect the firms’ financing patterns. We use the log of per capita gross domestic product (GDP) as a broad measure of economic development. Richer economies have more efficient institutions and a better compliance with the legal system in general, and with investor rights, accounting standards, and transparency rules (on the stock market) in particular. The inflation rate is an indicator of both the government’s management of the economy and widespread long-term contraction. It characterizes also the opportunity cost of holding money. Debt contracts may be specified in nominal terms. So we expect a negative correlation between the rate of inflation and firms’ indebtedness. Two other control variables for asset markets conditions are the real interest rate and the cost of equity. Finally, the volatility of inflation is a proxy for macroeconomic instability.

37. Note that all our firms are publicly listed.

38. The cost of equity $R_{eq}$ is $R_{eq} = (1 + g)/(P/E)$ where $g$ is the average rate of growth of future earnings and $P/E$ the current price-earnings ratio. For $g$ we use the average rate of growth of earnings over the period, and we use the $P/E$ ratio index in a given year provided by Datastream.
Financial System Characteristics

The financing patterns of firms, especially their access to external finance, depend on the characteristics of the financial system.\(^{39}\) This, in turn, affects the ability of firms to have a higher rate of growth than the one permitted by their internal resources.\(^{40}\) The stock market and banking sector variables provide a control group guaranteeing that our contractual savings variables are not simply a proxy for the level of development of the financial system.

**The Stock Market.** First, we measure the size of stock markets by stock market capitalization (in percentage of GDP). This variable has been widely used in the recent literature. The ability of the stock market to provide risk diversification opportunities and information also depends on its level of activity and liquidity (Levine and Zervos 1998). Greater liquidity will encourage investors to acquire stakes in risky firms and will enhance information acquisition by large investors (Holmstrom and Tirole 1993). Greater informational content in prices will increase the efficiency of capital allocation. Better public information may have a spillover effect on the long-term debt market by reducing initial informational asymmetries, as illustrated in the model. Activity on the stock market is measured by the turnover ratio, that is, the total value traded, in proportion to stock market capitalization.

**The Banking System.** Banks have a comparative advantage in acquiring private information on borrowers and in monitoring their actions. A sound and efficient banking sector is essential for firms, especially those that do not have access to capital markets. The use of short-term debt reduces the scope for opportunistic behavior, thus reducing the cost of monitoring. The implication for the debt maturity of firms, however, is not clear. A developed banking system implies lower monitoring costs in general. This will lead to an increase in the supply of short-term debt, but also in the supply of long-term debt, in the sense that more projects will be able to be financed by long-term debt. The

---


\(^{40}\) See Beck and others (2000) for a synthetic approach.

\(^{41}\) Greater liquidity will also make efficient restructuring decisions; see Maug (1998) for a theoretical argument.
overall impact may be negative or positive. Moreover, monitoring per se is not the only issue. The market structure of the banking sector (that is, the degree of competition among banks, and the indirect competition from other financial institutions) will have an impact on the lending behavior of banks. For instance, greater information disclosure on the stock market and in general easier outside options for firms will affect the lending behavior of banks: their ex post informational rent may be reduced, which may reduce their ex ante incentive to invest in information (see Stulz (2000)). On the other hand, greater information disclosure and better accounting standards associated with capital market development are likely to increase the supply of bank credit by limiting managerial slack. Finally, the development of nonbank financial intermediaries will probably not be neutral. This may increase competitive pressure on banks, leading them to specialize on their short-term debt comparative advantage. This competitive pressure may be direct or indirect. Contractual savings development may, however, complement the activity of the banking industry. This will be the case if these institutions act as suppliers of funds to the banking industry, instead of lending directly to firms. Because contractual savings do not face unexpected liquidity needs, they will reduce the scope for bank runs, thus limiting the term transformation risk in the banking industry. Such a mechanism would increase the incentive of banks to offer long-term loans. As a measure of the activity of the banking sector, we use the total credit to the private sector as a percentage of GDP.

**Contractual Savings Institutions.** We define several variables that proxy for the development and investment behavior of contractual savings institutions. The first variable, CSFAGDP, is defined as total contractual savings financial assets as a percentage of GDP. It measures the size of contractual savings institutions relative to the size of the economy. The second variable describes the size of contractual savings institutions relative to capital markets (CSFAMKT). It is defined as the ratio of contractual savings financial assets to market capitalization plus total bonds outstanding (with maturity greater than one year). There are two motivations for this variable: (a) it captures, although imperfectly, the relative importance of contractual savings as a provider of finance relative to the total supply of long-term finance; and (b) it partially corrects move-

42 We define also the variable CSSHGD, as contractual savings' equity investments, as a percentage of GDP.
ments in the price of shares that may introduce a spurious correlation between our firm-level variable and this explanatory variable (this is also true for the variable CSSHCAP defined below). Imagine, for instance, an exogenous rise in the prices of shares. The value of contractual savings assets and the stock market capitalization will increase, implying a correlation that has no economic meaning. Similarly, this may also introduce a negative correlation with firms’ debt-to-equity ratio. This effect is likely to be stronger when we measure firm equity by the market value of the firm. This is the reason why, as discussed in the previous paragraph, we prefer the book value of equity rather than the market value. Still, in principle, a negative correlation (but presumably weaker) may remain because firms are sensitive to their market value when they decide to issue new shares. Thus, we use both variables in order to get a rough idea of such price effects. Finally, the behavior of contractual savings institutions may significantly depend on their investments. For instance, they will have a greater incentive to be active investors in the stock market when they hold a large share of their assets in stocks. Conversely, explanations favoring corporate governance issues are less likely to be relevant in countries in which contractual savings hardly invest in the stock market. To account for such effects, we define the variable CSSHFA as the proportion of shares in the portfolio of contractual savings institutions. It is likely that the incentive for contractual savings institutions to actively exercise corporate governance on owned listed shares is positively correlated with CSSHFA. Therefore, this variable aims at capturing cross-country and time-series differences in the behavior of these institutions.

**Empirical Strategy**

Given that we use macroeconomic variables in our estimations, firm-level data are not appropriate. However, we still want to keep information from the within-country heterogeneity. For this reason, our analysis

---

43. Pagano, Panetta, and Zingales (1998) show, for instance, that initial public offerings (IPOs) are partly motivated by stock overvaluation in the industry in which the firm operates.

44. In the final set of estimations, we also report the results obtained with the variable CSSHCAP (contractual savings’ equity investments, as a percentage of stock market capitalization). This variable allows investigation into whether the size of the contractual savings’ stock holdings, relative to stock market capitalization, is an important factor.
is conducted at two different levels. First, at the country level, by taking the average values of firms' characteristics by country, and for each year. This gives us 229 observations. We use this country-level data set to illustrate our results (these results are not reported; please refer to Impavido, Musalem, and Tressel (2001)). Second, we confirm the robustness of the results by repeating the analysis at the 2 digit (SIC code) industry level by taking the average values of firms' characteristics by country and industry, for each year (tables 2, 3, and 4). Therefore, we obtain a panel data set (of approximately 6,000 observations) in which the unit is industry-country-year. OLS and fixed effects estimates are reported.

Empirical Results

The Strategy

We investigate the relationship between the development of contractual savings institutions and corporate financing patterns after controlling for firms' characteristics, macroeconomic factors, and standard financial system characteristics. In each case, we report pooled OLS and within estimations. Although endogeneity may be an issue in this type of analysis, in our case, the simultaneity bias can be expected to be lower for several reasons.45

The size and characteristics of the financial system may indeed evolve to respond to the aggregate demand for capital by the corporate sector and the public sector. Although each firm takes the size and activity of the banking sector and capital markets as given, the aggregate decisions of firms affect the size of the financial institutions. Moreover, shocks affect the financial sector and the corporate sector simultaneously. For instance, unexpected good news on profit opportunities will increase the demand for external finance by firms, and banks will also tend to offer more loans. Hence, it will increase simultaneously the size of the banking sector and firms' indebtedness. In the case of contractual savings, however, it seems more difficult to expect that their size will be significantly affected by firms' demand for capital, unless one is willing

45. In the pooled regressions, we include dummy variables for the countries having a book reserve system (Korea, Austria, Italy, and Germany) or centrally managed provident funds (Malaysia and Singapore).

46. See Demirgüç-Kunt and Maksimovic (1999) for a two-stage, least squares treatment of endogeneity of the banking sector size in a similar approach.
to argue that pension contributions and insurance premiums are significantly affected by the current business environment.

As already noted, however, endogeneity may arise because the value of contractual savings assets will move with stock market capitalization. We provide three controls for this source of simultaneity bias. First, firms' net worth is measured at book value. Second, the variable CSFAMKT should in principle partially correct for those price movements. Finally, the stock market capitalization variable should also capture the effects of such price movements. Portfolio decisions will, of course, depend on the relative returns of the different assets. For this reason, the asset allocation of pension funds may be endogenous. However, we expect the endogeneity problem to be limited also in this case because (a) price movements affecting the corporate financing patterns should be captured in the stock market capitalization variable; (b) investment regulations may be binding, especially in developing countries;\textsuperscript{47} and in many developed economies, implicit limits or strong (conservative) asset management traditions may be as important as relative returns in determining the allocation of assets;\textsuperscript{48} and (c) the results of Impavido and Musalem (2000) suggest that contractual savings development and asset allocation have an exogenous impact on capital markets development over the period studied.

\textit{Institutional Investors and Firms' Financing Patterns}

First, firms' characteristics are averaged by country, and an unbalanced panel is constructed. The OLS regressions (see Impavido, Musalem, and Tressel 2001) show that firms' capital structure and the development of contractual savings institutions are significantly correlated, after controlling for firms' characteristics (such as the maturity of assets, profitability, risk, or potential agency costs), macroeconomic factors (for example, inflation and level of development), and banking sector and stock market size and liquidity. Leverage (respectively, debt maturity) is negatively (positively) associated with the development of contractual savings institutions.

The results of pooled cross-country and cross-industry regressions (see table 2) lead to the same conclusion. After controlling for firms'  

\textit{(Text continues on page 207.)}

\textsuperscript{47} See, for instance, Srinivas and Yermo (1999).

\textsuperscript{48} For instance, in the case of Germany, it seems difficult to attribute the 2.77 percent of equity in total financial assets to low stock returns relative to other assets.
### Table 2. Contractual Savings Institutions Development and Firms’ Capital Structures

**Financial Assets, percent GDP: Pooled and Panel Estimates**

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Total Debt/Equity</th>
<th>Long-Term Debt/Equity</th>
<th>Short-Term Debt/Equity</th>
<th>Long-Term Debt/ Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled</td>
<td>Within</td>
<td>Pooled</td>
<td>Within</td>
</tr>
<tr>
<td><strong>Firms’ Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth opportunities</td>
<td>0.002</td>
<td>-0.016</td>
<td>-0.013</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td>(-0.65)</td>
<td>(-1.08)</td>
<td>(-1.15)</td>
</tr>
<tr>
<td>Net fixed assets (%)</td>
<td>0.58***</td>
<td>0.43***</td>
<td>0.54***</td>
<td>0.56***</td>
</tr>
<tr>
<td></td>
<td>(6.19)</td>
<td>(10.98)</td>
<td>(5.47)</td>
<td>(16.87)</td>
</tr>
<tr>
<td>Net sales (%)</td>
<td>0.015***</td>
<td>0.017***</td>
<td>-0.0004</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(2.62)</td>
<td>(4.2)</td>
<td>(-0.31)</td>
<td>(-0.44)</td>
</tr>
<tr>
<td>Size</td>
<td>0.13***</td>
<td>0.031</td>
<td>-0.008</td>
<td>-0.29***</td>
</tr>
<tr>
<td></td>
<td>(5.94)</td>
<td>(0.37)</td>
<td>(-1.0)</td>
<td>(-4.23)</td>
</tr>
<tr>
<td>Profitability</td>
<td>-0.07</td>
<td>-0.17***</td>
<td>-0.006***</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(-1.21)</td>
<td>(-3.75)</td>
<td>(-2.27)</td>
<td>(-1.38)</td>
</tr>
<tr>
<td>Volatility of earnings</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.001</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(-0.89)</td>
<td>(-0.66)</td>
<td>(-0.68)</td>
<td>(-1.06)</td>
</tr>
<tr>
<td><strong>Macroeconomic Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of equity</td>
<td>0.77</td>
<td>-0.13</td>
<td>0.76**</td>
<td>-0.031</td>
</tr>
<tr>
<td></td>
<td>(1.46)</td>
<td>(-0.32)</td>
<td>(1.84)</td>
<td>(-0.09)</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.04***</td>
<td>-0.004</td>
<td>-0.029***</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(-3.14)</td>
<td>(-0.17)</td>
<td>(-2.96)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Real lending interest rate (short-term)</td>
<td>-0.047***</td>
<td>0.037*</td>
<td>0.006</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(-2.90)</td>
<td>(1.77)</td>
<td>(0.48)</td>
<td>(1.31)</td>
</tr>
<tr>
<td>Volatility of inflation</td>
<td>0.04***</td>
<td>0.013</td>
<td>0.021***</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(3.13)</td>
<td>(0.25)</td>
<td>(2.13)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Log (GDP/capita)</td>
<td>-0.13**</td>
<td>-0.19</td>
<td>0.06</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(-1.70)</td>
<td>(-0.62)</td>
<td>(0.81)</td>
<td>(0.68)</td>
</tr>
</tbody>
</table>

(Table continues on the following page.)
TABLE 2. (CONTINUED)

Financial Assets, percent GDP: Pooled and Panel Estimates

<table>
<thead>
<tr>
<th>Financial System Development</th>
<th>Dependent Variables</th>
<th>Total Debt/Equity</th>
<th>Long-Term Debt/Equity</th>
<th>Short-Term Debt/Equity</th>
<th>Long-Term Debt/Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pooled</td>
<td>Within</td>
<td>Pooled</td>
<td>Within</td>
</tr>
<tr>
<td>Credit to private sector</td>
<td>0.006***</td>
<td>0.009***</td>
<td>0.0038***</td>
<td>0.0058**</td>
<td>0.0033***</td>
</tr>
<tr>
<td>Stock market capitalization</td>
<td>-0.0067***</td>
<td>-0.004**</td>
<td>-0.003***</td>
<td>-0.0026**</td>
<td>-0.003***</td>
</tr>
<tr>
<td>Stock market liquidity</td>
<td>-0.32**</td>
<td>0.005</td>
<td>-0.02</td>
<td>0.007</td>
<td>0.0008</td>
</tr>
<tr>
<td>Contractual savings</td>
<td>-0.45***</td>
<td>0.36</td>
<td>-0.13</td>
<td>0.18</td>
<td>-0.27***</td>
</tr>
<tr>
<td>development (financial assets, % GDP)</td>
<td>(-4.68)</td>
<td>(0.76)</td>
<td>(-1.72)</td>
<td>(0.46)</td>
<td>(-5.94)</td>
</tr>
<tr>
<td>Dummy Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector -country (2 digit SIC code)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Book reserve system</td>
<td>1.03***</td>
<td>0.3</td>
<td></td>
<td>0.96***</td>
<td></td>
</tr>
<tr>
<td>Centrally managed pension funds</td>
<td>-0.13</td>
<td>-0.16*</td>
<td></td>
<td>-0.028</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.096</td>
<td>0.04</td>
<td>0.05</td>
<td>0.024</td>
<td>0.087</td>
</tr>
<tr>
<td>No. of observations</td>
<td>6,728</td>
<td>6,728</td>
<td>6,728</td>
<td>6,728</td>
<td>6,728</td>
</tr>
<tr>
<td>No. of cross-section units</td>
<td>1,046</td>
<td>1,046</td>
<td>1,046</td>
<td>1,046</td>
<td></td>
</tr>
<tr>
<td>Fixed effects</td>
<td>2.34***</td>
<td>4.06***</td>
<td>2.14***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Financial Assets, percent of Market Capitalization: Pooled and Panel Estimates

### Dependent Variables

<table>
<thead>
<tr>
<th></th>
<th>Total Debt/Equity</th>
<th>Long-Term Debt/Equity</th>
<th>Short-Term Debt/Equity</th>
<th>Long-Term Debt/Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled Within</td>
<td>Pooled Within</td>
<td>Pooled Within</td>
<td>Pooled Within</td>
</tr>
<tr>
<td><strong>EXPLANATORY VARIABLES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Firms' Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth opportunities</td>
<td>0.0026 (-0.017)</td>
<td>-0.008 (-0.016)</td>
<td>0.022 (0.018)</td>
<td>-0.0036*** (-0.002)*</td>
</tr>
<tr>
<td></td>
<td>(0.08) (-0.64)</td>
<td>(-0.62) (-0.71)</td>
<td>(0.75) (0.11)</td>
<td>(-3.10) (-1.78)</td>
</tr>
<tr>
<td>Net Fixed assets (%)</td>
<td>0.61*** 0.45***</td>
<td>0.568*** 0.58***</td>
<td>0.031 -0.12***</td>
<td>0.02*** 0.014***</td>
</tr>
<tr>
<td></td>
<td>(6.38) (10.88)</td>
<td>(5.69) (16.85)</td>
<td>(0.89) (-4.77)</td>
<td>(6.23) (7.23)</td>
</tr>
<tr>
<td>Net sales (%)</td>
<td>0.016*** 0.018***</td>
<td>-0.0008 -0.002</td>
<td>0.015*** 0.021***</td>
<td>-0.0007*** -0.0006**</td>
</tr>
<tr>
<td></td>
<td>(2.47) (4.12)</td>
<td>(-0.51) (-0.69)</td>
<td>(2.52) (7.47)</td>
<td>(-3.58) (-3.01)</td>
</tr>
<tr>
<td>Size</td>
<td>0.14*** 0.037</td>
<td>-0.017 -0.36***</td>
<td>0.059*** 0.07</td>
<td>0.026*** -0.006</td>
</tr>
<tr>
<td></td>
<td>(5.87) (0.39)</td>
<td>(-0.176) (-4.3)</td>
<td>(5.31) (1.21)</td>
<td>(12.05) (-1.36)</td>
</tr>
<tr>
<td>Profitability</td>
<td>-0.08 -0.18***</td>
<td>-0.035** -0.052</td>
<td>-0.045 -0.13***</td>
<td>-0.00006 0.0023</td>
</tr>
<tr>
<td></td>
<td>(-1.31) (-3.91)</td>
<td>(-2.23) (-1.34)</td>
<td>(-0.78) (-4.36)</td>
<td>(-0.02) (1.08)</td>
</tr>
<tr>
<td>Volatility of earnings</td>
<td>-0.002 -0.005</td>
<td>-0.001 -0.005</td>
<td>-0.001 -0.0006</td>
<td>0.0003 -0.0013***</td>
</tr>
<tr>
<td></td>
<td>(-0.94) (-0.91)</td>
<td>(-0.70) (-1.25)</td>
<td>(-0.79) (-0.18)</td>
<td>(0.80) (-5.17)</td>
</tr>
<tr>
<td><strong>Macroeconomic Factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of equity</td>
<td>0.88 0.34</td>
<td>0.94** 0.39</td>
<td>0.11 -0.008</td>
<td>-0.057*** 0.004</td>
</tr>
<tr>
<td></td>
<td>(1.42) (0.52)</td>
<td>(1.97) (0.74)</td>
<td>(0.42) (-0.02)</td>
<td>(-2.67) (0.14)</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.025* 0.004</td>
<td>-0.029** 0.007</td>
<td>-0.007 -0.0079</td>
<td>-0.006*** -0.0029***</td>
</tr>
<tr>
<td></td>
<td>(-1.71) (0.16)</td>
<td>(-1.98) (0.33)</td>
<td>(-1.21) (-0.47)</td>
<td>(-4.49) (-2.33)</td>
</tr>
<tr>
<td>Real lending interest rate (short-term)</td>
<td>-0.049** 0.046*</td>
<td>0.013 0.027</td>
<td>-0.051*** 0.023</td>
<td>0.0027* -0.001</td>
</tr>
<tr>
<td></td>
<td>(-2.26) (1.79)</td>
<td>(0.76) (1.27)</td>
<td>(-3.82) (1.49)</td>
<td>(2.0) (-0.92)</td>
</tr>
<tr>
<td>Volatility of inflation</td>
<td>0.037*** 0.009</td>
<td>0.002* 0.0001</td>
<td>0.017*** 0.01</td>
<td>-0.006 -0.0006</td>
</tr>
<tr>
<td></td>
<td>(2.64) (0.17)</td>
<td>(1.82) (0.004)</td>
<td>(2.81) (0.30)</td>
<td>(-1.34) (-0.25)</td>
</tr>
<tr>
<td>Log (GDP/capita)</td>
<td>-0.226*** -0.50</td>
<td>0.15** 0.048</td>
<td>-0.32*** -0.17</td>
<td>0.09*** -0.016</td>
</tr>
<tr>
<td></td>
<td>(-2.61) (-1.44)</td>
<td>(2.08) (0.16)</td>
<td>(-5.14) (-0.79)</td>
<td>(10.51) (-1.01)</td>
</tr>
</tbody>
</table>

(Table continues on the following page.)
<table>
<thead>
<tr>
<th>Financial System Development</th>
<th>Total Debt/Equity</th>
<th>Long-Term Debt/Equity</th>
<th>Short-Term Debt/Equity</th>
<th>Long-Term Debt/Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled</td>
<td>Within</td>
<td>Pooled</td>
<td>Within</td>
</tr>
<tr>
<td>Credit to private sector</td>
<td>0.006***</td>
<td>0.012***</td>
<td>0.003***</td>
<td>0.0076**</td>
</tr>
<tr>
<td></td>
<td>(4.49)</td>
<td>(2.78)</td>
<td>(2.44)</td>
<td>(2.13)</td>
</tr>
<tr>
<td>Stock market capitalization</td>
<td>-0.008***</td>
<td>-0.00017</td>
<td>-0.0044***</td>
<td>-0.002**</td>
</tr>
<tr>
<td></td>
<td>(-5.59)</td>
<td>(0.08)</td>
<td>(-5.65)</td>
<td>(-0.13)</td>
</tr>
<tr>
<td>Stock market liquidity</td>
<td>-0.52**</td>
<td>0.008</td>
<td>-0.03</td>
<td>0.009</td>
</tr>
<tr>
<td>(Turnover ratio)</td>
<td>(-2.92)</td>
<td>(0.44)</td>
<td>(-0.17)</td>
<td>(0.61)</td>
</tr>
<tr>
<td>Contractual savings development</td>
<td>0.065</td>
<td>2.09***</td>
<td>0.18</td>
<td>1.27*</td>
</tr>
<tr>
<td>(financial assets, % CAP. MKT.)</td>
<td>(0.32)</td>
<td>(2.39)</td>
<td>(1.28)</td>
<td>(1.74)</td>
</tr>
<tr>
<td>Dummy Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector-country</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Book reserve system</td>
<td>1.19***</td>
<td>0.34</td>
<td>0.96***</td>
<td>-0.086**</td>
</tr>
<tr>
<td></td>
<td>(4.88)</td>
<td>(1.43)</td>
<td>(4.95)</td>
<td>(-6.87)</td>
</tr>
<tr>
<td>Centrally managed pension funds</td>
<td>-0.16</td>
<td>0.036</td>
<td>-0.028</td>
<td>-0.16***</td>
</tr>
<tr>
<td></td>
<td>(-1.24)</td>
<td>(0.46)</td>
<td>(-0.56)</td>
<td>(-8.36)</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.099</td>
<td>0.017</td>
<td>0.05</td>
<td>0.024</td>
</tr>
<tr>
<td>No. of observations</td>
<td>5,867</td>
<td>5,867</td>
<td>5,867</td>
<td>5,867</td>
</tr>
<tr>
<td>No. of cross-section units</td>
<td>943</td>
<td>1046</td>
<td>943</td>
<td>936</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>2.46***</td>
<td>4.06***</td>
<td>2.01***</td>
<td>11.3***</td>
</tr>
</tbody>
</table>

* = Significant at the 10 percent level.
** = Significant at the 5 percent level.
*** = Significant at the 1 percent level.

Note: t-statistics in parentheses.
characteristics averaged by industries in each country, for macroeconomic factors, and for financial system characteristics, the level of development of contractual savings institutions is negatively correlated with leverage and positively correlated with the maturity of debt. Moreover, it is positively correlated with debt maturity. Further inspection of the table shows that the coefficients on firms' characteristics are consistent with what we expected. Firms are more indebted and have more long-term debt when net fixed assets represent a larger share of total assets. Larger sales relative to total assets imply more debt and more short-term debt. More profitable firms tend to be less indebted, and growth firms have less long-term debt relative to total debt. Finally, riskier firms have a lower maturity of debt. The size of the banking sector is correlated positively with firms' leverage and negatively with debt maturity. This second point is consistent with the result of Demirgüç-Kunt and Maksimovic (1999). As expected, the stock market capitalization is negatively correlated with leverage. It is also positively correlated with debt maturity. One explanation sometimes proposed for this effect is that there are informational spillovers from the stock market, which reduces the asymmetries of information, hence increasing the supply of long-term debt.

Our results, however, are not robust to the inclusion of unobserved fixed effects at the industry level in each country. In table 2b, we perform the same regressions by using the variable CSFAMKT.49 The results suggests the previous variable CSFAGDP indeed introduces a spurious correlation (as discussed before) between the level of development of contractual savings and leverage. Now, leverage is positively correlated with the level of development of contractual savings. As suggested by the regressions on LTDTE and LTDTD, however, the mechanism seems to work through an increase in long-term debt relative to equity and long-term debt relative to short-term debt.

Overall, these two sets of regressions tend to support the hypothesis of a global impact of contractual savings development on leverage. Moreover, the development of contractual savings institutions seems to foster the use of long-term debt.

The absence of a strongly robust effect on the whole sample should not be totally surprising, given that contractual savings institutions, as we showed in the previous paragraph, have extremely different

49. We ran the regressions with CSSHCAP, with very similar results (not reported here).
investment behaviors from one country to another. We should expect a fall in leverage when contractual savings develop only if the cost of equity finance falls, which happens if the aggregate supply of equity increases (or for other reasons listed in the section, Empirical Results). How contractual savings institutions invest their resources should have a crucial impact. The next result confirms this hypothesis.

In table 3, we look at the impact of contractual savings portfolio choices on corporate financial decisions. We obtain a strong and economically significant effect on leverage. An increase in the proportion of financial assets invested in shares is associated with a decrease in corporate leverage. It leads also to a decrease in short-term debt relative to equity. This is robust to unobserved industries’ fixed effects. This set of results is consistent with the claim that the investment behavior of contractual savings institutions matters for corporate financing patterns. Their investment decisions have a significant impact on firms’ capital structure: for instance, the coefficients of the pooled and within estimates imply that if Korean contractual savings institutions had had the same investment behavior as in South Africa (where contractual savings are investing 44 percent of their financial assets in shares on average over the period, compared with 12 percent in Korea), the debt-to-equity ratio of Korean firms would have decreased from 4.9 to 4.6 in the pessimistic case, or to 3.9 in the optimistic case—hence a decrease of between 6 percent and 20 percent. Overall, these results strongly suggest that: any attempt to understand cross- and within-country variations in corporate financing patterns needs to assess the role of nonbank financial intermediaries, such as institutional investors; and policy interventions that remove binding constraints on portfolios may have sizeable effects on the corporate sector financing patterns.

Figure 1 plots the investment limits on equity investments (vertical axis) against the actual proportion of equity in pension funds and life insurance portfolios in 1998 for a subset of countries. As shown, investment limits seem not to be binding.50 Note, however, that the existence of latitude in equity investment per se does not mean that investment regulations have no impact on investments. It can be interpreted in two opposite ways: restrictions have no effects on the asset allocation; or restrictions (may) lead to very cautious portfolio management to avoid breaching them, even if potential returns are high. The existence

50. It is tight for Swedish life insurance companies.
FIGURE 1. EQUITY: PORTFOLIO RESTRICTIONS AND INVESTMENTS (1998)

(a) Pension funds

Investment limit on equity funds

<table>
<thead>
<tr>
<th>Country</th>
<th>Investment limit on equity funds</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITA</td>
<td>100</td>
</tr>
<tr>
<td>JPN</td>
<td>66.8</td>
</tr>
<tr>
<td>CAN</td>
<td>NLD</td>
</tr>
<tr>
<td>NLD</td>
<td>USA GBR</td>
</tr>
<tr>
<td>SWE</td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>CHL</td>
</tr>
<tr>
<td>FIN</td>
<td></td>
</tr>
<tr>
<td>DEU</td>
<td>MEX</td>
</tr>
</tbody>
</table>

Equity share of pension fund portfolio, 1998

(b) Life insurance

Investment limit on equity funds

<table>
<thead>
<tr>
<th>Country</th>
<th>Investment limit on equity funds</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLD</td>
<td>USA GBR</td>
</tr>
<tr>
<td>ARG</td>
<td></td>
</tr>
<tr>
<td>BRA</td>
<td>FIN</td>
</tr>
<tr>
<td>CHL</td>
<td></td>
</tr>
<tr>
<td>DEU</td>
<td></td>
</tr>
<tr>
<td>JPN</td>
<td>SWE</td>
</tr>
<tr>
<td>CAN</td>
<td>ITA</td>
</tr>
</tbody>
</table>

Equity share of life insurance portfolio, 1998

Sources: OECD (2000) and Davis (2001).

of a positive correlation between maximum limits and actual investments in equity may favor the second interpretation.

One issue is, however, difficult to address with the data available. In the recent years, there has been a trend in the internationalization of
# Table 3. Contractual Savings Portfolios and Firms' Capital Structures

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Firms' Characteristics</th>
<th>Long-Term Debt/Equity</th>
<th>Long-Term Debt/Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pooled Within</td>
<td>Pooled Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent Variables</td>
<td>Long-Term Debt/Equity</td>
<td>Long-Term Debt/Equity</td>
<td>Long-Term Debt/Equity</td>
</tr>
<tr>
<td></td>
<td>Pooled Within</td>
<td>Pooled Within</td>
<td>Pooled Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth opportunities</td>
<td>-0.02 (-0.56)</td>
<td>-0.047 (-1.48)</td>
<td>-0.0058 (-0.32)</td>
</tr>
<tr>
<td>Net fixed assets (%)</td>
<td>0.60*** (0.58)</td>
<td>0.43*** (10.17)</td>
<td>0.56*** (4.71)</td>
</tr>
<tr>
<td>Net sales (%)</td>
<td>0.018*** (2.68)</td>
<td>0.021*** (4.81)</td>
<td>0.00001 (-0.35)</td>
</tr>
<tr>
<td>Size</td>
<td>0.09*** (4.17)</td>
<td>0.14 (1.57)</td>
<td>-0.035 (-0.35)</td>
</tr>
<tr>
<td>Profitability</td>
<td>-0.077 (-1.26)</td>
<td>-0.18*** (-3.98)</td>
<td>-0.036*** (-2.77)</td>
</tr>
<tr>
<td>Volatility of earnings</td>
<td>-0.007 (-0.28)</td>
<td>-0.006 (0.75)</td>
<td>0.006 (0.11)</td>
</tr>
<tr>
<td>Macroeconomic Factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of equity</td>
<td>0.82 (1.36)</td>
<td>-0.22 (-0.55)</td>
<td>0.80* (1.75)</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.03** (-3.42)</td>
<td>-0.015 (-0.92)</td>
<td>-0.022*** (-2.79)</td>
</tr>
<tr>
<td>Real lending interest rate</td>
<td>-0.046*** (-3.14)</td>
<td>0.02 (1.20)</td>
<td>-0.006 (-0.63)</td>
</tr>
<tr>
<td>Volatility of inflation</td>
<td>0.022 (1.66)</td>
<td>-0.005 (-0.10)</td>
<td>0.014 (1.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial System Development</td>
<td>Log (GDP/capita)</td>
<td>-0.19*</td>
<td>-0.44</td>
</tr>
<tr>
<td></td>
<td>(-1.73)</td>
<td>(-1.25)</td>
<td>(0.45)</td>
</tr>
<tr>
<td></td>
<td>Financial System Development</td>
<td>Credit to private sector</td>
<td>0.006**</td>
</tr>
<tr>
<td></td>
<td>(-2.37)</td>
<td>(2.62)</td>
<td>(3.29)</td>
</tr>
<tr>
<td></td>
<td>Stock market capitalization</td>
<td>-0.007***</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(-4.97)</td>
<td>(-0.53)</td>
<td>(-3.35)</td>
</tr>
<tr>
<td></td>
<td>Stock market liquidity</td>
<td>-0.24</td>
<td>0.0079</td>
</tr>
<tr>
<td></td>
<td>(Turnover ratio)</td>
<td>(-1.09)</td>
<td>(0.52)</td>
</tr>
<tr>
<td></td>
<td>Contractual savings portfolio</td>
<td>-0.92***</td>
<td>-3.12***</td>
</tr>
<tr>
<td></td>
<td>(-3.48)</td>
<td>(-2.51)</td>
<td>(-4.13)</td>
</tr>
<tr>
<td></td>
<td>Dummy Variables</td>
<td>Sector-country</td>
<td>0.89***</td>
</tr>
<tr>
<td></td>
<td>(2 digit SIC code)</td>
<td>(3.21)</td>
<td>(0.08)</td>
</tr>
<tr>
<td></td>
<td>Book reserve system</td>
<td>-0.43***</td>
<td>-0.51***</td>
</tr>
<tr>
<td></td>
<td>(-3.77)</td>
<td>(-3.05)</td>
<td>(-1.37)</td>
</tr>
<tr>
<td></td>
<td>Adjusted R-squared</td>
<td>0.1</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>No. of observations</td>
<td>5,501</td>
<td>5,501</td>
</tr>
<tr>
<td></td>
<td>No. of cross-section units</td>
<td>904</td>
<td>904</td>
</tr>
<tr>
<td></td>
<td>Fixed effects</td>
<td>2.66***</td>
<td>4.17***</td>
</tr>
</tbody>
</table>

* = Significant at the 10 percent level.
** = Significant at the 5 percent level.
*** = Significant at the 1 percent level.

Note: t-statistics in parentheses.
stock markets and the diminished importance of the domestic stock market. Several questions are difficult to address here. For instance, what is the impact of foreign contractual savings on domestic stock markets? Conversely, what proportion of domestic contractual savings funds are invested abroad? We lack a sufficiently large country coverage to address these issues. Still, the domestic bias in investment decisions is likely to be important, whatever the reason.

In addition, the channels through which contractual savings institutions affect the corporate financing decisions cannot be disentangled on the basis of this first cross-country analysis. Moreover, as suggested by descriptive statistics (Impavido, Musalem, and Tressel 2001), we may simply capture cross-country differences in their overall financial structure (although such an argument cannot explain our fixed-effects results). The results displayed in the next section enlighten the channels through which contractual savings institutions affect corporate financing choices. They provide a basis for better-targeted policy interventions.

Financial Structure and Financial Channels

We use the classification of macroeconomic financial structures developed by Demirgüç-Kunt and Levine (1999). Countries are divided into two subgroups (see Impavido, Musalem, and Tressel 2001): economies with bank-based financial structures and economies with market-based financial structures. This classification has been constructed by using a large set of indicators for size, activity, and efficiency of the banking sector and the stock market. It provides a rough evaluation of whether savings are channeled to productive activities mainly through the banking system or the stock market. This is, therefore, a relevant classification for our purposes. In market-based economies, for instance, the contractual savings industry accounts for 46.3 percent of long-term capital markets size, and equity investments are 30.7 percent of total financial assets and 29 percent of stock market capitaliza-

---

51 A recent paper by Beck and others (2000) shows that the financial structure does not explain economic growth and the reliance on external financing after controlling for the level of financial development. Our results are not contradictory: we show that this classification does help to identify different channels through which corporate financing choices are affected by the development of contractual savings institutions.
tion. In bank-based economies, the same figures are, respectively, 22.3 percent, 12.3 percent, and 12.2 percent. Therefore, the contractual savings industry is less developed in countries classified as bank-based than in market-based countries. Moreover, pension funds and life insurance companies invest significantly less on the stock market in bank-based economies than in market-based economies.

The rationale for using this distinction is that it allows for disentangling the different impacts on firms' capital structure, and thus suggests that there is more than one way through which contractual savings institutions development modifies firms' financing choices.

Although we have no information on the maturity of debt instruments held by contractual savings institutions (except for four countries), we are able to break their assets between the two categories: bills and bonds (hereafter BB); and loans (hereafter LL), for a significant number of countries. In market-based economies, BB represents 42.6 percent of total financial assets and LL only 13.9 percent. In bank-based economies, the same figures are, respectively, 45 percent and 31.6 percent. It seems, therefore, that, on average, lower equity investments in bank-based economies are mostly explained by a higher proportion of loans in portfolio.

The relative importance of pension funds and life insurance companies differs in the two groups of countries. Pension funds account on average for 30 percent and 20.4 percent of total contractual savings financial assets, respectively, in market-based and bank-based economies. In particular, Anglo-Saxon countries and continental Europe exhibit strongly different contractual savings industries. Pension funds hold 70 percent, 54 percent, and 50 percent of contractual savings financial assets in the United States, the United Kingdom, and Australia, respectively. In Germany, Italy, and France, the figures are 12 percent, 37 percent, and less than 1 percent.

First, in Impavido, Musalem, and Tressel (2001), we display the conditional correlation between leverage and, respectively, contractual savings size and asset allocation in a pooled regression at the country level. After controlling for firms' characteristics, macroeconomic factors, and bank and stock market size, a significant correlation remains between leverage and contractual savings size. Contractual savings development, however, has a different impact on leverage in market-based or bank-based economies. In countries in which the stock market is the core of the financial system, the development of pension funds and insurance companies leads to a decrease in leverage. In bank-based financial systems the opposite effect seems to dominate:
the development of contractual savings implies an increase in leverage. Moreover, the proportion of equity investments in the portfolio is negatively correlated with leverage in market-based financial systems, whereas it seems to have no significant effect on leverage in bank-based economies.

The analysis is repeated for the debt maturity. Again, the impact of contractual savings development is strikingly different in bank-based economies and in market-based economies. In the latter, the development of contractual savings institutions implies a decrease in debt maturity, whereas in the former it implies an increase in the debt maturity.

The industry-level analysis (see table 4) confirms the results obtained at the country level. We report the coefficient on the contractual savings variable and its significance for each subgroup of countries. In market-based economies, there is a strongly significant impact of contractual savings portfolio choices on firms' financing patterns: an increase in equity investments by contractual savings leads to a decline in leverage, for our three variables. The effect is robust to unobserved industry-specific fixed effects within countries, and it is economically large. The impact of contractual savings development is somewhat weaker, although it affects leverage in a similar way. Debt maturity is also negatively correlated either to the level of development of contractual savings or the proportion of share investments in the portfolios of contractual savings. These results are consistent with the intuition. Because contractual savings are large in these countries on average, it is likely that their marginal effect on firms' financing patterns go through their investment choices rather than through an increase in their size. As they increase their equity holdings, firms tend to substitute equity finance for debt finance. These results suggest that banks and institutional investors are indirect competitors. The fall in the maturity of debt may be partly attributed to the fact that banks concentrate on their core activity, which is short-term lending.

In the case of bank-based economies, the channels through which firms' capital structures are affected are noticeably different. The dominant effect is the level of development of the contractual savings

52. We ran the regression by moving Korea from the market-based subgroup to the bank-based one; results were not affected.

53. More precisely, it seems that the characteristics of contractual savings portfolio are more important than the size of share holdings relative to stock market capitalization. This result favors a corporate governance explanation.
### Table 4. Market-Based and Bank-Based Financial Systems—Contractual Savings and Firms' Financing Choices

**Pooled and Panel Estimates**

**Summary of the Results**

#### Market-based Financial Systems

<table>
<thead>
<tr>
<th>Contractual savings</th>
<th>Total Debt/Equity</th>
<th>LT Debt/Equity</th>
<th>ST Debt/Equity</th>
<th>Long-Term Debt/Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS Within</td>
<td>OLS Within</td>
<td>OLS Within</td>
<td>OLS Within</td>
</tr>
<tr>
<td>Development (financial assets, % GDP)</td>
<td>-0.55***</td>
<td>0.007</td>
<td>-0.37***</td>
<td>0.053</td>
</tr>
<tr>
<td>Development (financial assets, % Capital market)</td>
<td>-0.28</td>
<td>-0.01</td>
<td>-0.30***</td>
<td>0.4</td>
</tr>
<tr>
<td>Development (shares, % stock market cap.)</td>
<td>-0.64***</td>
<td>0.23</td>
<td>-0.44***</td>
<td>0.45</td>
</tr>
<tr>
<td>Portfolio (Shares, % financial assets)</td>
<td>-0.96***</td>
<td>-4.06***</td>
<td>-0.64***</td>
<td>-1.93***</td>
</tr>
</tbody>
</table>

215

(Table continues on the following page.)
### TABLE 4. (CONTINUED)

**Pooled and Panel Estimates**

*Summary of the Results*

**Bank-Based Financial Systems**

<table>
<thead>
<tr>
<th></th>
<th>Total Debt/Equity</th>
<th>LT Debt/Equity</th>
<th>ST Debt/Equity</th>
<th>Long-Term Debt/Total Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS Within</td>
<td>OLS Within</td>
<td>OLS Within</td>
<td>OLS Within</td>
</tr>
<tr>
<td>Contractual savings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>development (financial</td>
<td>3.66***</td>
<td>2.78**</td>
<td>0.69</td>
<td>0.217***</td>
</tr>
<tr>
<td>assets, % GDP)</td>
<td>(3.42)</td>
<td>(2.49)</td>
<td>(1.14)</td>
<td>(0.217)</td>
</tr>
<tr>
<td>Contractual savings</td>
<td>3.3***</td>
<td>2.85***</td>
<td>0.21</td>
<td>0.41***</td>
</tr>
<tr>
<td>development (financial</td>
<td>(4.26)</td>
<td>(3.53)</td>
<td>(0.49)</td>
<td>(10.28)</td>
</tr>
<tr>
<td>assets, % Capital market)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contractual savings</td>
<td>1.88***</td>
<td>1.12</td>
<td>0.49</td>
<td>0.04</td>
</tr>
<tr>
<td>development (shares, %</td>
<td>(2.47)</td>
<td>(1.25)</td>
<td>(1.19)</td>
<td>(1.19)</td>
</tr>
<tr>
<td>stock market sap.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contractual savings</td>
<td>2.27</td>
<td>-6.9</td>
<td>0.79</td>
<td>-0.013</td>
</tr>
<tr>
<td>portfolio (Shares, %</td>
<td>(1.11)</td>
<td>(-1.11)</td>
<td>(0.32)</td>
<td>(3.84)</td>
</tr>
<tr>
<td>financial assets)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = Significant at the 10 percent level.
** = Significant at the 5 percent level.
*** = Significant at the 1 percent level.

Notes: t-statistics in parentheses. Control variables include: firm characteristics, macroeconomic factors, and financial system characteristics.
industry, whereas the asset allocation hardly affects firms' capital structures (still we find evidence of a positive impact on the maturity of debt). The no-correlation result with the portfolio variable makes sense: because contractual savings investment in equity is no more than 12 percent of stock market capitalization, a change in their behavior (measured by CSSHFA) is very unlikely to affect significantly the aggregate corporate financing choices.

The level of development of the contractual savings industry has a strong positive effect on leverage and a positive effect on the maturity of debt. These results suggest that the channel through which contractual savings affect the corporate financing patterns does not go through the stock market. Indeed, contractual savings development is associated with an increase in debt finance—and an increase in debt maturity. As explained above, it is very unlikely that this can be explained by the higher investments in bonds in bank-based economies than in market-based economies. Rather, the explanation is likely to be related to loans: either they lend directly to the productive sector, or they are complementary to the banking sector. More specifically, by reducing the risk of liquidity in the banking system, they may increase the incentive for banks to provide more long-term loans in proportion to total loans.

**Conclusion**

In this paper, we have analyzed the relationship between the development, and asset allocation, of contractual savings institutions and firms' capital structures after controlling for firms' characteristics, macroeconomic factors, and standard financial system characteristics. We have shown that the development of contractual savings institutions, as well as their portfolio decisions, is significantly associated with firms' financing patterns across and within countries. The empirical results are consistent with contractual savings institutions having a comparative advantage in supplying long-term finance to the corporate sector.

We have identified different channels through which contractual savings affect the financing decisions of firms. In bank-based economies, the development of contractual savings is associated with an increase in firms' leverage and maturity of debt. In market-based economies, instead, the asset allocation affects firms' leverage: an increase in the proportion of shares in the portfolio of contractual savings is associated with a decrease in firms' leverage. These results sug-
gest that there might be an efficiency gain at the firm level: an increase in the array of external financing possibilities is associated with increased maturity of firms' liabilities. In other words, when contractual savings institutions are underdeveloped, firms cannot obtain enough long-term finance.

Increased maturity of corporate sector liabilities should increase its resilience to various shocks (such as refinancing risks and bankruptcy risks). The impact works through several possible channels. In market-based economies, the main effect seems to work through the stock market and equity finance. In bank-based economies, it seems to work through the supply of loans. More analysis, however, is needed to identify the precise channels through which contractual savings institutions interact with the financial system.

Finally, the policy implications of the paper are clear. If demographic, institutional, and political preconditions for pension reforms (or reform of the insurance industry) are met, policymakers should pay particular attention to financial sector development policies that enhance the efficiency of the contractual savings industry as a major provider of noncaptive funds. Regulation, in particular for equity investments, may have a large impact, as suggested by our preliminary results, when portfolio limits affect actual investments. In addition, policy intervention should be based on a precise evaluation of the interaction between institutional investors and other components of the financial system (especially banks).

54. See, for instance, Vitas (2000).
References


Davis, E. Philip. 2001. “Portfolio Regulation of Life Insurance Companies and Pension Funds.” The Pensions Institute, Birbeck College, University of London.


Gregorio Impavido, Alberto R. Musalem, and Thierry Tressel


Public Expenditures and Risk Reduction

Shantayanan Devarajan and Jeffrey S. Hammer

Abstract

Governments in developing countries spend money on goods and services that have an impact on people's exposure to risk. This paper presents a simple approach to valuing such expenditures from the perspective of risk reduction. There are two types of expenditures: public provision of insurance, such as for health care or crop yields; and policies aimed at other objectives that change peoples' risk profile, such as transfer programs, tax and subsidy policies, and infrastructure investments. Several examples of each type show that incorpo-
rating the risk-reducing perspective significantly alters the value of public expenditures, which indicates that these considerations should be included in standard public expenditure analysis.

That public spending rises with a country’s per capita income is well known. In the United States, for example, public spending was 7.5 percent of gross domestic product (GDP) in 1913 and is 33 percent today. Governments in present-day developed countries spend about twice as much as developing countries. Less well known is that government spending on goods and services is the same in developed and developing countries; the difference is almost entirely the result of transfer payments, which are about 22 percent of GDP in the industrial world (Tanzi and Schuknecht 1997). Many of these transfer payments—unemployment insurance, pensions, health insurance, and guaranteed loans—have the characteristic that they are aimed at mitigating risk in the private economy.

In this paper, we explore how the existing framework for evaluating government spending on goods and services, welfare economics (Samuelson 1954; Musgrave 1959), can be extended to incorporate the government’s various risk-reducing activities. Because governments do not typically classify their expenditures by their risk-altering characteristics, our approach will be more conceptual than empirical. We illustrate our points with some simple examples and models designed to capture the risk-reducing properties of various public expenditures. We show that adopting a risk-reducing perspective has implications for the costs and benefits of certain public expenditures and taxes that are different from the standard analysis, which indicates directions of change in the composition of public spending that are welfare-enhancing.

In the first section of the paper, Analytical Framework, we present an approach to evaluating the welfare consequences of policies that influence the size and distribution of risks that people face. We also show how the approach can provide insight into the positive question of why governments of developed countries spend more on risk-related expenditures than governments of developing countries. In the next section, Applying the Framework, we apply the framework to the normative question of evaluating the benefits and costs of common programs associated, directly or indirectly, with the reduction of risk, including crop insurance, medical care, income support, flood control, education, and loans. The last section offers some concluding remarks.
Analytical Framework

The framework for evaluating public expenditures aimed at reducing risk begins with the metric for valuing the reduction of risk to the individual, which is the familiar von Neumann-Morgenstern framework. Risk aversion is modeled in this framework by a utility of income function that rises at a decreasing rate. As a result, people will generally prefer a certain outcome to a risky one with the same expected value. A job at $20,000 per year is better than taking a 50 percent chance on getting one at $40,000 with a 50 percent chance of no income at all. How much that is worth depends on how much greater the difference in utility is between $20,000 and zero than the difference between $20,000 and $40,000. There is an amount of money that one is willing to pay to assure an income of $20,000 (minus that payment) as opposed to taking the risk. This is called the risk premium and the amount of income left over after paying the premium is called the certainty equivalent income to the risky situation.

Formally, this can be expressed as $U(W - V) = EU(W + \Sigma \epsilon_i)$ where $U(\cdot)$ is the utility function of income (strictly speaking, wealth) denoted $W$, and $V$ is the maximum amount one would pay to have a certain income relative to the variable one. The expectations operator $E$ takes the average of utility when wealth is risky, and $\Sigma \epsilon_i$ is the sum of all risky components of wealth, each normalized to have zero mean. This expression says that there is a value $V$ that makes the individual indifferent between the situation with a certain income, $W - V$, and the situation in which that person faces all risks. The risky component is written as a sum of potentially many different “shocks” to income in which only their sum—their net impact on income—is of ultimate concern to the individual.

Even if we can “explain” the higher government expenditure in developed countries by showing the value of reducing risks (box 1), it does not follow that all those expenditures are justified. Public expenditures in general are justified only when market failures or distribu-
tional concerns exist, and this is true for risk-reducing public expenditures, too. After briefly sketching out the foundations of this approach to the analysis of public expenditures, we turn therefore to an examination of potential failures in risk markets, and proceed to explore the implications for public policy in some special cases.

The framework for evaluating government spending on goods and services is based on the rationale for public intervention in the economy, which in turn is derived from the fundamental theorems of wel-
Box 1: Why Do Rich Countries Spend More on Reducing Risk Than Poor Countries?

At first glance, the fact that public spending on risk reduction rises with income seems counterintuitive, because a common assumption is that people’s aversion to risk declines with income, such that the risk premium (and therefore the benefits from government spending to reduce risk) would be higher in poorer countries. The countervailing effect is that the magnitude of the shocks to income is much greater in rich countries. Many of the risks that public programs mitigate are related to income. If someone earning $100,000 loses a job, the absolute value of the loss is considerably greater than if the initial income was $20,000. Can this feature explain the large variation in public spending on risk reduction across countries and over time? With constant relative risk aversion, if losses are strictly proportional to income, then so will be the premium, in which case this feature alone cannot explain the variation in public spending. If, however, the losses are more than proportional to income, the premium (as a percentage of income) rises quite dramatically with income (figure 1). If, for example, the level of income that one is left with after a typical shock to income rises with income, but only with an elasticity of 0.8, we observe that the risk premium rises from zero to nearly 18 percent of income at levels of around $6,000. That this gap of 18 percent also happens to be close to the difference in public spending on transfers between developed and developing countries suggests that such reasoning may be an explanation for the difference.

Figure 1. Risk Premium as a Share of Income — Illustrative Parameters

Premium (in percent of income)
fare economics. If the conditions of the first welfare theorem were to hold, there would be no need for a government, because the unfettered market would reach a Pareto-efficient allocation. If there is a concern for equity, the second welfare theorem shows how, with a suitable redistribution of initial endowments, the desired Pareto-efficient allocation can be achieved by the private market. Hence, the rationale for public intervention must be associated with one or more of the conditions of the welfare theorems not being met. The most common ones are the existence of externalities, public goods, non-competitive markets, and various elements of imperfect information (often collectively referred to as “market failure”) on the one hand, and the inability to redistribute endowments to achieve equity objectives on the other. This simple point alone can be a powerful tool in scrutinizing public expenditures. The largest item in the Indian government’s agriculture budget, for example, is a fertilizer subsidy. Forty years ago, the subsidy was justified on the grounds that it was a new technology so unknown and inherently risky that individual farmers might not have an incentive to adopt it. Today, the market-failure rationale for the subsidy has all but disappeared (Pradhan and Pillai-Essex 1994).

The existence of a market failure only indicates a rationale for government intervention; it does not necessarily imply a need for public expenditure. The textbook case of an externality is the polluting factory, which emits toxic chemicals into a stream and inflicts a cost to downstream users of that stream. Although the competitive equilibrium in this case will not be Pareto optimal, the solution is typically to levy a pollution tax on the factory, rather than to initiate a public expenditure program.

Finally, for cases where there is some market failure, and where public expenditure is the most appropriate instrument, there remains the issue of how important the market failure is. Because governments have limited resources, we need to have a sense of the quantitative benefits and costs of these different expenditure programs to allocate public resources rationally. The quantitative assessment is made up of two components: (a) the difference between social and private benefits (in the price dimension), and (b) the net addition of service (in the quantity dimension). In evaluating these benefits and costs, we need to keep in mind that most cases of market failure are ones where a private market exists, but does not provide the socially optimal level of output. For example, many believe that education carries with it a positive externality, insofar as society attaches a value to having a literate
and numerate population, beyond the benefit increasing the wage the individual receives from education. Yet education is a private good, so the benefit of public provision of education (assuming that provision is the best instrument), then, is only the external effect of the additional educational attainment over and above what the private sector would have achieved in the absence of public intervention. Because education and many other public services are nontraded goods, the calculation of net benefits should take into account the extent to which public provision crowds out the private sector. If the government was providing education, but the private sector could still provide more (with perfectly elastic supply), the public education would completely crowd out private education, which would make the net benefit of this public program zero (Devarajan, Squire, and Suthiwart-Narueput 1997; Hammer 1997).

Although quantitative analyses of the benefits of public expenditure programs (in the welfare-theoretical sense developed here) are hard to come by, some suggestive evidence exists. Hammer, Nabi, and Cercone (1995) evaluate the impact on infant mortality of the Malaysian government's expenditures on public medical personnel and immunization. They find that government spending on doctors at the margin has no significant effect on infant mortality, whereas spending on services such as immunization, which have clear external effects, is highly significant. Spending on public medical personnel was simply crowding out private medical personnel, which left the net effect not significantly different from zero. Similar results for health care have been found by Alderman and Lavy (1996), for income transfers by Cox and Jimenez (1995), and for secondary education by Jimenez and Lockheed (1995).

Finally, the theory of the second best is often invoked in justifying and evaluating public expenditures. If there is a distortion in the economy, government intervention, and possibly government expenditure in some other (undistorted) market, may be warranted because it can affect welfare in the distorted market. For example, if a failure in the

---

2. Some claim education to be a “public good” on these grounds, but this does not accord with standard definitions. A public good is nonexcludable, meaning that you cannot charge for it even in principle, because nonpayers cannot be excluded from benefiting. A public good is also nonrivalrous, meaning that one person’s use of the good does not reduce the amount available for others. Although underutilized classrooms may fall into this category, usually teachers’ time and classroom seats are limited.
credit market prevents young people from obtaining student loans, public support to education may be justified. Note, however, that two conditions have to be met. First, the market in which intervention is being considered must be linked to a truly distorted market. Second, removing the original distortion must be more difficult or costly than this “second-best” approach. As to the first, the mere fact that government policies change conditions in related markets is not in itself a justification. Such effects could be in the form of a “pecuniary externality” where the impact of a policy is solely through the workings of competitive markets. There may be distributional consequences; for example, universal primary education supported by government could well raise the wages of teachers (or all people who are potential teachers), but if the supply of such factors is competitive, the existence of such effects poses no difficulties or particular issues for policy analysis. If markets are incomplete, even pecuniary externalities with competitive markets could have effects on efficiency (Greenwald and Stiglitz 1986, de Meza and Gould 1992).

Of course, when serious market failures are associated with these affected activities, the activities need to be taken into account. For example, a project, such as a road that indirectly increases steel output, would not have to take into account the changes in steel or of the coal or labor used in its production if they were all competitive markets. If, however, steel production caused pollution, the value of the reduction of pollution would be a further cost of the project that would have to be valued. This example also illustrates the second condition. Appropriate pollution control policies applied directly to the steel industry would obviate the need for the road project to worry about steel production. Only when such policies are unavailable—for example, for technical or political reasons—is this interconnectedness important (Sen 1972).

As the discussion on evaluating public expenditures makes clear, the fact that governments affect the risk profile, and hence welfare, of private agents is not sufficient justification for the existence of a public expenditure program to mitigate risk. Many markets associated with the bearing of risk, however, are characterized by market failures. In some cases, the markets may simply not exist. In others, private agents will supply a suboptimal level of risk reduction. Consequently, there is a role for government, both in attempting to correct these market failures directly, and—where that may not be feasible—in addressing risk-market failures through intervention in other markets.
Applying the Framework

Several important failures in risk and risk-related markets can be discussed with reference to the framework outlined in the previous section. The most common one in the literature is the frequent absence of insurance markets. The simple model of individual decisionmaking under risk specified above implies that there will be a demand for insurance—a willingness to pay the quantity $V$ above and beyond the actual expected cost of assuring wealth $W$. A firm that can pool all risks and ensure a payment to all customers to make their income $W - V$ can collect $V$ as profit. Competition should drive this profit down to the actual cost of providing the insurance itself, so that people will end up paying this cost, which is less than $V$, and gaining consumer surplus from the difference.

Many reasons explain why such a market will fail to emerge or will supply insurance in far less than optimal amounts. They fall under the general categories of adverse selection and moral hazard (Rothschild and Stiglitz 1976). Adverse selection occurs when there is asymmetric information between buyers and sellers of insurance. For example, an individual may know if he is a bad health risk, but an insurance company may not be able to detect this. Consequently, insurance companies offer health insurance reflecting the average risk of the population. At this price, however, assuming no quantity constraints, only those with a higher-than-average risk will purchase insurance. As a result, the lower-than-average risk population leaves the market, saddling the insurance company with a riskier population than they expected. If the company raises its premium, even more people leave the market, and eventually the market dries up.

Rothschild and Stiglitz (1976) show that, for a market to exist, insurance companies will have to offer price-quantity packages, thereby limiting the amount of insurance at a given price that an individual can buy. This kind of quantity rationing, however, as they show, can be Pareto inefficient.

Moral hazard is a situation where an individual, having purchased insurance, may have an incentive to undertake suboptimal levels of risk-reducing activity. For instance, purchasers of theft insurance may not lock their doors, even though society would be better off if they did. Perhaps the most graphic example is that of arson—when people burn down their own houses to collect on fire insurance.

The existence of moral hazard and adverse selection can prevent the insurance market from appearing at all. The complete absence of the
market imposes costs on people of the full amount of $V$, although this fact alone does not justify government intervention—let alone government expenditure—in risk markets. The first question to ask is whether, by intervening, the government can do better.

That someone, such as an insurance company, has the ability to pool or otherwise bear the risks that at least some individuals would prefer not to is a basic insight into the value that government can bring to the market. Efficient markets will result in those who either do not care as much (are less risk averse) or who have such risk-reducing options as diversification opportunities available to them actually having more risk shifted onto them from the more risk-averse, less protected consumers. In exchange for absorbing more risk, they are paid some of the risk premium, $V$, that the risk-averse individuals gain in the bargain. Government may be in the position to bear this risk itself better than some individuals. The government would then do the pooling. It is not clear, however, how publicly provided insurance gets around the problem of moral hazard. Mandatory, publicly provided insurance can get around the problem of adverse selection. Alternatively, the government may choose to regulate insurance markets to correct some of the existing failures. In all these cases, the main thrust of the policy would be to shift risks, and the value of doing so would be $V$ per affected person.

Explicit insurance is not the only way that people deal with exposure to risk. In many circumstances, people have opportunities to reduce their own exposure through diversification of various sorts. The classic example forms the basis of the contemporary theory of finance. The value of any security is not simply its expected return, but is related to the degree to which it is correlated with the rest of the market and therefore serves to reduce the risk of holding portfolios. In our notation, a premium is to be paid to any one asset $e_i$ if it can reduce the variation of the sum of all returns—the investor's net variance.

People have other means to help deal with risk. In traditional societies, the extended family provides an insurance policy of sorts. Hard times may result in intrafamily transfers with either explicit or implicit repayment arrangements; that is, they may be gifts or loans. The credit market itself may serve as an insurance mechanism if people use it to borrow or draw down savings in bad years and pay back or build up savings in good. However, as will be seen shortly, credit markets themselves are often faulty for reasons similar to insurance markets, especially for consumption loans. The degree to which they are faulty will determine the value of policies that reduce the risk that one would borrow against.
In sum, the valuation of mitigating risk needs to be in comparison to the net exposure $\Sigma e_i$ after diversification or other protective activities are undertaken. Savings on any real costs associated with the protection, however, would be another benefit from the program. For example, agricultural households are sometimes noted to have more livestock or other, relatively liquid, productive assets than would be justified by considerations of profitability alone. The increase in farm profits from shedding such unprofitable activities, caused by having to handle less risk or having more efficient means of handling those risks, would be a benefit of an insurance program for, say, crops or health, or even unemployment. Other costs associated with protection include delayed schooling of children (Alderman and Paxson 1992), inefficient (in terms of expected income) crop mixes (Morduch 1994), mobility-reducing insurance arrangements (Bannerjee and Newman 1993), and nepotism in labor markets (Hoff and Sen 2001).

The actual calculation of certainty equivalent incomes, or the risk premium that could be obtained from people, requires specifying an explicit functional form for utility. This introduces a highly subjective element into the calculation because this is not a directly observable function. Further, there is no reason to believe it is common across people, nor even that the degree of risk aversion on the margin is equal, unless markets are working so well as to allow the equalization of marginal risk across people. If such markets did exist, however, there would be no particular justification of government intervention at all. The most careful calculation would try to approximate the willingness to pay for a particular degree of risk reduction for different types of people and add up across types (differing by income, risk aversion, and degree of wealth at risk).

Finally, in addition to providing insurance, governments use a variety of other instruments to address problems of risk. For instance, governments may attempt to mitigate the risk of price fluctuations facing farmers by agreeing to buy farm output at a fixed price, even when the world price is varying. In what follows, therefore, we examine two forms of public expenditures associated with risk reduction: (a) public provision of insurance and (b) other public expenditures that alter the risk profile facing individuals.

**Government Provision of Insurance**

Government policies can affect various different components that go into the calculation of the risk premium. Sometimes governments
attempt to provide insurance directly when the market does not. Two common areas where this occurs are in health and crop insurance.

**Health Insurance.** Although direct provision of services is more common in the developing world, many countries have instituted explicit health insurance as a means to help people deal with the financial consequences of medical care. The issue of health insurance is a complicated one to be sure—witness the recent debates in the United States and most other Organisation for Economic Co-operation and Development (OECD) countries. Here we only want to highlight the issue of valuation of the benefits of health insurance. From the perspective of correcting market failures, the benefit that the public can obtain over and above the laissez-faire equilibrium can be substantial. As mentioned, insurance markets for medical services are likely to be seriously distorted. In the early part of this century, the insurance industry in the United States considered medical care an uninsurable service because of the severe problems of adverse selection in voluntary markets and in the potential for abuse in terms of moral hazard (Arrow 1985). In the developing world, this situation still holds with very little private insurance existing even where medical care itself is largely private (Lewis and Chollet 1997).

To a large extent, evaluations of health insurance have focused on the benefits of medical services rather than on the benefits of insurance per se. By ignoring risk-reducing aspects, many discussions of health insurance and the relative merits of services to be covered by public schemes have been seriously flawed. The benefits of publicly provided health insurance should be the willingness to pay for insurance services that are not available because of the market failure reasons stated above. As a result, the value of public coverage depends at least as much on the probability of illness and the size of the expenses avoided by the policy as on the medical benefits of the treatments covered.

For example, if there is no insurance, what happens when a person falls ill with a condition that is treatable? The person could either choose to take the treatment or decide that it is too expensive and suffer with the condition. If he chooses to take the treatment, the value of public coverage of that condition is no longer related to the medical value of the treatment because the person is treated with or without public support. The value that public policy brings to this case is purely financial and is the willingness to pay, ex ante, for insurance against that disease condition. If the standard (constant relative risk aversion) utility function is used to analyze this situation, the value of
insurance will be: \( V = Y - U^{-1}(pU(Y - C) + (1 - p)U(Y)) \) where \( Y \) is income, \( p \) is the probability of illness, and \( C \) is the cost of the treatment. Note that health effects of the treatment do not appear in the valuation. This value must be higher than the administrative cost associated with processing the insurance. Otherwise there is no gain to be had from insuring the service at all, and it would be better to have people pay out of pocket when they need it.

If the person would not purchase the treatment out of pocket because it was too expensive, we might still ask if the person would have purchased actuarially fair insurance for the treatment if it had been available. The answer to this question is no longer independent of the health benefits that the treatment provides. A person would be indifferent between buying insurance and not buying it if the following equality holds:

\[
\begin{align*}
    pU(H_1, Y - pC) + (1 - p)U(H_0, Y - pC) &= pU(H_2, Y) + (1 - p)U(H_0, Y) \\
    (1)
\end{align*}
\]

where \( H_0 \) is health status when not sick at all, \( H_1 \) is health status after treatment when sick, and \( H_2 \) is health status when sick and left untreated. The left-hand side is expected utility if a person is insured and getting treatment that improves his or her health status from \( H_2 \) to \( H_1 \), and the right-hand side is the expected utility of refusing to insure and taking the risk of suffering with health status \( H_2 \) if the person gets ill. All this is contingent on \( U(H_1, Y - C) < U(H_2, Y) \) because we have assumed that this treatment would not have been purchased out of pocket. The value of providing insurance in this case is the difference between the left- and right-hand sides of the above inequality.

Figure 2 shows the above relations graphed in the space of cost of treatment and health benefits of treatment. For the case of treatments that would be purchased out of pocket, curve OA is drawn with a health status of \( H_1 \) when illness occurs because it is assumed that treatment will be taken. The line segment DE is the combination of \( H_1 \) and \( C \) that solves equation (1). The vertical line at \( C = B \) is the level of treatment costs such that the administrative costs of insurance exceed the value of the insurance itself. The figure is thus divided (by solid lines) into four areas. In area I, treatment would be paid for out of pocket, but people would prefer to insure against it. In area II, people would pay for treatment out of pocket, but would not bother to buy insurance because such treatments are too cheap to cover the administrative costs of insurance (aspirin for headaches is a good example). In area III, people would neither buy the treatment out of
pocket nor demand actuarially fair insurance for it. In area IV, people would not buy the treatment out of pocket, but would pay for insurance for it. This represents a catastrophic loss for direct purchase, but is rare enough to have a sufficiently low expected cost to be worth the insurance value.

For comparison, the ray OC has been superimposed on the graph. These points share a common “cost-effectiveness ratio,” or a constant health benefit per dollar spent on a medical treatment. This has been proposed as a criterion for public intervention in health care (Jamison and others 1993) and as a criterion for inclusion in an insurance package, public or private (Gold and others 1996). As illustrated here, treatments sharing a common cost-effectiveness ratio fall into all four areas. Thus, cost-effectiveness ratios provide no information whatsoever concerning the value of provision when insurance markets are absent—the market failure that justifies public coverage of the private benefits of health care. Further, within areas I and IV, where insurance

---

3. The actuarially fair costs of insurance should, strictly speaking, have included the administrative costs, $A$, and be equal to $Y - pC - A$.

4. The external benefits would be evaluated separately.
would be demanded if available, the loss imposed by the absence of insurance rises with the cost of treatment. The cost effectiveness ratio, on the other hand, worsens with higher costs and thus moves in the opposite direction from the true valuation of public provision.

**Crop Insurance.** Crop insurance is another area in which governments have sometimes provided a direct insurance instrument that private insurers would not. The reasons why such insurance policies would not be written by the private sector are again the potentially large problems of adverse selection and moral hazard. Moral hazard is a particular problem because there are many actions that a farmer could take that are not easily (that is, without very high cost) observable to the insurer, and that determine crop output along with truly random events, such as rainfall and other farm-specific risks. Effort and purchased inputs are two examples. A cotton insurance program in India ran into difficulty in part because some farmers would stop applying inputs (late in the production cycle) when it appeared that output would not be much higher than insured-for levels. Detailed characteristics that determine land quality would lead to adverse selection by those who know their land to be poor. There might also be an interaction of the two problems if those who knew themselves to be the type who would exploit the moral hazard problem would also disproportionately sign up for the program.

For all the reasons that private markets would not support crop insurance markets, the public sector has had a similarly bleak history of providing the service. Hazell, Pomereda, and Valdes (1986) cite numerous problems that have plagued public crop insurance programs. Often, a goal of such programs is to be financially sustainable. The reasoning is that the service provided is genuinely valuable and can be covered with cesses on agricultural output. That these programs typically cannot be sustained without continual subsidies illustrates a problem that should be balanced with the identification of a market failure in the private sector. In many cases of seeming failure in risk and information-related markets, there may be no advantage that governments can bring to the problem to improve matters. Although the maximum potential of providing insurance can be calculated from the reduction of risks that people might like to avoid, it is not always the case that governments can improve upon the allocation of the market. If there is nothing that the government could know that a private insurer could not, the free market allocation may be "constrained Pareto-optimal."
That a market is constrained-Pareto-optimal means that the government cannot do any better than the private sector by intervention in the market with the information failure. As a result of the theory of the second best, however, it is still possible that there are other instruments directed at complementary or substitute markets that can improve welfare, a topic to which we now turn (Greenwald and Stiglitz 1986).

**Other Public Expenditures**

Other than providing insurance directly, governments intervene in less direct ways. Some policies are intended to reduce risk by changing particular elements of a risky component of income. In this section, we examine the impact of price stabilization schemes, transfer programs, government guarantees, and public expenditures on investments and consumption.

**Price Stabilization.** One common form of this is through commodity price stabilization schemes (Newbery and Stiglitz 1981). Countries often try to protect producers or consumers from wide fluctuations in the prices of basic commodities. Although they are often simply a transfer program in disguise, these stabilization schemes are publicly advocated as a way of reducing risk. The value of the stabilization plan to a producer depends on how the price variations translate into income. In turn, this will depend on the degree of diversification of farm production (monocultural areas being at greatest exposure to price risk) or of farm family income (farm families often have members in nonagricultural activities, sometimes as migrants to cities, as a hedge against low farm incomes), access to credit, and the nature of the market for farm output. As to the last consideration, if the commodity whose price is being stabilized is not traded internationally, as may be the case for basic staples, prices would ordinarily rise in times of low production and fall in times of good production. For a wide range of demand elasticities, this market mechanism provides substantial smoothing of farm revenues. Indeed, it is possible (Newbery and Stiglitz 1981) for stabilization of prices to destabilize incomes by removing the negative correlation of price and sales. To the extent that price stabilization leads to income stabilization, the value of the scheme can be approximated by the formula: 
\[ V = -\alpha \times \Delta \sigma^2 / 2 \]
where \( \alpha \) is the coefficient of relative risk aversion and \( \Delta \sigma^2 \) is the change in the variance of income.
Transfer Programs. Another type of government policy that has significant implications for risk reductions is transfer programs for income support. Usually they are introduced for reasons completely different from risk reduction per se with the exception of unemployment insurance. Unemployment insurance is one area where it is clear that private markets are likely to be limited because of the extreme problem of moral hazard and adverse selection inherent in a voluntary program. There are many reasons for an individual to know his or her own probability of getting fired better than an insurance company would. Indeed, when combined with the moral hazard problem—people may choose to be unemployed if insured at high rates—people are certain to know more about their own inclinations to abuse the policy in this way than would the company. So, except for unusual, job-specific assets that might be covered by a specialty insurance contract (such as Lloyds of London's insuring a pianist against broken fingers), unemployment is not a good candidate for private insurance. Its benefit, though, may be estimated by combining the concerns for risk using the method above with models of incentive effects of labor supply.

Again it is important to evaluate the benefits of programs relative to private adjustments to the problem. Although private markets for unemployment insurance are likely to have serious problems, many arrangements in labor markets are clearly motivated by concerns over risk-sharing. Lifetime employment guarantees (explicit or implicit), as well as different quantity and wage adjustments as appear in macroeconomic models, are examples. The calculation of benefits is unlikely to be particularly persuasive in advocating (or contesting) the introduction of unemployment insurance because this is particularly a politically charged area. In the design of different elements of the program, however, length of time covered, job search requirements, and so forth may have quite different risk-reduction characteristics and may be evaluated one against the other differently.

Other transfer programs have risk-reducing characteristics even if that is not their main justification. In the framework above, safety net provisions, progressive income taxes, and other redistributive policies can induce a negative correlation of government transfers with random shocks to income. We might think of the policy as one that makes the net-of-tax-and-transfer income a function of the random shocks that make up income as in \( W(\sum x_i) \) where \( W'<0 \). The transfer program will have its main benefit in the promotion of equity and its main costs associated with incentive effects for work. However, there could be a
further benefit on efficiency grounds because of the risk-reducing nature of the policy. As an example: a linear income tax system (with a lump sum transfer component) would make after-tax income a linear function of pretax income with slope $b(1 - t) < 1$, where $t$ is the tax rate. The variance of income for any one individual is reduced to $b^2$ times the original variance. From the above formula, the value of this reduction in variance is $\alpha \times (1 - b^2)\sigma^2/2$. Depending on the distributions of risk aversion and the risky element of income across people, $\sigma$, this can be a substantial sum. For a tax rate of 30 percent, a typical level of $\alpha$ of 2 and the risky component of income having a variance of 10 percent of income, this comes to about 5 percent of income as a welfare gain. As this example shows, even nonprogressive (linear) income taxes can have this welfare-increasing property. The main distinction is between lump sum taxes and taxes, such as agricultural output taxes, that vary with income (Hoff 1991).

Public Investment. Governments often justify public investment—even in private goods, such as steel plants or textile factories—on the grounds that the returns on these investments are risky and governments are better able to take such risks. We now evaluate this argument in light of the framework developed above.

The notion that governments are better able to take risks stems from the proposition of Arrow and Lind (1970). They argued that, because the government is able to spread the risk of a public project across a large number of individuals (namely, taxpayers), the amount each individual bears is minuscule. The government then should behave as if it were risk-neutral, even if the individuals in society were risk-averse. The corollary is that the government should undertake risky projects (with a positive expected value) that private firms, being risk-averse, would not.

For several reasons, the Arrow-Lind theorem may not apply to developing countries. First, the risks associated with some public projects may not be easily spread across the population. Large irrigation projects, such as the Aswan Dam in Egypt or the Mahaweli Scheme in Sri Lanka, could have an impact on the fortunes of the whole country. To the extent that public projects are producing public goods, a case can be made that by definition the returns on the project are not independent of other risks faced by individuals. If the project's output is nonexcludable and nonrival, then the risks associated with the project are also nonexcludable and nonrival. When a dam bursts, the costs cannot be easily proportioned across the affected population (Fisher 1973).
Second, although governments in richer countries with wide tax bases are in a position to diversify their risk, smaller countries with narrower tax bases may not be. Meanwhile, the private sector may be represented by a multinational corporation with access to many more risk-diversifying instruments. This is especially true if the investment is in a private good, where the justification for government involvement is weak to begin with.

Third, the risk associated with a project may not be some exogenous factor, such as an earthquake or flood, but political risk. Especially for large infrastructure projects with high up-front investment costs, the major risk facing a private investor is that the government may nationalize the enterprise, or impose foreign exchange restrictions on multinationals’ expatriating their profits. In this case, the most appropriate response of government would be to provide a guarantee against these events (because it is the government that controls them). Of course, a government’s unilateral provision of a guarantee may not be credible, so a neutral third party, such as the World Bank, could provide the guarantee and monitor it. The increasingly widespread use of these guarantee instruments in developing countries is testimony to the importance of political risk in infrastructure investment decisions. This is an example of governments’ addressing the appropriate market failure, namely, that of political risk insurance. Technological progress has turned many kinds of infrastructure into private goods, so there is no longer a compelling case for public provision of intercity roads, power, and telecommunication. Yet there remains the problem of political risk, which hinders private investment in this sector. The solution is not for governments to undertake the infrastructure investment, but to address the risk market failure by entering into contracts with guarantees against political risk.

Public Consumption. As mentioned above, in poor countries, public health insurance schemes are less common than direct public provision of health care. The reasons for this pattern are complex. They are likely to include the relative monitoring and regulation requirements of an arm’s length insurance program versus management of services. In any case, whether insurance is publicly provided, the fear of catastrophic loss is a fact of life everywhere, and the ability to pay out of pocket for expensive medical treatment may be quite limited even when actuarially fair insurance would be affordable.

Much of the health care budgets of developing countries are devoted to hospital care. This has generally been criticized by donor
countries as being “inefficient” and inequitable in comparison to providing lower-level primary health care. However, given that people may be willing (and therefore able) to pay for relatively cheap services, but unable to pay for the catastrophic financial burden of hospital services, there is likely to be an efficiency argument for subsidy to expensive medical care. Here, the benefit of public expenditure in health should take into account the improvement in welfare from risk reduction, as well as any direct benefit of the services. This will be important for the higher-end services.

This observation does set up a potential conflict between efficiency and equity objectives. It is true, as the donor community points out, that hospital services tend to be disproportionately used by higher-income people and that the political power of urban elites has distorted investment priorities toward hospital services as a result. Although some reconciliation of the two goals is possible (such as fees for outpatient service for everyone and strict controls on referrals for inpatient services, which would limit free access to those with medical need rather than social influence), a residual conflict will still exist between the subsidy for high-end services and using health expenditures for redistribution goals.

As an application of this approach, we consider a recent project in Indonesia (see figure 3). Among other things, the project provides emergency referral and transportation for pregnant women with complications during delivery. The cost of such services is quite high, but the probability of surviving the delivery is greatly enhanced.

Because several of the variables in the calculation are either subjective or difficult to determine, a sensitivity analysis was performed on the calculation of risk premiums with regard to probability of need, risk aversion, and cost of the service. A complete evaluation would depend on the joint distribution of these parameters, and would require the summation of each family’s willingness to pay to get at the overall insurance value. Nevertheless, the calculation for a single household with typical parameter values is revealing. It shows that for services that could potentially cost a substantial fraction of the family’s wealth, the value of risk reduction per se can be in the neighborhood of 40 percent of the expected cost of the service. Evaluations that neglect this effect can, therefore, be far off.

Government Guarantees. As mentioned above, credit markets are often operating suboptimally because of information problems, which leaves open the possibility of government improvement. One kind of invest-
ment that is particularly prone to such problems is in human capital, or "education" to the layman. Two features of the market for loans to finance education make it particularly subject to distortion from uncertainty. First, as in consumption loans to smooth fluctuations in earnings, loans for education are generally unsecured. In the absence of slavery, lenders usually have no feasible way to impose collateral requirements—people are particularly mobile. This makes the market for loans for education (especially higher education because it is more expensive both in fees and in financing consumption over the period) especially risky from the lender's perspective, thus reducing loans for that purpose. On the borrower's side, there is another consideration. For the same reason—that you cannot put yourself up as collateral—you also cannot sell shares in yourself. As a result, the decision to enter a particular, specialized field represents a decision to "plunge" into a market. Diversifying your human capital investments is not possible as it is for financial investments. This can have two possible consequences.
First, it may simply result in people's bearing considerably more risk than they would prefer, which would lead to a direct loss in welfare. Second, to the extent that this effect is stronger in some fields—those with strong specialization and those that are subject to large swings in net demand (engineering, perhaps, because it is often mentioned as a "cobweb" market)—this may lead to people shying away from these subjects toward more generalist professions (law, perhaps). In developed countries, especially, a large part of the assets that people hold are in the form of education, skills, and experience.\(^5\)

Policy implications for these market failures are not always easy to define. The absence of a functioning credit market holds open the possibility of student loan programs as a means of substituting for the market. It does not argue particularly well for a tuition subsidy as is common for developing countries, though it may provide a third or fourth best solution. Even a loan program, however, does not solve the diversification problem. For this, there may be some benefit in a risk-sharing arrangement with government—perhaps a surcharge on income taxes for those who attend university in lieu of standard repayment schedules.

Concluding Remarks

In this paper, we have argued that there is a case for incorporating the value of risk reduction when evaluating government expenditure, provided that the expenditure meets the standard welfare-economics criteria for government intervention in the economy. Using a series of examples, we showed areas where government spending on risk reduction could improve welfare, either by alleviating a risk market failure, or by reducing uncertainty in otherwise distorted markets. We also gave a few illustrative calculations of the risk reduction benefits of public expenditures and pointed to cases where their neglect could lead to serious underestimation.

\(^5\) Interestingly, another large part is in the form of real estate, a house being one's largest asset. This, too, is difficult if not impossible to diversify and is subject to very variable returns which depend on regional markets. To the extent that returns to skills are also associated with a dominant regional market, these two large parts of a portfolio are covariate; witness the fate of steel or automobile workers in the "rust belt" of the United States who lost both jobs and equity in their homes as industry declined. See also Shiller (1993).
Lest our points be interpreted as a blanket call for increased government spending in these areas, we mention two important caveats. First, the design of optimal policies to correct information-based distortions has to be approached with care—it may stretch the institutional capabilities of the government. For instance, the sophisticated bidding scheme at the U.S. Federal Communications Commission, designed by some eminent game theorists to maximize welfare, was still seen to be prone to collusion among bidders. Similarly, our off-the-cuff proposal to tax university graduates still has the effect of subsidizing the consumption value of education—for which there is no particular public interest.

The second problem raises a deep, fundamental issue of incentives. Although there is demand for reducing exposed risk, especially for the poor, getting rid of all risk leads to no reason to do anything at all; this is the essence of the moral hazard problem in the first place. Hence, when taken to the limit, a fundamental tradeoff may exist between avoiding risk and fostering economic efficiency.

References


When society expands access to public services, such as health care, primary education, and clean water, how much do poor people really benefit?

In the wrenching transition to a market economy in the former Soviet Union, is it better to have a hundred friends or a hundred rubles?

Do patients end up paying for their ignorance of medical matters—and if so, how much do they pay?

Does social assistance displace private sources of health-care financing?

What is the best way to reduce water pollution in rapidly growing India—by command and control, market mechanisms, or negotiation?

This volume answers these questions and more as it investigates some of the key dimensions of human development and growth. It pulls together eight of the most noteworthy papers from the second World Bank Economists' Forum—a gathering of economists from across the World Bank—that explored the topic of "Enabling Investment, Empowering the Poor."