THE WORLD BANK ECONOMIC REVIEW

Volume 13 September 1999 Number 3

A SYMPOSIUM ON EDUCATION REFORMS

ISSN 0258-6770

THE WORLD BANK ECONOMIC REVIEW

The World Bank Economic Review is a professional journal for the dissemination of World Bank–sponsored research that informs policy analyses and choices. It is directed to an international readership among economists and social scientists in government, business, and international agencies, as well as in universities and development research institutions. The *Review* emphasizes policy relevance and operational aspects of economics, rather than primarily theoretical and methodological issues. It is intended for readers familiar with economic theory and analysis but not necessarily proficient in advanced mathematical or econometric techniques. Articles will illustrate how professional research can shed light on policy choices. Inconsistency with Bank policy will not be grounds for rejection of an article.

Articles will be drawn primarily from work conducted by World Bank staff and consultants. Before being accepted for publication by the Editorial Board, all articles are reviewed by two referees who are not mem-bers of the Bank's staff and one World Bank staff member; articles must also be recommended by at least one external member of the Editorial Board.

The *Review* may on occasion publish articles on specified topics by non-Bank contributors. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages in length to the Editor.

The views and interpretations expressed in this journal are those of the authors and do not necessarily represent the views and policies of the World Bank or of its Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of the data included in this publication and accepts no responsibility whatsoever for any consequences of their use. When maps are used, the boundaries, denominations, and other information do not imply on the part of the World Bank Group any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

Comments or brief notes responding to *Review* articles are welcome and will be considered for publication to the extent that space permits. Please direct all editorial correspondence to the Editor, *The World Bank Economic Review,* The World Bank, Washington, D.C. 20433, U.S.A.

The World Bank Economic Review is published three times a year (January, May, and September) by the World Bank. Single copies may be purchased at $12.95. Subscription rates are as follows:

	Individuals	*Institutions*
1-year subscription	US$30	US$50
2-year subscription	US$55	US$95
3-year subscription	US$70	US$130

Orders should be sent to: World Bank Publications, Box 7247–7956, Philadelphia, PA 19170–7956, U.S.A. Subscriptions are available without charge to readers with mailing addresses in developing countries and in socialist economies in transition. Written request is required every two years to renew such subscriptions.

Manufactured in the United States of America
ISBN 0-8213-4362-9; ISSN 0258-6770

This journal is indexed regularly in *Current Contents/Social & Behavioral Sciences, Index to International Statistics, Journal of Economic Literature, Public Affairs Information Service, and Social Sciences Citation Index*®. It is available in microform through University Microfilms, Inc., 300 North Zeeb Road, Ann Arbor, MI 48106, U.S.A.

THE WORLD BANK ECONOMIC REVIEW

Volume 13 September 1999 Number 3

A SYMPOSIUM ON EDUCATION REFORMS

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 409-13

Evaluating Education Reforms: Four Cases in Developing Countries

Elizabeth M. King and Peter F. Orazem

This symposium features four studies of education reforms and their impact on enrollment and learning. Three are part of a research project funded by the World Bank's Development Research Group and its Research Support Budget to evaluate innovations in the education systems of selected developing countries. Two of the articles focus on Latin America, where decentralization reforms have been in place since the early 1990s. El Salvador has implemented a program that involves community education councils in the operation of public schools, and Colombia ran a voucher program that subsidized poor students, enabling them to attend private secondary schools. The third article analyzes a government subsidy program in Pakistan that encourages communities to establish nongovernmental schools that enroll girls. The symposium's fourth study evaluates a pilot project in the Philippines that uses different school inputs to improve student enrollment and performance in primary school.

Together, these four studies make an important point for development: by evaluating ongoing policies and programs, policymakers can learn what works and what does not work under specific circumstances. Learning requires some planning on the part of program managers and policymakers. It involves systematically collecting and analyzing information on communities, schools, and students, so as to identify changes in outcomes that can be attributed to the program.

In most developing countries the central government provides basic education. The frequently cited reasons for central control are that the technical expertise to establish and enforce educational standards, to plan and manage budgets, and to hire and train teachers is not broadly available. Further, local capacity to finance and operate schools is uneven across communities. Central funding, with centralized management and centralized norms, is believed to ensure better quality education and more equal access to education. Yet centralized management has its own blind spots: learning occurs behind classroom doors and away from the direct gaze of government officials. Consequently, at the risk of losing some of the advantages of centralized control, a growing number of countries have

Elizabeth M. King is with the Development Research Group at the World Bank, and Peter F. Orazem is with the Department of Economics at Iowa State University. Their e-mail addresses are eking@worldbank.org and pfo@iastate.edu.

been experimenting with ways to transfer responsibility and authority away from the center. Fiske (1996:*v*) describes this movement as a global phenomenon:

> Nations as large as India and as tiny as Burkina Faso are doing it. Decentralization has been fostered by democratic governments in Australia and Spain and by an autocratic military regime in Argentina. It takes forms ranging from elected school boards in Chicago to school clusters in Cambodia to vouchers in Chile.

One type of decentralization reform is school autonomy reform, which shifts management responsibility and resources directly to the school. The hope is that bringing decisionmaking power and accountability closer to those who teach and manage schools will make schools more efficient in allocating and using resources and more effective in instructing students. The hope is also that making those who teach and manage schools more directly accountable to students, parents, and communities will, in turn, establish local incentives that reduce the need for centralized control and supervision. The reform must change the relationships among the actors in the education system—government officials, school principals, teachers, parents, and even students—and must affect what teachers do in the classroom.

El Salvador's Community-Managed Schools Program (*Educación con Participación de la Comunidad,* EDUCO) has been expanding education in rural areas by enlisting and financing community management teams to operate schools. These teams are made up of parents, who are elected by the community. They are required to follow a centrally mandated curriculum and maintain a minimum student enrollment, but they have the power to hire and fire teachers and to equip and maintain the schools. In this issue Jimenez and Sawada examine the impact of this program on several educational outcomes. They find that, compared with traditionally managed schools, EDUCO schools have lower teacher and student absenteeism and comparable student achievement, holding the characteristics of students constant.

When publicly funded schools rely on nongovernmental management or when private schools receive public funding, the separation between public and private education becomes blurred. Increasingly, governments are relying on partnerships with the private sector to meet educational needs for which government resources alone are inadequate. Private schools already enroll a large number of students in many developing countries, often with the assistance of government subsidies. Governments have pushed private schools to increase their capacity in several ways—by paying for construction but relying on private groups to build and manage private schools (as in the Philippines), by subsidizing part of the construction and maintenance costs of private schools and assigning some public school teachers to teach in private schools (as in Indonesia), by financing a large proportion of the salaries of private school teachers (as in Bangladesh), by providing tax incentives to private education foundations (as in Colombia and the

Philippines), and by establishing a student voucher scheme (as in Chile and Colombia).

If the government wants to induce private schools to locate in areas that have few schools, directly funding communities, nongovernmental organizations, or education foundations to build schools that will be privately managed may be more effective than indirectly funding schools through student vouchers. How private schools and communities respond to these incentives depends on the elasticity of school supply in the target area. The more elastic is the private supply of educational services, the more attractive is it for the government to expand educational services by subsidizing private schools. To our knowledge, no one has estimated such elasticities. However, supply seems to be more elastic in urban areas than in rural areas, implying that a larger government subsidy is needed to induce private investment in rural areas than in urban areas. Even so, unless the elasticity is zero, private school subsidies represent a potential strategy for expanding school capacity.

In the city of Quetta in Balochistan, Pakistan, the government has initiated a pilot project that subsidizes the establishment of private girls' schools. Parents in ten neighborhoods were given the financial resources and technical assistance to contract a school operator to open a private school in their neighborhood. The amount of financial resources available was tied to the number of girls that the new school could attract. This financial assistance proved to be much less than the cost of opening and operating a government school in the area. In their analysis of this program Kim, Alderman, and Orazem find that all ten neighborhoods attracted bids from school operators and that the enrollment of both girls and boys rose significantly in response to the creation of the new private schools. The success of this program bodes well for expanding primary schooling by directly subsidizing private schools.

Several countries have been experimenting with voucher programs, which transfer resources directly to parents to help pay private school tuition. The pro-voucher literature argues that since school resources are tied to parental demand, which presumably responds to school quality, voucher programs force public and private schools to provide quality education efficiently. Schools that offer poor-quality education or use resources inefficiently will have to improve or face bankruptcy. Opponents of voucher programs dispute the efficiency and quality claims, suggesting that vouchers further segregate access to schools by wealth. To date, there is limited empirical evidence to guide the debate over which view is correct.

For developing countries the efficiency concerns may be secondary to the more basic problem that demand for schooling exceeds supply. When poor students are prevented from continuing in school because of overcrowded public schools or the lack of schools, the private sector may offer a way to expand school capacity. Vouchers can be targeted to the poor to prevent benefits from leaking to the wealthy and to induce private schools to locate or expand in areas that cater to the poor. In this issue King, Orazem, and Wohlgemuth examine how schools and municipalities responded to Colombia's national voucher system. They demon-

strate that municipalities were more likely to participate if excess demand for schooling was present but modest and if local private schools already had the capacity to absorb additional students. The schools that accepted voucher students tended to have fees and average student test scores that were in the lower range among private schools, although the design of the voucher program may have prevented the lowest-price and lowest-quality schools from participating. Most important, the targeted vouchers appear to have increased secondary school enrollment among poor children, allowing very little leakage to rich households or to schools catering to the wealthy.

Although many countries are reforming their education systems, few are systematically evaluating the impact of those reforms. This is the case in both industrial and developing countries. Yet impact evaluations are powerful tools for policymakers. They provide the information needed to terminate or improve ineffective policies or programs. And this information may be especially important for evaluating innovative reforms or programs that do not have well-charted histories.

Often the impact of a specific reform or program cannot be isolated from the impact of coincident changes in the policy or economic environment. We can measure the impact if we know how the participants (or treatment population) would have fared without the program. This is the counterfactual state, and the selection and observation of appropriate control populations to characterize this state are important features of evaluation strategies.

The Philippine study demonstrates the benefits of monitoring and evaluating education programs when the use of randomized control designs is feasible. In the Philippines the central government continues to play a key role in supplying primary education, as decentralization has not yet progressed to the extent that it has in the other countries. This article illustrates that even within the public sector, the government can benefit from routine monitoring and evaluation activities. Tan, Lane, and Lassibille examine four experimental interventions in selected low-income areas, using pre- and post-intervention data collected from program and control schools. The authors find that providing teachers with learning materials and encouraging parents to get more involved in the schooling of their children were more effective than a school feeding program in reducing dropout rates and increasing learning.

The four articles in this symposium illustrate that it is not always easy, however, to determine whether or not a particular policy or program has attained its objectives. The authors apply a variety of impact evaluation strategies depending on the nature of the reform, the stage at which the evaluation began, and the availability of appropriate data. Although developing an evaluation strategy at the inception of the reform or program makes evaluation a little easier, doing so is often not possible. In the absence of baseline surveys, evaluations may have to rely on more advanced statistical methods that may be more difficult for nonstatisticians to understand and undertake.

The ability to rank interventions is a valuable policy tool, as countries search for effective strategies to improve schooling outcomes within constrained education budgets. For the World Bank the practice of evaluating the most innovative components of its investment projects both improves the quality of its portfolio and enriches the knowledge base that it can share with its member countries.

References

Fiske, Edward. 1996. *Decentralization of Education: Politics and Consensus*. Washington D.C.: World Bank.

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 415-41

Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program

Emmanuel Jimenez and Yasuyuki Sawada

*This article examines how decentralizing educational responsibility to communities and schools affects student outcomes. It uses the example of El Salvador's Community-Managed Schools Program (*Educación con Participación de la Comunidad, EDUCO*), which was designed to expand rural education rapidly following El Salvador's civil war. Achievement on standardized tests and attendance are compared for students in* EDUCO *schools and students in traditional schools. The analysis controls for student characteristics, school and classroom inputs, and endogeneity, using the proportion of* EDUCO *schools and traditional schools in a municipality as identifying instrumental variables. The article finds that enhanced community and parental involvement in* EDUCO *schools has improved students' language skills and diminished student absences, which may have long-term effects on achievement.*

Central governments in developing countries usually play a major role in allocating educational resources. Even when authority is delegated to subnational levels, such as provinces or municipalities, individual school administrators and parents play only a limited part. This kind of centralized structure may work best for regulating and administering large systems uniformly, but it may also be ineffective and expensive when school needs differ widely across communities and when there are diseconomies of scale. Moreover, a centralized system can stifle the initiative of those who are most critical in affecting school outcomes—teachers, principals, and parents.

Despite the compelling case for school-based management, there is relatively little empirical evidence documenting its merits in developing countries.[1] The main reason is that these administrative arrangements have only recently been

1. Two exceptions are James, King, and Suryadi (1996) for Indonesia, and Jimenez and Paqueo (1996) for the Philippines. Both studies conclude that community-based involvement improves efficiency.

Emmanuel Jimenez is with the Development Research Group at the World Bank, and Yasuyuki Sawada is with the Department of Advanced Social and International Studies at the University of Tokyo. Their e-mail addresses are ejimenez2@worldbank.org and sawada@stanfordalumni.org. This project has been financially supported by the Development Research Group and the Research Support Budget (RPO 679-18 and 682-08) of the World Bank. The authors gratefully acknowledge the comments and support from the Evaluation Unit of the Ministry of Education of El Salvador, which collected the data. They also thank Marcel Fafchamps, Paul Glewwe, Elizabeth King, Takashi Kurosaki, Martin Ravallion, Laura Rawlings, Fernando Reimers, Diane Steele, the anonymous referees and editors, and participants in seminars at the World Bank, Stanford University, the University of the Philippines, International Child Development Centre (United Nation's Children's Fund, Florence), and the Institute of Developing Economies (Japan) for useful discussions and comments.

implemented (World Bank 1994, 1995, 1996). One celebrated example is El Salvador's Community-Managed Schools Program (*Educación con Participación de la Comunidad*, EDUCO). EDUCO is an innovative program for both preprimary and primary school designed to decentralize education by strengthening the direct involvement and participation of parents and community groups.

A prototype of today's EDUCO schools emerged in the 1980s, when public schools could not be extended to rural areas because of El Salvador's civil war. Some communities took the initiative to organize their own schools, which an association of households administered and supported financially. Although these early attempts were constrained by the low income base in rural areas, they demonstrated communities' strong inherent demand for education and desire to participate in the governance of their schools. In 1991 El Salvador's Ministry of Education (MINED), supported by aid agencies such as the World Bank, decided to use the prototype that communities themselves had developed as the basis of the EDUCO program.

Today, EDUCO schools are managed autonomously by community education associations (*asociaciones comunales para la educación*, ACEs), whose elected members are parents of the students. Reimers (1997) describes community associations as being composed of literate members of the community who are given basic training in school management. These associations meet periodically with teachers and also provide them with teaching materials. In EDUCO schools the ACEs are in charge of administration and management; MINED contracts them to deliver a given curriculum to an agreed number of students. The ACEs are then responsible for hiring (and firing) teachers, closely monitoring teachers' performance, and equipping and maintaining the schools. The partnership between MINED and the ACEs is expected to improve school administration and management in that the ACEs can better gauge local demand. In the future MINED intends to introduce community management to all traditional schools.

The EDUCO program was conceived as a way to expand educational access quickly to remote rural areas. Initial evidence indicates that it has accomplished this goal (El Salvador, MINED 1995; Reimers 1997). The question that remains is whether this expansion has come at the expense of learning. Professional administrators in the center are less involved in the day-to-day running of schools, which are now in the hands of local communities. But many of the parents in these communities have an inadequate education themselves. Thus it remains to be seen whether moving away from traditional, centralized programs and toward greater community and parental involvement also improves students' learning.

This article assesses the impact of EDUCO schools. We estimate school production functions using three measures of educational outcomes for third-grade students.[2] Two of the measures are standardized test scores in mathematics and

2. This study is part of a larger effort by the World Bank to distill the lessons of decentralized education (see World Bank 1996). Eventually, we want to determine whether all students in EDUCO schools achieve better educational outcomes at comparable costs relative to their counterparts in traditional public schools. This article has a more limited objective: it uses school production functions to compare three measures of educational outcomes among third-grade students only.

language. These may be good indicators of educational outcomes. However, they may also be relatively unresponsive in the short run to changes in school governance. Thus we also use an indicator that can be considered more of an intervening variable in determining student achievement but is likely to exhibit a short-run response: the number of school days that a student has missed.

As with all comparisons of educational achievement, the key is to quantify how much of the differential in academic achievement can be explained by differences in household background, the schools' quantitative inputs, and, most important, organizational factors attributable to intangible differences in the way that traditional and decentralized schools are run.[3] We also address parents' endogenous school choice by explicitly considering how the government selected which municipalities would be the first to have EDUCO schools.

I. The Conceptual and Empirical Framework

Educational outcomes are products of the complex interactions of agents who participate in the schooling process. Students' characteristics and motivation are key, but so are the actions of individual parents, parent groups (such as parent-teacher associations), teachers, and administrators from the school level up to the education ministry. In addition, agents not directly connected to the educational system can affect educational outcomes if they influence the environment in which students learn. For example, decisions about road infrastructure in a locality could afford access to certain types of schools, or the provision of electricity in a municipality could enable students to study at night.

The Basic Model

It would be impossible to model the structural relationships that capture the behavior of each relevant agent.[4] Instead, we postulate a simple reduced-form model of educational outcomes (Y). Most studies measure educational output by using students' achievement scores, attendance rates, repetition rates, decision to continue in school, or dropout rates. These variables are thought to capture prospects of future earnings in the labor market. In this article we focus on two components of Y: scores on standardized achievement tests (S) and days absent from school (A).

Studies of education production functions have had mixed success in explaining S.[5] Aside from measurement and estimation issues, outcomes may be determined by endogenous choices. For example, some of the explanatory policy variables that determine S, such as type of school, may be systematically related to unobservable characteristics, which themselves may not be random across observations. This could lead to bias. As explained below, we attempt to correct for

3. See Levin (1997) for a good review of these intangibles.

4. McMillan (1999) presents an interesting model of the interaction of parental and school preferences in determining educational outcomes.

5. See Hanushek (1995) for a review.

this problem by modeling and estimating the choice of school type and then using that estimate in the production function to control for participation. It is often difficult to identify such models. But we are able to use the participation rule that the Salvadoran authorities used in choosing where to place EDUCO schools as the identifying restriction that directly affects choice but not outcomes.

It may take time for a policy change such as decentralization to affect school performance, which tends to be a cumulative measure. We thus also consider an important intervening variable that eventually influences student outcomes: absence from school (A). Students may be absent for a number of reasons, some of which, such as illness, have nothing to do with decentralization. But other reasons may be tied to school organization and management. Students (or their parents) may not be motivated to ensure regular attendance because the quality of schooling is poor or because parents do not feel involved in the education process. Also contributing to student absence is teacher absence, an important reason why students do not attend school in El Salvador. If teachers are absent, classes are usually canceled, since there is no tradition of using substitute teachers. Although teachers are sometimes absent for legitimate reasons, such as sickness, more often they are simply not fulfilling their duty. Teacher absence is an issue in many countries besides El Salvador:

> Lack of motivation and professional commitment produce poor attendance and unprofessional attitudes towards students. Teacher absenteeism and tardiness are prevalent in many developing countries . . . absenteeism is especially acute in rural areas. Students obviously cannot learn from a teacher who is not present, and absenteeism among teachers encourages similar behavior among students. In some countries . . . parents react to high rates of teacher absenteeism by refusing to enroll their children in school. (Lockheed and Vespoor 1991: 101.)

Teacher absence could be minimized if teachers were appropriately monitored. We would expect that, in a decentralized school, parental involvement would mitigate such behavior.

We assume that the components of $Y = [S\ A]$ can be independently estimated. A will likely affect S, and we assume an implicit recursive process $S = S(A)$, in which the residuals from the different equations are independent of each other, and the matrix of coefficients of endogenous variables is triangular. Each structural equation can thus be estimated by ordinary least squares (OLS), equation by equation (Greene 1997).

A simple model for the ith student in the nth school in the mth community is

$$Y_{inm} = f(\mathbf{X}_{inm}, \mathbf{C}_m, D_{inm}) \tag{1}$$

where $\mathbf{X}$ is a vector of student and household characteristics, $\mathbf{C}$ is a vector of community variables for municipality m, and D is the type of school, either a decentralized EDUCO school or a traditional school. In this model the type of school is assumed to determine most of the school characteristics that affect stu-

dent outcomes. This model is the ultimate reduced form—it assumes that the effect that a school's observed characteristics, such as class size and teacher characteristics, have on achievement is fully determined by the school's management structure (that is, whether it is a decentralized EDUCO school or a traditional school) and the characteristics of the students and parents who participate in decisions concerning the school.

We can often observe the effects of management structure through differences in school and classroom inputs, such as teacher-pupil ratios, teacher remuneration, or the educational background of teachers and administrators. But even if we were to enter as many observable school characteristics as we could in equation 1, the type of school may still be significant because it captures unobserved managerial inputs (Levin 1997). Indeed, in reviewing 96 studies on the effects of five educational inputs on student performance in developing countries, Hanushek (1995) concludes that there are no clear and robust technical relationships between key school inputs and student performance.[6] Thus differences in resources might not be important determinants of school outputs, implying that schools in developing countries are paying for inputs that have little consistent effect on student performance. We distinguish the unobserved effect of community participation from the other unobserved effects of management by explicitly taking into account differences in the level of community involvement. Accordingly, we also derive an alternative model:

$$Y_{inm} = f(\mathbf{X}_{inm}, \mathbf{C}_m, D_{inm}, \mathbf{Z}_{nm}, P_{nm}) \tag{2}$$

where $\mathbf{Z}$ is a vector of observed school and classroom characteristics, and P is the intensity of community participation. Since $\mathbf{Z}$ and P vary by school rather than by student, equation 2 expresses the achievement of the ith student in the nth school. To simplify notation, we drop the school and community subscripts in the rest of this article.

Empirical Specification

Linearizing and adding a stochastic term, which represents a well-behaved measurement error term, to equation 1, we derive the following regression formula:

$$Y_i = \mathbf{X}_i\beta + \mathbf{C} + D_i\alpha + u_i. \tag{3}$$

D takes a value of 1 if the ith student attends a decentralized EDUCO school and 0 if the student attends a traditional, centralized school. By assumption, $E(u_i) = 0$ and $Var(u_i) = \sigma_u^2$. We add school and classroom characteristics and the intensity of community participation to derive the empirical version of equation 2.[7]

6. Hanushek (1995) does, however, suggest that a minimal level of basic school resources, such as textbooks and facilities, is important to student achievement.

7. To simplify notation, we do not add the error terms associated with the school and municipal-level variables. We handle the school variables by using a program participation model and the municipal-level variables by using a municipality-level fixed-effects model.

Observed household and student characteristics reflect the ability of parents to provide a supportive environment for their children. If capital markets were perfect, then life-cycle consumption and human capital investments could be determined independently. Parents would simply borrow to finance the home inputs needed to maximize their children's learning. But since credit markets are far from perfect in El Salvador, the economic circumstances of the household become important. In this article we use asset variables to control for the attributes that are hypothesized to be positively correlated with schooling outcomes (homeownership and the availability of electricity, sanitary services, and piped water). In addition, we control for parents' education, which may also directly affect living standards and preferences for children's education.

We cannot measure students' innate ability directly. However, student characteristics that may be important include gender, since parents or teachers my treat boys and girls differently; age, since older students, while more mature and more likely to score higher, may be self-selected as underachievers and left behind by their cohort; and number of siblings, since the greater the number, the less time parents have to devote to any one child, that is, there are resource competition effects.

We capture community characteristics, C, by municipality-level fixed effects. In El Salvador municipalities are the next administrative level below the department level. There is substantial variance in the distribution of resources across municipalities, which could affect students' access to ancillary services, such as the availability of electricity needed to study, which, in turn, could affect schooling outcomes.

Endogenous Program Participation

A key estimation issue is endogenous program participation. Endogeneity may arise because parents choose which type of school their children attend (conditional on their choosing to send them to school, since we do not have information on children who are not in school).[8] If attendance at an EDUCO school is systematically based on unobserved characteristics that could also influence student achievement, then the OLS estimates of the effect of EDUCO would be biased. That is, α in equation 3 may not accurately measure the value of attending an EDUCO school.

The direction of the bias is ambiguous. If the important unobserved characteristics are students' motivation to learn and parents' commitment to education, and these variables are positively correlated with participation in EDUCO, then comparing outcomes, even after holding constant for observed characteristics,

8. Although EDUCO schools were targeted to areas with limited primary school coverage, parents would still have had a choice of whether or not to send their children to school. They could have had their children commute, albeit over long distances (child fosterage for schooling is not uncommon in developing countries; see Ainsworth 1992 and Glewwe and Jacoby 1994). Or, they could have changed residences (Salvadoran migration rates are high). Unfortunately, the school-based nature of the sample prevents us from including nonattendance as an option.

would overestimate the effect of EDUCO. This bias, however, may be mitigated by the fact that EDUCO targets economically disadvantaged communities.

To take the possibility of bias into account, we explicitly model program participation (that is, whether or not a student enrolls in an EDUCO school rather than in a traditional school). Using a familiar method for obtaining so-called treatment effects, we then estimate this model to obtain the parameters needed to correct equation 3.[9]

WHAT DETERMINES PROGRAM PARTICIPATION? We assume that governments set priorities regarding which municipalities will receive an EDUCO school. Households then use that information to choose the type of school that maximizes their indirect lifetime utility, V. Parents make this choice by weighing the benefits and costs of an EDUCO school relative to other types of schools. The benefits of EDUCO depend on households' perceptions of the virtues of a decentralized program. Some of these preferences can be captured by measurable household characteristics, $\mathbf{X}$, but others are unobserved.

The cost of an EDUCO school relative to a traditional school depends on relative direct costs, such as tuition payments, books, and other fees. The most important components of cost are largely the same for both types of schools: all schools and books are free in first through sixth grade. But there are differences in the other direct costs. EDUCO students do not pay a registration fee, do not buy uniforms, and receive a basic package of school supplies, such as pencils, rulers, and markers. Students in traditional rural schools must bear all of these costs.[10] However, EDUCO parents must devote a substantial amount of time to the school by providing school meals and by building, maintaining, and administering the school.[11]

The principal cost differential between EDUCO and traditional schools comes from differences in access, given the relative paucity of schools in rural areas. We do not have information on households' schooling options (such as the distance from households to feasible EDUCO or traditional schools) because our data are school-based, not household-based. However, we assume that a household is more likely to choose an EDUCO school if the government considers the municipality a priority for the program and thus an EDUCO school is available in the community. The government gives priority to municipalities considered to be neediest according to a classification system developed by MINED and the Ministry of Health.

Municipalities' uneven access to social services has always been a serious issue in El Salvador. However, poverty is more widespread in smaller municipalities,

9. See Greene (1997: 981–82) for a clear discussion of this estimation strategy.

10. We are grateful to Diane Steele of the World Bank for this information, which she received from a phone interview with MINED staff.

11. We do not have data on the magnitudes of these costs. We assume in this article that these cost differentials are roughly offsetting for decentralized and traditional schools. We will verify this assumption with data from surveys that were fielded only in 1999.

which usually lack the financial and institutional capacity to administer and manage social services. The EDUCO program was developed in 78 of the country's poorest municipalities. It started in 1991 with six ACEs in three departments; by the end of 1992, the program had extended to all 14 departments.

The key variables in the targeting system are the incidence of severe malnutrition (the percentage of undersize children in the municipality), the repetition rate, the percentage of overage students, and the net enrollment rate. Higher values for the first three variables, and a lower value for the last, make a municipality a higher priority. In the next section we discuss how this prioritization affects our choice of instruments.

THE FORMAL MODEL OF PROGRAM PARTICIPATION. In the model a household chooses the type of school that yields the highest level of indirect utility, V_j.[12] There are two options: $j = D$ if the household chooses a decentralized EDUCO school or $j = T$ if the household chooses a traditional rural school. V_j depends on the relative benefits and costs of attending an EDUCO school as perceived by parents. Parents choose EDUCO if, for the ith student:

$$D_i^* = V_{Di} - V_{Ti} > 0 \tag{4}$$

where D^* is a latent variable that describes the likelihood that a child is in an EDUCO school. It is determined by:

$$D_i^* = \mathbf{W}_i \boldsymbol{\omega} + \varepsilon_i \tag{5}$$

$$D_i = 1 \text{ if } D_i^* > 0; \text{ 0 otherwise.}$$

In equation 5, $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma_\varepsilon^2$, $\boldsymbol{\omega} = [\delta\ '\pi']'$, and $\mathbf{W}_i = [\mathbf{X}_i\ \mathbf{R}_i]$, which is a vector capturing the benefits and costs of attending an EDUCO school. These benefits and costs are proxied by household characteristics and $\mathbf{R}$, a vector of school density variables (the percentages of EDUCO and traditional schools in all primary schools in a municipality). Equation 5 can be estimated as a probit model under the assumption that ε_i is normally distributed.

The essence of the endogenous participation problem is that the errors in equations 3 and 5 are correlated, that is, $Cov(u_i, \varepsilon_i) \neq 0$, leading to bias.[13] If we assume that u_i and ε_i are jointly normally distributed, the expected value of the outcome variable for EDUCO and traditional schools would be:

$$E(Y_i | D_i = 1) = \mathbf{X}_i\beta + \mathbf{C}_m\gamma + \alpha + \sigma_{u\varepsilon}\lambda_{Di} \tag{6}$$

$$E(Y_i | D_i = 0) = \mathbf{X}_i\beta + \mathbf{C}_m\gamma - \sigma_{u\varepsilon}\lambda_{Ti} \tag{7}$$

12. The basic structure of the model follows the standard treatment of program participation (Greene 1997: 981–82). It is also closely related to the econometric model of self-selection (Willis and Rosen 1979 and Cox and Jimenez 1991). Readers who are not interested in the technical discussion of EDUCO participation can proceed to the next section.

13. Note that $E(u_i | D_i^* > 0) = E(u_i | \mathbf{W}_i\omega + \varepsilon_i > 0) \neq 0$.

where λ_{D_i} and λ_{T_i} are selection terms estimated from Mills ratios.[14] The difference in expected performance between EDUCO participants and nonparticipants, conditional on having chosen a type of school, can be obtained by subtracting equation 7 from equation 6:

$$(8) \qquad E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0) = \alpha + \sigma_{u\varepsilon}(\lambda_{D_i} + \lambda_{T_i})$$

where α is the coefficient of the EDUCO intercept and is usually referred to as the "true" program effect (see Maddala 1983).

Thus if we define $e_i = u_i - \sigma_{u\varepsilon}\lambda_{D_i} D_i + \sigma_{u\varepsilon}\lambda_{T_i}(1 - D_i)$, a term whose expectation is 0 for each of the cases $D = (1, 0)$, the following regression would yield unbiased estimators:

$$(9) \qquad Y_i = \mathbf{X}_i\beta + \mathbf{C}_m\gamma + D_i\alpha + \sigma_{u\varepsilon}[\lambda_{D_i} D_i - \lambda_{T_i}(1 - D_i)] + e_i.$$

If we omitted the selection correction terms (in brackets) from this regression, the difference in equation 8 would be equal to what is usually estimated as the least squares coefficient on the treatment dummy variable. But this expression would overestimate or underestimate the treatment effect, depending on the direction of the participation bias.

To estimate this model, we employ a two-step method.[15] In the first step we estimate equation 5 as a probit model and then use the results to calculate the inverse Mills ratios λ_D and λ_T. In the second step we use the estimated inverse Mills ratios to form the participation terms in equation 9. We then estimate equation 9 with municipal dummies to capture regional fixed effects.

If the error terms in the probit and outcome equations are negatively correlated, that is, if $\sigma_{u\varepsilon} < 0$ (this would occur if an unobserved variable, such as student motivation, negatively affected the likelihood of attending an EDUCO school but positively affected student achievement), then equation 6 implies that the predicted score of a student drawn randomly from the population would be underestimated in the case of EDUCO schools if we use sample mean scores. This can be easily verified by the relationship, $E(Y_i \mid D_i = 1) < \mathbf{X}_i\beta + \mathbf{C}_m\gamma + \alpha$ if $\sigma_{u\varepsilon} < 0$. A similar calculation can be done for equation 7 in the case of traditional schools.

In a linear model, estimation and parameter identification are possible only if the vectors $[\mathbf{X}\ \mathbf{C}_m]$ and $\mathbf{W}$ have no elements in common and are linearly independent. However, in the model above, even if $[\mathbf{X}\ \mathbf{C}_m]$ and $\mathbf{W}$ are identical, equation 9 is estimable. This is because the first-stage estimation results are entered as a nonlinear function in the second stage (equation 9). The nonlinearity helps to identify the model.

14. Assuming joint normality between u_i and ε_i, $E(u_i \mid \mathbf{W}_i\omega + \varepsilon_i > 0) = E(u_i \mid D_i^* > 0) = \sigma_{u\varepsilon}\lambda_{D_i}$, where $\lambda_{D_i} = \phi(\mathbf{W}_i\omega) / \Phi(\mathbf{W}_i\omega)$, the inverse Mills ratio. Similarly, $E(u_i \mid D_i^* < 0) = E(u_i \mid \mathbf{W}_i\omega + \varepsilon_i < 0) \neq 0$, which we can rewrite as $E(u_i \mid D_i^* < 0) = -\sigma_{u\varepsilon}\lambda_{T_i}$, where $\lambda_{T_i} = \phi(\mathbf{W}_i\omega) / [1 - \Phi(\mathbf{W}_i\omega)]$.

15. An alternative way to estimate program participation on unobservables is to use a maximum likelihood method without focusing on Mills ratios. However, this method is more burdensome computationally.

II. Data Description

MINED collected the data in October 1996 with the assistance of the World Bank and the U.S. Agency for International Development (USAID). The survey covered 162 of the country's 262 municipalities. These municipalities share responsibility with the central government for delivering social services.

Since EDUCO was introduced only in 1991, it was not possible to compare the scores on achievement tests given in 1996—only five years later—of EDUCO students who were about to finish their primary education and students in traditional schools. Instead, MINED decided to compare outcomes for third graders. MINED designed the sampling scheme so that the survey was nationally representative. Moreover, the sample was selected so as to consider four types of schools: pure EDUCO, pure traditional, mixed, and private. We dropped students from private schools and traditional public urban schools from the sample, since they are not comparable to EDUCO students. Mixed schools have both EDUCO sections run by ACEs and traditional sections. Some EDUCO programs rented space from traditional schools. The small number of students in these mixed schools attend either EDUCO or non-EDUCO classes located in traditional schools. Since the administration and management of mixed schools are different from the administration and management of pure EDUCO and traditional schools, and thus we should control for unknown management and school-level cross-effects, we could not include mixed schools in pure school samples. Nor could we isolate them as a separate category because of the small sample size. To ensure the robustness of our results, we based our estimations on pure schools only.[16] This left us with 605 students in 30 EDUCO schools and 101 traditional schools.

The survey comprises five questionnaires, one each for students, parents, school directors, teachers, and parent associations. The students' questionnaire requests information about students' relationship with their guardians, type of school, gender, and achievement test results. The parents' data include information on family background and living standards, such as parents' education level, the household's living standard, and asset ownership, as well as detailed socioeconomic information on students, including age, schooling, and health status. The questionnaire for the school director consists of questions about the director, student enrollment, the quality and quantity of teachers, school facilities, and finances. The data collected from teachers include their educational background, years of experience, and salaries, as well as information about the classroom, such as the availability of school materials and frequency with which members of the community association visit the classroom. Lastly, the community and parent association questionnaire contains qualitative information on how the association is organized and how members participate in administration and management of the school. The information on EDUCO schools was collected from

16. The results with both mixed and pure samples, which are not reported here, are consistent with the results for pure schools only.

ACEs, and the information on traditional schools was collected from a counterpart parent organization, the *Sociedad de Padres de Familia.*

Dependent Variables

MINED administered the achievement tests in October 1996 with the assistance of the Intercultural Center for Research in Education (El Salvador, MINED 1997). The tests were given nationally in the third, fourth, and sixth grades, but because EDUCO students had reached only the third grade when the data were collected, we use only the third-grade results in the analysis. Also, we focus only on scores for the mathematics and language tests, ignoring the social studies, science, health, and environment components.

The mathematics test is composed of 30 questions covering ten key subjects—that is, three questions for each subject. Students have mastered a subject if they have answered two of three questions correctly. The language test includes 36 questions covering nine subjects—that is, four questions on each. Students have mastered a subject if they have answered three of four questions correctly. In our sample the average student was able to master 3.70 of 10 subjects in math, but only 1.75 of 9 subjects in language (table 1). These results are not out of line when compared with national averages (El Salvador, MINED 1997).

Of greater interest are the average values for EDUCO and traditional schools. Students in EDUCO schools score marginally lower than students in traditional schools in both subjects, although the differences are not statistically significant (table 1). Our main concern is whether this similarity persists when we control for participation and student, school, and community characteristics.

In addition to test scores, we examine another dependent variable from the parents' questionnaire. This is the response to the following question: "In the past four weeks, how many days of school did the child miss?" As mentioned earlier, we interpret student absence as an important intervening variable that eventually influences educational outcomes. Since we hold constant for student illness, we believe that this variable captures motivational factors. Comparing

Table 1. *Means and Standard Deviations of Variables by Type of School*

Variable	*All schools*	*EDUCO schools*	*Traditional schools*
Output variables			
Math achievement test score	3.70	3.59	3.73
	(2.54)	(2.77)	(2.47)
Language achievement test score	1.75	1.73	1.76
	(1.71)	(1.85)	(1.67)
Days absent from school in past four weeks	0.95	0.95	0.95
	(0.10)	(0.11)	(0.10)
Child and household variables			
Gender (female = 1)	0.51	0.51	0.51
Child's age	10.58	11.01	10.44
	(1.76)	(1.97)	(1.66)

(Table continues on the following page)

Table 1 *(continued)*

Variable	*All schools*	*EDUCO schools*	*Traditional schools*
Child lives without parent(s)[a]	0.14	0.16	0.13
Child had respiratory illness or flu in the past two weeks[a]	0.60	0.63	0.59
Number of siblings (ages 4–15)	2.01	2.11	1.98
	(1.54)	(1.50)	(1.56)
Mother began basic education[a]	0.53	0.50	0.54
Mother's education missing[a]	0.08	0.06	0.09
Father began basic education[a]	0.39	0.38	0.40
Father's education missing[a]	0.04	0.03	0.04
Own house[a]	0.72	0.68	0.73
Electricity available[a]	0.58	0.28	0.67
Sanitary service available[a]	0.18	0.06	0.22
Water available[a]	0.06	0.01	0.08
School variables			
Teacher–pupil ratio (school level)	0.04	0.05	0.03
	(0.056)	(0.09)	(0.041)
Sanitation or latrine available at school[a]	0.93	0.89	0.94
Electricity available at school[a]	0.68	0.30	0.80
Piped water available at school[a]	0.32	0.12	0.38
Teacher and classroom variables			
Teacher finished university education[a]	0.46	0.75	0.37
Years of teacher experience	7.83	4.37	8.89
	(6.44)	(2.71)	(6.87)
Monthly base salary of teacher (thousands of colones)	3,035.21	2,919.23	3,070.71
	(523.38)	(269.40)	(574.84)
Teacher receives bonus[a]	0.64	0.74	0.61
All students have math textbook[a]	0.61	0.58	0.62
Math textbook information missing[a]	0.11	0.25	0.07
All students have language textbook[a]	0.59	0.59	0.59
Language textbook information missing[a]	0.12	0.28	0.07
Teacher instructs multigrade classroom[a]	0.24	0.39	0.20
Multigrade information missing[a]	0.01	0.04	0.00
Number of books in classroom library	74.32	114.63	61.98
	(197.59)	(272.84)	(166.42)
Classroom library information missing[a]	0.47	0.24	0.54
Community participation variable			
Number of parent association visits to classroom in the past month	2.52	5.65	1.56
	(4.82)	(6.59)	(3.63)
Regional school distribution			
Percentage of pure EDUCO schools in all primary schools within municipality	0.21	0.75	0.04
	(0.34)	(0.29)	(0.11)
Percentage of pure traditional schools in all primary schools within municipality	0.69	0.15	0.86
	(0.37)	(0.28)	(0.20)
Inverse Mills ratio	0.00	0.24	0.07
	(0.33)	(0.47)	(0.22)
Number of observations	605	142	463

Note: Standard deviations are in parentheses.

a. Binary variable equals 1 if response is "yes," 0 otherwise.

sample means themselves indicates that, on average, students in both EDUCO and traditional schools missed 0.95 days in the four weeks before the survey.

Explanatory Variables

The means of the explanatory variables show the following. Both EDUCO and traditional schools have an equal number of girls and boys. A fairly large portion of students live without their parents, the proportion being slightly higher for EDUCO students. EDUCO students also have more siblings and are older, although the differences are not significant.

Parents of traditional school students have more education than parents of EDUCO students. Fifty-four percent of mothers or female guardians of traditional students have had some basic education compared with 50 percent for EDUCO students. The gap also holds for fathers (40 and 38 percent). These differences in education are reflected in the asset variables. Fewer EDUCO parents are homeowners or have access to electricity, sanitary services, and running water, suggesting that EDUCO students come from poorer backgrounds than traditional school students.

The socioeconomic characteristics of students are consistent with the characteristics of schools. While teacher-pupil ratios, access to textbooks, and the availability of sanitary facilities are similar in both types of school, fewer EDUCO schools have access to electricity or piped water. However, more EDUCO teachers have finished university education, although they have less teaching experience. The EDUCO teaching corps consists of relatively young recent graduates who receive a bonus for teaching in the program. Another difference is that EDUCO parent associations visit classrooms more than once a week, which is three to four times more often than their traditional counterparts.

The overall picture, then, is one of poor communities that have succeeded in mobilizing parents to become more involved in their children's education, despite their lower standard of living. What we want to know is how much of the differences in outcomes are due to EDUCO.

Identification

We account for possible program endogeneity by explicitly modeling the likelihood of participation in EDUCO and using that information to correct the production function. The main challenge with such corrections is specifying the identifying restriction that allows us to estimate the model.

We include the percentages of EDUCO and traditional schools in all primary schools in each municipality to capture the relative cost of access to each type of school. Arguably, these percentages affect the likelihood that a student will attend an EDUCO school without directly affecting the education production functions at the student level. To isolate general community effects on achievement from the cost-of-access effect, we also include municipal fixed effects in the educational output equation (equation 3). Although the percentages of EDUCO schools are linear combinations of these municipal dummies, we achieve identification because the probit participation equation has a nonlinear functional form.

In order to test the robustness of our results, we also estimate specifications that do not rely exclusively on functional form for identification. For example, instead of using the proportion of EDUCO schools in each municipality, we include the variables that the government uses to prioritize program placement: the extent of malnutrition, the proportion of overage students, repetition rates, and net enrollment rates. Because the government exogenously determines the prioritization formula, the variables included can be used to identify the participation equation. We do not include them in the achievement equations, since, to the extent that local geographic conditions affect achievement, the municipal fixed effects capture their influence. Because our basic qualitative results do not change with these specifications, and to conserve space, we do not report the results here. They are, however, available from the authors.

III. Empirical Results: Student Achievement

The first step of the analysis is to estimate the determinants of participating in EDUCO to correct for possible endogeneity. The most significant variables are mother's education, household assets, and the geographical variables that capture the cost of EDUCO schools relative to that of traditional schools. Mother's education, homeownership, and the availability of water are negatively correlated with EDUCO participation (table 2). Students from households that are better off have a higher likelihood of attending a traditional school. As expected, the availability of an EDUCO school within each municipality significantly increases the probability of enrolling in an EDUCO school.

The next question is whether an EDUCO student (captured by an EDUCO dummy variable) achieves different test scores than a traditional school student. The regressions, which include the participation correction, use math and language achievement as dependent variables, and student and community characteristics (the latter captured by municipality fixed effects) as explanatory variables (table 3). The negative coefficient of the Mills ratio indicates that the error terms of the participation and achievement equations are negatively correlated. This means that EDUCO students have unobserved characteristics that are negatively correlated with achievement test scores.

EDUCO's unconditional effect on language test scores is positive and significant, while its effect on math performance is positive and not significant (table 4). Thus the program has not lessened child learning (after correcting for participation). In fact, it has improved performance in language. However, our measure of EDUCO's advantage in language may be imprecise. The estimate of the EDUCO coefficient is sensitive to the specification of the participation equation—it becomes insignificant when we use in the first stage the municipal prioritization variables instead of the proportion of EDUCO schools in each municipality.[17]

17. However, the qualitative results described in the text hold. The results are available from the authors.

Table 2. *Probit Analysis of School Choice*

Variable	*Coefficient*
Child and household variables	
Gender (female = 1)	−0.14
	(0.54)
Child's age	0.11
	(1.39)
Child lives without parent(s)[a]	0.05
	(0.12)
Child had respiratory illness or flu in the past two weeks[a]	−0.47
	(1.72)*
Number of siblings (ages 4–15)	0.04
	(0.43)
Mother began basic education[a]	−0.54
	(1.75)*
Mother's education missing[a]	−1.29
	(2.07)**
Father began basic education[a]	−0.37
	(1.18)
Father's education missing[a]	−0.48
	(0.63)
Own house[a]	−0.59
	(2.08)**
Electricity available[a]	−0.17
	(0.61)
Sanitary service available[a]	0.19
	(0.46)
Water available[a]	−3.00
	(1.88)*
Regional school distribution (proxies for cost variables)	
Percentage of pure EDUCO schools in all primary schools within municipality	5.96
	(7.14)***
Percentage of pure traditional schools in all primary schools within municipality	−1.66
	(2.63)***
Constant	−1.19
	(1.11)
Log likelihood	−62.42
Pseudo R^2	0.81

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.
Note: School choice is the dependent variable, which equals 1 for an EDUCO school and 0 for a traditional school. *t*-statistics are in parentheses.
a. Binary variable equals 1 if response is "yes," 0 otherwise.

Some of EDUCO's effects can be explained by observed differences in school inputs and community participation. In order to see the extent of these differences, we estimate a model that includes school, classroom, and community participation effects (table 5). EDUCO's impact is lessened with the addition of these independent variables, indicating that some of the differences in test scores can be explained by differences in school inputs and the degree of community in-

Table 3. *Municipality Fixed-Effects Regressions of Student Achievement*

	Mathematics		*Language*	
Variable	*(1)*	*(2)*	*(1)*	*(2)*
EDUCO variables				
EDUCO school present[a]	0.45		2.17	
	(0.33)		(2.32)**	
EDUCO school built in 1991–94[a]		0.74		2.16
		(0.46)		(1.91)*
EDUCO school built in 1995[a]		1.68		2.91
		(1.05)		(2.62)***
EDUCO school built in 1996[a]		–0.37		1.73
		(0.26)		(1.72)*
Year missing[a]		–0.64		2.27
		(0.35)		(1.77)*
Child and household variables				
Gender (female = 1)	–0.69	–0.69	0.01	0.02
	(3.19)***	(3.18)***	(0.08)	(0.12)
Child's age	0.19	0.19	0.04	0.04
	(2.91)***	(2.88)***	(0.80)	(0.80)
Child lives without parent(s)[a]	0.38	0.35	0.43	0.42
	(1.04)	(0.97)	(1.73)	(1.68)
Child had respiratory illness or flu in the past two weeks[a]	0.33	0.32	0.16	0.14
	(1.42)	(1.39)	(0.98)	(0.90)
Number of siblings (age of 4–15)	–0.05	–0.05	–0.02	–0.02
	(0.65)	(0.65)	(0.40)	(0.37)
Mother began basic education[a]	–0.09	–0.05	0.06	0.07
	(0.35)	(0.20)	(0.32)	(0.39)
Mother's education missing[a]	–0.06	–0.06	0.33	0.30
	(0.13)	(0.14)	(1.11)	(1.01)
Father began basic education[a]	–0.05	–0.04	0.19	0.20
	(0.19)	(0.16)	(1.16)	(1.21)
Father's education missing[a]	0.54	0.43	–0.46	–0.49
	(0.91)	(0.72)	(1.10)	(1.19)
Own house[a]	–0.14	–0.17	0.13	0.13
	(0.53)	(0.65)	(0.72)	(0.70)
Electricity available[a]	0.07	0.06	0.01	0.006
	(0.24)	(0.21)	(0.03)	(0.03)
Sanitary service available[a]	0.55	0.49	0.25	0.22
	(1.76)*	(1.56)	(1.13)	(0.99)
Water available[a]	–0.31	–0.25	–0.35	–0.32
	(0.61)	(0.50)	(1.01)	(0.91)
Inverse Mills ratio	–0.46	–0.27	–1.16	–1.05
	(0.56)	(0.33)	(2.03)**	(1.80)*
Constant	1.82	1.85	0.51	0.50
	(2.17)**	(2.18)**	(0.88)	(0.84)
Number of observations	605	605	605	605
Number of municipalities	90	90	90	90
R^2	0.0242	0.0126	0.0002	0.0001

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.
Note: Dependent variable is score on mathematics or language test. *t*-statistics are in parentheses.
a. Binary variable equals 1 if response is "yes," 0 otherwise.

Table 4. *Summary of* EDUCO *Effects on Student Achievement*

Subject	*Without school inputs or community participation variables*	*With school inputs but without community participation variables*	*With school inputs and community participation variables*
Mathematics	0.45	0.40	−0.77
	(0.33)	(0.27)	(0.47)
Language	2.17	1.57	0.74
	(2.32)**	(1.51)	(0.65)

**Significant at the 5 percent level.
Note: *t*-statistics are in parentheses.

volvement. Community involvement is captured by the coefficient on the number of visits that members of the parent association made to classrooms. The coefficient is consistently positive and significant for the basic model with EDUCO dummy variables. This suggests that active community participation is crucial for improving students' achievement in EDUCO schools. An additional classroom visit per week could increase mathematics and language test scores 3.8 and 5.7 percent, respectively.[18] Teacher monitoring by members of parent associations could also improve the quality of education, particularly in EDUCO schools.

We try to distinguish between cohort years by including dummy variables for when the EDUCO program began: prior to 1995, in 1995, or in 1996. Our hypothesis is that the EDUCO effect may be stronger for schools that were built earlier, since they may have learned how to operate the system better. An alternative hypothesis is that newer schools would have better outcomes if there were a "Hawthorne" effect—that is, if the staff and students of newer schools were more motivated and ready to undertake reforms, the kind of enthusiasm that may wane over time. The coefficients for EDUCO are greater for entrants in 1995 and, in fact, are significant and positive for specifications without the community participation variable. This result is consistent with a Hawthorne effect. Still, most of the coefficients are not statistically significant. We can conclude from the OLS results, then, that the EDUCO program has not had a deleterious effect on student achievement, despite its rapid expansion.

Looking at household background, we find that girls perform significantly worse than boys on the mathematics test. In contrast, there are no differences across gender in language. The coefficients on parents' education are not statistically significant, possibly because this variable is likely to be highly correlated with some of the asset variables, and children from households with greater assets or access to infrastructure tend to have better outcomes. For example, performance in mathematics increases almost 15 percent of the mean if students come from households where sanitation is available. It is not surprising that homeownership is not significant—even poor rural families tend to own their own homes in El Salvador.

18. By estimating separate regressions, we find that there are significant effects when the EDUCO dummy is interacted with the participation variable. We do not present these results here.

Table 5. *Municipality Fixed-Effects Regressions of Student Achievement with School Inputs and Participation*

	Mathematics				*Language*			
Variable	*(1)*	*(2)*	*(3)*	*(4)*	*(1)*	*(2)*	*(3)*	*(4)*
EDUCO variables								
EDUCO school present[a]	0.40	–0.77			1.57	0.74		
	(0.27)	(0.47)			(1.51)	(0.65)		
EDUCO school built in 1991–94[a]			–1.93	–1.87			0.82	0.83
			(0.97)	(0.94)			(0.59)	(0.60)
EDUCO school built in 1995[a]			3.21	5.50			3.26	3.85
			(1.75)*	(1.59)			(2.55)	(1.59)
EDUCO school built in 1996[a]			–0.49	–0.11			0.43	0.53
			(0.28)	(0.06)			(0.35)	(0.42)
Year missing[a]			–4.41	–4.82			0.45	0.34
			(1.49)	(1.60)			(0.22)	(0.16)
Child and household variables								
Gender (female = 1)	–0.57	–0.53	–0.51	–0.51	0.07	0.10	0.11	0.10
	(2.61)***	(2.40)**	(2.31)**	(2.33)**	(0.47)	(0.67)	(0.69)	(0.68)
Child's age	0.17	0.17	0.18	0.18	0.02	0.02	0.02	0.02
	(2.57)**	(2.62)***	(2.66)***	(2.65)***	(0.45)	(0.50)	(0.52)	(0.52)
Child lives without parent(s)[a]	0.42	0.42	0.40	0.39	0.46	0.46	0.44	0.43
	(1.17)	(1.17)	(1.11)	(1.08)	(1.85)*	(1.85)*	(1.73)*	(1.72)*
Child had respiratory illness or flu in past two weeks[a]	0.23	0.22	0.20	0.20	0.09	0.08	0.06	0.06
	(1.01)	(0.94)	(0.88)	(0.86)	(0.56)	(0.48)	(0.39)	(0.38)
Number of siblings (ages 4–15)[a]	–0.04	–0.03	–0.02	–0.02	–0.02	–0.01	–0.01	–0.01
	(0.57)	(0.38)	(0.29)	(0.33)	(0.34)	(0.14)	(0.16)	(0.17)
Mother began basic education[a]	–0.02	–0.02	–0.02	–0.02	0.07	0.08	0.08	0.08
	(0.09)	(0.09)	(0.07)	(0.06)	(0.43)	(0.43)	(0.46)	(0.47)
Mother's education missing[a]	–0.14	–0.12	–0.07	–0.07	0.34	0.35	0.33	0.33
	(0.32)	(0.28)	(0.15)	(0.16)	(1.09)	(1.13)	(1.06)	(1.05)
Father began basic education[a]	–0.07	–0.07	–0.06	–0.05	0.15	0.15	0.16	0.16
	(0.31)	(0.30)	(0.25)	(0.23)	(0.90)	(0.92)	(0.96)	(0.97)

Father's education missing[a]	0.57	0.55	0.54	0.53	–0.37	–0.38	–0.42	–0.42
	(0.94)	(0.92)	(0.89)	(0.89)	(0.89)	(0.92)	(1.00)	(1.00)
Own house[a]	–0.19	–0.21	–0.23	–0.23	0.05	0.04	0.04	0.04
	(0.74)	(0.80)	(0.89)	(0.88)	(0.26)	(0.19)	(0.21)	(0.21)
Electricity available[a]	–0.02	–0.06	–0.15	–0.18	–0.01	–0.03	–0.06	–0.07
	(0.07)	(0.18)	(0.50)	(0.57)	(0.03)	(0.15)	(0.28)	(0.31)
Sanitary service available[a]	0.59	0.54	0.55	0.57	0.28	0.24	0.24	0.24
	(1.84)*	(1.68)*	(1.73)*	(1.78)*	(1.25)	(1.09)	(1.05)	(1.07)
Water available[a]	–0.26	–0.24	–0.24	–0.24	–0.39	–0.38	–0.36	–0.36
	(0.50)	(0.47)	(0.48)	(0.47)	(1.10)	(1.07)	(1.01)	(1.01)
School variables								
Teacher–pupil ratio	–27.39	–19.66	–28.18	–33.42	5.77	11.30	9.84	8.50
	(1.16)	(0.82)	(1.17)	(1.33)	(0.35)	(0.68)	(0.59)	(0.49)
Sanitation or latrine available[a]	0.42	0.38	0.32	0.35	0.18	0.15	0.24	0.25
	(0.52)	(0.47)	(0.39)	(0.42)	(0.31)	(0.26)	(0.42)	(0.43)
Electricity available[a]	0.16	0.34	0.95	1.09	0.27	0.39	0.57	0.60
	(0.30)	(0.60)	(1.48)	(1.64)*	(0.69)	(1.00)	(1.27)	(1.30)
Piped water available[a]	–0.19	–0.22	–0.13	–0.09	–0.21	–0.23	–0.19	–0.18
	(0.37)	(0.43)	(0.24)	(0.18)	(0.58)	(0.64)	(0.50)	(0.47)
Teacher and classroom variables								
Teacher finished university education[a]	–0.57	–0.80	–0.66	–0.48	–0.15	–0.31	–0.06	–0.02
	(1.33)	(1.79)*	(1.29)	(0.87)	(0.50)	(1.00)	(0.17)	(0.04)
Years of teacher experience	0.06	0.05	0.05	0.05	0.03	0.02	0.03	0.03
	(1.35)	(1.16)	(1.15)	(1.27)	(1.07)	(0.87)	(1.08)	(1.11)
Monthly base salary of teacher (thousands of colones)	–0.78	–0.72	–0.81	–0.87	–0.53	–0.49	–0.58	–0.59
	(1.71)*	(1.56)	(1.70)*	(1.80)*	(1.67)*	(1.52)	(1.75)*	(1.77)*
Teacher receives bonus[a]	0.53	0.51	0.53	0.55	0.49	0.48	0.46	0.47
	(1.16)	(1.12)	(1.09)	(1.13)	(1.54)	(1.51)	(1.36)	(1.37)

(Table continues on the following page)

Table 5. *(continued)*

Variable	Mathematics (1)	Mathematics (2)	Mathematics (3)	Mathematics (4)	Language (1)	Language (2)	Language (3)	Language (4)
All students have math textbook[a]	0.68	0.73	0.44	0.35	0.05	0.09	0.01	–0.01
	(1.01)	(1.08)	(0.63)	(0.49)	(0.11)	(0.18)	(0.03)	(0.02)
Math textbook information missing[a]	2.29	1.90	–5.89	–8.42	0.40	0.13	–2.54	–3.19
	(1.27)	(1.05)	(1.48)	(1.64)*	(0.32)	(0.10)	(0.92)	(0.89)
All students have language textbook[a]	–0.75	–0.76	–0.95	–1.02	–0.16	–0.16	–0.25	–0.27
	(1.14)	(1.16)	(1.42)	(1.52)	(0.34)	(0.36)	(0.55)	(0.58)
Language textbook information missing[a]	–0.87	–0.54	7.82	10.65	0.28	0.52	3.55	4.27
	(0.47)	(0.29)	(1.90)*	(1.94)*	(0.22)	(0.40)	(1.24)	(1.12)
Teacher teaches in multigrade classroom[a]	0.75	0.93	0.68	0.55	0.49	0.62	0.53	0.50
	(1.30)	(1.58)	(1.12)	(0.88)	(1.23)	(1.51)	(1.25)	(1.14)
Multigrade information missing[a]	–0.01	–0.75	0.40	1.33	1.72	1.19	2.52	2.75
	(0.00)	(0.33)	(0.16)	(0.48)	(1.10)	(0.75)	(1.45)	(1.43)
Number of books in classroom library (thousands of books)	0.91	0.90	0.70	0.63	1.54	1.53	1.47	1.45
	(0.89)	(0.87)	(0.68)	(0.61)	(2.14)**	(2.13)**	(2.05)**	(2.01)**

Classroom library information missing[a]	0.40	0.36	0.40	0.45	0.30	0.27	0.40	0.42
	(0.73)	(0.67)	(0.69)	(0.76)	(0.79)	(0.73)	(0.99)	(1.01)
Community participation variable								
Number of parent association visits to classroom in past month		0.14		−0.11		0.10		−0.03
		(1.72)*		(0.78)		(1.77)*		(0.29)
Inverse Mills ratio	−0.58	−0.20	−0.05	−0.12	−1.09	−0.82	−0.73	−0.75
	(0.68)	(0.22)	(0.06)	(0.13)	(1.84)*	(1.34)	(1.20)	(1.22)
Constant	3.97	3.45	4.13	4.47	1.12	0.74	0.83	0.92
	(1.71)*	(1.47)	(1.75)*	(1.87)*	(0.69)	(0.45)	(0.51)	(0.55)
Number of observations	605	605	605	605	605	605	605	605
Number of municipalities	90	90	90	90	90	90	90	90
Overall R^2	0.0153	0.0129	0.0173	0.0175	0.0186	0.0010	0.0108	0.0077

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.
Note: Dependent variable is mathematics or language test score. *t*-statistics are in parentheses.

Children with more siblings perform worse on both math and language tests, although the coefficients are not statistically significant. This result may indicate that parents devote less time to their children's individual needs. Older children do better in math than younger ones, even though they are in the same grade. However, age does not matter in determining language scores.

The EDUCO effect can be mediated through school and classroom-level indicators and through the intensive involvement of parent associations. To capture these effects, we include school and classroom-level characteristics and a community participation variable in the regressions.[19] The EDUCO effect is less than that in regressions without school-level variables, indicating that a significant portion of the difference between EDUCO and traditional schools can be captured by differences in observable school characteristics and differences in community involvement (see table 4). The EDUCO coefficient, however, is still statistically insignificant. The basic results for the effects of socioeconomic characteristics do not change.

Most of the school-level variables are not significantly different from zero. The two exceptions are teachers' base salary, which has a negative coefficient, and the availability of a classroom library, which is positively related to language achievement scores. EDUCO teachers receive a piece-wage rate, which the ACEs determine, while teachers in traditional schools have a fixed-wage scheme (World Bank 1995). The results for teachers' base salary might capture the inefficiency of fixed wage schemes in traditional schools. The positive effect on language scores of having a classroom library is also consistent with past evaluation of EDUCO (World Bank 1995: 19–20). It may be that classroom libraries help teachers to complete their lesson plans and stimulate students' interest and reading habits, both of which improve language scores.

Most notably, the community participation variable has a positive and statistically significant coefficient for the basic specification with the EDUCO dummy variable (see column 2 of table 5). This finding indicates that the intensity of community involvement is significantly related to students' academic achievements. Community participation might have a positive peer effect or work to monitor teachers.

IV. Empirical Results: Student Absence

Parents' negative perceptions of education are an important issue, and this is true not only in El Salvador. In many rural areas throughout the developing world uneducated parents underestimate the value of education and thus do not send their children to school. The reasons may be cultural, social, or economic. The main problem, however, seems to be that parents are given poor incentives or poor information. Teacher absenteeism is also a chronic problem in the public schools

19. We enter them linearly and interact them with the EDUCO dummy, since EDUCO may change school characteristics. We do not report the regressions with the interaction terms here. They are available on request.

of many developing countries. Although excuses are sometimes legitimate, such as sickness, more often teachers are simply derelict. When teachers are absent, classes are usually canceled, since there is no tradition of using substitute teachers.

Our hypothesis is that in a decentralized setting parents are better able and motivated to send children to school and to monitor teacher behavior. In fact, parents are more likely to send their children to school if they attend an EDUCO school, and teacher absenteeism is less prevalent in EDUCO schools, further reducing student absence (World Bank 1997).

The dependent variable in our regressions is the number of school days that the child missed in the past month. To control for student absence because of health problems, we add an additional indicator variable that equals 1 if a child suffered from a respiratory illness or flu in the past two weeks and 0 otherwise. The principal result is that the coefficients on the EDUCO dummy variable are consistently negative and statistically significant, especially if we control for participation bias (table 6). A student in an EDUCO school is less likely to be absent even after we hold constant household, school, and participation characteristics.[20]

An important finding is that mother's education has a negative and statistically significant coefficient in all specifications. Having an educated mother reduces student absence approximately 70 percent. This implies that the mother's positive perception of education drawn from her own experiences contributes to better child attendance.

We also differentiate the EDUCO dummy variables by year. The dummy variable for 1991–94 is insignificant—a third-grade student attending an EDUCO school during those years is just as likely to be absent as a student from a traditional school (table 6). However, the EDUCO coefficients are negative and significant for schools that were built in 1996 in the regressions with the participation correction. The EDUCO effect is stronger for newer schools, which is consistent with the Hawthorne effect described earlier. Thus the negative effect of a decay in early enthusiasm of staff and students seems to dominate the positive effect of experience on performance.

The coefficient of the participation correction term is positive, although not significant. This means that unobserved characteristics of EDUCO students make them more likely to miss school.

V. Conclusions

El Salvador's EDUCO program has been remarkably successful in expanding educational opportunities for the poor living in rural areas. Decentralization has

20. Teachers are absent more days in traditional schools (1.4 days a month) than in EDUCO schools (1.16 days a month), implying that parent association monitoring works. This finding contrasts with that of Reimers (1997), who conducted interviews in 140 schools in 1993. That study was done on a earlier vintage than the schools in this study and did not correct for other variables that may affect outcomes. Sawada (1998) also extensively investigates the transmission mechanism from community participation to better educational outcomes in EDUCO schools by estimating teacher compensation and effort functions, as well as input demand functions.

Table 6. *Municipality Fixed-Effects Regressions of Student Absence*

Variable	*(1)*	*(2)*	*(3)*	*(4)*
EDUCO variables				
EDUCO school present[a]	−3.01		−3.93	
	(1.79)*		(1.91)*	
EDUCO built in 1991–94[a]		−3.11		−3.55
		(1.53)		(1.41)
EDUCO built in 1995[a]		−2.64		−0.93
		(1.32)		(0.21)
EDUCO built in 1996[a]		−3.20		−4.23
		(1.76)*		(1.84)*
Year missing[a]		−3.89		−5.89
		(1.69)*		(1.55)
Child and household variables				
Gender (female = 1)	−0.08	−0.08	−0.03	−0.03
	(0.29)	(0.30)	(0.10)	(0.13)
Child's age	0.17	0.17	0.16	0.17
	(2.10)**	(2.07)**	(1.98)**	(2.04)**
Child lives without parent(s)[a]	−0.12	−0.12	−0.16	−0.21
	(0.27)	(0.28)	(0.36)	(0.46)
Child had respiratory illness or flu in past two weeks[a]	−0.29	−0.29	−0.26	−0.26
	(1.02)	(1.02)	(0.89)	(0.90)
Number of siblings (ages 4–15)	−0.11	−0.11	−0.08	−0.08
	(1.23)	(1.23)	(0.86)	(0.86)
Mother began basic education[a]	−0.70	−0.68	−0.67	−0.66
	(2.24)**	(2.17)**	(2.13)**	(2.09)**
Mother's education missing[a]	−0.23	−0.22	0.13	0.16
	(0.44)	(0.40)	(0.23)	(0.28)
Father began basic education[a]	0.41	0.41	0.36	0.35
	(1.38)	(1.38)	(1.21)	(1.18)
Father's education missing[a]	−0.58	−0.61	−0.68	−0.75
	(0.79)	(0.81)	(0.91)	(0.99)
Own house[a]	−0.17	−0.20	−0.24	−0.26
	(0.53)	(0.60)	(0.74)	(0.79)
Electricity available[a]	−0.39	−0.39	−0.75	−0.79
	(1.07)	(1.08)	(1.93)*	(2.02)**
Sanitary service available[a]	−0.20	−0.22	−0.29	−0.29
	(0.52)	(0.56)	(0.71)	(0.72)
Water available[a]	−0.58	−0.57	−0.68	−0.68
	(0.93)	(0.90)	(1.06)	(1.06)
School variables				
Teacher–pupil ratio			27.18	23.95
			(0.90)	(0.76)
Sanitation or latrine available[a]			−1.87	−1.79
			(1.84)**	(1.73)*
Electricity available[a]			1.38	1.73
			(1.95)**	(2.06)**
Piped water available[a]			0.09	0.25
			(0.13)	(0.37)

Table 6. *(continued)*

Variable	*(1)*	*(2)*	*(3)*	*(4)*
Teacher and classroom variables				
Teacher finished university education[a]			0.63	1.03
			(1.13)	(1.48)
Years of teacher experience			0.001	0.004
			(0.03)	(0.08)
Monthly base salary of teacher (thousands of colones)			1.08	0.90
			(1.88)*	(1.47)
Teacher receives bonus[a]			-0.17	-0.29
			(0.29)	(0.48)
All students have math textbook[a]			1.31	1.09
			(1.54)	(1.23)
Math textbook information missing[a]			1.50	-3.55
			(0.66)	(0.55)
All students have language textbook[a]			-0.73	-0.82
			(0.88)	(0.96)
Language textbook information missing[a]			-0.23	5.21
			(0.10)	(0.76)
Teacher teaches in multigrade classroom[a]			0.88	0.60
			(1.19)	(0.76)
Multigrade information missing[a]			-2.93	-0.98
			(1.02)	(0.28)
Number of books in classroom library (thousands of books)			-0.24	-0.30
			(0.19)	(0.23)
Classroom library information missing[a]			-0.50	-0.26
			(0.74)	(0.35)
Community participation variable				
Number of parent association visits to classroom			0.07	-0.05
			(0.74)	(0.26)
Inverse Mills ratio	0.96	1.02	1.48	1.59
	(0.93)	(0.98)	(1.34)	(1.43)
Constant	1.62	1.68	-2.30	-1.89
	(1.55)	(1.59)	(0.78)	(0.63)
Number of observations	605	605	605	605
Number of municipalities	90	90	90	90
R^2	0.0022	0.0024	0.0067	0.0074

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.
Note: Dependent variable is the number of school days student missed in past month. *t*-statistics are in parentheses.
a. Binary variable equals 1 if response is "yes," 0 otherwise.

also been instrumental in getting families and communities more involved in their children's schooling. But has the program delivered more? This article has evaluated EDUCO by determining if it has raised achievement scores and lowered student absence.

The average scores of EDUCO students on standardized mathematics and language tests are lower than those of students attending traditional schools. This is

not surprising since EDUCO students come from disadvantaged backgrounds. What is interesting is that, after we control for background and correct for participation bias in the samples, these differences disappear. In fact, the average performance of EDUCO students on language tests is slightly better than that of traditional school students. The similarity in outcomes holds regardless of how long a school has participated in the EDUCO program, although newer schools do show an advantage that is not significantly different from zero.

There is considerable variance in performance even after holding constant for type of school. The most important socioeconomic variables that have a positive effect on student achievement are being male, coming from a family with access to sanitary services, being older, and having fewer siblings. At the school level, the availability of a classroom library has a positive effect on achievement. Most important, coefficients of the parent participation variable are positive and statistically significant in the two specifications that include an EDUCO dummy variable. Test results are significantly and positively related to the number of visits by ACEs or their equivalent.

The number of days students miss is negatively related to intangible EDUCO effects. In a decentralized setting parents are more motivated to send children to school and better able to monitor teachers. Parents' education, especially mother's education, has a positive impact on student attendance. Moreover, teachers tend to miss fewer days when monitored by parent associations. This, in turn, implies that students miss fewer days.

We conclude that the rapid expansion of rural education through EDUCO's decentralized mechanism has not lowered achievement levels in El Salvador, even in the most needy parts of the country. In fact, it has improved language scores. The important transmission mechanism is likely to be the community monitoring system. Parents' deeper involvement in their children's education may result in improved attendance and pressure providers to deliver observable inputs. Although teachers, parents, and parent associations are not given direct incentives to raise standardized test scores in mathematics and language, the EDUCO program has had an overall positive impact. Thus parents' participation in school-based management seems to be an appropriate way of improving education in poor communities.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Ainsworth, Martha. 1992. "Economic Aspects of Child Fostering in Côte d'Ivoire." Living Standard Measurement Study Working Paper 92. World Bank, Washington, D.C. Processed.

Cox, Donald, and Emmanuel Jimenez. 1991. "The Relative Effectiveness of Private and Public Schools: Evidence from Two Developing Countries." *Journal of Development Economics* 34(1):99–121.

El Salvador, MINED (Ministerio de Educación). 1995. EDUCO *Learns and Teaches.* San Salvador: Algier's Impresores.

———. 1997. "Informe de evaluación del rendimiento en 3o, 4o, 6o grado de educación básica en lenguaje, matemática, estudios sociales, y ciencia, salud, y medio ambiente basado en la aplicación nacional de pruebas de octubre de 1996." Dirección Nacional de Evaluación e Investigación, San Salvador, February. Processed.

Glewwe, Paul, and Hanan Jacoby. 1994. "Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana." *Journal of Human Resources* 29(3):841–64.

Greene, William H. 1997. *Econometric Analysis.* 3d ed. Upper Saddle River, New Jersey: Prentice Hall.

Hanushek, Eric A. 1995. "Interpreting Recent Research on Schooling in Developing Countries." *The World Bank Research Observer* 10(2):227–46.

James, Estelle, Elizabeth King, and Ace Suryadi. 1996. "Finance, Management, and Costs of Public and Private Schools in Indonesia." *Economics of Education Review* 15(4):387–98.

Jimenez, Emmanuel, and Vicente Paqueo. 1996. "Do Local Contributions Affect the Efficiency of Public Primary Schools?" *Economics of Education Review* 15(4):377–86.

Levin, Henry M. 1997. "Raising School Productivity: An X-efficiency Approach." *Economics of Education Review* 16(3):303–11.

Lockheed, Marlaine, and Adriaan Verspoor. 1991. *Improving Education in Developing Countries.* New York: Oxford University Press for the World Bank.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge, U.K.: Cambridge University Press.

McMillan, Robert. 1999. "Parental Pressure and Competition: An Empirical Analysis of Public School Quality." Department of Economics, Stanford University, Stanford, Calif. Processed.

Reimers, Fernando. 1997. "The Role of the Community in Expanding Educational Opportunities: The EDUCO Schools in El Salvador." In James Lynch, Celia Modgil, and Sohan Modgil, eds., *Education and Development: Tradition and Innovation, Equity and Excellence.* vol. 2. London: Cassell.

Sawada, Yasuyuki. 1998. "Community Participation, Teacher Effort, and Educational Outcomes: A Case Study of El Salvador's EDUCO Program." Department of Economics, Stanford University, Stanford, Calif. Processed.

Willis, Robert, and Sherwin Rosen. 1979. "Education and Self-Selection." *Journal of Political Economy* 87(5):57–36.

World Bank. 1994. "El Salvador: Community Education Strategy: Decentralized School Management." Country Report 13502-ES. Washington, D.C. Processed.

———. 1995. "Staff Appraisal Report: El Salvador Basic Education Modernization Project." IBRD Report 14129-ES. Washington, D.C. Processed.

———. 1996. "Impact Evaluation of Education Projects Involving Decentralization and Privatization." Working Paper on Impact Evaluation of Education Reforms 0. El Salvador Evaluation Team, Washington, D.C. Processed.

———. 1997. "El Salvador's EDUCO Program: A First Report on Parents' Participation in School-based Management." Working Paper on Impact Evaluation of Education Reforms 4. El Salvador Evaluation Team, Washington, D.C. Processed.

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 443-65

Can Private School Subsidies Increase Enrollment for the Poor? The Quetta Urban Fellowship Program

Jooseop Kim, Harold Alderman, and Peter F. Orazem

This study evaluates a program designed to stimulate girls' schooling through the creation of private girls' schools in poor urban neighborhoods of Quetta, Pakistan. Enrollment growth in these randomly selected neighborhoods is compared to enrollment growth in otherwise similar neighborhoods that were randomly assigned to a control group. The analysis indicates that the program increased girls' enrollment around 33 percentage points. Boys' enrollment rose as well, partly because boys were allowed to attend the new schools and partly because parents would not send their girls to school without also educating their boys. This outcome suggests that programs targeted at girls can also induce parents to invest more in their boys. The success of the program varied across neighborhoods, although success was not clearly related to the relative wealth of a neighborhood or to parents' level of education. Thus the program offers tremendous promise for increasing enrollment rates in other poor urban areas.

Private schooling, often postulated to improve school quality, may also be a means to leverage public funds in order to provide access to schooling at rates faster than are possible with public funds alone. This article measures the impact on enrollment of a program designed to encourage the creation of new private girls' schools in Quetta, the capital city of Balochistan Province in Pakistan. The analysis represents a unique opportunity to apply experimental design methods to evaluate an educational policy innovation. By randomizing the implementation of the pilot program, we are able to generate robust estimates of the impact of the program on enrollment. We avoid the bias that often arises in such assessments

Jooseop Kim is senior researcher at the Korea Institute for Vocational Education and Training, Peter F. Orazem is with the Department of Economics at Iowa State University, and Harold Alderman is with the Development Research Group at the World Bank. Their e-mail addresses are jskim@krivet.re.kr, halderman@worldbank.org, and pfo@iastate.edu. This study was supported by the research project Impact Evaluation of Education Projects Involving Decentralization and Privatization under RPO 679-18 of the World Bank, supervised by Elizabeth King. The authors are indebted to current and past members of the Balochistan Primary Education Directorate, including Ijaz Ahmed Malik, Mohammed Ishaque, Quaratul Ain, and Bill Darnell, for their kind hospitality and research support. They are also grateful to Sultan Mahmood Niazi of the Balochistan Education Foundation and to Brian Spicer and Fahim Akbar of the Balochistan Management Information System for their help in designing and implementing the data collection. Finally, the authors thank their task managers from the South Asia Division, Mae Chu Chang, Guilherme Sedlacek, and Ivar Andersen, for offering many helpful comments and providing an unusually supportive research environment amid ongoing operational support; the referees for making valuable suggestions; and Donna Otto for preparing the final manuscript.

when individuals or groups who participate in a program are those who are best suited to benefit from it.

I. The Quetta Girls Fellowship Program

Primary school enrollment rates in Pakistan are lower than in other countries at the same level of economic development, including Bangladesh, India, and Nepal. Nationally, the gross enrollment rate is 58 percent: 69 percent for boys and 42 percent for girls. This gender gap is even wider in the province of Balochistan, where 62 percent of boys and only 29 percent of girls are enrolled in school.[1] The government of Pakistan has set a target of achieving universal primary enrollment by 2006. Meeting this goal would require girls' enrollment to more than double nationally and more than triple in Balochistan.

There is evidence that supply constraints partially account for low school enrollment and achievement in Pakistan, especially in rural areas and in poor urban neighborhoods.[2] However, in Pakistan, as in many other countries, the government's ability to increase school capacity is constrained by inadequate public budgets. Expansion is also circumscribed because the government generally constructs, rather than rents, school capacity and requires recipient neighborhoods to provide land for their new government schools. Many poor urban neighborhoods have developed as squatter communities that have poorly defined property rights, limiting their ability to donate land.

In the case of educating girls, the problem is not the lack of schools per se. Cultural prohibitions in many communities mean that educational opportunities for girls are frequently restricted in the absence of gender-specific programs. If Pakistan is to achieve universal primary enrollment for girls, more segregated girls schools or coeducational schools with female teachers will be needed. Given the limitations on increased government provision, one strategy is to make private girls schools more available in poor neighborhoods. Private schools do not face the same problem of land acquisition and may be less constrained financially than government schools. Consequently, the government may devote less time and expense to increasing school capacity for girls if it partially funds the expansion of private schools rather than fully funds the expansion of government schools.

Recognizing these concerns, the Balochistan Education Foundation launched the Urban Fellowship Program in Quetta in February 1995. The purpose of this pilot project was to determine whether establishing private schools in poor neighborhoods was a cost-effective means of expanding primary education for

1. These statistics are based on 1996 data provided by the Pakistan Education Management Information System.

2. Alderman and others (1996) find that differences in school availability account for 30–40 percent of the gap in cognitive skills between boys and girls in rural Pakistan.

girls in Quetta's lower-income neighborhoods. Recent evidence from the Pakistan Integrated Household Survey suggests that about 77 percent of girls who start school finish the primary cycle. It was thought that if the program could get these poor girls to start school, many would persist long enough to attain literacy.

The Urban Fellowship Program encouraged private schools, which were controlled by the community, to establish new facilities by paying subsidies directly to the schools. Schools were assured of government support for three years. The initial subsidy was 100 rupees (about $3) per month per girl enrolled. There was an upper scholarship limit of 10,000 rupees (100 girls at 100 rupees per girl) per month. This subsidy was sufficient to cover typical tuition at the lowest-priced private schools. In addition, each school received 200 rupees per girl to defray start-up costs. The subsidy was reduced in the second year and again in the third year. By the fourth year schools were expected to be largely self-sufficient through fees and private support, although they would still be eligible to apply to the Balochistan Education Foundation for additional grants. It is possible that uncertainty regarding the schools' long-term sustainability discouraged some parents from enrolling their children. Nevertheless, even if the results presented here are a lower bound, they are substantial. We discuss the issue of sustainability in the concluding section.

Fellowship schools were allowed to admit boys provided that they made up less than half of total enrollment. Boys had to pay tuition at least equal to, and often greater than, that of girls. The grant was calculated only for enrolled girls; schools received no additional subsidy for enrolling boys. Schools were required to keep class sizes at or below 50 boys and girls per classroom and had to hire at least one teacher for each classroom.

To implement the program, the Balochistan Education Foundation contracted the Society for Community Support of Primary Education in Balochistan (SCSPEB), a nongovernmental organization (NGO), to conduct an initial census of each site to ensure that there were a sufficient number of girls in the target age range (four to eight years) and to inform parents of the program. The SCSPEB had several years of experience in implementing primary school projects, particularly school promotion efforts in rural communities. The goal was to create a partnership between neighborhood parents and the school operator. The SCSPEB first conducted a meeting of parents to see if they were interested in attracting a private school to their neighborhood. The parents were asked to form a committee, which would represent the neighborhood in negotiations with potential school operators. With the assistance of the SCSPEB, the parent committee developed a proposal detailing the neighborhood's need for a school, the resources it was willing to provide the school (land, buildings, equipment), and any other requirements an operator was expected to satisfy. Experienced school operators were invited to submit proposals in response. The parent committees were allowed to select their school operator from among the proposals or to choose to run the school themselves.

II. Evaluation Strategy

Because government resources are limited and the need to expand enrollment is so great, the government of Balochistan needed an accurate measure of the program's success and its prognosis for expansion. The evaluation problem is to find an unbiased estimator of the impact of the fellowship program.

Use of Randomized Assignment

In this study the outcome variable of interest is school enrollment. However, the approach is general; the use of random assignment is not specific to the evaluation of this educational program or to educational programs per se.

Denoting school enrollment in the treatment and control neighborhoods as R_T and R_N, respectively, ideally we would like to estimate $\alpha = R_{Tit} - R_{Nit}$ for any individual i at time t.[3] However, we cannot estimate α directly, since a child cannot be simultaneously in both the treatment group and the control group.

One way to get an unbiased estimator of α is to use changes in the outcome variable over time. This approach, termed a reflexive evaluation, can be written:

$$E^R(\alpha) = E(R_{Tt}) - E(R_{T0}). \tag{1}$$

The reflexive estimator measures the expected effect of the program as the gap between the enrollment rate after the program was implemented, $E(R_{Tt})$, and the enrollment rate before the program was implemented, $E(R_{T0})$. The underlying assumption of this method is that the period t outcome in the treatment neighborhood without the program would be identical to the observed outcome before the program. In effect, the treatment group in the base period (before the intervention) serves as a control for the treatment group after implementation.[4]

However, reflexive evaluations are sensitive to trends that may be nationwide, but erroneously attributed to the intervention. Thus an alternative approach is to use a control group to derive estimates of the counterfactual state. The difference in outcomes between the treatment and control groups is then used as an estimate of α. This estimator could be either the mean difference, defined as

$$E^M(\alpha) = [E(R_{Tt}) - E(R_{Nt})], \tag{2}$$

or the difference-in-differences, defined as:

$$E^D(\alpha) = [E(R_{Tt}) - E(R_{Nt})] - [E(R_{T0}) - E(R_{N0})]. \tag{3}$$

The mean-difference estimator (equation 2) measures the expected effect of the program as the observed difference in outcomes between the treatment group

3. In this application the treatment neighborhoods are those targeted for the private school promotion and subsidy. A child is considered a member of the treatment group if he or she resides in the treatment neighborhood, whether or not the child is enrolled in a fellowship school. However, the child must be exposed to the possibility of enrolling in a fellowship school.

4. Grossman (1994) classifies a randomly assigned counterfactual group as a "control group" and a nonrandomly assigned counterfactual group as a "comparison group."

and the control group after the program was implemented. This method assumes that the control group perfectly matches the treatment group, which is often achieved by randomly assigning groups to the two populations (Newman, Rawlings, and Gertler 1994).

The difference-in-differences estimator (equation 3) measures the expected effect of the program as the difference between the outcome in the treatment group, $E(R_{Tt})$, and the control group, $E(R_{Nt})$, after program implementation adjusted by the difference between the two groups before implementation. This method assumes that the difference in outcomes between the two groups before the program was introduced would remain constant over time if it were not for the program. Thus the difference in outcomes between the two groups after the program was introduced reflects the initial difference as well as the difference brought about by the program. Differencing the differences yields an estimate of the program effect. If randomization is successful, there will be no difference between equations 2 and 3 because if the two groups are identical at the outset, the term in the second bracket of equation 3 will equal zero.

For reasons discussed below, randomization may be violated. If that happens, it becomes necessary to control for differences between the treatment and control groups that could also influence outcomes. To illustrate, consider a general model of individual enrollment choice in year t:

$$R_{it} = \mathbf{X}_{it}\boldsymbol{\beta}_t + U_{it} \tag{4}$$

In equation 4, $\mathbf{X}_{it}$ is a vector of observed characteristics, U_{it} is an error term, and $\boldsymbol{\beta}_t$ is a vector of parameters to be estimated. We can derive a covariate post-test estimate of α from a cross-sectional regression in some period t after the program is implemented, assuming that the αs are invariant across individuals:

$$R_{it} = \mathbf{X}_{it}\boldsymbol{\beta}_t + d_t\alpha + U_{it} \tag{5}$$

where d_t is a dummy variable indicating residence in a fellowship school neighborhood. With random assignment into the treatment group, we can assume that d_t is independent of the unobserved variables U_{it}, so that $E(U_{it} \mid d_t) = 0$. Under this assumption we obtain an unbiased estimate of α. Conversely, if d_t is correlated with the unobserved factors—as, for example, when assignment into the treatment groups is based on unobserved individual or community interest in educaton—then the estimate of α will be biased.

If the data set includes repeated observations of individuals, an alternative way to estimate the effect of the program using econometric analysis is to estimate equation 5 in terms of differences in the variables between the base period and some period, t, after the intervention has taken place.

$$R_{it} - R_{i0} = (\mathbf{X}_{it} - \mathbf{X}_{i0})\boldsymbol{\beta}_0 + d_t\alpha + U_{it} - U_{i0}. \tag{6}$$

Gender Differences in Enrollment Response

Parents may make different schooling investments in girls and boys either because costs differ or because parents obtain different benefits from educating girls and boys. Differences in costs may not just be differences in fees. Gender-specific costs may reflect differences in school access, the opportunity cost of a boy's or girl's time, or any disutility parents suffer because of cultural pressures against sending their daughters to school.

International experience indicates that returns in terms of proportional increases to household and wage productivity generally do not differ appreciably by gender even when average wages or patterns of labor participation do (Behrman and Deolalikar 1995; Schultz 1995). However, differences in how parents assess returns to schooling may also reflect how they benefit from the human capital of their daughters. Their assessment may reflect marriage and residency patterns as well as gender differences in remittances across generations. Nevertheless, Alderman and King (1998) indicate that for many purposes, including that of the present study, reduced-form models of investment in schooling are identical, regardless of whether the gender difference in education reflects differences in costs or differences in how parents value the gains from schooling boys and girls.

The girls fellowship program raises the cost of schooling in terms of fees (rupees) but lowers it in terms of travel time. In addition, it may also lower the cultural disutility of educating girls, which has the same impact as lowering the rupee price. Since all of the preexisting schooling options are still available to the family, the net impact of adding the new choice should be an increase in schooling for girls.[5]

The impact of the fellowship program on boys' enrollment is, however, ambiguous. Lowering the price of girls' schooling may lead to a substitution of girls' time for boys' time, with girls spending more time outside the home and boys devoting more time to home production. But there are at least two reasons why the fellowship program may have a positive impact on boys' schooling. First, the program creates a new low-priced private school that can accept boys. Second, boys' education may increase as their sisters go to school for a very practical reason: parents may want their sons to escort their sisters to and from school. Thus boys' and girls' education may be complementary goods.

If factors that affect enrollment, such as income, the cost of schooling, and the disutility of sending girls to school, differ systematically across treatment and control neighborhoods, they may affect the apparent impact of the program. Thus the econometric specification of equations 5 and 6 includes three measures of schooling costs: the fees charged in the preexisting neighborhood schools; av-

5. Becker (1981: ch. 6) develops a model of human capital investment in children that highlights parents' incentives to equalize wealth among their children. His model suggests that neutral parents invest more in less fortunate children so that all of their children are equally well off. Of course, there is considerable evidence that parents favor boys over girls. Kim, Alderman, and Orazem (1998) analyze the impact of reducing the price of girls' schooling on boys' and girls' enrollment when parents favor boys over girls at equal schooling prices.

erage distance to schools, measured in minutes of travel time (a proxy for transport costs); and the child's age, which may determine the opportunity cost of the child's time. The specification also includes the father's and mother's education, which are assumed to influence how a household values returns to schooling. In addition, the preference for education may depend on the child's birth order (parents may have a preference for educating the eldest child, particularly the eldest boy) and on citizenship (refugees may have different expectations of returns to education).[6] These variables, along with household income and the child's gender, make up the vector of exogenous variables we use in our analysis.

III. Data

Only a modest number of treatment groups was available because only 10 pilot sites were initially funded. For political expediency the government opted to place one neighborhood school in each of 10 urban slum areas of Quetta, ensuring that all major ethnic groups received at least one school. To accommodate this plan, the sample includes a degree of stratification, under which randomization is based on neighborhoods within each slum area.

A second problem is that no recent census of the population had been conducted from which to define treatment and control populations. The most recent census was 14 years old. The population of Quetta is estimated to have grown at about 7 percent a year since then, mainly within the neighborhoods that made up the target population. Consequently, the designers of the program chose an area frame sampling strategy to define the treatment and control neighborhoods.

They designed the area frame as follows. On a map of Quetta they outlined each of the 10 slum areas, selecting three sites in each, literally points on the map. They then randomly chose one of these areas to be the treatment neighborhood for the creation of a private school. Because all neighborhoods selected to participate accepted the invitation, the issue of self-selection was moot. The other neighborhoods became controls. The only criterion for the treatment neighborhood was that it could not already have a government girls' school. Although the control sites could have had a government girls' school, none of them did.

Given the small number of pilot sites, it is useful to see if the treatment and control populations differ in the factors that might also cause enrollment outcomes to differ. Such tests also help to determine which household characteristics contribute to program participation.

The baseline data collected in the treatment and control sites include information on households' socioeconomic characteristics, parents' education, and the educational attainment and current enrollment status of all children in the household. All of the households in treatment neighborhoods were surveyed in the summer of 1994, when the scholarship program was being promoted and before

6. Intrahousehold allocation of schooling is discussed in Parish and Willis (1993) and Butcher and Case (1994).

any fellowship schools were opened. The baseline survey of households in the control neighborhoods was conducted in July 1995. Because most of the data on household socioeconomic status do not change over a short period—no major economic transformation occurred in 1994 or 1995—the difference in the timing of the surveys should not be problematic. Information on the enrollment status of children in control neighborhoods was obtained for 1995 and retrospectively for the previous year.[7] Enrollment data were subsequently collected in 1996 in both treatment and control neighborhoods. The Balochistan Education Management Information System supervised all data collection and training of surveyors to ensure comparability of the data.

In this article we measure the program's impact using the estimators described in equations 1–6. By doing so, we can ascertain whether the results are robust. Moreover, we apply each of these estimators in two ways. First, we measure the change in enrollment for children in the target range of five to eight years. Second, we measure enrollment rates longitudinally for children ages four to seven in the initial year of the fellowship program.

IV. Results

The treatment sample includes 1,310 children, 781 girls and 529 boys (table 1). The control sample includes 1,358 children, 697 girls and 661 boys. The dependent variable in table 1 is a dummy variable that equals 1 if the child was enrolled in school. The other variables are exogenous variables believed to affect parents' enrollment choices for their children. Most of the variables come directly from the questionnaire. However, distance to school and annual fees are neighborhood averages of the children in school. We estimate household income using the number of adults in the household, their educational attainment, and a set of household assets. Details on the estimate of household income are contained in the appendix.

We test for statistical significance of the differences between the treatment and the control groups in two ways.[8] First, in order to determine if the randomization yielded observationally equivalent treatment and control populations, we test for the equality of means of the endogenous and exogenous variables. Second, we estimate enrollment equations using the baseline data. These equations test the null hypothesis of the equality of behavioral coefficients in the enrollment choice models for the treatment and control neighborhoods.

7. Collecting data retrospectively raises the possibility of recall bias, although parents should be able to remember whether their children were in school a year earlier. To verify this, we use multiple methods to evaluate the change in enrollment in the treatment neighborhoods. We find that the conclusions are not sensitive to differences in evaluation method.

8. Newman, Rawlings, and Gertler (1994) point out that researchers rarely test for statistical significance of the differences, so that probabilities of receiving a program may not be equal for individuals or communities in many of the evaluation studies in developing countries, especially those with few observations in the treatment group.

Table 1. *Summary Statistics of Baseline Data and Tests of the Equality of Means between Treatment and Control Groups*

	Girls			*Boys*		
Variable	*Treatment*	*Control*	*t-value*[a]	*Treatment*	*Control*	*t-value*[a]
Endogenous						
Enrollment rate	0.366	0.300	2.67	0.486	0.398	3.03
	(0.482)	(0.459)	[1,468]	(0.500)	(0.490)	[1,180]
Exogenous						
Household income	7,108	6,808	1.03	7,005	6,592	1.41
	(7,157)	(3,011)	[1,476]	(6,815)	(2,847)	[1,188]
Age	6.026	6.001	0.19	6.040	6.003	0.44
	(1.403)	(1.429)	[1,476]	(1.426)	(1.444)	[1,188]
Mother's highest grade	0.619	0.395	2.08	0.623	0.414	1.74
	(2.243)	(1.844)	[1,466]	(2.208)	(1.918)	[1,183]
Father's highest grade	3.405	3.079	1.27	3.635	2.723	3.38
	(4.745)	(4.882)	[1,417]	(4.579)	(4.548)	[1,162]
Birth order	2.832	3.004	2.21	3.074	2.965	1.27
	(1.474)	(1.510)	[1,476]	(1.447)	(1.482)	[1,188]
Citizenship	0.868	0.835	1.79	0.877	0.814	2.98
	(0.339)	(0.371)	[1,476]	(0.329)	(0.389)	[1,188]
Distance to school	17.77	17.81	0.05	16.93	16.42	0.62
	(9.443)	(9.991)	[491]	(9.338)	(9.394)	[515]
Annual fees	244.3	187.0	1.19	531.3	391.7	1.73
	(536.0)	(502.5)	[480]	(1,036.8)	(765.1)	[505]
Joint test[b]			9.0			27.2
Number of observations	781	697		529	661	

Note: Sample includes girls and boys ages four to eight in the baseline year. The baseline data were collected in 1994 for the treatment group and in 1995 for the control group. The numbers in parentheses are the standard deviations corrected for cluster effects using Huber's method. The numbers in square brackets are the degrees of freedom. The degrees of freedom differ because of missing information in the surveys.

a. The null hypothesis is that the mean of the variable in the treatment group is equal to that in the control group. If the z-value is smaller than 1.96, the null hypothesis cannot be rejected at the 5 percent significance level.

b. Corrected F-statistics with degrees of freedom (8, 1,376) for girls and (8, 1,149) for boys. The null hypothesis is that the means of the eight exogenous variables are jointly equal across the treatment and control neighborhoods. For both boys and girls the test statistic exceeds the critical value of 1.94 at the 5 percent significance level.

Source: Authors' calculations.

Baseline enrollment rates for both sexes are significantly higher in the treatment group than in the control group (columns 4 and 7 of table 1).[9] In addition, birth order and mother's education differ significantly between girls in the treatment and control neighborhoods, although the differences in means are small numerically. For boys, citizenship and father's education are significantly higher

9. It is unclear why girls' enrollment rates are 6 percentage points higher in the treatment neighborhoods, although we do not believe that the 10 fellowship school sites were strategically selected. Of the girls in school, 39 percent attended private school and 61 percent attended government boys' schools. The large proportion in private school is not unusual, especially given the limited availability of government schools. Alderman, Orazem, and Paterno (1996) also find that poor households in Lahore, Pakistan, used private schools extensively.

Table 2. *Baseline Probit Analysis of the Probability of Enrollment*

	Girls and boys			*Girls*			*Boys*		
Variable	*Treatment*	*Control*	*Difference*[a]	*Treatment*	*Control*	*Difference*[a]	*Treatment*	*Control*	*Difference*[a]
Household income per 10,000 rupees	0.138	0.422	2.03	0.171	0.572	2.58	0.037	0.218	0.25
	(2.362)	(2.879)		(2.377)	(2.870)		(0.346)	(0.954)	
Age	1.820	2.235	0.09	1.611	2.623	0.02	2.176	1.927	0.14
	(5.226)	(6.323)		(3.612)	(4.864)		(3.674)	(3.986)	
Age squared	–0.101	–0.140	0.89	–0.089	–0.174	0.47	–0.119	–0.119	0.99
	(3.621)	(5.014)		(2.508)	(4.127)		(2.513)	(2.884)	
Mother's highest grade	0.051	0.094	1.50	0.067	0.118	0.73	0.007	0.072	1.95
	(2.443)	(3.422)		(2.500)	(2.649)		(0.197)	(1.963)	
Father's highest grade	0.023	0.065	7.68	0.027	0.084	6.88	0.025	0.050	1.09
	(2.369)	(6.634)		(2.271)	(5.997)		(1.498)	(3.500)	
Birth order	–0.029	–0.036	0.07	–0.017	–0.020	0.00	–0.036	–0.053	0.04
	(0.918)	(1.214)		(0.416)	(0.461)		(0.717)	(1.251)	
Citizenship	0.693	0.335	1.20	0.628	0.214	1.36	0.762	0.538	0.00
	(5.207)	(2.556)		(3.590)	(1.079)		(3.545)	(2.888)	
Girl	–0.419	–0.541	0.05						
	(4.878)	(5.340)							
Number of observations	1,231	1,324		725	677		506	647	
Joint test[b]			29.9			23.3			13.7
Pseudo R^2	0.277	0.295		0.230	0.331		0.358	0.293	

Note: The numbers shown in parentheses are z-values corrected for cluster effects using Huber's method.

a. Test of the difference in coefficients between treatment and control neighborhoods, corrected for cluster effects using Huber's method.

b. Joint chi-square test of the null hypothesis of equality of coefficients across treatment and control neighborhoods. All results reject the null hypothesis of equality.

Source: Authors' calculations.

in the treatment group, although, again, the differences in means are small numerically. The joint test that the means of the exogenous variables are equal across all variables is easily rejected for both boys and girls. Therefore, we can reach a statistical conclusion that the treatment and control samples are not identical, a problem that we address in the analysis below.

A second way in which the treatment and control neighborhoods may differ is in parents' decisionmaking processes. To check this, we estimate a probit model of school enrollment based on equations 8–9 (table 2). The estimated parameters for the control and treatment groups exhibit the same signs and are qualitatively similar to results obtained in other studies of enrollment. The coefficient on household income is positive in both samples. Mother's and father's educational attainment positively influence their children's enrollment. Enrollment increases with age, but at a diminishing rate. First-born children have a higher probability of enrolling than their younger siblings, but the coefficient is not significant. Native Pakistanis also have a higher probability of enrolling than noncitizens. After pooling the treatment and control data, we can also estimate the effects of the average distance to school and average annual fees. Both have negative coefficients, except for a positive but insignificant effect of annual fees on boys' schooling.

The coefficients for the two groups are not statistically different, except for father's educational level in the girls' enrollment equation.[10] This result suggests that parents' decisions about education are similar in the treatment and control neighborhoods. Despite significant differences in characteristics between the control and treatment groups (as reported in table 1), we can still measure the change in enrollment due to the program by measuring the difference in enrollment rates between treatment and control groups, holding constant the differences in the exogenous variables.

Comparisons of Mean Enrollment Rates

Using enrollment rates for boys and girls before and after the program intervention (table 3), we can apply the three methods based on equations 1–3 (table 4). We report both age-specific and cohort-specific effects. The age-specific analysis looks at the enrollment of children ages five to eight in a specific year, while the cohort-specific analysis follows the enrollment of a fixed group of children ages four to seven in 1994.[11]

The age-specific results are similar for the three methods. All imply that the fellowship program had a positive effect on the enrollment of girls in the target age group as well as on that of boys. Applying the same methods to two years of data yields even larger estimates of the enrollment effects.

10. See Kim, Alderman, and Orazem (1998) for details of the statistical tests.

11. The cohort-specific enrollment rates in 1994 are lower than the 1994 average for the age-specific enrollment rates. The reason is that the age-specific groups are on average one year older in 1994. By 1996 the enrollment rates in the cohort-specific groups are higher than in the age-specific groups because, by then, the cohort-specific groups are on average one year older than the age-specific groups.

Table 3. *Enrollment Rates Before and After the Program*
(percent)

	Age-specific				*Cohort-specific*			
	Treatment		*Control*		*Treatment*		*Control*	
Outcome measure	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Enrollment rate before program (E_0)	56.33	45.29	51.06	34.86	38.75	34.06	36.55	29.03
Enrollment rate in 1995 (E_{95})	64.29	63.93	49.68	38.37	64.29	63.93	49.68	38.37
Enrollment rate in 1996 (E_{96})	76.15	71.30	43.50	36.20	85.50	78.36	59.87	45.97

Note: The age-specific analysis records the enrollment of children ages five to eight in the specified year, while the cohort-specific analysis follows the enrollment over time of a fixed group of children ages four to seven in the base year.

Source: Authors' calculations.

Table 4. *Age- and Cohort-Specific Effects of the Fellowship Program on Enrollment Rates*

Estimation method	*Mathematical expression*	*Age-specific*		*Cohort-specific*	
		Boys	*Girls*	*Boys*	*Girls*
Measure of effect using means					
Reflexive (1994–95)	$E^R(\alpha) = E(R_{Tt}) - E(R_{T0})$.	8.0	18.6	25.5	29.9
		(0.42)	(0.44)	(0.43)	(0.44)
Reflexive (1994–96)	$E^R(\alpha) = E(R_{Tt}) - E(R_{T0})$.	19.8	26.0	46.8	44.3
		(0.51)	(0.53)	(0.52)	(0.54)
Difference-in-differences (1994–95)	$E^D(\alpha) = [E(R_{Tt}) - E(R_{Nt})] - [E(R_{T0}) - E(R_{N0})]$.	9.3	15.1	12.4	20.5
		(0.53)	(0.54)	(0.54)	(0.54)
Difference-in-differences (1994–96)	$E^D(\alpha) = [E(R_{Tt}) - E(R_{Nt})] - [E(R_{T0}) - E(R_{N0})]$.	27.4	24.8	23.4	27.4
		(0.73)	(0.70)	(0.74)	(0.71)
Mean-difference (1994–95)	$E^M(\alpha) = [E(R_{Tt}) - E(R_{Nt})]$.	14.6	25.6	14.6	25.6
		(0.65)	(0.67)	(0.65)	(0.67)
Mean-difference (1994–96)	$E^M(\alpha) = [E(R_{Tt}) - E(R_{Nt})]$.	32.7	35.1	25.6	32.4
		(0.59)	(0.65)	(0.60)	(0.66)
Measure of effect using regression					
Covariate post-test (1995 cross-sectional)	$R_{it} = \mathbf{X}_{it}\boldsymbol{\beta}_t + d_f\alpha_t + U_{it}$	22.4	33.4	22.4	33.4
		(0.04)	(0.03)	(0.04)	(0.03)
Covariate post-test (1996 cross-sectional)	$R_{jt} = \mathbf{X}_{jt}\boldsymbol{\beta}_t + d_f\alpha_t + U_{jt}$	38.4	42.7	26.8	39.9
		(0.07)	(0.05)	(0.05)	(0.04)
First-difference, time-invariant β (1994–95)	$R_{it} - R_{i0} = d_f\alpha_t + U_{it} - U_{i0}$			29.2	36.7
				(0.08)	(0.07)
First-difference, time-invariant β (1994–96)	$R_{it} - R_{i0} = d\alpha_t + U_{it} - U_{i0}$			8.8	26.4
				(0.10)	(0.08)
First-difference, time-varying β (1994–95)	$R_{it} - R_{i0} = \mathbf{X}_i(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}) + d_t\alpha_t + U_{it} - U_{i0}$			42.8	46.9
				(0.11)	(0.08)
First-difference, time-varying β (1994–96)	$R_{it} - R_{i0} = \mathbf{X}_i(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-2}) + d_t\alpha_t + U_{it} - U_{i0}$			24.2	28.1
				(0.14)	(0.10)

Note: Numbers in parentheses are standard errors corrected for cluster effects using Huber's method.
Source: Authors' calculations.

The cohort-specific analysis has the advantage of enabling us to control for unobservable effects that are specific to the individual and that might also be correlated with program outcomes. However, because enrollment increases with age, at least initially, some of the enrollment growth in the cohort-specific analysis will reflect a maturity effect. This effect will bias the reflexive method estimates upward. Indeed, the implied 46.8 percent increase in boys' enrollment, and 44.3 percent increase in girls' enrollment between 1994 and 1996 (the last two columns of table 4) are much higher than the corresponding estimates in the age-specific analysis.

The estimates generated from the difference-in-differences and mean-difference methods eliminate the maturity effect by assuming that it is common across neighborhoods. Consequently, the measured program effects using these methods are smaller than the reflexive estimates and are more comparable to the estimates generated with the age-specific sample. All of the results show large gains in both boys' and girls' enrollment following the opening of the fellowship schools. Most estimates show slightly higher enrollment gains for girls than for boys. Looking across the age-specific and cohort-specific estimates, we can conclude that girls' enrollment rose 25–35 percent as a result of the program and that boys' enrollment rose a few percentage points less.

Comparisons Using Regression Analysis

Because the treatment and control neighborhoods have different characteristics that are believed to affect parents' educational choices, a simple comparison of unconditional means could yield biased estimates of the program effect. We use an alternative method based on equation 5 on the same samples.[12] The program effects based on the covariate post-test using cross-sectional data are reported in table 4; the full probit regression results are in table 5. The enrollment rate in fellowship neighborhoods rose 33.4 percent for girls and 22.4 percent for boys in the first year of the program (table 4). After two years enrollment in the fellowship neighborhoods had risen 42.7 percent for girls and 38.4 percent for boys using the age-specific analysis. The gain was slightly less using the cohort analysis. These results are consistent with the results based on community means.

Considering that the fellowship schools were established in February 1995 and that survey data were collected in July of that year, the response of parents in target areas was nearly instantaneous. This supports the view that there was excess demand for primary education in these poor areas. Moreover, the fellowship program grew more successful year by year. For girls the estimated program effect increased almost 10 percent in 1996 relative to the effect in 1995. Boys' enrollment rates grew 16 percent in the year after implementation.

Another possible source of bias in our estimates of the program's effect is unobserved heterogeneity in children that is correlated with the program's out-

12. We cannot use first-difference methods for the age-specific analysis because enrollment decisions for younger cohorts can be observed only after the fellowship schools are in existence.

Table 5. *Post-Test Probit Analysis of Probability of Enrollment Using Cross-Sectional Data*

	1995[a]		*1996, cohort-specific*[b]		*1996, age-specific*[c]	
Variable	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*
Treatment dummy	0.334	0.224	0.399	0.268	0.427	0.384
	(10.148)	(5.143)	(9.679)	(5.511)	(8.488)	(5.495)
Household income per 10,000 rupees	−0.001	−0.003	0.012	0.072	0.034	0.128
	(0.022)	(0.080)	(0.333)	(1.513)	(0.724)	(1.872)
Age	0.141	0.276	0.229	0.936	0.615	1.330
	(0.652)	(1.197)	(0.797)	(3.416)	(1.970)	(3.925)
Age squared	−0.008	−0.016	−0.011	−0.057	−0.036	−0.083
	(0.496)	(0.890)	(0.570)	(3.113)	(1.546)	(3.268)
Mother's highest grade	0.016	0.030	0.029	0.011	0.027	0.018
	(0.040)	(2.330)	(1.505)	(0.867)	(1.822)	(1.231)
Father's highest grade	0.013	0.003	0.030	0.011	0.035	0.020
	(3.383)	(0.707)	(6.293)	(2.433)	(6.656)	(3.523)
Birth order	−0.008	−0.026	−0.016	−0.020	−0.0002	−0.031
	(0.720)	(2.042)	(1.214)	(1.516)	(0.016)	(1.904)
Citizenship	0.152	0.225	0.143	0.201	0.187	0.173
	(3.040)	(4.362)	(2.374)	(3.501)	(2.783)	(2.465)
Distance to school	−0.008	0.003	−0.029	−0.027	−0.035	−0.036
	(1.074)	(0.358)	(3.190)	(2.347)	(3.361)	(2.511)
Annual fees per 1,000 rupees	−0.443	−0.030	−0.170	−0.362	−0.316	−0.618
	(3.640)	(0.241)	(1.088)	(2.535)	(1.719)	(2.723)
Number of observations	1,031	830	845	700	764	650
Pseudo R^2	0.141	0.100	0.312	0.215	0.350	0.380

Note: The coefficients reported here are dF / dX, where F is the dependent variable and X is the independent variable, not actual coefficients. Since the dependent variable is a discrete variable, dF / dX is not identical to actual coefficients. The numbers shown in parentheses are z-values corrected for cluster effects using Huber's method. Dummy variables for each neighborhood are included.

a. Children are ages five to eight in 1995. Dependent variable is enrollment status in 1995.

b. Children are ages five to eight in 1995. Dependent variable is enrollment status in 1996.

c. Children are ages five to eight in 1996. Dependent variable is enrollment status in 1996.

Source: Authors' calculations.

come. If cross-sectional differences in individual fixed effects are contributing to measured program effects, then we can remove the fixed effects by differencing the dependent variable.

We conduct the first-difference analysis under the assumption that the coefficients of the regressors are time-invariant, as in equation 6 (table 6). The dependent variable is the change in enrollment status before and after implementation of the program. The coefficient on the treatment dummy measures the effect of the program on enrollment choice. The last two specifications of the first-difference analysis allow the coefficients on the individual and neighborhood effects to vary over time.

The results of these tests corroborate the results presented above in that the coefficient representing the program effect is significantly positive and larger for girls than for boys. However, the estimated program effect is larger after one

Table 6. *First-Difference Probits for the Change in Enrollment Decision*

	1994–95		*1994–96*		*1994–1995*		*1994–96*	
Variable	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*
Treatment dummy	0.367	0.292	0.264	0.088	0.469	0.428	0.281	0.242
	(5.518)	(3.591)	(3.165)	(0.909)	(5.833)	(3.755)	(2.931)	(1.723)
Change in age squared	–0.077	–0.082	–0.047	–0.046	–0.071	0.032	0.079	0.073
	(4.785)	(4.447)	(5.006)	(4.502)	(0.343)	(0.137)	(0.641)	(1.323)
Age in 1994 squared					–0.001	–0.022	–0.047	–0.080
					(0.040)	(0.525)	(1.055)	(1.686)
Income per 10,000 rupees					–0.151	–0.009	–0.009	–0.005
					(2.680)	(0.122)	(1.309)	(0.588)
Mother's highest grade					–0.007	0.016	–0.009	–0.030
					(0.374)	(0.652)	(0.380)	(1.020)
Father's highest grade					0.004	–0.028	0.043	0.014
					(0.458)	(2.751)	(4.561)	(1.226)
Birth order					–0.029	–0.050	–0.008	–0.021
					(1.152)	(1.667)	(0.250)	(0.616)
Citizenship					0.006	0.093	0.243	0.212
					(0.047)	(0.717)	(1.515)	(1.318)
Distance to school					–0.001	0.027	–0.054	0.029
					(0.051)	(1.407)	(1.936)	(1.272)
Annual fees per 1,000 rupees					–0.755	–0.103	–0.588	0.765
					(2.424)	(0.299)	(1.623)	(1.813)
Number of observations					1,055	861	863	725
Pseudo R^2					0.04	0.04	0.09	0.05

Note: The coefficients reported here are dF / dX, not actual coefficients. Children in the sample were ages four to seven in 1994.
Source: Authors' calculations.

year than after two years, in contrast to the cross-sectional results. The reason for this discrepancy is unclear, although it may be related to the fact that first-difference regressions control for fixed effects. Enrollment rates were initially higher in the fellowship neighborhoods, and children who were in school before the fellowship schools opened do not contribute to the measured enrollment effect in the first-difference analysis.

Also, the opening of the fellowship schools may have encouraged parents to send their children to school at a younger age, and the smaller effect over time reflects the first-time enrollment of older children in the control neighborhoods. In fact, some of the later enrollment growth in control neighborhoods may have been related to the fellowship program if the promotion of children's education in fellowship neighborhoods spilled over to control neighborhoods. Nevertheless, the estimated two-year effect on enrollment growth is still large. Controlling for fixed effects lowers the estimated effect 12 to 30 percent, leaving the estimated enrollment impact at 24.2 percent for boys and 28.1 percent for girls.

Given the apparent success of the fellowship schools in increasing enrollment, the question remains as to how much the children are learning. Assessment efforts are in their infancy in Balochistan, but the Balochistan Education Management Information System did pilot an achievement test to several third-grade classes, including some from fellowship schools. The results are not definitive because the samples were small, but they show no significant differences in outcome between fellowship and government schools. Still, any concrete assessment of the relative quality of fellowship and government schools will require further tests.

V. Discussion

The fellowship program certainly increases enrollment. But is it cost-effective when compared with alternative policy options? One way to look at this question is in terms of the cost per student enrolled. In 1996 the recurrent cost per student in government primary schools in Balochistan was 2,500 rupees (World Bank 1997). This amount is nearly twice the subsidy per girl offered to the fellowship schools in the first year of operation and is far more than three times the subsidy per student.[13] The disparity is not due to cost recovery from students, since in the first year of operation students were asked to pay only between 10 and 25 rupees per month. The difference comes mainly from lower teacher salaries.

In addition to these recurrent costs, the Primary Education Department spent approximately 1,500 rupees per student in contracting a local NGO to help communities establish schools and train teachers. This amount also covered the monitoring of schools during their first two years. Since the fellowship program did

13. Ideally, we would want to evaluate the cost of educating a student until graduation, but the project is too new to ascertain this.

not build schools, these costs were its main start-up expenses. In the future NGO costs will be lower, since the cost of monitoring activities necessary to conduct this evaluation will not be needed in expanding the pilot program. Nevertheless, even the upper-bound estimated start-up cost of 1,500 rupees per student is far less than the estimated cost of 600,000 rupees for the construction of a government primary school, which typically has two classrooms. Assuming 50 students per classroom, this represents an initial investment of 6,000 rupees per student. Thus there is a substantial difference in the cost of establishing a government school relative to a fellowship school.

Instead of looking at the average cost per student, we can look at the marginal cost of increasing enrollment. We consider two alternative policies needed to match the enrollment increase that resulted from the fellowship program: income transfers to poor households and construction of new schools. Our estimates are based on estimated elasticities of enrollment choice with respect to income and distance to school.

Income has only a moderate impact on participation in the program. Consequently, the benefits of the program are not strongly skewed to upper-income households. The moderate income response also implies that it would take a sizable income transfer to achieve the same impact on enrollment as the program. In particular, the income response in our estimates implies that a direct subsidy of 3,471 rupees per household would be needed to raise the probability of girls' enrollment 25 percent (table 7). This is roughly 2.5 times the initial recurrent cost of a fellowship school—1,400 rupees a year per girl. As boys' enrollment is less income-sensitive, a similar increase in the probability of boys' enrollment would require an income transfer of 15,030 rupees, compared with the negligible marginal cost of increasing boys' enrollment through the girls' fellowship program.[14]

The overall impact of the fellowship program might also be influenced by the fact that it reduces the distance to schools. Unfortunately, in our sample there is insufficient variance in distance to schools to directly estimate this influence. Using the 1996 coefficient of distance for girls, –0.03 (see table 5), we estimate that the distance to private schools would have to be cut in half to increase enrollment by the same amount as the project. Halving the distance to schools in a two-dimensional environment implies a fourfold increase in the number of schools.

An additional question is whether the success of the fellowship program depended on the attributes of the neighborhoods in which the new schools were instituted. There are large differences in success rates across neighborhoods, with the increase in girls' enrollment varying from 8 to 67 percentage points. Increases in boys' enrollment also differ across neighborhoods, from a drop of 2 percent-

14. The result for Quetta is similar to results for both sexes in low-income neighborhoods of Lahore, where a 10 percent increase in household income causes a 1.2 percent increase in the enrollment rate in private schools (Alderman, Orazem, and Paterno 1996). Thus in Lahore, a city in which overall primary school enrollment rates are more than 90 percent, an income transfer of 14,808 rupees would be required to raise enrollment 25 percent for both sexes.

Table 7. *Estimates of Alternative Ways to Raise Enrollment to the Target Level*

	Elasticities		*Change required to meet target level (25 percent)*	
Alternatives	*Girls*	*Boys*	*Girls*	*Boys*
Direct subsidy to household	0.503	0.115	3,471 rupees per household (50 percent)	15,030 rupees per household (150 percent)
Decrease distance to school	0.320	0.732	13.48 minutes (78 percent)	5.71 minutes (34 percent)

Note: Children in the sample were ages 4 to 7. The numbers in parentheses are the amount as a percentage needed to meet target effect. For example, a direct subsidy to the household that leads to a 50 percent increase in household income may raise girls' enrollment rates 25 percent.

Source: Authors' calculations.

age points to an increase of 61 percentage points. Are the fellowship schools more successful in neighborhoods where households are not as poor, better educated, or unique in other observable ways? For the most part success appears to be unrelated to neighborhood attributes.[15] The attributes of neighborhoods with above-average enrollment gains do not differ much from those with below-average gains. For example, average parental education levels are similar in the most and least successful neighborhoods, and average income levels are actually lower in the more successful neighborhoods. One intriguing result is that neighborhoods with the largest increases in girls' enrollment also have the largest increases in boys' enrollment, which is consistent with our presumption that boys' and girls' schooling are complementary goods.

To the extent that the schools benefit some types of children more or less than others, it appears that enrollment of younger children rises more than that of older children. This is a natural consequence of the fact that older children are more likely to be at an age at which parents planned to remove them from school anyway. First-born children are more positively affected, presumably because parents favor the eldest child. Nevertheless, the joint test of uniform effects across all children cannot be rejected at standard significance levels, suggesting a high probability of success from expanding the program to other poor neighborhoods, regardless of residents' socioeconomic attributes.

A final concern is whether these schools are sustainable. Ideally, we would like to know if they will continue to operate for a generation or more. Unfortunately, we cannot answer this question given the time frame of our research. There are, however, encouraging indications that the program is viable. First, there is a demand for such schools from urban neighborhoods, including other cities in the province. The program has expanded from 11 schools with slightly more than 2,000 students to 40 schools with 10,000 students. All of the schools that opened in urban areas between 1995 and 1998 remain open. Moreover, the enrollment

15. This discussion is based on Kim, Alderman, and Orazem (1998: 21–22, app. 3).

of girls in the original schools covered in this study increased 15 percent between 1996 (the last year in our household sample) and 1997, despite a nominal increase in fees of 15 percent—a small real increase of about 3 percent.[16]

Still, the schools are not financially independent. The Balochistan Education Foundation provides about 20 percent of the subsidy given to these schools in their first year of operation. This commitment—financed from an endowment that is distinct from the Primary Education Department's budget—may be intended to address the issue of equity. There is no direct indication that it is needed to ensure the viability of the schools.

VI. Summary

We have used several evaluation methods based on experimental design to measure the effect of the Quetta Urban Fellowship Program on the enrollment of boys and girls in poor neighborhoods. Regardless of how the impact is measured, we find that the fellowship program raised enrollment for both boys and girls (see table 3). Most estimates show that the effect was larger for girls than for boys. We can conclude that the estimated program effects are robust to differences in assumptions about possible biases arising from measured and unmeasured differences between treatment and control neighborhoods.

Before the project was implemented, it was not clear whether girls' low enrollment rates were due to cultural barriers that cause parents to keep their daughters out of school or to an inadequate supply of girls' schools. The urban fellowship experiment provides strong evidence that subsidizing the establishment of primary schools for girls can sharply increase girls' enrollment. In addition, even though the fellowship was given only to girls, boys' enrollment in those neighborhoods also increased sharply. This suggests that boys' education and girls' education are complementary goods: by encouraging parents to send their girls to school, the program had collateral benefits of raising boys' enrollment rates.

The measured change over two years yields mixed evidence on whether the advantage for enrollment growth in fellowship neighborhoods relative to control neighborhoods continued to increase. However, even if the initial enrollment gain decreased in subsequent years, the increase in enrollment after two years was still around 25 percentage points. This is a substantial improvement over the baseline enrollment rate of 45 percent for girls who are five to eight years old. School success appears not to depend on neighborhood income or other observable socioeconomic variables, suggesting that expanding the program to other poor neighborhoods is also likely to be successful.

Future work will be required to assess the long-term effects of the fellowship program. In particular, the sustainability of the schools and the enrollment ef-

16. The school with the largest fee increase has the largest percentage increase in enrollment, perhaps reflecting the endogeneity of fee structures.

fects after the subsidies expire must be evaluated. The short-term success of the fellowship program does not guarantee long-term success when the financial burden of supporting the schools is fully borne by the neighborhoods. School outcomes must also be assessed. The ultimate success of the fellowship program depends on whether children attain literacy.

Appendix. Details on the Estimates of Household Income

It is difficult to derive income estimates for households in Pakistan. The relative importance of production for home consumption, informal labor market arrangements, barter trade, and other economic activity occurring outside formal markets complicate efforts to measure income. The budget for this project did not include resources sufficient to carry out a careful analysis of income for each household. However, the Pakistan Integrated Household Survey (PIHS) conducted such a detailed survey of household income and socioeconomic attributes in 1991. The PIHS allows us to predict household income based on a regression of income on easily observed household attributes. In this study we collected information on these attributes and then used the PIHS estimates to generate predicted income.

The PIHS income equation is reported in table A-1. The specification follows Alderman and Garcia (1996), who estimate income and expenditure equations for 217 households in a single district in Balochistan. Their estimates can serve as independent validation of the income estimates we derive from the PIHS data. They are less useful for our purpose than is the PIHS because their data are from 1986 and include only rural households. The PIHS has sufficient urban observations to estimate an income equation for urban households, and it is closer to our 1994 base period. The variables in the income equation include the number of adult men and women, the number of men and women with primary-, secondary-, and tertiary-level schooling, and the value of household assets. Alderman and Garcia find that this income specification generates predicted values that perform well in explaining household savings, loans, and nutrition status.

In general, the PIHS income estimates are sensible. Households with more capital assets, more human capital, and more adult males have higher incomes. The results correspond reasonably well in sign with those in Alderman and Garcia. More important, the two studies generate equivalent estimates of relative household income. The correlation in predicted income based on the PIHS compared with the Alderman-Garcia estimates is 0.82. The higher variance in income in the treatment neighborhoods is a result of three wealthy households residing in those neighborhoods. When those households are removed, the treatment and control neighborhoods have similar means and variances in estimated income.

Table A-1. *Income Equations*

Variable	*Alderman and Garcia*	*Pakistan Integrated Household Survey*
Intercept	5,999	3,303
	(2.61)	(4.64)
Number of males ages 16 or older	938	1,219
	(0.92)	(3.73)
Number of males ages 6–16	1,691	—
	(2.09)	
Number of females ages 16 or more	-709	-188
	(-0.54)	(-0.57)
Number of females ages 6–16	1,009	—
	(0.64)	
Number of children age 5 or younger	2,820	—
	(2.99)	
Number of males with primary schooling	6,140	-1,171
	(2.95)	(-2.55)
Number of males with secondary schooling	2,279	-364
	(1.69)	(-0.92)
Number of males with more than secondary schooling	6,435	147
	(1.41)	(0.96)
Number of females with primary schooling	6,707	-406
	(1.85)	(-0.69)
Number of females with middle schooling or more	7,758	889
	(1.35)	(3.68)
Rainfed land	110	n.a.
	(2.34)	
Irrigated land	665	n.a.
	(4.93)	
Acres of orchards	4,065	n.a.
	(2.57)	
Value of livestock	0.335	n.a.
	(1.05)	
Value of vehicles	0.171	0.012
	(8.55)	(2.48)
Value of machinery and tools	0.125	0.007
	(1.27)	(1.88)
R^2	0.747	0.03
Number of observations	217	2,112

— Not available.
n.a. Not applicable.
Note: Numbers in parentheses are standard errors.
Source: Authors' calculations.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Alderman, Harold, Jere Behrman, David Ross, and Richard Sabot. 1996. "Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy." *Journal of Human Resources* 32(1):229–54.

Alderman, Harold, and Marito Garcia. 1996. *Poverty, Household Food Security, and Nutrition in Rural Pakistan.* Research Report 96. Washington, D.C.: International Food Policy Research Institute.

Alderman, Harold, and Elizabeth King. 1998. "Gender Differences in Parental Investment in Education." *Structural Change and Economic Dynamics* 9(4):453–68.

Alderman, Harold, Peter Orazem, and Elizabeth M. Paterno. 1996. "School Quality, School Cost, and the Public/Private School Choices of Low-Income Households in Pakistan." Impact Evaluation of Education Reform Working Paper 2. World Bank, Development Research Group, Washington, D.C. Processed.

Becker, Gary S. 1981. *A Treatise on the Family.* Cambridge, Mass.: Harvard University Press.

Behrman, Jere, and Anil Deolalikar. 1995. "Are There Differential Returns to Schooling by Gender? The Case of Indonesian Labor Markets." *Oxford Bulletin of Economics and Statistics* 57(1):97–117.

Butcher, Kristen, and Anne Case. 1994. "The Effect of Sibling Sex Composition on Women's Education and Earnings." *Quarterly Journal of Economics* 109(3):531–62.

Grossman, Jean Baldwin. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *The World Bank Research Observer* 9(2):159–80.

Kim, Jooseop, Harold Alderman, and Peter Orazem. 1998. "Can Private School Subsidies Increase Schooling for the Poor? The Quetta Urban Fellowship Program." Impact Evaluation of Education Reforms Working Paper 11. World Bank, Development Research Group, Washington, D.C. Processed.

Newman, John, Laura Rawlings, and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *The World Bank Research Observer* 9(2):181–201.

Parish, William, and Robert Willis. 1993. "Daughters, Education, and Family Budgets: Taiwan Experiences." *Journal of Human Resources* 28(4):863–98.

Schultz, T. Paul. 1995. *Investment in Women's Human Capital.* Chicago: University of Chicago Press.

World Bank. 1997. "Pakistan: Toward a Strategy for Elementary Education." Report 16670-PAK. World Bank, South Asia Region, Washington, D.C. Processed.

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 467-91

Central Mandates and Local Incentives: The Colombia Education Voucher Program

Elizabeth M. King, Peter F. Orazem, and Darin Wohlgemuth

In decentralized education systems programs that promote central mandates may have to be devolved to local governments, communities, and providers. When participation by local governments and providers is voluntary rather than compulsory, the determinants of program placement are important in predicting potential benefits to individuals. This article analyzes incentives for municipalities and private schools to participate in Colombia's voucher program. It finds that the demand for secondary education relative to the capacity of public schools and the availability of spaces in private schools in the municipality were key predictors of municipal participation, whereas the number of underserved students had a nonlinear effect on participation. Schools whose educational quality was moderate and charged moderate tuition fees were the most likely to participate; the program was less attractive to schools whose quality and fees were high and to schools whose quality and fees were low.

The debate regarding the use of vouchers for private schools, whether in industrial or developing countries, centers on issues of equity and efficiency (see Levin 1991, 1992; Henig 1994; and CERI, OECD 1994 for a summary of the arguments). Proponents claim that vouchers provide the poor with a way out of overcrowded or low-quality schools by allowing them to enroll in private schools that they would otherwise not be able to attend. Furthermore, if competition for voucher students impels both private and public schools to improve, then vouchers make the delivery of education per public dollar spent more efficient.[1] Proponents also see vouchers as a means of transferring some control over educational resources from the central or local government to parents and students. In countries where educational policy decisions are heavily

1. Moreover, there is consistent evidence that children perform better in private schools than in public schools in developing countries, even when accounting for selection. See Jimenez, Lockheed, and Paqueo (1991) and Lockheed and Jimenez (1994) for reviews of this literature.

Elizabeth M. King is with the Development Research Group at the World Bank, and Peter F. Orazem and Darin Wohlgemuth are with the Department of Economics at Iowa State University. The authors' e-mail addresses are eking@worldbank.org, pfo@iastate.edu, and darinw@iastate.edu. The authors are grateful to Marybell Gutiérrez, Martha Laverde, Carlos Pardo, Laura Rawlings, Carlos Torres, and Fabio Sánchez Torres for assistance in locating and accessing data. They also thank the referees for numerous helpful comments and suggestions, Berk Ozler and Dan Levy for expert assistance in organizing the data, and Donna Otto and Diana McLaughlin for preparing the manuscript. This study was funded jointly by the World Bank's Development Research Group and Research Support Budget under RPO 679-18.

centralized, vouchers represent a dramatic shift in the locus of decisionmaking, with great potential to improve efficiency.

Opponents counter that vouchers rob public schools of much-needed resources, that vouchers are used by richer students who would have paid for private school anyway, and that private schools can pick and choose among students, leaving public schools with the worst students. Winkler and Rounds (1993) find that parents select private schools based on the characteristics of the students already enrolled. As a result private schools tend to have a continuing advantage in attracting students with higher socioeconomic status. Opponents also counter that unless an efficient information system is developed to publicize the relative quality of different private schools, parents and students will make uninformed choices—a situation not likely to induce schools to improve their performance.

These efficiency issues are important for assessing the role of vouchers in industrial countries; however, they may be only secondarily so in developing countries, where the education choice is not whether to attend public school or private school, but whether or not to attend school at all, given the absence of a nearby school. Expanding access to any school—not widening the choice of school—is probably the best argument in favor of vouchers in these countries. Partnerships with the private sector, which could lead to huge gains in enrollment, are a pragmatic response to strained or inadequate public school capacity, especially in poor rural areas or in overpopulated cities. High population growth rates will only exacerbate the problem over time. Without private schools, many children will not be able to attend school at all. Indeed, the proportion of secondary students in private schools in developing countries is roughly double the median in industrial economies (James 1993).

Previous research on vouchers has concentrated almost exclusively on evaluating the benefits of giving parents and students greater choice by comparing students' performance in public and private schools. Witte's (1996) review of the literature contends that there is no evidence of increased achievement in private schools, once we control for private schools' nonrandom selection of students with potentially higher abilities. In fact, Hoxby (1996), simulating the effect of vouchers on student achievement, argues that vouchers would improve test scores averaged across all schools, but would lower test scores in private schools. Average scores would rise because of improved public school efficiency, necessitated by greater competition for students. However, Murnane, Newstead, and Olsen (1985), Sander (1996), Neal (1997), Sander and Krautmann (1995), and Witte (1996) find that attending Catholic school improves student achievement, even after controlling for selection, and raises the probability of attending college. Hoxby's (1996) and Lankford and Wyckoff's (1992) simulations based on school choice models predict that vouchers would increase private school enrollment, but would not affect private school students who were poor or whose parents had less education.

The evaluation of Chile's voucher program, one of the oldest voucher programs in the developing world, has been the subject of debate as well. Winkler

and Rounds (1996) and Gauri (1998) find that students with higher socioeconomic status were significantly more likely to enroll in better-performing schools than were students from lower-income families. Rodríguez (1988, as cited by Gauri) concludes that private subsidized or voucher schools in Santiago outperformed the public centralized schools that still existed at that time. Parry (1993, as cited by Gauri) regresses schools' average scores in a national standardized test on average characteristics of the students and finds that, although municipal schools scored higher than voucher schools when parents had relatively little education, voucher schools scored higher when parents were relatively more educated.

This research, however, generally ignores the supply response of providers, assuming that students are able to find space in a private school with attributes equal to those of the average private school. In reality, private schools may refuse to participate in a voucher program, as may local school boards, if participation is voluntary. Hence, the subset of participating providers may differ significantly from the full population of private schools. The potential benefits of the program, whether in terms of enrollment or student learning, will then depend on the willingness of local communities and schools to participate.[2]

In this article we examine the extent to which a central government is able to attain a national goal of expanding enrollment, especially among the poor, even as it transfers power to local governments and private providers. Voluntary program participation and cost-sharing schemes between the central government and local governments are gaining currency in developing countries and are consistent with shifts toward more decentralized delivery systems. We study Colombia's national voucher program for secondary education, which the government launched in late 1991 and terminated in 1997.

True to Colombia's broad decentralization reforms of the early 1990s (Montenegro 1995; Hanson 1995), municipalities and schools participated in this voucher program voluntarily, and municipalities assumed part of the cost of the program. Two key assumptions of the program were that public schools, especially those in large cities, were overcrowded and that private schools had excess capacity. Vouchers for use in private schools were viewed as a means to expand enrollment in secondary schools at relatively low cost, while reducing the

2. The absence of a broad-based voucher program has made it impossible to estimate the supply response of private schools and municipalities to such a system. Witte (1996: 170) argues, "Nearly all quantitative estimates of [voucher effects on] school selection and the effects of school choice on performance are based on extrapolations from the current system of education. However, a broad based voucher system . . . might create such a different market for education that estimates based on the current arrangements would be meaningless."

To develop reasonable estimates of the effects of a voucher program, it is important to observe how schools respond to the existence of a voucher program, but "current experiments with vouchers . . . are simply too small to provide evidence of market reactions by either the public or private sectors" (Witte 1996: 172). Witte contends that the crucial information on how schools respond to vouchers will be available only when a large-scale voucher program is implemented. Until then, researchers will not be able to establish the validity of simulations based on existing schools.

enrollment pressure on public schools.[3] In addition, encouraging more private provision without directly subsidizing specific schools was seen as a way to elicit better performance from private schools. Vouchers could also improve the quality of public schools by reducing overcrowding, thereby easing the pressure on resources in public schools.

We examine whether the decision of municipalities and private schools to participate in Columbia's voucher program was consistent with stated national objectives. In particular, did municipalities' excess demand for secondary education or private schools' desire to reach poorer households significantly influence participation?

I. Colombia's Voucher Program

Colombia launched its national voucher program near the end of 1991 as part of a broader transformation, begun in the late 1980s, to decentralize the organization and management of its education system. Hanson (1995) traces the impetus for decentralization to the growing awareness among Colombia's elite of the need to establish the legitimacy of the government and its institutions as a means to deal with the country's increasing violence. The strategy for establishing legitimacy hinged on giving people participatory control over public institutions. By the late 1980s several government institutions, including the educational system, had begun to decentralize. The new constitution of 1991 codified and integrated these initiatives.

Together with other reforms, the voucher program was meant to address deficiencies in the performance of the public education system, especially the low transition rate from primary to secondary school among the poor. In 1992 only 51 percent of 13–19-year-old urban youths belonging to the poorest quintile were enrolled in school, compared with 75 percent of those in the richest quintile (World Bank 1994). The shortage of space in public schools, especially in large cities, where demand was thought to be greatest, was seen as a real problem.

The Ministry of Education initially targeted the country's 10 largest cities for participation in the program. Adoption was ultimately voluntary, although the Ministry may have pressured cities to join. In 1991 the Ministry of Education held a meeting with the heads of departments (states) to announce the program. The meeting elicited subsequent letters of intention to participate from departments and municipalities, as well as statements about which municipalities were not able or likely to join. The program grew nationally and during its most active year, 1995, had 217 participating municipalities in 27 of the country's 30 departments. In 1995 the government awarded about 90,000 vouchers to students in 1,800 private schools. Voucher students made up 8 percent of all students in

3. Limited evidence supports the conclusion that vouchers raise enrollment. Ribero and Tenjo (1997) find that in Bogotá, where there were too few vouchers to meet demand, enrollment rates for students who received vouchers were 12 percentage points higher than for students who qualified for but did not receive vouchers.

private secondary schools. Student applicants had to have graduated from primary school (completed fifth grade) and been admitted to a participating private school. Awardees could renew their vouchers in the subsequent year only if they were promoted to the next grade. Dropouts and repeaters automatically lost their vouchers.

Although participation in the program was widespread, most of the vouchers were issued in large urban areas, where private schools are concentrated. Ten departments absorbed more than 70 percent of all vouchers issued, with the capital city of Bogotá alone taking 13 percent.[4] Participation was contingent on the municipality's willingness to cofinance and administer the program. The municipality provided 20 percent of the funds for the vouchers issued in its area, and the central government provided the remaining 80 percent.[5]

By design, Colombia's program avoided two common criticisms of voucher programs. First, the program did not threaten the resources available to existing public schools. The government assured public schools that current levels of funding would not decrease. This promise eliminated competition over finances, although public and private schools still had to compete for voucher students, especially because municipalities were able or willing to fund only a limited number of vouchers. In fact, each year the demand for vouchers exceeded the supply in nearly all participating municipalities. Conversations with officers from the administering agency, the Colombian Institute for Education Credit and Training Abroad (ICETEX), suggested that anywhere from 20 percent (in the department of Atlántico) to 90 percent (in Antioquia) of qualifying applicants received vouchers. In many cases where supply exceeded demand, a lottery was used to select beneficiaries. Because this lottery randomized the selection of students for the program, it provided a valuable mechanism for assessing the program's impact on individual students. Angrist and others (1999) examine this.

Second, only the poor qualified for vouchers, countering the claim that vouchers amount to a net subsidy for the wealthy at the expense of the poor. The targeting criterion used was based on a neighborhood stratification scheme that ranks neighborhoods on a scale of 1–6, from poorest to richest. A national poverty map, derived from five poverty indicators and used to distribute other transfers, established the socioeconomic status of different neighborhoods.[6] Only stu-

4. See King and others (1997) for more details about the program.

5. The agreement included additional conditions that did not relate directly to the program but were part of the decentralization reform. Municipalities had to agree to the terms of Law 160, which transferred responsibility for maintaining public schools from the central to the municipal government, and had to maintain a system of accounts that satisfied nationally prescribed standards.

6. The poverty map index, *Necesidades Básicas Insatisfechas,* was computed on the basis of five indicators: the proportion of households living in inadequate homes (such as homes without walls), the proportion of households without an adequate water supply or sanitation services, the proportion of households living in overcrowded quarters (defined as an average of more than three people per room), the proportion of households with high economic dependency (defined using the ratio of all household members to employed household members and the educational attainment of the household head), and the proportion of households with children between the ages of 6 and 12 years who were not enrolled in school (Colombia, Department of National Planning 1994).

dents residing in neighborhoods ranked 1 or 2 were eligible to receive a voucher. Morales-Cobo (1993) and Ribero and Tenjo (1997) conclude that the program's geographical targeting mechanism was accurate in delivering vouchers to poor students.[7] In order to establish a student's socioeconomic status, and thus eligibility, the program required that each applicant present a national identification card or utility bills to verify residence.

The voucher covered the cost of tuition—the yearly entrance fee plus monthly fees—for students in sixth to eleventh grade, subject to an upper limit. ICETEX, which administered the program for the Ministry of Education, set the maximum value of the voucher each year. In 1995 the voucher was worth a maximum value of Col$145,307 (Colombian pesos) or about $180 (U.S. dollars), with the actual value of each voucher depending on the prevailing tuition fee in the school in which the voucher was to be used.[8] This upper limit met or exceeded the annual fees of lower-priced schools, but covered less than half of the cost of the highest-priced schools.

Since the voucher did not make all private schools affordable to poor parents, interest was greatest in the lower-cost schools. In our sample of schools the voucher covered only one-fifth of the annual fees of the highest-priced private school. In theory, parents could have used the voucher to pay for part of the fee and paid the balance themselves. However, ICETEX administrators discouraged this practice, fearing that private schools might raise their fees in response, thus transferring part of the effective subsidy from the students to the schools.

II. Municipal Program Participation

In this section we introduce a model that estimates the probability that a municipality will participate in the voucher program. Underlying the model is the assumption that the municipality is responsible for providing secondary schooling, subject to a fiscal constraint. To fulfill this responsibility, the municipality can build public schools or subsidize private schools.

In 1991 Colombia's central government transferred the responsibility for maintaining school facilities to departments and municipalities (Hanson 1995). Each municipality thus inherited a supply gap in secondary schooling, which we call underserved students, $S_U = S_D - S_S$, where S_D is demand for secondary schooling

7. Morales-Cobo (1993) finds that, at least in Bogotá, the program reached its intended beneficiaries. Ribero and Tenjo (1997) find similar success, except that students in neighborhoods ranked 3 received 9–17 percent of the vouchers, depending on the particular municipality. Because residents of these neighborhoods still had below-average incomes, Ribero and Tenjo confirm that the program was relatively well targeted. Under this system of geographical targeting, leakage to unintended beneficiaries is likely, since low-income neighborhoods may have better-off residents. The higher cost of finer targeting, however, justifies a certain degree of leakage.

8. The voucher's maximum value was adjusted annually according to the estimated national inflation rate. The same adjustment was made to the voucher for each participating school, irrespective of changes in fees in those schools. In 1994, for example, the mean value of the voucher ranged from Col$59,700 in the department of Choco to Col$119,100 in Quindío.

and S_S is the sum of secondary enrollment capacity in public and private schools. If $S_U \leq 0$, then secondary schooling capacity exceeds demand, and there is no need to expand supply. If $S_U > 0$, then the municipality must decide how to expand secondary school capacity, conditional on having the funds to finance that expansion.[9]

Model

If the municipality chooses to participate in the voucher program, it faces a cost, v, per student for the voucher plus administration and advertising costs of $a(t)$ per student. We assume that these administrative and advertising costs decrease in the taste for private school, t, because there would be less need to motivate parents of underserved students to participate in the voucher program if there were a strong tradition of private education in the municipality. As an alternative to the voucher system, the municipality has the option of increasing public school capacity.[10] This strategy has very high fixed costs relative to the voucher program. However, as the number of underserved students increases, the average cost of this option, $C(S_U)$, declines. If it falls below $v + a(t)$ over the range $(0, S_U)$, the municipality's cost-minimizing choice would be to increase public school capacity.

The municipal choice is illustrated in figure 1. Assume initially that the average cost of providing a voucher is $v + a(t_0)$. The average cost of expanding public school capacity is given by $C(S_U)$. If the number of underserved students is positive but below S_U^0, and there is no capacity constraint on private schools, the municipality will opt for the voucher program, it being the least-cost means of adding secondary school capacity. Beyond S_U^0 and up to S_U^1, the municipality will reject the voucher program in favor of providing additional space in public schools. But municipal choices are not limited to one or the other. If $S_U > S_U^1$, the municipality will provide more public schools up to capacity S_U^1 and then vouchers for $S_U - S_U^1$.

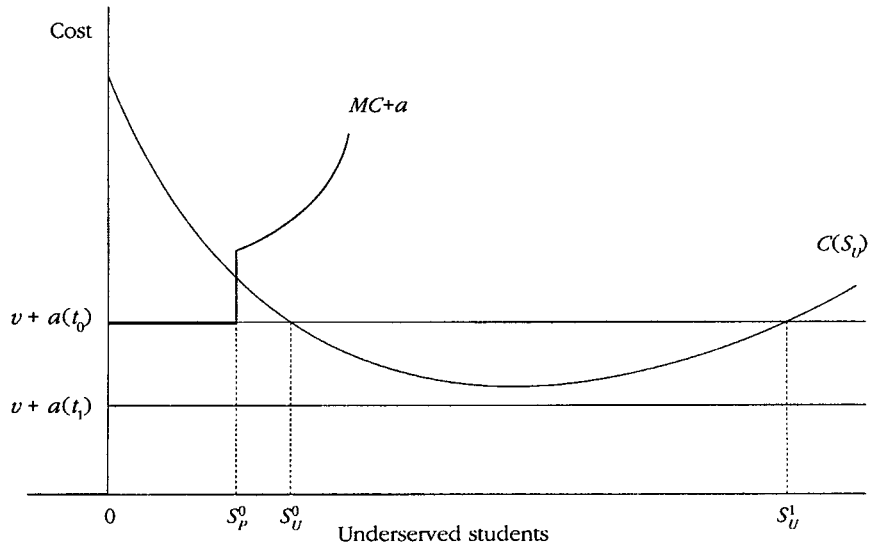
Thus far we have assumed that the municipality knows $v + a(t)$ and $C(S_U)$ with certainty. More realistically, there will be uncertainty regarding the average cost of expanding capacity. Let u_V be a random additive error to the average cost of the voucher, and let u_C be a random additive error to the average cost of school construction. Let V^G be the number of vouchers issued by the municipal government. The probability that the municipality will participate in the voucher program is

$$P(V^G > 0) = P[(S_U > 0) \wedge \{[v + a(t) - C(S_U)] < (u_c - u_v)\}] \tag{1}$$

9. Ministry of Education and departmental officials informed us that the central government had targeted some municipalities for participation, but targeted municipalities could and did turn down the invitation. Furthermore, municipalities that the central government did not target could and did enter the program. Therefore, it is reasonable to model municipal choice as a local decision.

10. Of course, municipalities could reject vouchers and any other role in secondary education. Thus municipalities that opted not to participate in the voucher program did not necessarily expand secondary school capacity by other means.

Figure 1. *The Average Cost of Vouchers Compared with the Average Cost of Increasing Public School Capacity When There Is No Private School Constraint*



Note: S_U is the number of underserved students, and $C(S_U)$ is the average cost of the voucher program. S^0_P is the number of primary school students. v is the per student cost of the voucher, and $a(t)$ are the administrative and advertising costs t is the taste for private school.

If S_U is less than 0, $P(V^G > 0) = 0$. In addition, assume that at $S_U = 0$, $C(0) + u_C > v + a_G(t) + u_V$, so that the average cost of increasing public school capacity exceeds the cost of the voucher program at the lowest values of S_U. As a consequence, $P(V^G > 0)$ rises initially as S_U increases from negative to positive values. However, in its most general form, $C(S_U)$ is a convex function, so that $C'(S_U) < 0$ and $C''(S_U) > 0$. Thus as S_U rises above 0, $C(S_U)$ falls, but at a decreasing rate. We would then expect the probability of a municipality participating in the voucher program to vary with S_U according to $dP/dS_U > 0$, $d^2P/dS_U^2 < 0$, and $d^3P/dS_U^3 > 0$.

As familiarity with or taste for private education increases from t_0 to t_1 in figure 1, the unit cost of vouchers falls to $v + a(t_1)$. In this case the voucher program dominates expansion of public school capacity for all levels of S_U, when the average cost of adding public school spaces is given by $C(S_U)$. In general, the lower is $v + a(t)$, the smaller is the range of students (S_U^0, S_U^1) for which $C(S_U) < v + a(t)$, and the higher is the probability that the municipality will opt for the voucher program. In terms of equation 1, $dP/dt > 0$.

Another assumption we have made is that the supply of space in private schools is perfectly elastic, given the voucher price of v. If, instead, space in private schools is limited, then the private capacity to absorb students will also influence the municipality's decision to participate. The marginal cost of expanding private schools will increase when existing capacity is exceeded. If, for example, excess

capacity of existing private schools is S_P^0, the supply curve will have a discontinuous jump at that point, and the number of vouchers offered by the municipality will be less than if supply were perfectly elastic. Therefore, if S_P is excess capacity in existing private schools, we expect $dP/dS_P > 0$.

We have also assumed thus far that municipalities do not face budget constraints. However, as a condition of participation, municipalities had to demonstrate that they had the fiscal capacity to take over the maintenance of public schools. Municipalities with little ability to raise public revenues, B, would be constrained from participating, so that $dP/dB > 0$.

The model of a municipality that minimizes costs suggests that the probability of participation in the voucher program as defined by equation 1 can be operationalized as:

$$P(V^G > 0) = f[\underset{+}{C(S_U)}, \underset{+}{S_P}, \underset{+}{t}, \underset{+}{B}, \underset{-}{v}, \varepsilon] \tag{2}$$

where $\varepsilon = u_C - u_V$, and the expected signs of the partial derivatives of $P(V^G > 0)$ with respect to the explanatory variables are indicated.

Data

We estimate equation 2 using data for 923 municipalities in Colombia, 208 (or 23 percent) of which opted to participate in the program. The 923 represent all of the municipalities in the 28 of 30 departments for which we had the necessary data. We collected program administrative data from ICETEX central and regional offices over 1995–97, and we found data on the characteristics of municipalities from other existing databases (table 1). Since the variables that we are treating as exogenous in the equation could themselves change in response to a municipality's decision to participate or not in the program, we measure all such variables as of 1991, before the voucher program was implemented.

The number of underserved students is the difference between the number of primary school students, S_{1-5}, and the number of secondary school students, S_{6-11}, measured before implementation of the voucher program. Given the goal of universal secondary enrollment, primary enrollment measures the population of potential secondary students. The total number of private and public secondary school students is a measure of existing secondary school capacity. We divide the difference between the number of potential secondary school students and the number of current secondary school students by the number of public secondary school teachers. This measure suggests how many additional students would need to be placed per public classroom in order to cover all potentially underserved students. The resulting measure is

$$S_U = \frac{S_{1-5} - S_{6-11}}{T_{6-11}^G}, \tag{3}$$

Table 1. *Sample Statistics for Participating and Nonparticipating Municipalities*

Variable	*Total*	*Nonparticipants*	*Participants*
Endogenous variable			
Participation decision	0.22	0	1
	(0.41)		
Exogenous variables			
Underserved students (S_U)[a]	12.16	11.27	15.31
	(24.84)	(23.60)	(28.65)
Basic needs index (N)	0.60	0.63	0.51
	(0.21)	(0.20)	(0.20)
Underserved students interacted with basic needs index ($S_U * N$)	8.03	7.72	9.14
	(17.93)	(17.64)	(18.93)
Ratio of secondary private teachers to primary students	0.0052	0.0025	0.015
	(0.012)	(0.0097)	(0.015)
Proportion of private primary students	0.052	0.028	0.14
	(0.10)	(0.07)	(0.13)
Per capita taxes paid (Col$10,000)	0.29	0.22	0.44
	(0.67)	(0.39)	(0.66)
Proportion of primary schools that are rural	0.30	0.31	0.29
	(0.34)	(0.35)	(0.31)
General information			
Pupil-teacher ratio			
Private secondary	15.16	13.06	16.42
Public secondary	18.46	17.49	21.90
Number of students			
Primary	3,537	1,131	12,107
Secondary	3,022	792	10,963
Number of municipalities	923	715	208

Note: Values given are means. Standard deviations are in parentheses.

a. Underserved students = (number of primary students – number of secondary students)/number of secondary public teachers.

Source: Authors' calculations based on program administrative data from ICETEX central and regional offices, 1995–97, and SABER data, 1992–93.

where T^G_{6-11} is the number of full-time public secondary school teachers. Larger values of S_U will make it harder for existing public schools to absorb the additional students. Consistent with that presumption, measured S_U was 36 percent higher in participating than in nonparticipating municipalities.

Because vouchers were targeted toward poor students, we use the basic needs index (see footnote 6). This index, N, measures the proportion of the municipal population that is considered poor according to five different poverty indicators. Fifty-one percent of the population in participating municipalities was considered poor, compared to 63 percent in nonparticipating municipalities.

It is possible that existing public schools may have been to provide spaces for poor children in some municipalities. Thus we interact N with S_U to generate a proxy measure of needy underserved students per public secondary school teacher. This measure is 18 percent higher on average in participating than in nonpartici-

pating municipalities. Consequently, before the voucher program began, participating municipalities had a smaller proportion of needy households but a larger proportion of needy underserved students relative to existing public school capacity.

We measure the capacity of private schools to absorb additional students by T^{P}_{6-11} / S_{1-5}, where T^{P}_{6-11} is the number of secondary school teachers (and, presumably, classrooms) in existing private schools. Before the voucher program was established, this ratio was six times higher in participating than in nonparticipating municipalities. In addition, the proportion of primary school students in private school was 0.14 in participating municipalities, five times higher than the proportion in nonparticipating municipalities.

We use two measures of government capacity to raise revenue. The first is the proportion of poor people in the municipality, as measured by the needs index. Poorer municipalities have less capacity to raise revenue, although the average poverty indexes for participating and nonparticipating municipalities were virtually identical. The second measure is 1991 per capita income taxes paid in the municipality. This was two times higher in participating than in nonparticipating municipalities. Participating municipalities also had larger populations of school children. Primary enrollment was nearly 11 times higher in participating municipalities. However, the proportion of rural schools, as designated by the Ministry of Education, was nearly identical across the two groups.

Estimation and Results

The most important parameters in the probit model of voucher participation (equation 2) pertain to the measure of underserved students (table 2). We expect S_U to affect participation in a nonlinear fashion, initially raising and then lowering the probability of participating. We first include a cubic form of S_U. Although the sign pattern corresponds to our expectations, the third-order term is not significant. The results reported in table 2 use a quadratic approximation of S_U. At sample means the marginal effect of this variable is positive. The elasticity implies that a 10 percent increase in the number of underserved students per secondary school teacher would increase the probability of municipal participation by 2.8 percent.

We also trace out the nonlinear effect of S_U using a spline function. The coefficients on the dummy variables representing whether the measure of underserved students was positive, in the upper 50 percent, upper 30 percent, upper 20 percent, and upper 10 percent of the distribution of S_U or NS_U are cumulative (columns 2 and 4 of table 2).[11] That is, other things constant, the total effect of being in the upper 30 percent of the distribution of S_U is the sum of the coefficients on

11. Municipalities that attracted students from surrounding towns show negative values of S_U. The distributional information on S_U and NS_U (table 1) shows that they both turn positive at the thirtieth percentile, so the first dummy variable represents the upper 70 percent of the distribution of underserved students.

Table 2. *Probit Estimates for Municipal Participation in the Voucher Program*

Variable	*1*	*2*	*3*[a]	*4*[a]
$S_U/100$	1.779***		2.667***	
	(3.800)		(3.503)	
	[0.285]		[0.285]	
$(S_U/100)^2$	-0.743***		-1.600**	
	(-2.436)		(-2.277)	
$S_U \geq 0$		-0.0288		-0.034
		(-0.174)		(-0.206)
		[-0.001]		[0.0004]
Municipality is in the fiftieth percentile of S_U distribution		0.327		0.368
		(1.900)		(2.109)
		[0.040]		[0.239]
Municipality is in the seventieth percentile of S_U distribution		0.268		0.249
		(1.360)		(1.262)
		[0.066]		[0.037]
Municipality is in the eightieth percentile of S_U distribution		0.196		0.103
		(0.922)		(0.472)
		[0.075]		[0.025]
Municipality is in the top ninetieth percentile of S_U distribution		-0.105		0.058
		(-0.486)		(0.262)
		[-0.082]		[-0.031]
Secondary private teachers/primary students/100	0.342***	0.341***	0.341***	0.343***
	(8.718)	(8.651)	(8.691)	(8.687)
	[0.265]	[0.264]	[0.264]	[0.268]
Proportion of private primary students	4.091***	4.011***	4.061***	4.006***
	(7.476)	(7.291)	(7.384)	(7.251)
	[0.315]	[0.309]	[0.313]	[0.311]
Proportion of primary schools that are rural	-0.125	-0.161	-0.093	-0.124
	(-0.634)	(-0.797)	(-0.481)	(-0.633)
	[-0.056]	[-0.072]	[-0.042]	[-0.056]
Per capita taxes paid (Col$10,000)	0.0363	0.0566	0.0272	0.024
	(0.448)	(0.640)	(0.340)	(0.296)
	[0.015]	[0.024]	[0.012]	[0.010]
Needs index	-0.966***	-0.960***	-1.224***	-1.368***
	(-3.642)	(-3.577)	(-4.278)	(-4.445)
	[-0.856]	[-0.853]	[-1.087]	[-1.222]
Constant	-0.87***	-0.963***	-0.716***	-0.736***
	(-4.856)	(-4.732)	(-3.837)	(-3.315)
Sample size	923	923	923	923
Log-likelihood	-358.097	-355.594	-358.196	-355.195
Pseudo R-squared	0.2729	0.2780	0.2727	0.2766

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

Note: S_U, the number of underserved students, equals (number of primary students – number of secondary students)/number of secondary public teachers. z-statistics are in parentheses. Elasticities are in brackets and are computed at the median of the range for the dummy variables.

a. S_U is interacted with the needs index.

Source: Authors' calculations based on program administrative data from ICETEX central and regional offices, 1995–97, and SABER data, 1992–93.

S_U when S_U is positive, in the upper 50 percent of the distribution, and in the upper 30 percent of the distribution. The results suggest a rising probability of participation between the fiftieth and ninetieth percentiles.

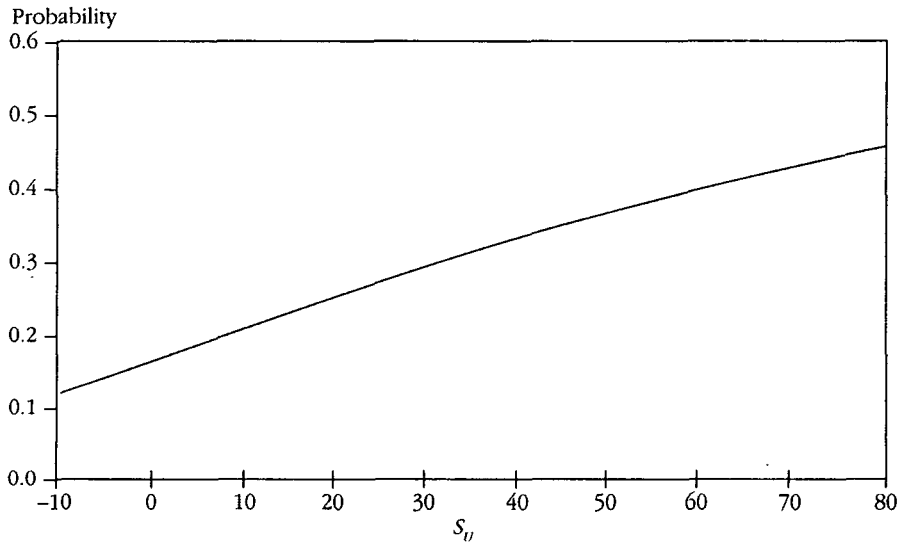
The relationship between $P(V^G > 0)$ and S_U as captured by the equations in columns 1 and 2 are shown in figures 2 and 3. In the quadratic representation (figure 2) the probability rises throughout the sample range of underserved students. The spline function (figure 3) peaks at just over 20 underserved students per teacher, although we cannot reject the hypothesis that the probability of participation is equal for municipalities that fall in the upper half of the range of underserved students.

The model of program participation suggests that municipalities would be more likely to participate if private schools had excess capacity and if the population were already familiar with private schools. Both of these predictions are borne out by the estimates. In all four specifications the ratio of private secondary school teachers to current primary students (a measure of the number of classrooms available for future secondary students) has a strong positive effect on municipal participation. The elasticity suggests that a 10 percent increase in private capacity raises the probability of municipal participation by about 2.6 percent. The elasticity of municipal participation with respect to private schools' share of primary students is 0.32. We speculate that municipalities whose populations are already familiar with private schools would need to exert less effort to promote voucher applications.

We anticipate that municipalities with more limited fiscal resources would be unable to meet the financial requirements of the voucher program. Our results show that higher per capita tax payments increase the probability of voucher participation, although the coefficients are never precisely estimated and imply very small elasticities. Municipalities with a higher proportion of poor people (as measured by the needs index) are significantly less likely to participate, with an elasticity of −0.86. We also expect that the proportion of primary schools that are rural would be associated with lower costs of building new schools and thus a weaker incentive to participate in the voucher program. All four specifications show a smaller probability of participation in more rural municipalities, but the elasticities are extremely low.

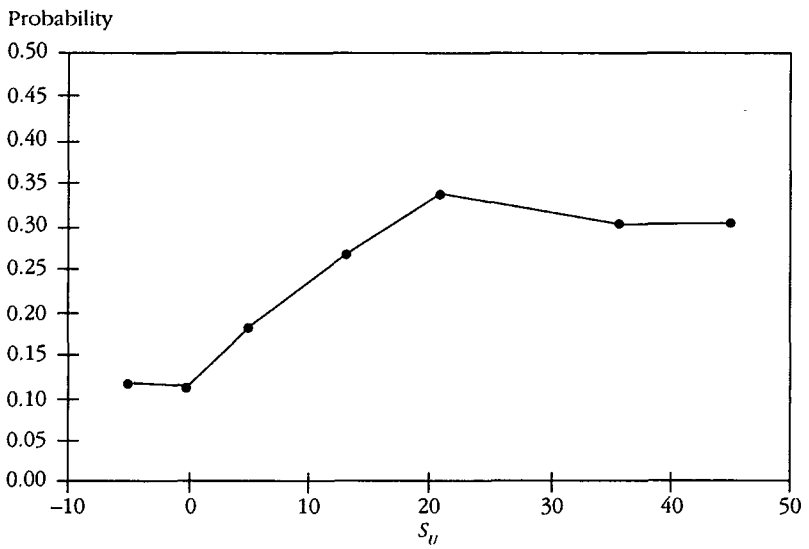
Overall, the results in table 2 are strongly consistent with a model of municipal cost minimization given an obligation to provide secondary schooling to underserved students. To determine if participation was influenced by central government pressure on certain municipalities, we estimate the equation again after eliminating the 10 largest municipalities that were initially invited to participate. The coefficients and significance levels change very little, and signs do not change at all. Thus it seems that municipalities were in fact most likely to use vouchers to address the needs of their students if private schools had excess capacity and if the cost of the voucher program was lower than the cost of expanding public secondary schools.

Figure 2. *Probability of Municipal Participation in the Voucher Program, Quadratic Specification*



Note: S_U = (primary students – secondary students)/secondary public teachers.
Source: Authors' calculations.

Figure 3. *Probability of Municipal Participation in the Voucher Program, Spline Specification*



Note. S_U = (primary students – secondary students)/secondary public teachers.
Source: Authors' calculations.

III. School Program Participation

A school could participate in the voucher program only if its municipality agreed to participate. All private schools were eligible to join. The number of private schools that participated increased steadily between 1992 and 1997. By mid-1995, 1,795 private schools were accepting voucher students. In this section we examine the school's participation decision both theoretically and empirically.

Model

We assume that private schools compete on both quality, q, and price, f, but that public schools offer uniform quality, q_G, at near-zero cost, f_G.[12] Given the available public school, parents will select the ith private school when

$$U(q_i, f_i; \mathbf{Z}) \geq U(q_G, f_G; \mathbf{Z}) \tag{4}$$

where $U(.)$ is parents' indirect utility, and $\mathbf{Z}$ are factors that enter utility and are separable from school quality and price. We assume that school quality raises utility ($U_q' > 0$), school fees lower utility ($U_f' < 0$), and private school fees exceed public school fees ($f_i > f_G$). Therefore, parents must believe that the quality of private schools exceeds that of public schools ($q_i > q_G$) for the condition in equation 4 to hold.

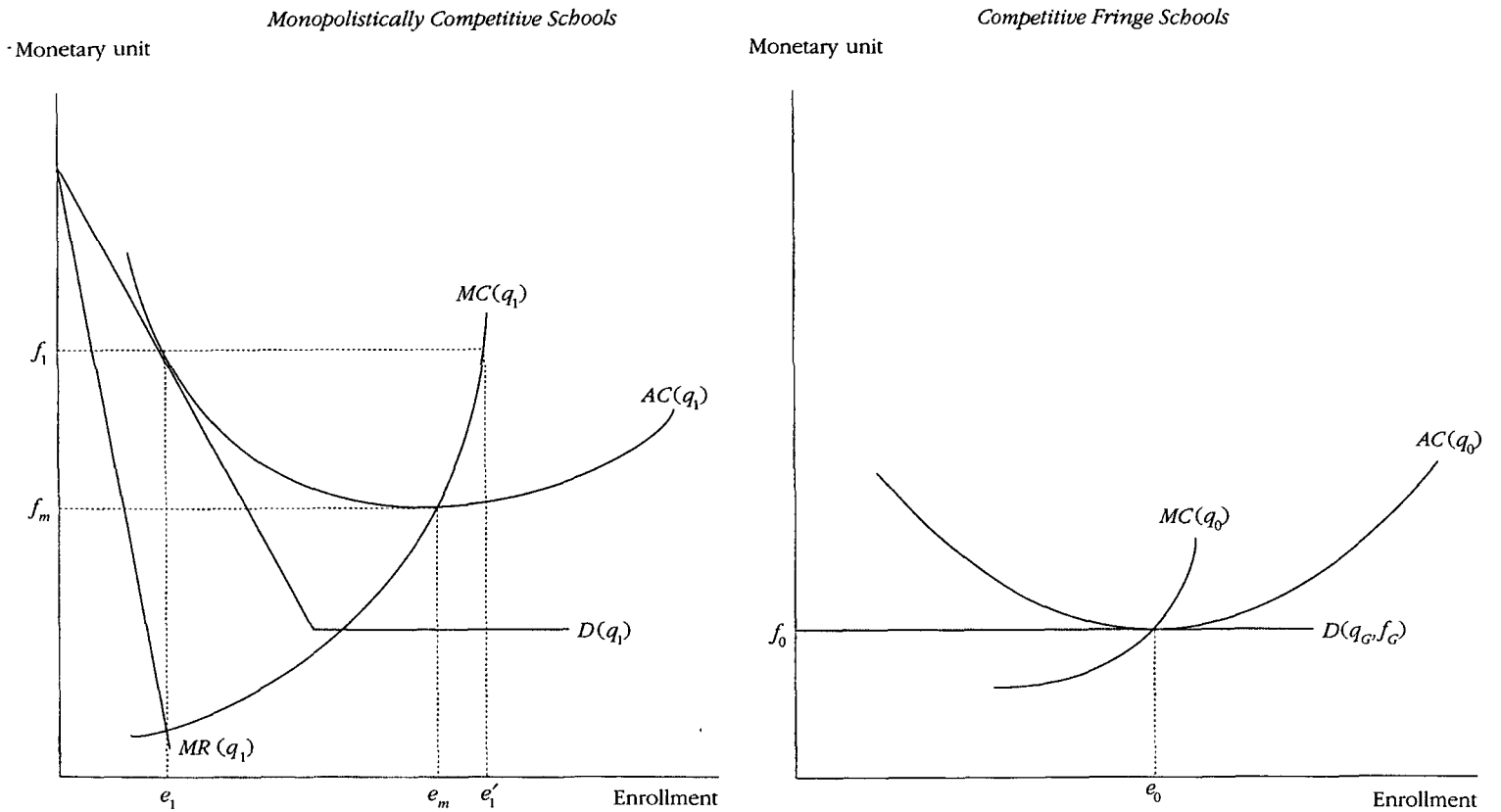
Since barriers to entry for new private schools appear to be low, we assume that private schools will earn no economic profits in the long run. Thus $f_i = AC(q_i)$, the school's average cost of producing quality q_i. Consequently, school fees must increase with school quality to enable the school to break even.[13] There will be a level of school quality, q_0, and fees, f_0, $= AC(q_0)$, that will satisfy condition 4 with equality. These levels will define the competitive fringe of private schools. For competitive schools, f_0 and q_0 are dictated by the market, conditional on the availability of public schools of quality q_G charging f_G.

We illustrate the private school market in figure 4. Competitive fringe schools must provide at least quality q_0 to lure parents away from public schools and must charge fees of f_0 to break even (the right panel of figure 4). Schools could also offer quality above q_0 and then compete on both price and quality as in a monopolistically competitive market. In equilibrium such a school offers quality $q_1 > q_0$ (the left panel of figure 4). Enrollment demand increases as the school lowers its fees and becomes perfectly elastic as the fees approach f_0. However, the

12. This specification contrasts with Manski's (1992) simulation model, which assumes that all private schools charged the same price and offered the same quality. However, Manski (1992: 360) points out that his assumption is too restrictive and that "a more realistic model would permit private schools to set different tuition levels, with associated differences in the quality of the schooling that they offer."

13. In Colombia the correlation coefficients between private tuition fees and average mastery levels in math and language examinations of ninth-grade students are 0.55 and 0.52, respectively. Presumably, there are other school products, such as students' civic awareness or physical safety, that distinguish schools and are not captured by these test scores.

Figure 4. *Representative School Demand and Cost Curves for Minimum- and Intermediate-Quality Private Schools*



Note: f represents school fees, and q represents school quality.

school's optimum fee is f_1. Because entry is easy, the school cannot earn economic profits in the long run, so $f_1 = AC(q_1)$ at an enrollment level of e_1. If all schools decide to offer quality above q_0, then all will be monopolistically competitive.

The voucher program may alter school incentives, depending on school quality and fees. Let v_M represent the maximum value of the voucher allowed. If $v_M < f_0$, no private school could recoup its marginal cost if it offered the minimum quality.[14] If $v_M = f_0$, fringe schools would be indifferent between participating and not participating. Because they can enroll as many students as they want at f_0, they have no need to acquire additional students through the voucher program. In fact, schools on the minimum-quality fringe have no capacity to absorb additional students. Voucher students would have to displace nonvoucher students to leave the school with enrollment e_0 and no economic profits.

Even if $v_M > f_0$, schools in existence when the program was initiated could not benefit from the program. Rules prevented schools from raising fees in the first year of participation. In subsequent years schools were allowed to raise fees only according to an officially sanctioned adjustment for inflation. Thus minimum-quality schools could not raise fees above f_0, which meant that they could not profitably raise enrollment beyond e_0.

For the monopolistically competitive schools whose quality is higher than q_0, the voucher program may provide an incentive to add students. These schools have excess capacity because they can lower their average costs by enrolling more students. If the voucher is equal to f_1, such a school will expand enrollment to e_1' and earn higher economic profits than it did before the voucher program. Even if $v_M < f_1$, the school could profit from the voucher program by lowering its fees to v_M and expanding enrollment according to $v_M = MC(q_1)$, provided v_M is higher than $f_m = \min \mathrm{AC}(q_1)$. If this school participates, it will increase its enrollment to at least e_m.[15]

Monopolistically competitive schools will not participate if $v_M < f_m(q)$. Because f_m increases as quality increases, the highest-quality schools will be those for which $v_M < f_m$. For these schools the marginal cost of adding a student exceeds the maximum value of the voucher. Of course, high-quality private schools that already offer need-based scholarships might still participate. The voucher program would allow these private schools to admit poor students using a cofinancing scheme that reduces the burden to their own scholarship program.

14. We assume that schools were not allowed to charge parents more than the amount of the voucher, and, in fact, schools were discouraged from demanding that parents "top up" the voucher. Since vouchers were targeted to the poorest segments of the population, many families did not have much more to offer.

15. In the long run schools will not be able to earn positive economic profits from the vouchers. New schools will enter. If parents can easily obtain and understand information on school quality, a new set of voucher schools will emerge that select quality such that $v_M = \min \mathrm{AC}(q) > f_0$. Therefore, in the long run the voucher program should increase the number of intermediate-quality schools. However, the poorest-quality schools will exit because they are now inferior to voucher schools and are excluded from the voucher subsidy.

The simple model illustrated in figure 4 can incorporate cofinancing. Costs of many private schools are partially offset by monetary or in-kind assistance from sponsoring religious or nonsectarian organizations. These transfers lower the schools' average costs by the value of the donation, d, per student. We can also adjust the model to incorporate school administrative costs, a_S. In order to participate in the voucher program, schools had to attract students, elicit applications from families in the lowest income groups, and cooperate with the regional program agency to verify the continued attendance of each voucher student.

The donations and administrative costs are observed with error, u_S. The probability that a school offers a positive number of vouchers is

$$(5) \qquad P(V_i^S > 0) = P\{[v_m - f_m(q_i) - a_s + d > u_s] \wedge [f_i > f_0]\}$$

Equation 5 states that the ith private school will participate when the maximum value of the voucher plus any per-pupil external support covers the average cost of a voucher student, provided that the school is not a minimum-quality private school charging f_0.

Equation 5 can be operationalized by noting the implied participation incentives for a continuum of schools sorted according to fees and quality from lowest to highest. Schools that offer the lowest quality, q_0, and charge the lowest fees cannot profit from expanding enrollment because rules prevent them from raising their fees in response to the voucher program. Schools that offer intermediate quality and fees can profitably expand enrollment as long as the value of the voucher exceeds their minimum average costs. Eventually, as we move to schools with progressively higher fees and quality, the voucher will fail to cover the marginal cost of an additional student. This suggests a probit equation of the form

$$(6) \qquad P(V_i^S > 0) = q(\underset{+}{f_i}, \underset{-}{f_i^2}, \underset{+}{q_i}, \underset{-}{q_i^2}, \underset{+}{d}, \underset{-}{a_s}, u_s)$$

where the expected signs are included below the explanatory variables.[16] The theory suggests a nonlinear relationship between participation and school fees or quality so that schools with intermediate fees and quality will be the most likely to participate. Because fees and measured quality are not perfectly correlated, we incorporate both in the analysis. Holding the two constant, schools will be more likely to participate if they have access to external support and if they face lower costs of administering the voucher program.

Data

The Ministry of Education collected data on schools and students as part of its *Sistema Nacional de Evaluación de la Calidad de la Educación* (SABER). The Ministry collected the data in late 1992 or early 1993, the first year in which it was

16. Note that v_M is the same for all schools, and so it cannot explain variation in the probability of participation across schools. Its influence is captured in the constant term.

expanding the voucher system nationally. At that time municipalities could have been participating in the program for one year, but most had not yet adopted the program. The data on students cover only seventh and ninth graders, none of whom qualified to receive the vouchers, since the program was restricted to new sixth graders in 1992–93. Since schools were not allowed to change their fees in the first year of the program, the data on schools and students predate the influence of the vouchers on school or household attributes.

The sample consists of 71 participating private schools and 77 nonparticipating private schools. The ratio of participation is larger in our sample than in the population of private schools overall because the SABER sampling design is skewed toward more urban areas.[17] Consequently, the results should be interpreted as reflecting the decision process of private urban schools.

Monthly fees for schools that participated in the voucher program averaged about 60 percent of the fees in nonparticipating private schools (table 3). The nonparticipating schools were also of higher quality, as indicated by higher average scores on the nationwide school-leaving exam given to eleventh-grade students. Average test scores in the public schools were also higher than those in participating private schools, albeit by only a small and insignificant margin. There were other indications that the quality of program schools was lower on average than that of nonprogram schools. Participating schools were more likely to offer nonacademic (vocational or technical) tracks and had higher pupil-teacher ratios than did nonparticipating schools. In terms of quality, participating private schools were more similar to public schools than to nonparticipating private schools.

Promotional costs of administering the voucher program are likely to differ across schools. Schools that cater to poorer families will have an advantage in identifying potential voucher recipients. Using counts of household durable goods to measure household wealth, we find that the wealthiest households tended to be in nonparticipating schools. Public schools had the lowest average value of this indicator.

Estimation and Results

Table 4 reports three sets of probit estimates of the school participation function. Since only schools in participating municipalities could join the voucher program, we first investigate whether a school's participation decision was related to the municipality's participation decision. We introduce a correction for possible selection bias in equation 6, using the inverse Mills ratio derived from the estimated municipal participation equation (see table 2). The results, which we do not report here, indicate that the probability of municipal participation is

17. Also, the SABER sample includes only schools that educate students through eleventh grade, the terminal year of secondary education. Many schools, primarily in rural areas, educate students only through ninth grade.

Table 3. *Means and Standard Deviations of Variables Used in School Participation Function*

Variable	*Participating private schools*	*Nonparticipating private schools*	*Public schools*
Monthly fee per student (Colombian pesos)	7,777.1	13,161.5	626.9
	(4,643.1)	(11,140.29)	(1,582.8)
Average school-leaving test score	47.6	52.8	49.2
	(5.1)	(7.4)	(5.1)
Nonprofit school	0.51	0.53	n.a.
	(0.50)	(0.50)	n.a.
Offers academic education	0.75	0.92	0.71
	(0.44)	(0.27)	(0.46)
Pupil-teacher ratio	21.9	17.6	20.5
	(9.7)	(8.8)	(7.5)
Household assets of students	9.9	10.9	9.1
	(1.3)	(1.6)	(1.2)
Sample size	71	77	112

n.a. Not applicable.

Note: Standard deviations are in parentheses.

Source: Authors' calculations based on program administrative data from ICETEX central and regional offices, 1995–97, and SABER data, 1992–93.

not correlated with the regressors in the school participation equation.[18] Thus we can interpret the results of the school participation equation as holding for private schools generally and not just for those in participating municipalities.

The main variables of interest are tuition fees and average test scores. Both have the predicted quadratic sign pattern: the probability of participation first rises and then falls as the variables increase in value. This pattern holds regardless of whether fees and test scores enter the specification jointly or separately. We take this result as strong support for the presumption that schools make participation decisions at least partially according to the profit-maximizing model described above.

The implied impact of fees and test scores on school participation is traced out in figure 5. Participation is highest in the lower half of the range of values for both fees and test scores. The peak for fees corresponds almost exactly to the maximum value of the voucher and drops off sharply thereafter. Almost no schools with fees in the upper half of the tuition range opted to participate. Participation was more broadly dispersed across school quality as measured by test scores, with many schools in the upper half of the range participating.

These results suggest that the participation decision was tied to a school's prior fee structure, so that higher-quality schools still participated if they charged

18. The corresponding probit coefficients (and z-statistics) for the inverse Mills ratio for two specifications similar to those shown in table 4 are 0.332 (0.92) and 0.218 (0.60). We also employ an alternative procedure, using the variables in table 2 as instruments for the unknown selection process. None of the school participation coefficients was sensitive to alternative specifications in terms of sign, significance, or magnitude.

Table 4. *Probit Estimates of School Participation Function*

Independent variable	*1*	*2*	*3*
Monthly fee per student ($\times 10^3$)	0.240	0.275	
	(2.39)	(2.71)	
	[0.25]	[0.21]	
Monthly fee squared ($\times 10^6$)	−0.011	−0.021	
	(−2.56)	(−3.01)	
School-leaving test score	0.561		0.703
	(1.55)		(2.07)
	[6.76]		[11.34]
Test score squared	−0.006		−0.008
	(−1.70)		(−2.31)
Nonprofit school	0.624	0.480	0.512
	(2.18)	(1.90)	(1.87)
	[0.17]	[0.12]	[0.19]
Offers academic stream	−0.546	−0.620	−0.482
	(−1.60)	(−1.81)	(−1.49)
	[−0.26]	[−0.28]	[−0.29]
Pupil-teacher ratio	0.007	0.013	0.009
	(0.53)	(1.00)	(0.66)
	[0.08]	[0.13]	[0.12]
Average assets of students	−0.142	−0.206	−0.095
	(−1.03)	(−2.08)	(−0.76)
	[−0.79]	[−1.05]	[−0.71]
Constant	−12.217	−1.115	−14.81
	(−1.41)	(−1.16)	(−1.81)
Sample size	139	146	139
Log-likelihood	−73.2	−79.9	−78.2
Pseudo R-squared	0.240	0.209	0.188

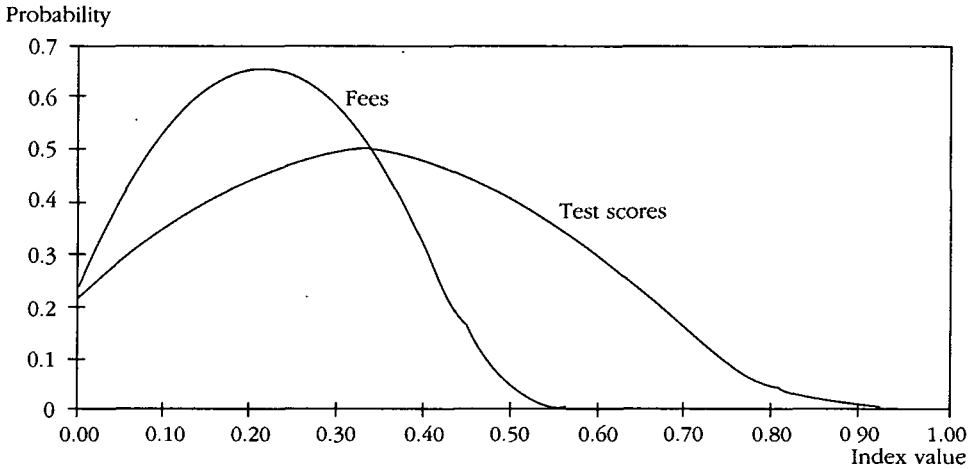
Note: z-statistics are in parentheses, and elasticities, evaluated at the mean, are in square brackets. The specification also includes a dummy variable indicating if the school fee is missing (9 percent of the sample).

Source: Authors' calculations based on program administrative data from ICETEX central and regional offices, 1995–97, and SABER data, 1992–93.

relatively low fees. Such schools were likely to depend on external support, which enabled them to provide quality at low cost. Nonprofit schools were more likely to participate, as were schools that served students with below-average wealth. Apparently, vouchers did encourage higher-quality schools catering to poorer families to expand their services.

Other variables supported the general finding that schools in the lower half of the quality distribution were more likely to participate. Schools offering nonacademic, vocational programs and schools with higher pupil-teacher ratios were

Figure 5. *Estimated Probability of School Participation over the Range of School Fees and Test Scores*



Note: The horizontal axis is normalized to range from 0 to 1 where 1 corresponds to the highest test score or fees in the sample and 0 corresponds to the smallest. If f_{max} is the largest fee in the sample and f_{min} the smallest, the fee index is defined as $(f_i - f_{min})/(f_{max} - f_{min})$. Similarly, the test score index is $(q_i - q_{min})/(q_{max} - q_{min})$. The probabilities are estimated with all other variables fixed at their sample means.
Source. Authors' calculations.

more likely to admit voucher students. These coefficients were not precisely estimated, however.

The results show that the attributes of participating schools did not reflect the average of all private schools. Peak participation occurred in schools at the twentieth percentile in the range of fees and at the thirty-fifth percentile in the range of test scores. Therefore, assuming that private school averages apply to voucher students generally overstates the potential gains of the program.

IV. Conclusions

Colombia's national voucher program demonstrates how a central government can effectively mobilize local government resources and private providers to alleviate constraints to public provision of education. At the end of five years the program had given vouchers to more than 100,000 secondary school students throughout the country. However, even with financial incentives from the central government, national programs do not necessarily beget universal adoption. Only 25 percent of Colombia's municipalities joined the voucher program. Adoption was most likely in municipalities where existing private schools could expand capacity, where a large proportion of students were already enrolled in private school, and where there was a limited number of underserved students. Municipalities with a very large number of underserved students or whose existing private schools had relatively little capacity opted not to participate in the

voucher program, presumably because the cost of participating would have exceeded the cost of building additional public schools.

Only about half of all private secondary schools participated in the program. Participation was not random; the characteristics of participating schools differed from those of private schools overall. Schools that responded to the program typically served students from poorer households, and voucher recipients generally came from the poorest socioeconomic strata of the country. The lowest- and highest-quality schools were less likely to participate than intermediate-quality schools, suggesting that private schools did not exploit voucher recipients by offering inferior quality at a publicly subsidized price. The rule preventing schools from raising tuition fees in response to the voucher program may have limited the participation of the lowest-quality schools.

Colombia terminated its national voucher program in 1997, with the youngest cohort of new recipients awarded vouchers in 1996. Although the program was successful in many dimensions—implementing a fairly accurate targeting system, attracting thousands of students, gaining support from municipalities and schools with characteristics consistent with program objectives—several factors conspired against it. For one, the monitoring system, even for a geographically targeted program, proved to be quite costly. To guard against "ghost" voucher awardees, four times a year the regional offices of the administrative agencies obtained the signature of each voucher student prior to releasing funds. The program's success in expanding the number of voucher students increased its total cost, if not its average cost. Delays in compiling students' signatures and in obtaining the appropriate officials' signatures at the designated disbursement junctures delayed payments to schools and created difficulties for school operators and principals. This tardiness worsened over the years, primarily because of increasing difficulty in securing the necessary central funds on time. The frequent turnover of leadership at the Ministry of Education slowly eroded support for the program. By 1997 the department of Antioquia had piloted and adopted a different form of subsidy to private schools. Also by 1997 the Ministry had announced the end of the voucher program and was discussing the elements of another demand-based system of subsidies.

In this article we raised one set of questions about a voucher program that allows voluntary participation at the municipal and school level. Other aspects of Colombia's voucher program deserve attention, but these require student-level data that were not available to us at the time of this study. Several questions warrant further research. Did the program increase net enrollment in the participating areas, or did it simply allow students to transfer from public schools to private schools? Was the program successful in keeping its focus on the poor beyond its first years? Controlling for individual characteristics, did voucher students learn better than public school students? Did voucher students have higher graduation rates than public school students? Because schools lost a voucher if a student failed or left, schools had an incentive to retain voucher students. Although Colombia's 1991 program no longer exists, answers to these questions

would provide useful (and rare) evidence about the potential benefits—as well as the pitfalls—of similar programs in the future.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth M. King, and Michael Kremer. 1999. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." Massachusetts Institute of Technology, Harvard University, and the World Bank. Processed.

CERI, OECD (Centre for Educational Research and Innovation, Organisation for Economic Co-operation and Development). 1994. *School: A Matter of Choice.* Paris.

Colombia, Department of National Planning. 1994. "Sistemas de indicadores sociodemográficos para Colombia: Pobreza en el país y por departamentos." Boletín 5. Bogotá. Processed.

Gauri, Varun. 1998. *School Choice in Chile: Two Decades of Educational Reform.* Pittsburgh: University of Pittsburgh Press.

Hanson, Mark. 1995. "Democratization and Decentralization in Colombian Education." *Comparative Education Review* 39(1):101–19.

Henig, Jeffrey R. 1994. *Rethinking School Choice: The Limits of the Market Metaphor.* Princeton: Princeton University Press.

Hoxby, Caroline M. 1996. "The Effects of Private School Vouchers on Schools and Students." In H. F. Ladd, ed., *Holding Schools Accountable: Performance-Based Reform in Education.* Washington, D.C.: Brookings Institution.

James, Estelle. 1993. "Why Do Different Countries Choose a Different Public-Private Mix of Educational Services?" *Journal of Human Resources* 28(Summer):571–92.

Jimenez, Emmanuel, Marlaine E. Lockheed, and Vicente Paqueo. 1991. "The Relative Efficiency of Private and Public Schools in Developing Countries." *The World Bank Research Observer* 6(July):205–18.

King, Elizabeth M., Laura Rawlings, Marybell Gutiérrez, Carlos Pardo, and Carlos Torres. 1997. "Colombia's Targeted Education Voucher Program: Features, Coverage, and Participation." Impact Evaluation of Educational Reforms Working Paper 3. World Bank, Washington D.C. Processed.

Lankford, Hamilton, and James Wyckoff. 1992. "Primary and Secondary School Choice among Public and Religious Alternatives." *Economics of Education Review* 11(December):317–37.

Levin, Henry M. 1991. "The Economics of Educational Choice." *Economics and Education Review* 10(2):137–58.

———. 1992. "Market Approaches to Education: Vouchers and School Choice." *Economics of Education Review* 11(4):279–85.

Lockheed, Marlaine, and Emmanuel Jimenez. 1994. "Public and Private Secondary Schools in Developing Countries." HRO Working Paper 43. Human Resources and Operations Policy Vice Presidency, World Bank, Washington, D.C. Processed.

Manski, Charles F. 1992. "Educational Choice (Vouchers) and Social Mobility." *Economics of Education Review* 11(December):351–69.

Montenegro, Armando. 1995. "An Incomplete Educational Reform: The Case of Colombia." HCO Working Paper 60. Human Capital Development and Operations Policy Vice Presidency, World Bank, Washington, D.C. Processed.

Morales-Cobo, Patricia. 1993. "Demand Subsidies: A Case Study of the PACES Voucher Program." Department of Economics, University of the Andes, Bogotá.

Murnane, Richard J., Stuart Newstead, and Randall J. Olsen. 1985. "Comparing Public and Private Schools: The Puzzling Role of Selectivity Bias." *Journal of Business and Economic Statistics* 3(January):23–35.

Neal, Derek. 1997. "The Effects of Catholic Secondary Schooling on Educational Attainment." *Journal of Labor Economics* 15(January):98–123.

Parry, Taryn Rounds. 1993. "Theory Meets Reality in the Great Voucher Debate." University of Georgia. Processed.

Ribero, Rocio M., and Jaime G. Tenjo. 1997. "Evaluación del programa de becas paces." University of the Andes, Bogotá. Processed.

Rodríguez, Jorge G. 1988. "School Achievement and Decentralization Policy: The Chilean Case." *Revista de Análisis Económico* 3(1):1.

Sander, William. 1996. "Catholic Grade Schools and Academic Achievement." *Journal of Human Resources* 31(Summer):540–48.

Sander, William, and Anthony C. Krautmann. 1995. "Catholic Schools, Dropout Rates, and Educational Attainment." *Economic Inquiry* 33(April):217–33.

West, Edwin G. 1996. "Education Vouchers in Practice and Principle: A World Survey." Human Capital Development Working Paper 64. World Bank, Washington, D.C. Processed.

Winkler, Donald, and Taryn Rounds. 1996. "Municipal and Private Sector Response to Decentralization and School Choice." *Economics of Education Review* 15(4):365–76.

Witte, John F. 1996. "School Choice and Student Performance." In Helen F. Ladd, ed., *Holding Schools Accountable: Performance-Based Reform in Education.* Washington, D.C.: Brookings Institution.

World Bank. 1994. *Poverty in Colombia.* Washington, D.C.

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 493-508

Student Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments

Jee-Peng Tan, Julia Lane, and Gerard Lassibille

Policymakers in most developing countries are concerned about high dropout rates and poor student learning in primary education. The government of the Philippines initiated the Dropout Intervention Program in 1990–92 as part of its effort to address these issues. Under this program, four experimental interventions were randomly assigned to 20 schools in selected low-income areas. Pre- and post-intervention data were collected from these schools, as well as from 10 control schools, in order to evaluate the program's impact on dropout behavior and student learning. The economic justification for replication appears to be strongest for the interventions that provided teachers with learning materials, which helped them to pace lessons according to students' differing abilities, and that initiated parent-teacher partnerships, which involved parents in the schooling of their children. The justification was weakest for the school feeding intervention. In addition to the results specific to the Philippines, this research demonstrates the feasibility of monitoring and evaluating interventions in the education sector in other developing countries, including the use of randomized control designs.

Most developing countries now recognize that investing in education, particularly primary education, provides an essential bedrock for economic and social development. In past decades governments emphasized expanding enrollment, but as coverage rose, the problems of low completion rates and inadequate student learning came to the fore (see Lockheed and Verspoor 1991 for a comprehensive treatment of these issues). Policymakers need information on the costs as well as on the impact of different methods of improving schooling outcomes. Unfortunately, however, the literature on the quantitative relationship between inputs and outcomes in education is sparse, and most developing countries have only a nascent capacity to conduct their own context-specific research and evaluation.[1]

1. See Harbison and Hanushek (1992) for a summary of results from 96 studies on the relationship between school inputs and learning based on data from developing countries. Quantitative studies on the relationship between school inputs and dropout behavior and between school inputs and grade repetition are much more rare. Recent examples include Hanushek and Lavy (1994), Gomes-Neto and Hanushek (1994), and Chuard and Mingat (1996).

Jee-Peng Tan is with the Human Development Network at the World Bank, Julia Lane is with the Urban Institute, and Gerard Lassibille is with the Centre National de la Recherche Scientifique in Dijon, France. The authors' e-mail addresses are jtan@worldbank.org, jlane@ui.urban.org, and gerard.lassibille@u-bourgogne.fr. The authors thank Ms. Lidi Santos and her staff at the Bureau of Elementary Education of the Philippines Department of Education, Culture, and Sports for facilitating access to the data, the World Bank's Research Department for partial financial support, Alain Mingat for helpful discussions on the analytical method, and three anonymous referees for comments and suggestions on earlier drafts.

This article documents an evaluation effort in the Philippines intended to guide policymaking in primary education. Almost all children in the country enter first grade, but not all of them reach the end of the primary school cycle. Data for the early 1990s suggest that noncompleters represent about 25 percent of each entering cohort of first graders. Further, many children leave school having learned only a fraction of the primary school curriculum (Miguel 1993). Achievement tests administered in the 1991 Household and School Matching Survey, for example, show that pupils in grades two to six mastered less than half of the curriculum they were taught (Tan, Lane, and Coustère 1997).

As part of its strategy to improve primary education, the Philippine government implemented the Dropout Intervention Program (DIP) in the context of a World Bank–financed elementary education project. The program comprised four experimental interventions. Each of these was implemented in five schools during the 1991–92 academic year, for a total of 20 schools. As its name suggests, the DIP focused primarily on reducing dropout rates, but it was also expected to improve student learning.

To determine whether or not the pilot interventions should be replicated, the government randomly assigned them to schools in select low-income communities and collected both pre- and post-intervention data over two school years. The government also collected data on schools that were not part of the program to provide a benchmark for assessing the impact of the interventions. The resulting data set is rare in a developing country.[2] Thus while one goal of this paper is to shed light on elementary education policy in the Philippines, a broader aim is to demonstrate that project evaluation in developing countries is both feasible and worthwhile.

I. The DIP Interventions

The DIP consisted of four experimental interventions: school feeding; multi-level learning materials, which are pedagogical materials for teachers; school feeding combined with parent-teacher partnerships; and multi-level learning materials combined with parent-teacher partnerships.[3] On a per-student basis school feeding is very expensive, the use of multi-level learning materials is considerably cheaper, and parent-teacher partnerships entail minimal additional costs because they involve mostly an adjustment in the way parents interact with teachers. The substantial differences in costs of the three interventions make it especially important to compare them in terms of both benefits and costs.

2. For recent examples of the evaluation of social sector programs based on randomized control designs, see Newman, Rawlings, and Gertler (1994). See Glewwe, Kremer, and Moulin (1997) for a recent application to education.

3. Parent-teacher partnerships envision a more active role for parents than they are commonly assigned, especially in developing countries. See Epstein (1991) for a discussion of how teachers' interactions with parents can improve student achievement.

Under the school feeding intervention all pupils in beneficiary schools received a free school meal while classes were in session. Because of substitution effects, this intervention may not increase pupils' food intake, as Jacoby (1997a, 1997b), for example, suggests. This problem is inherent to all feeding programs. Since we have no way to quantify the amount of substitution between food provided at home and food provided under the DIP, any change associated with the intervention must necessarily be interpreted as its net rather than its gross impact.

Under the multi-level learning materials intervention, all teachers in the beneficiary schools received pedagogical materials designed to help them pace their teaching according to the differing abilities of their students. Prior to implementation of the DIP, teachers attended a week-long training course on the use of the materials. Parent-teacher partnerships comprised a series of regular (usually monthly) group meetings throughout the school year between school staff and parents. The authorities chose to implement parent-teacher partnerships in combination with one of the other two interventions (rather than on its own) because the other interventions provided a way to attract parents to the meetings and provided the substantive focus for meetings.

The DIP project team, which was part of the Bureau of Elementary Education, followed a three-stage procedure in selecting schools for the interventions. They first identified five regions of the country and, within each region, two districts that met the official definition of a low-income municipality (the municipality had to meet at least three of five poverty criteria relating to education, health, housing, unemployment, and household consumption). The sample schools were located in 10 provinces: Mindoro Oriental and Palawan in Southern Luzon Region, Camarines Sur and Sorsogon in Bicol Region, Ilioilo and Negros Occidental in Western Visayas Region, Northern Samar and Western Samar in Eastern Visayas Region, and North Cotabato and Maguindanao in Central Mindanao Region. In one district the treatment choices were packaged as no intervention, multi-level learning materials, or multi-level learning materials combined with parent-teacher partnerships, while in the other district they were packaged as no intervention, school feeding, or school feeding combined with parent-teacher partnerships. The decision of which of the two intervention packages to assign to each site in the region was made by the toss of a coin.

Next, in each district the project team selected three schools that met the following criteria: each school offered all grades of instruction in the elementary cycle, with one class of pupils per grade; had a high dropout rate, based on administrative records; was not located in an area with security risks; and did not offer any school feeding services. Each school was typically the only school in its locality. Finally, by random drawing, the three schools in each district were assigned to the control group or to one of the two intervention options. The process generated a sample of 20 intervention schools and 10 control schools. One school from the control group was eventually dropped because of logistical difficulties in collecting data.

The use of random assignment yields evaluation results that are both convincing to researchers and easy for policymakers to understand (Burtless 1995). Heckman and Smith (1995) point out that selection bias may remain a problem in randomized trials because people in treatment groups may opt out of the treatment and those in the control group may compensate for their exclusion from the experiment. In the DIP evaluation, schools in the treatment group could not select into or out of the assigned interventions, and schools in the control group could not substitute other types of educational interventions to compensate for not being in the treatment group. However, because resources were scarce, the project team found it necessary to strike a balance between the priority of addressing pressing needs in poor schools and the advantage for program evaluation of having complete randomization in the placement of interventions. In the end the program team decided to target the interventions to needy districts and schools. Although randomized selection was not used to identify the two districts in each region nor the schools in each district, it was the basis for assigning the two intervention packages to the sites in each region and for assigning treatment or control status to schools.

II. Data Collection

The interventions were implemented in the 1991–92 school year, but data collection began in 1990–91 to generate baseline information (table 1). Data were collected from all pupils in all grades in each sample school. Because the schools had only one section per grade, being poor rural schools, the resulting data set has information on the full population of students.[4]

The data include the characteristics of the schools and the classroom environment (including teacher characteristics), as well as information about the pupils: family and personal background; scores on grade-specific tests in mathematics, English, and Filipino, with one set of tests administered at the start of the school year and a second set administered at the end; and transition to the next school year. The survey also attempted to record students' daily attendance throughout the school year, but the data proved unreliable because of poor record-keeping. The data on transition to the next grade comprised two kinds of information: the school management's year-end decision to promote the pupil to the next grade or retain him or her in the same grade, and whether or not the pupil actually returned the next school year.

The data are two-year records of pupils' transition through school for those who remained in the sample for both years of the study—that is, those who entered first through fifth grade in year one and who did not drop out. (There were very few transfers to other schools because most of the schools in the project sites were the only ones in their locality). The transition record is truncated for

4. We account for this feature of the sampling procedure in the multivariate regression analyses below by allowing for a school- or teacher-level structure in the variance-covariance error term.

Table 1. *Sample Composition in the Philippine Dropout Intervention Program, 1990–92*

Intervention	Number of schools	Number of pupils[a] 1990–91	1991–92
Full sample	29	4,267	3,953
Program intervention			
No intervention	9	1,356	1,279
School feeding			
Alone	5	751	695
With parent-teacher partnership	5	858	792
Multi-level learning materials			
Alone	5	673	634
With parent-teacher partnership	5	629	553

a. Includes only pupils in grades one to five with data on personal and family background who advanced to the next school year.

Source: Survey data from the 1990–92 Dropout Intervention Program.

sixth graders in year one and first graders in year two. Sixth graders had left primary school by the second year, and their schooling career was not tracked; as a result, their data are for year one only. First graders in year two have no data for year one, since they were not yet in primary school.

As with any complicated effort to collect data that involves many actors, the data that eventually became available had some shortcomings. Unlike the dropout and transition data, the data on student performance were collected for one year only. The original intention was to gather longitudinal data, but unforeseen coding problems prevented that.[5] Thus only the achievement data for year one—comprising scores at the start and at the end of the year—could be linked to the data on student background. Fortuitously, in year two the same achievement tests were administered to the new cohort of entering first graders. Adding these pupils to the first graders from year one produced a data set containing the information needed to evaluate the impact of the interventions on student learning among first graders. Since schooling outcomes in first grade turn out to be especially relevant to elementary education policy in the Philippines—that is, the dropout problem is concentrated in the first grade—the lack of suitable data for the other grades proved to be a less serious flaw than appeared at first sight.

To confirm that the DIP interventions were in fact randomly assigned across schools, we compared dropout rates and student learning as well as students' socioeconomic background in treatment schools and control schools prior to the implementation of DIP. The results suggest that in terms of the outcome variables—dropout rates and year-end test scores—the treatment schools are not significantly different from the control schools (table 2). The random assignment

5. The identification codes on the achievement files from the second year lacked sufficient detail to permit secure matches to the data from the first year. The files on attendance and transition status were collected using a separate procedure and did not suffer from this flaw.

Table 2. *Pupils in Control and Treatment Schools Prior to Implementing the Dropout Intervention Program*

Variable	Control schools	Schools that received the school feeding intervention: Alone	Schools that received the school feeding intervention: With parent-teacher partnership	Schools that received the multi-level learning materials intervention: Alone	Schools that received the multi-level learning materials intervention: With parent-teacher partnership
Outcome variables					
Mean dropout rate (percent)	9.56	8.58	7.02**	9.29	10.01
Mean *z*-score on year-end test[a]	0.02	0.01	0.07	–0.10	–0.10
Student characteristics					
Percent repeating current grade	0.21	0.28*	0.23	0.24	0.27
Percent attended preschool	19.8	14.6*	10.6**	21.1	13.5**
Percent whose father is a farmer	47.2	53.6*	47.9	36.2**	44.9
Percent from non-Tagalog-speaking homes	48.9	52.4*	57.1**	33.1**	32.0**
Mean years of mother's schooling	6.0	5.8*	5.5**	6.3**	6.1
Mean number of brothers and sisters	5.2	5.1	4.7**	5.4	4.7*
Mean *z*-score on entering test[a]	0.06	–0.14*	0.19**	0.15	–0.41*

* Deviation from the control group is statistically significant at the 5 percent level.

** Deviation from the control group is statistically significant at the 1 percent level.

a. Test scores are for first graders only. They are expressed in units of standard deviation from the sample mean.

Source: Authors' calculations.

process thus appears to have been valid, implying that a simple analysis of the differences between the mean impacts on the control and treatment groups would capture average treatment effects.

The schools are less similar, however, with regard to pupil characteristics. In particular, children in schools that received the school feeding program appear to be systematically less well off than children in the control schools. The presence of such differences is not surprising, given the relatively small number of schools in the sample. Below we use multivariate methods to control for these differences in evaluating the impact of the DIP interventions.

III. The Impact of the DIP Interventions

Dropout rates decline in all sample schools between the pre- and post-treatment years (table 3). However, the decline is statistically significant only in the schools that received multi-level learning materials, with or without parent-teacher

partnerships. To isolate the pure intervention effect, we compute the difference in the change in the dropout rate over time between each treatment group and the control group and then perform *t*-tests on the resulting difference-in-difference estimates. In the schools with a feeding program, for example, dropout rates decline 2.9 percentage points between year one and year two compared with a decline of 1.2 percentage points in the control group. The *t*-test on this difference-in-difference estimate (1.7 percentage points) suggests that it is not statistically significant. In contrast, in both treatment schools with multi-level learning materials (with or without parent-teacher programs) the decline in the dropout rate is statistically significant at the 10 percent level or better.

With regard to student learning in first grade, the change between the pre- and post-treatment years is particularly striking for schools that received multi-level learning materials and initiated parent-teacher partnerships, but the change is not statistically significant. Likewise, the difference-in-difference estimate, while positive and relatively large, is also not statistically significant. For the other treatment groups the change between the baseline and treatment years is more modest, and none of the estimates of the program's impacts is statistically significant. The lack of statistical significance is not surprising, however, given that we have mean test scores on only five classes of pupils for each intervention.

Table 3. *Impact of the Dropout Intervention Program on Schooling Outcomes between 1990–91 and 1991–92*

		School feeding		*Multi-level learning materials*	
Variable	*Control schools*	*Alone*	*With parent-teacher partnership*	*Alone*	*With parent-teacher partnership*
Dropout rates					
Percentage change	−1.2	−2.9	−2.8	−4.8	−6.4
P-value	0.328	0.104	0.110	0.004***	0.005***
Difference-in-difference estimate[a]	n.a.	−1.7	−1.6	−3.6	−5.2
P-value	n.a.	0.440	0.465	0.080*	0.028**
Student achievement[b]					
Change in *z*-score	0.11	−0.01	0.04	−0.17	0.47
P-value	0.787	0.989	0.910	0.809	0.240
Difference-in-difference estimate[a]	n.a.	−0.12	−0.07	−0.28	0.36
P-value	n.a.	0.839	0.902	0.705	0.500

n.a. Not applicable.
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.
a. The difference-in-difference estimate refers to the difference between the treatment and control groups in the change in dropout rates or test scores.
b. Data are for first graders only. *z*-scores are test scores expressed in units of standard deviation from the sample mean.
Source: Authors' calculations.

These results suggest that the impact of the DIP interventions is ambiguous, positively affecting dropout behavior, but not influencing student learning. Given the systematic differences in student characteristics between the treatment and control groups, as well as the small samples involved, such a pattern is expected. Including more test sites for each of the DIP interventions would have brought more clarity, but it also would have required many more resources than the government was willing or able to allocate to the exercise. Indeed, in most developing countries large-scale experiments are rarely affordable as routine procedures for policy analysis. We can nonetheless exploit the data generated from the DIP's experimental design by conducting a multivariate analysis to control for the systematic differences between the control and intervention groups. For this analysis we use pupils rather than schools as the unit of observation.

Multivariate Analysis

As above, we focus on dropout behavior and student learning as the relevant outcomes for evaluating the DIP interventions. Following the literature (for example, Hanushek and Lavy 1994), we postulate that the probability that child i drops out of school s at time t (DP_{ist}) depends on the child's personal characteristics (PC_i) and family background (FB_i), the learning environment (LE_{st}), and the characteristics of the community (CC_i) in which he or she lives. Each DIP intervention j ($INTER_j$) can affect both dropout behavior and student learning. The school feeding intervention, for example, lowers the cost of schooling, thus boosting the incentives for parents to keep children in school. At the same time, it can improve children's general health and attentiveness, thereby stimulating academic progress and lowering the chances of dropping out. The provision of multi-level learning materials aims to improve the effectiveness of the pedagogical process. Thus it could potentially heighten children's interest in school as well as enhance their learning. Both would reduce the probability of dropping out. Finally, parent-teacher partnerships expand parents' involvement in the schooling of their children, thereby minimizing the influence of adverse social and academic factors on dropout behavior.

More formally,

$$(1) \qquad DP_{ist} = \beta_0 + \beta_1 PC_i + \beta_2 FB_i + \beta_3 LE_{st} + \beta_4 CC_i + \beta_5 INTER_{jt} + \varepsilon$$

In measuring the impact of the DIP on student learning, we again follow the literature (for example, Harbison and Hanushek 1992) in postulating that child i's academic performance in school s at time t, AP_{ist}, is a function of his or her initial achievement, AP_{ist-1}, personal characteristics (PC_i), and family background (FB_i), as well as the learning environment (LE_{st}), and community characteristics (CC_i):

$$(2) \quad AP_{ist} = \delta_0 + \delta_1 AP_{ist-1}, + \delta_2 PC_i + \delta_3 FB_i + \delta_4 LE_{st} + \delta_5 CC_i + \delta_6 INTER_{jt} + \varepsilon$$

Two important econometric issues—discussed fully in Angrist and Krueger (1999) and Vella (1998)—arise in the estimation of equation 2. The first is the

presence of a lagged dependent variable, which, while providing an important control factor, may be correlated with the error term. This problem can be addressed by choosing instrumental variables that are correlated with the lagged variable but not with the error term. We choose the lagged values of scores on other tests as instruments. Of course, the reduction in bias comes at the expense of a loss of efficiency, and the reliability of this approach depends on the validity of the instruments. In our case the r-squared values of the correlation between the instruments and the lagged values range from 0.43 to 0.50, suggesting some loss of efficiency and a consequent downward bias in the t-statistics.

The second econometric issue is that of selection bias associated with dropping out. Since the weakest students are those who are most likely to drop out, the analysis of student learning is performed on a censored sample. Although the dominant method of correcting for this problem is to apply a Heckman two-step correction by constructing an index based on the probability of censoring, there is some dissatisfaction with this approach, as Hamermesh (1999) describes.[6] The index is based on the assumption of a normal distribution, and, as Hamermesh notes, the results are extremely sensitive to distributional assumptions. Further concerns are often raised about the choice and adequacy of identifying variables in the first step, although in our case we are fortunate in having detailed information on variables in equation 1 that affect the cost of education but are not directly associated with a child's academic performance. In particular, we have data on the distance to school, whether or not the father is a farmer (since an important opportunity cost of schooling in poor rural communities is children's contribution to farm work), and whether or not the student is the oldest child in the family.

A number of alternative approaches have been proposed to deal with selection bias—nonparametric and semiparametric methods, as well as nonindex-oriented models—although no consensus has yet emerged as to which is preferred. Consequently, we estimate and report the results from three separate procedures.[7] As a basis for comparison, we first report the results of a simple regression of year-end test scores against the intervention variables and control factors, with no correction for selection bias. Then we use a nonindex instrumental variable approach, following Krueger (1997), in which we simply assign to students who have dropped out their academic ranking based on their initial test score. Finally, we follow the conventional Heckman approach, which includes the Mills ratio as an additional regressor in equation 2. It should be noted that in the Heckman approach the standard errors are biased because it is impossible to estimate them correctly when simultaneously applying that procedure and using instrumental variables to correct for the problem of lagged dependent variables.

6. Noteworthy, however, is Vella's (1998) finding that the Heckman approach performs quite well compared with other approaches.

7. In all three approaches we correct for the problem of having a lagged dependent variable by using the instrumental variables procedure described above.

To estimate equation 1 we use data for pupils from the first to fifth grades in order to increase the sample size for analyzing what in statistical terms is still a relatively rare event. We could not include pupils from the sixth grade because their dropout record is incomplete for reasons explained earlier. We represent a pupil's characteristics and family background as a vector of commonly used variables, such as the child's sex and mother's education.[8]

Regression Results

The full dropout regression model, with controls for the complete range of background factors, achieves a reasonable overall goodness-of-fit, with a Hosmer-Lemeshow chi-squared statistic of 13.28 (column 2 of table 4). We also estimate a simplified model using only the interventions as regressors in order to see whether the interventions were correlated with the background variables (column 3 of table 4).[9] Not surprisingly, the simplified regression model as a whole has no explanatory power; it is nonetheless noteworthy that all of the coefficient estimates are comparable to the corresponding estimates in the full model. Moreover, in both regressions only the intervention involving the use of multi-level learning materials has a measurable effect on dropout behavior. The positive impact of this intervention is consistent with the results based on sample means (see table 3). Those results also suggest that interventions combining the use of multi-level learning materials with parent-teacher partnerships have a positive impact.

We then estimate equation 2 for the three different subjects—mathematics, Filipino, and English. Students were given tests in these subjects at both the start and end of the school year (table 4).[10] The first feature of the results is that no intervention consistently improves student learning across all three subjects, a finding that jibes with the data in table 3, which are based on average performance across the three subjects. For both Filipino and English the coefficients on the intervention involving multi-level learning materials combined with parent-teacher partnerships are statistically significant in all three regressions. The coefficients are comparable in magnitude, especially in the regressions for English.

For English the coefficients on the school feeding intervention, whether alone or combined with the parent-teacher partnerships, are statistically significant only in the two regressions that control for selection bias (the second and third columns in each subject block). For mathematics the coefficient on school feeding combined with parent-teacher partnerships is also statistically significant in both regressions adjusted for selection bias. Overall, the findings suggest that the DIP interventions are better at helping students learn languages than mathematics. Further, the interventions involving the use of multi-level learning materials, as

8. Father's education is almost perfectly collinear with mother's; we use the latter because the mother is more likely to provide help with homework.

9. We thank an anonymous referee for suggesting this specification.

10. In addition to the variables mentioned above, we also include teacher and school fixed effects. For each subject we use incoming test scores on the other two subjects to instrument the incoming test score.

currently designed and implemented, appear to produce more consistent results than do the other components.

IV. Policy Implications

To assess the policy implications of the DIP, we need to consider both its cost and impact. We can assess these qualitatively with available cost data and our interpretation of the regression results discussed above (table 5). The underlying cost data are drawn from implementation records kept by the Bureau of Elementary Education. The school feeding program costs an average of P946 (pesos) per beneficiary, with a range between P621 and P1,054. The multi-level learning materials program involved only pedagogical materials, which cost an average of P90 per child, and the parent-teacher partnerships involved monthly meetings that cost an average of about P33 per child (in direct costs). To be perfectly comparable, these costs should be adjusted for the opportunity cost of time—that of teachers supervising the school feeding program, that of parents and teachers in parent-teacher partnerships, and that of people who train teachers to use the multi-level learning materials. The cost of the multi-level learning materials program should also be adjusted to reflect the fact that the pedagogical resources it provides have a typical lifetime of more than one year. Without making the adjustments explicit, however, the existing cost data already point to an obvious ranking of the interventions, with school feeding at the high end, followed by multi-level learning materials and parent-teacher partnerships at the low end.

The impact of the interventions on dropout behavior and student learning range from nonexistent (0) to promising (++++). The ranking reflects our interpretation of the results from the difference-in-difference estimates and the regression analysis. If none of the analyses shows an appreciable impact, we categorize the intervention as having zero impact; if the results are consistently positive across all or most of the estimation methods and the magnitude of the impact is relatively large, we label the intervention promising and assign it three or four pluses to reflect the degree of consistency; and if the results are sporadically positive, we categorize the intervention as having a weak impact, assigning it only one or two plus signs.

Given these cost and benefit criteria, the combination of multi-level learning materials and parent-teacher partnerships appears to be the most cost-effective. In contrast, the school feeding intervention, at least in the form implemented in DIP, seems to be a weak candidate for replication. This does not imply that a more targeted program—directed, for example, at only the most malnourished and underprivileged children—would not be cost-effective. That possibility cannot be confirmed, however, with the data available here. Note, though, that the impact on student learning of multi-level learning materials in combination with parent-teacher partnerships is probably limited to instruction in languages.

Table 4. *Regression Results of the Impact of the Dropout Intervention Program*

			Year-end test scores (first graders only)								
			Math			*Filipino*			*English*		
				Correction for selection bias			*Correction for selection bias*			*Correction for selection bias*	
	Probability of dropping out (first to fifth graders)		*No correction for selection bias*	*Using nonindex instrumental variable approach*	*Using Heckman's approach*	*No correction for selection bias*	*Using nonindex instrumental variable approach*	*Using Heckman's approach*	*No correction for selection bias*	*Using nonindex instrumental variable approach*	*Using Heckman's approach*
Independent variable	*(1)*	*(2)*									
Intervention variables[a]											
School feeding	–0.254	–0.255	0.241	0.248	0.121	0.317	0.160	0.031	0.317	0.323	0.009
	(0.56)	(1.26)	(0.77)	(2.72)**	(1.36)	(1.80)	(1.82)	(0.32)	(1.80)	(3.63)**	(3.73)**
Multi-level materials	–0.428	–0.458	–0.092	–0.045	–0.008	0.647	0.234	0.178	0.647	0.548	0.543
	(1.71) *	(1.99)*	(0.18)	(0.38)	(0.07)	(1.66)	(2.05)*	(1.42)	(1.66)	(4.71)**	(0.66)
School feeding with parent-teacher partnerships	–0.311	–0.319	0.370	0.347	0.277	0.458	0.114	0.058	0.458	0.442	0.544
	(1.40)	(1.63)	(0.84)	(3.74)**	(3.08)**	(1.63)	(1.28)	(3.31)**	(1.63)	(4.89)**	(1.66)*
Multi-level materials with parent-teacher partnerships	–0.410	–0.367	0.217	0.210	0.081	0.870	0.225	0.309	0.870	0.754	1.048
	(1.15)	(1.56)	(1.50)	(1.83)	(0.76)	(3.12)**	(2.02)*	(2.65)**	(3.12)**	(6.64)**	(8.83)**
Initial test scores (instrumented)[b]			0.510	0.607	0.520	0.373	0.618	0.473	0.373	0.485	0.341
			(8.69)**	(18.07)**	(15.16)**	(5.11)**	(18.51)**	(13.98)**	(5.11)**	(13.83)**	(9.99)**
Pupil is a girl	–0.172		0.153	0.116	0.171	0.290	0.224	0.246	0.290	0.240	0.181
	(1.61)		(2.64)*	(2.95)**	(4.43)**	(4.57)**	(5.89)**	(4.46)**	(4.57)**	(6.24)**	(2.08)*
Attended preschool	–0.211		–0.151	–0.173	–0.096	–0.014	–0.045	0.468	–0.014	–0.039	0.500
	(1.08)		(2.12)*	(2.96)**	(1.61)	(0.25)	(0.79)	(3.57)**	(0.25)	(0.69)	(1.18)
Mother's years of education	–0.072		0.013	0.016	0.007	0.001	0.010	0.004	0.001	0.000	–0.006
	(3.22)**		(1.16)	(2.11)*	(0.87)	(0.15)	(1.32)	(0.44)	(0.15)	(0.03)	(2.22)*
Father is a farmer	–0.070		0.026	0.024	–0.035	–0.103	–0.002	0.143	–0.103	–0.122	–0.081
	(0.52)		(0.26)	(0.47)	(0.74)	(0.86)	(0.04)	(1.28)	(0.86)	(2.44)*	(1.52)
Non-Tagalog speaker	–0.140		–0.031	0.000	0.028	–0.005	0.183	0.014	–0.005	0.009	0.034
	(0.86)		(0.26)	(0.01)	(1.04)	(0.03)	(3.33)**	(2.85)**	(0.03)	(0.16)	(0.54)

Eldest child	0.092										
	(0.70)										
Repeated previous grade	0.327										
	(2.68)**										
Child has active personality[c]	–0.133										
	(0.62)										
Family income (pesos per year)	–0.109										
	(1.24)										
Distance to school (kilometers)	0.450										
	(0.88)										
Distance squared	–0.125										
	(0.97)										
Hosmer-Lemeshow χ^2	13.28	0.00									
(p-value)	(10.25)	(1.00)									
Inverse Mills ratio					–0.087			–0.351			–0.471
(standard errors)					(0.178)			(0.122)			(0.099)
Number of observations	8,229	8,229	1,676	1,676	1,676	1,676	1,676	1,676	1,676	1,676	1,676
R-squared			0.41	0.41		0.43	0.45		0.42	0.44	

* Statistically significant at the 10 percent level.

** Statistically significant at the 5 percent level.

*** Statistically significant at the 1 percent level.

Note: The t-statistics (in parentheses) are consistent with standard errors adjusted for group-specific heteroskedasticity using the Huber-White correction procedure. For the dropout regression the standard errors are adjusted for clustering on schools. For the test score regressions the standard errors are adjusted for clustering on teachers. The dropout regressions include dummy variables for grades.

a. The intervention variables are defined as dummy variables that take on the value of 1 when the school attended by the child is a recipient of the indicated intervention and zero otherwise.

b. Tests scores at start of the school year, instrumented by scores on the other two subjects.

c. As assessed by pupil's teachers.

Source: Authors' calculations.

Table 5. *Policy Implications of the Dropout Intervention Program Evaluation*

Intervention	*Costliness of intervention*	*Impact of intervention*[a]	
		On dropout behavior	*On test scores*
School feeding	High	0	+ +
Multi-level learning materials	Low	+ + + +	+
Parent-teacher partnership with school feeding	High	0	+ +
Parent-teacher partnership with multi-level learning materials	Low	+ +	+ + + +

a. A rating of 0 indicates no impact, one or two pluses indicates a weak impact, and three or four pluses indicates a strong impact.

Source: Authors' calculations.

V. Conclusions

The DIP represents an important effort by the Philippine government to experiment with new ways to address problems in elementary education. To evaluate the program, the government collected pre- and post-treatment data on pupils in test schools and in control schools. We used these data here to assess how the DIP interventions affect dropout behavior and learning. Data coding problems limited the analysis of student achievement to first graders only.

Dropout rates and student achievement in the control and treatment schools are comparable in the baseline year, suggesting that the random assignment process worked as expected. However, the schools differed in the background characteristics of pupils, so that a simple comparison of mean differences between the control and treatment schools before and after implementing the DIP was not sufficient to establish the true impact of the interventions. The fact that the sample included only five classes of pupils per intervention also hampered the analysis. Both deficiencies prompted us to use multivariate analysis to complement the comparison of means in the control and treatment schools.

The evaluation period is admittedly too short to reach firm conclusions, but the preliminary findings reported in the paper offer a good basis for assessing the economic justification for replicating the DIP interventions. Taken as a whole, they imply that, of the four interventions implemented, the argument for replication is strongest for multi-level learning materials combined with parent-teacher partnerships, and weakest for the school feeding program, at least as it was implemented in the DIP.

It is important to note that if improved student learning is an objective, the combination of multi-level learning materials and parent-teacher partnerships is only one of many potential interventions (such as offering preschool education, expanding teacher training, improving classroom conditions, and supplying more student textbooks and workbooks). Thus although our evaluation offers some support for replicating one of the DIP experiments, it by no means implies that we have finished our search for ways to address dropout and student learning prob-

lems in the Philippines. Further, given that the experiment was effective mainly in promoting student performance in Filipino and English, other measures clearly need to be considered if improved performance in mathematics is also desired.

The search for cost-effective strategies to improve schooling outcomes is an issue in all countries. Often, however, such work is hampered by the absence of a routine system for assessing alternative investment options. The fact that the DIP was implemented and evaluated augurs well for the future, the problems encountered notwithstanding. It shows that the institutional capacity to evaluate social experiments properly exists or can be nurtured within ministries of education or related agencies. The task of building research capacity, particularly when the institutions involved are outside of academia, is undeniably difficult. The benefits are probably worth the effort, however, because the scope for mistakes in the choice of investment decisions is wide in the absence of quantitative information about education processes. And these mistakes are costly, not only in financial terms, but also in terms of hindering children's education.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Angrist, Joshua, and Alan Krueger. 1999. "Empirical Strategies in Labor Economics." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*. Amsterdam: North-Holland Press.

Burtless, Gary. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *Journal of Economic Perspectives* 9(2):63–84.

Chuard, Dominique, and Alain Mingat. 1996. "Analysis of Dropout and Student Learning in Primary Education in South Asia." Asian Development Bank, Manila. Processed.

Epstein, Joyce L. 1991. "Effects on Student Achievement of Teacher Practices of Parent Involvement." In Steven Silvern, ed., *Advances in Reading/Language Research*. Vol. 5: *Literacy through Family, Community, and School Interaction*. Greenwich, Conn.: JAI Press.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 1997. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Development Economics Research Group, World Bank, Washington, D.C. Processed.

Gomes-Neto, Joao Batista, and Eric A. Hanushek. 1994. "The Causes and Consequences of Grade Repetition: Evidence from Brazil." *Economic Development and Cultural Change* 43(1):117–48.

Hamermesh, Daniel S. 1999. "The Art of Labormetrics." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*. Amsterdam: North-Holland Press.

Hanushek, Eric A., and Victor Lavy. 1994. "Schooling Quality, Achievement Bias, and Dropout Behavior in Egypt." Living Standards Measurement Study Working Paper 107. Development Economics Research Group, World Bank, Washington, D.C. Processed.

Harbison, Ralph W., and Eric A. Hanushek. 1992. *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. New York: Oxford University Press.

Heckman, James J., and Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9(2):85–110.

Jacoby, Hanan. 1997a. "Is There an Intrahousehold 'Flypaper Effect'? Evidence from a School Feeding Program." Jamaica School Feeding Program. Processed.

———. 1997b. "Self-Selection and the Redistributive Impact of In-Kind Transfers." *Journal of Human Resources* 32(2):233–49.

Krueger, Alan. 1997. "Experimental Estimates of Educational Production Functions." NBER Working Paper 6051. National Bureau of Economic Research, Cambridge, Mass. Processed.

Lockheed, E. Marlaine and Adriaan M. Verspoor. 1991. *Improving Primary Education in Developing Countries*. Washington D.C.: Oxford University Press for the World Bank.

Miguel, Marcelina M. 1993. "Variables That Affect Student Achievement in the Philippines." Bureau of Elementary Education, Department of Education, Culture, and Sports, Manila. Processed.

Newman, John, Laura Rawlings, and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *The World Bank Research Observer* 9(2):181–201.

Tan, Jee-Peng, Julia Lane, and Paul Coustère. 1997. "Putting Inputs to Work in Elementary Education: What Can Be Done in the Philippines?" *Economic Development and Cultural Change* 45(4):857–79.

Vella, Francis. 1998. "Estimating Models with Sample Selection Bias: A Survey." *Journal of Human Resources* 33(1):127–69.

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO 3: 509-38

Calm After the Storms: Income Distribution and Welfare in Chile, 1987–94

Francisco H. G. Ferreira and Julie A. Litchfield

After rising during most—but not all—of the 1960–85 period, inequality in Chile seems to have stabilized since around 1987. Following the stormy period of economic and political reforms of the 1970s and 1980s, no statistically significant Lorenz dominance results could be detected since 1987. Scalar measures of inequality confirm this picture of stability, but suggest a slight change in the shape of the density function, with some compression at the bottom being "compensated for" by a stretching at the top. As inequality remained broadly stable, sustained economic growth led to substantial welfare improvements and poverty reduction, according to a range of measures and with respect to three different poverty lines. Poverty mixed stochastic dominance tests confirm this result. All of these findings are robust to different choices of equivalence scales.

Two "storms" have recently raged over the distribution of income in Chile. The first, and most important, was caused by a series of structural reforms of the economy—which started in 1974 and were largely completed by the late 1980s—and by successive changes in political regime—which had important implications for, among other things, the regulation of labor markets. The economic reforms included trade liberalization, privatization of state-owned assets, deregulation of various markets, and reforms in the structure of taxes, subsidies, and benefits. They have been extensively discussed elsewhere, and are well beyond the scope of this paper. (See Edwards and Edwards 1987 and Scott 1996 for excellent summaries.) The political changes were fundamentally the military coup d'état of 1973, which installed General Pinochet as president, and the restoration of democracy in 1990, with the election of President Aylwin.

The second storm, closer to the "tea cup" variety, has raged in academic and policy circles, as the effects of the Chilean model on poverty and inequality have been hotly debated. This storm also had two separable components that, for

Francisco H. G. Ferreira is with the Economics Department at the Catholic University of Rio de Janeiro, and Julie A. Litchfield is with the Economics Department at the University of Sussex. Their e-mail addresses are fferreira@econ.puc-rio.br and j.a.litchfield@sussex.ac.uk. The authors are grateful to Dante Contreras, Indermit Gill, Jesko Hentschel, Osvaldo Larranaga, Chris Scott, Tony Shorrocks, Miguel Szekely, Alberto Valdes, Michael Walton, two anonymous referees, and participants at conferences in Santiago and London for their very useful comments and contributions. The authors particularly wish to thank Iris Delgado at MIDEPLAN and Juan Carlos Feres at CEPAL for being so generous with their vast knowledge of the details of the CASEN data.

convenience, we will name "Chile versus Stolper-Samuelson" and "See, All is Not Well After All." The former component originates from the finding that, as Chile liberalized its trade regime, the ratio of the wages of skilled workers to those of unskilled workers rose, rather than fell (Robbins 1994). If the Chilean and the world economies could be approximated by a model in which unskilled and skilled labor were the only two factors of production, and the other Hecksher-Ohlin assumptions held (notably constant technology and the absence of non-tradable goods), then this finding would violate the predictions of the Stolper-Samuelson theorem, which establishes the link between goods and factor prices in a Hecksher-Ohlin world. If Chile, like other developing countries, had relatively abundant supplies of unskilled labor, opening the country to trade should have increased the returns to this factor relative to the returns to skilled labor—the opposite of what Professor Robbins finds.

The second strand, "See, All is Not Well After All," draws on Robbins's findings, but percolates beyond academia to Chilean politics and society more broadly. The *Concertacion* governments of Presidents Aylwin and Frei made "growth with equity" their paramount objective. Failure to promote equity—indeed, the observation of actual increases in inequality—can be seized on as evidence of their failure to deliver on their stated objectives. Income distribution statistics became increasingly important in the Chilean political and social debate of the mid-1990s, to the point where the Catholic Church's *Conferencia Episcopal de Chile* issued an open letter in January 1996 entitled, "Is Chile an Equitable Country?" Its concluding section began by stating that, "The current distribution of income in our country should be cause for scandal among Christians . . . The distance between the rich and the poor has grown in Chile in an alarming fashion. A caring nation cannot accept this reality. Neither can a modern country tolerate these differences." (*Comision Nacional de Justicia y Paz* 1996:24).

Indeed, although there was no single *national* household survey that *regularly* collected detailed information on incomes *from all sources* prior to 1985, the dominant view is that inequality in Chile did rise substantially throughout the 1960s, 1970s, and 1980s (except for a brief decline in 1970–73).[1] This information is largely based on the long time series of the Universidad de Chile Household Survey data, which covers only Greater Santiago. Riveros (1983) was the first to note the rising trend in the Gini coefficient, focusing on the period from 1958 to 1982. This was corroborated by Robbins (1994), and by Montenegro (1996), for wage incomes, although both authors find that the trend changed in 1990. More recently, using the only national data available, Londono and Szekely (1997) confirm these findings for total household income across the entire country. Data points satisfying these requirements prior to 1990 are available only for 1971, 1980, and 1989. Londono and Szekely (1997) report that the Gini coefficient rose over those three years from 0.47 in 1971, to 0.53 in 1980, and to 0.59 in 1989. They, too, find a reversal beginning in 1990, with the Gini falling to 0.57 in 1994.

1. For a dissenting view see Marcel and Solimano (1994), who emphasize the limited variation in income distribution across different presidential mandates in Chile since the 1950s.

Nevertheless—although Robbins (1994) finds a decline in the skilled-unskilled wage ratio from 1990 to 1992, Montenegro (1996) finds that the wage Gini for greater Santiago fell from 0.57 in 1987 to 0.46 in 1996, and Londono and Szekely (1997) report the aforementioned small decline in the national total income Gini between 1989 and 1994—there does not yet seem to be a consensus on the post-1990 part of the story. Robbins (1995) himself plays down the post-1990 decline, attributing it to the rise in the number of graduates from "low-quality" private universities. The bishops who wrote the 1996 letter also clearly doubt that there has been an improvement since the turn of the decade. Having become accustomed to three decades of rising inequality, it is as if Chilean society (and some foreign economists) refuse to believe that a reversal is possible, and wonder what economic mechanisms might underlie it. Meller (1996) suggests that a reduction in wage disparities might be due, after all, to Stolper-Samuelson at work. Robbins (1996) and Wood (1997) remain unconvinced.

In part, room for this controversy is generated by the absence of a thorough, definitive analysis of the best available data since 1987. Such an approach should be based on as solid a treatment of the data as possible, taking into account issues such as regional price variations and differences in family composition. It should also overcome the ambiguity inherent to inequality analysis based on scalar measures: income share ratios might fall, while Ginis rise; coefficients of variation might suggest increasing dispersion, while the mean log deviation suggests declines. And it should deal explicitly with the statistical nature of inequality analysis: some of the changes observed, particularly in the 1990s, have been so small that one wonders whether or not they are statistically significant.

This article aims to fill that gap, and to contribute to the debate on the dynamics of personal income distribution in Chile, by establishing which facts and conclusions are indeed empirically robust. It provides a comprehensive description of the levels of and changes in poverty and inequality in Chile from 1987 to 1994, drawing primarily on a detailed analysis of four household survey microdata sets—the *Caracterización Socioeconómica Nacional* (CASEN) surveys of 1987, 1990, 1992, and 1994 (these are discussed briefly in section I below, and described in more detail in appendix A). We address the ambiguity of scalar measurement of welfare, poverty, and inequality by relying on stochastic dominance techniques. The article incorporates the need for statistical testing of hypotheses about changes in income distribution by applying the Howes (1993) intersection-union test for statistical significance of stochastic dominance. It also addresses the issue of robustness with respect to the assumptions made about different needs across households and economies of scale within households, by presenting all results for equivalized income, as well as for per capita income.

I. Concepts, Data, and Methodology

In studying the evolution of the Chilean distribution of income during 1987–94, this article presents results concerning three distinct, but related concepts:

social welfare, poverty, and inequality. Social welfare is perhaps the vaguest of the three: it seeks to capture the level of well-being of a society or population. Because of information constraints, it is usually proxied in household survey studies by income or by consumption expenditures.[2] Many have argued for more encompassing measures, which would include a value for leisure or other non-monetary dimensions of quality of life, such as the availability of public goods, whether environmental in nature or not.[3] Although including these dimensions would be ideal in principle, we are constrained in practice—by data availability and reliability—to relying on either consumption expenditures or income for detailed distributional analysis.

Since the CASEN surveys—widely regarded in Chile as the best available sources of information on households since their creation in 1985—use income as the welfare concept, we will here as well. Hence, the social welfare measures we will be concerned with are nondecreasing functions of current income and are symmetric (they do not discriminate among recipients along any dimension other than income). We will not present values for any specific social welfare measures in this study, relying instead on the theorems of stochastic dominance (discussed below) to investigate changes in social welfare.

In seeking to measure welfare, whose current income should one consider? The income distributions conventionally studied in the Chilean literature are household income per capita per household distributions; which is to say that the income unit is household income per capita, and the recipient unit is the household. Using this recipient unit creates a number of inconveniences. For example, higher deciles have smaller shares of the population than lower deciles because household size declines with household income per capita in Chile (see MIDEPLAN 1992:34).[4] Or, the headcount poverty measure is less than the proportion of the

2. There are two reasons why consumption expenditures are usually preferred to income as a welfare indicator. First, if capital markets work at all, expenditures are a better proxy for permanent income. Second, there is a growing consensus among practitioners that income data from household surveys are less reliable than expenditure data because of a number of problems related to misreporting and mismeasurement. See Deaton (1997) and Chauduri and Ravallion (1994) for discussions. Chile's last available expenditure survey, however, is the *Encuesta de Presupuestos Familiares* of 1987–88. Apart from being old and unable to provide any information on changes during the recent high-growth period, it covered only the Metropolitan Region of Santiago.

3. Becker (1965) proposes a "full income" concept that includes an imputed value for leisure. Sen (1981) proposes a concept of entitlements that aims to capture some of the benefits of public goods, in addition to the value of income. Since 1990, the United Nations Development Programme (UNDP) has been computing a Human Development Index (HDI), which is meant to complement income-based indicators by incorporating life expectancy, adult literacy, and median years of schooling, in addition to per capita income. In Chile life expectancy at birth rose from 71.0 to 72.0 years between 1982 and 1992. Over the same period adult literacy rose from 91.1 percent to 94.6 percent; median years of schooling rose from 5.7 to 7.6. Combining all of these improvements in non-money-metric dimensions of welfare with the income increases discussed below, Chile's HDI registered a 13.5 percent increase from 1982 to 1992, to 0.851. This places the country in the UNDP's "high human development" category. For details see UNDP (1996).

4. Strictly speaking, a decile is a separator. There are nine deciles, and decile i is the income that separates tenth i from tenth $i+1$ in the distribution. Like other quantiles (such as percentiles or quintiles), decile is widely misused to refer to the actual tenth (or hundredth, or fifth) of the distribution. Rather than

population (in persons) that is poor. We therefore adopt the individual as the recipient unit throughout: all of our distributions are vectors of individuals. Their imputed income is one of two income concepts that we adopt: household income per capita, to allow for comparisons with previous studies and for ease of direct interpretation, and household income per equivalent adult, to take account of differences in needs arising from distinct household sizes and compositions. Below we discuss the specific equivalence scale that we use.

Poverty can be thought of as the negative of a welfare function defined over a censored distribution. It is negative in the sense that a rise in poverty, everything else remaining constant, is a decline in welfare. More important, poverty measures are defined over a censored distribution in the sense that they measure the welfare of those below a certain threshold in the overall distribution.[5] That threshold is the poverty line, which, in this article, as in many others, is defined as an absolute income level (fixed over time in real terms) that is judged just sufficient to provide an individual with minimum nutritional and other requirements. Details are discussed in section III.

Inequality, like welfare, is defined over complete distributions of an indicator. Unlike welfare, however, it is independent of the mean of those distributions, concerning itself only with their second moment. Inequality measures are defined over mean-normalized income distributions. They are generally required to satisfy the Pigou-Dalton transfer principle, which demands that the measure rise (or at least not fall) in response to a mean-preserving spread.

There are, of course, many such inequality measures, which may validly rank the same two distributions in opposite ways, if the measures are more or less sensitive to distances in different parts of the distribution. The best remedy to this ambiguity is to rely on stochastic dominance, which is discussed below. But we will also present four different inequality measures, to provide information on changes in inequality from a number of different perspectives. These measures are the Gini coefficient (which is most sensitive to incomes in the middle of the distribution), the mean log deviation (which is most sensitive to incomes at the bottom of the distribution), the Theil index (whose sensitivity is constant across the distribution), and a transform of the coefficient of variation (which is most sensitive to incomes at the top of the distribution). Their formulas are given in section II.

We now turn briefly to the CASEN data sets for the years 1987, 1990, 1992, and 1994, on which the analysis below is based.[6] The CASEN surveys are nationally and regionally representative household surveys, conducted by the Chilean Ministry of Planning (MIDEPLAN), through the Department of Economics of the

confuse the reader by departing from the usual misnomer, decile (and its finer and coarser analogues) will be used to mean tenth (or hundredth, or fifth) in this article.

5. Most poverty measures, including all those used in this study, satisfy the focus axiom, which requires them to be invariant with respect to any change in income levels above the poverty line.

6. A first CASEN survey was conducted in 1985, but it is widely reported to be less comparable with subsequent surveys, in addition to being less reliable and of inferior quality.

Universidad de Chile. They have the dual objectives of generating a reliable portrait of socioeconomic conditions across the country and monitoring the incidence and effectiveness of the government's social programs and expenditures. Given these ends, questions are asked pertaining both to the household and to the individuals within the household. Topics covered include demographics; characteristics of the dwelling; access to utilities and public services; educational attainment (if currently enrolled, detailed questions are asked about the school, method of education financing, benefits, and so on); health conditions; health insurance; health services used and benefits received; occupation and employment; and income. The questions on income are designed to permit the distinction between labor income in cash, labor income in kind (agricultural and nonagricultural), income from capital, rental income, imputed rent, employment-related transfers (such as occupational, invalidity, or widow's pensions), and entitlement transfers (such as the basic pensions, PASIS, or the family allowance, SUF). Details on the sampling methodology employed by the Universidad de Chile, as well as on the adjustments made to the raw data by the United Nations Economic Commission for Latin America and the Caribbean (CEPAL), are provided in appendix A.

The income variable from the CASEN data used as the welfare indicator in this study is total household income, further adjusted in two ways. First, we deflate the income vector by a regional price index, with Santiago as the base location. Traditionally, nominal income has been deflated only by a common national consumer price index, taking no account of regional variations in price levels, which, as table A-1 (in appendix A) indicates, can be considerable in some cases. This is at odds with most current views of best practice in the analysis of income distribution. If average price levels vary considerably across space, nominal currency units have different real purchasing power along that dimension, just as they do over time. The rationale for regional price deflation is exactly analogous to that for temporal price deflation: in comparing an income value for 1994 with one for 1987, one wants to ensure that any variation in average price levels is taken into account, so that a "real" peso buys the same over time. Likewise, in comparing incomes in Santiago with those in Iquique, one wants to adjust them to take into account any variation in average price levels, so that a real peso buys the same across space.[7]

The need for such price deflation is reinforced by Chile's geography: the extreme southern and northern regions have substantially higher average prices than do the regions closer to Santiago. Our regional price deflation is based on the only source of prices outside Santiago, the National Statistical Institute's (INE) *Anuario de Precios* survey of 16 cities.[8] Given the variation in price levels from year to year, we use an

7. See Deaton (1997) and Ravallion and Bidani (1994) on the need for regional price deflation and alternative methodologies.

8. There are no systematic surveys of prices in any rural areas in Chile. Although in the past rural prices were assumed to be lower than prices in urban areas, we find no justification for this arbitrary markdown. We thus generalize urban prices to the whole region.

average of the index from 1985 to 1994 (the values are given in table A-1) . We then deflate regionally adjusted incomes over time by the national consumer price index for November (the survey month) of the relevant years, as given in CEPAL (1995:24). All income values reported in this article, therefore, are expressed in 1994 Santiago pesos (except for incomes in appendix B).

To ensure transparency about the effects of regional price deflation, we present the basic poverty and inequality results of the paper for the undeflated distribution of per capita income in tables B-1, B-2, and B-3 in appendix B. A comparison of those tables with tables C-1, C-2 (in appendix C), and 7, which contain the analogous results for the deflated per capita distribution, reveals that regional price deflation entails a slight reduction in measured mean and median incomes (since it acknowledges the fact that measured prices in the regions are higher on average), a commensurately slight increase in poverty statistics, and only negligible changes in the values of the inequality measures. Income shares are also largely unaltered, except for the bottom two deciles, whose income shares seem to be overstated in the absence of regional deflation. Trends in all variables are also basically unchanged.[9]

The second adjustment we make is to introduce the concept of household income per equivalent adult, by adopting an equivalence scale. We do this to capture the changes in measured inequality and poverty that arise when we take into account the different needs of households with different compositions (such as a household with four adults compared to a household with two adults and two small children), as well as economies of scale that arise from sharing fixed housing or other costs. There are a number of different approaches to deriving an equivalence scale, and there is no single accepted dominant method. Rather than attaching excessive importance to the specific values of our chosen coefficients, we sought to provide a reasonably reliable alternative to the per capita income concept, which is well known to constitute an extreme assumption in terms of differences in needs (it recognizes none), as well as in terms of economies of scale (it assumes these absent).

Our chosen scale is a revised version of the equivalence scale for Chile, calculated by Contreras (1995), using the Rothbarth adult-goods method. Contreras estimates his scale excluding all households with a single adult from the sample and taking two-adult households as the reference type. He finds that adult-good expenditures are restored to the level of a childless couple when incomes for families with one child in a particular age category are raised by a particular percentage amount. The age categories and amounts are given in table 1.[10]

9. Note, however, that it would be wrong to infer from this exercise that regional price deflation is an innocuous or irrelevant procedure. This is not entirely true even for country aggregates, but its real importance would arise in the context of interregional comparisons, such as a poverty profile. Although we do not make such comparisons here, the aggregate results we obtain would be compatible with them.

10. The entries in table 1 are bottom-line approximations. Although they capture the results at an appropriate level of confidence, they do not do justice to the complexity of the estimation method and do not take into account the different standard errors associated with different age categories. See Contreras (1995) for these and other details.

Table 1. *Proportional Compensating Variations for Children in Chile*

Child age (years)	*No child*	*0–4*	*5–10*	*11–15*
Cost increase (percent)	0	15	20	40

Source: Authors' calculations.

Since we must also cover those households made up by a single individual, and in order to take into account some economies of scale within the household, we assume the cost of a single adult to be 60 percent of that of a couple. (This is roughly equivalent to saying that the cost of the second adult is 70 percent of the cost of the first adult.[11]) Our equivalence scale is therefore given by:

$$(1) \quad Y_i = X_i / M_i, \text{ with } M_i = 1.2 + 0.8\,(N_{aa} + N_{11-15}) + 0.4\,N_{5-10} + 0.3\,N_{0-4},$$

where Y_i is household i's equivalized income, X_i is total household income prior to equivalization, M_i is the equivalence scale applied to the household, N_{aa} is the number of additional adults in the household, N_{11-15} is the number of children ages 11–15 (inclusive) in the household, N_{5-10} is the number of children ages 5–10 (inclusive) in the household, and N_{0-4} is the number of children ages 0–4 (inclusive) in the household. Note that this formulation maintains households with two adults as the reference group. Their household income will be divided by two. An additional child in the 11–15 age category "costs" an extra 40 percent, as before. An additional child in the 5–10 age category costs an extra 20 percent, as before. And an additional child in the 0–4 age category costs an extra 15 percent, as before. The second adult accounts for 40 percent of the couple's total costs.

By introducing the equivalence scale and the regional price adjustments directly to the incomes to be analyzed, we ensure consistency in the assumptions underlying the inequality and poverty analyses. Also, since all incomes are effectively expressed in 1994 Santiago pesos, there is no need for regional poverty lines. Lines expressed in 1994 Santiago pesos are the appropriate comparators for all incomes. Similarly, there is no need to develop different poverty lines for different household types; it suffices to compare household income per equivalent adult with an individual adult poverty line.[12] The advantage of this approach over introducing those concepts through different poverty lines, in addition to simplicity, is that the inequality analysis now incorporates regional price and equivalence scale adjustments.

The analysis of inequality and welfare changes in the next section relies on equivalized household income, based on the equivalence scale described above. In order to both preserve comparability with previous studies and demonstrate

11. This is a common assumption, adopted, for instance, in the construction of the Organisation for Economic Co-operation and Development's (OECD's) equivalence scale (see OECD 1982).

12. However, the absolute values of the poverty measures do, of course, depend on the type of reference household chosen in defining the scale.

the robustness of the main results to the adoption of the scale, we replicate the analysis for household income per capita in appendix C.

II. Levels of and Changes in Chilean Inequality and Social Welfare

Before presenting the detailed results, it is worthwhile to highlight the broad picture that arises from the data. This picture can be characterized by three stylized facts. First, the entire distribution function shifts to the right over time, with people in the same relative positions earning higher incomes in later years. Although this is not strictly the case for every segment of the distribution in every year, it is the case in general.[13] This is clearly the result of economic growth.

Second, the dispersion of the distribution seems to remain broadly stable as it moves to the right over this period. If anything, there is some indication of a slight reduction in overall inequality, although there are no significant unambiguous changes in inequality between any of the years surveyed by CASEN.[14] This result would suggest that the benefits of economic growth were distributed in a pattern roughly similar to that of the existing income distribution. Third, to the extent that there are any discernible changes in the shape of the density function, within the broad context of stability, these appear to be a slight compression in the lower tail and a slight increase in dispersion in the upper tail. That is, inequality among the poor fell, while inequality among the very rich, and between the very rich and those just poorer than them, seems to have increased.

As stated in the previous section, the four scalar inequality measures we use in this study are the Gini coefficient, the mean log deviation, the Theil index, and a transform of the coefficient of variation (half of its square, to be precise). In their formulas below we use the following standard notation: y_i is the income of individual i, $i \in (1, 2, ..., n)$; the subscript j denotes any individual other than i; n is the number of individuals in a given distribution; and $\mu(y)$ is the arithmetic mean of the distribution.The Gini coefficient is given by:

$$G = \frac{1}{2n^2\mu(y)} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|. \tag{2}$$

The other three measures are all members of the generalized entropy class of inequality indexes, which satisfy a number of desirable properties, such as symmetry, population replication, scale invariance, and decomposability (see Cowell 1995 for details). The general formula for the parametric class is given by:

$$E(\alpha) = \frac{1}{\alpha^2 - \alpha}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\mu(y)}\right)^{\alpha} - 1\right]. \tag{3}$$

13. The first-order welfare dominance results reported below specify the instances in which it was true for the entire distribution.

14. That is, there are no statistically significant Lorenz dominances between any two years in our sample. See below.

Using l'Hopital's rule, we can obtain $E(0)$, the mean log deviation:

$$E(0) = \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\mu(y)}{y_i}\right). \tag{4}$$

Similarly, the Theil index corresponds to $E(1)$, which is given by:

$$E(1) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\mu(y)}\log\left(\frac{y_i}{\mu(y)}\right). \tag{5}$$

The fourth measure we use is $E(2)$, which can be expressed as:

$$E(2) = \frac{1}{2n\mu(y)^2}\sum_{i=1}^{n}[y_i - \mu(y)]^2. \tag{6}$$

Table 2 lists mean and median incomes, as well as the four inequality measures, for the household income per equivalent adult distribution in each of the four years we analyze.

The impact of economic growth can be seen immediately through the sharp upward trend in mean and median incomes. The large differences between the mean and the median, which persist over the period, are an indication of the skewness of the distribution. The four measures of inequality confirm the high level of inequality in Chile, by international standards.[15]

In terms of temporal evolution, the Gini coefficient, which is not especially sensitive to the top or the bottom of the distribution, changes very little over the period. The Theil index, $E(1)$, behaves slightly differently, displaying a sine pattern. It also does not suggest a strong trend in either direction. The other two measures suggest the small changes referred to above: the mean log deviation, $E(0)$, which is particularly sensitive to low incomes, falls a little more markedly than the Gini throughout. $E(2)$, a transform of the coefficient of variation that picks up differences in the upper tail with greater weight, rises monotonically from the beginning to the end of the period. Overall, there appears to be limited change in inequality, but with greater distances at the top compensating for smaller distances at the bottom. Table 3 allows us to investigate a more disaggregated picture, listing decile shares for household income per equivalent adult. The first row, taken as a memo item from table 2, provides the overall average $\mu(y)$. To

15. Although the average Gini coefficient for Chile in this period—with respect to distributions of household income per capita (see appendix C) for comparability—was 0.5539, a recently compiled international inequality database indicates that the average Gini in the 1980s (1990s) was 0.3323 (0.3375) in industrial countries and high-income developing countries; 0.2501 (0.2894) in Eastern Europe; 0.3501 (0.3188) in South Asia; 0.3870 (0.3809) in East Asia and the Pacific; 0.4045 (0.3803) in the Middle East and North Africa; 0.4346 (0.4695) in Sub-Saharan Africa; and 0.4975 (0.4931) in Latin America and the Caribbean (see Deininger and Squire 1996). Not only is Chile in a different league than industrial countries, or indeed than countries in Asia, but its Gini coefficient is higher than the Latin American average.

Table 2. *Monthly Household Income Per Equivalent Adult and Descriptive Statistics*

Statistic	*1987*	*1990*	*1992*	*1994*
Mean income[a]	67,232	75,007	90,797	93,981
Median income[a]	36,265	42,455	50,212	53,196
Gini coefficient	0.5468	0.5322	0.5362	0.5298
$E(0)$	0.5266	0.4945	0.4891	0.4846
$E(1)$	0.6053	0.5842	0.6151	0.5858
$E(2)$	1.3007	1.3992	1.5050	1.5634

a. In 1994 Santiago pesos.
Note: See text for definition of statistics.
Source: Authors' calculations.

Table 3. *Decile Income Shares of Household Income Per Equivalent Adult* (percent)

Income category	*1987*	*1990*	*1992*	*1994*
Mean income (1994 Santiago pesos)	67,232	75,007	90,797	93,981
Decile 1	1.34	1.39	1.52	1.43
Decile 2	2.41	2.57	2.60	2.57
Decile 3	3.17	3.33	3.38	3.36
Decile 4	3.97	4.19	4.16	4.18
Decile 5	4.88	5.14	5.04	5.14
Decile 6	6.04	6.28	6.16	6.33
Decile 7	7.66	7.92	7.73	7.93
Decile 8	10.24	10.39	10.16	10.55
Decile 9	15.71	15.51	14.82	15.76
Decile 10	44.58	43.28	44.43	42.73
Top percentile	12.02	12.35	13.68	12.41

Source: Authors' calculations.

derive the decile-specific mean income $\mu_d(y)$, simply note that $\mu_d(y) = 10s_d\mu(y)$, where s_d denotes the income share of decile d.

If one looks at the evolution of decile-specific means, calculated as above, one obtains a remarkable confirmation of the gains to all deciles from economic growth. In fact, every decile has seen its average income rise in every subperiod, with only two exceptions: the bottom and top deciles in 1994. This sustained increase in real incomes across the distribution, over a period of seven years, is an achievement most countries would be proud of.

As for the exceptions, the fall in mean income for the first decile in 1994 has been the subject of considerable debate in Chile since the data first became available. This study confirms that the decline also took place if the measure used is household income per equivalent adult. There seems to be little question that it was due, at least in part, to the decline in the overall rate of GDP growth, which was 11.8 percent in the second semester of 1992 and 4.3 percent in the same

period of 1994. In particular, this cyclical deceleration—brought about largely by contractionary monetary policy aimed at curbing inflationary pressures—caused unemployment to increase from 4.8 percent to 6.5 percent.[16] The unemployment rate was much higher in the poorest quintile, relative to the other quintiles, rising there from 18 percent to 22 percent (see Cowan and De Gregorio 1996). Since labor earnings are such an important component of the incomes of the poor, the reduction in demand for unskilled labor, which is behind the rise in unemployment, is bound to have contributed to the recorded decline in their overall incomes.[17] This means, of course, that the sustained increases in social welfare achieved from 1987 to 1992 did not continue unambiguously through 1994, despite continued GDP growth. Similarly, some (low) poverty lines would indicate an increase in poverty from 1992 to 1994, as indeed is the case for some poverty measures with respect to the indigence line (reported in the next section).

Table 3 also sheds light on inequality, as depicted by decile shares. The overall impression is once again of a stable (mean-normalized) distribution, with changes in decile shares being generally small in proportion to the shares themselves. Nevertheless, there is some evidence of a trend of compression at the bottom of the distribution and increased dispersion at the top, certainly until 1992.[18] For the first three years in the sample the shares of the bottom three deciles rise, while those for deciles 8 and 9 fall. The top decile shows no trend, but there is some indication that incomes at the very top are climbing faster than others, with the share of the richest 1 percent of the population rising over the period.

However, 1994 does represent a break in this trend. The decline in mean incomes at the top and the bottom must imply a reduction in shares for deciles 1 and 10. Since the overall mean continued to grow, we do in fact see a reduction in the shares of the first three deciles. Those who gained were the middle classes, broadly defined as deciles 4 to 9. We must emphasize, however, that despite an absolute loss in income in the bottom decile—which does have implications for poverty and welfare—it would be wrong to conclude, as many commentators have, that inequality increased unambiguously from 1992 to 1994. Decile shares provide merely a (somewhat) disaggregated view of the distribution, rather than an accurate yardstick of inequality. There is no Lorenz dominance of 1992 over 1994, and, indeed, three of our four measures actually fall in that interval. From the evidence presented in this study, we could claim that inequality (defined so as to be consistent with the Pigou-Dalton transfer principle) worsened only if we choose $E(2)$ as the only measure that matters.

In closing this section, we turn to stochastic dominance analysis. As the preceding discussion illustrates, inequality comparisons of different distributions

16. These figures are from the three months ending with November (the CASEN survey month) of both years, according to Cowan and De Gregorio (1996).

17. See also Beyer (1995) for an interesting discussion of the patterns of employment and labor force participation in the first quintile.

18. The increased dispersion at the top of the distribution is suggested both by the rising share of the top percentile (shown in table 3) and by the sustained rise in the value of $E(2)$ (shown in table 2).

depend on the specific measures employed, and ambiguities are often inevitable. The concept of stochastic dominance, which originates from the analysis of financial risk, was introduced to the field of income distribution analysis to help establish when we can make unambiguous comparisons of inequality or social welfare.

Distribution A displays first-order stochastic dominance over distribution B if its cumulative distribution function $F_A(y)$ lies nowhere above and at least somewhere below that of B, $F_B(y)$. For any income level y, fewer people earn less than y in distribution A than in distribution B. If that is the case, a theorem due to Saposnik (1981) establishes that any social welfare function that is increasing in income will record higher levels of welfare in A than in B.

Distribution A displays second-order stochastic dominance over B if its deficit function—the integral of the distribution function $G(y_k) = \int_0^{y_k} F(y)dy$—lies nowhere above (and somewhere below) that of B. It is a weaker concept than its first-order analogue and is in fact implied by first-order dominance. Shorrocks (1983) has shown that if second-order stochastic dominance holds, any social welfare function that is increasing *and* concave in income will record higher levels of social welfare in A than in B.

The dominance criteria described above are alternative concepts suitable for comparing welfare. To measure inequality, we must abstract from the mean and concentrate on the dispersion of the distribution. For this purpose mean-normalized second-order dominance—also known as Lorenz dominance—is the appropriate concept. Distribution A is said to Lorenz-dominate distribution B if the Lorenz curve associated with A lies nowhere below, and at least somewhere above, that associated with B. A Lorenz curve is a mean-normalized integral of the inverse of a distribution function: $L(p) = \frac{1}{\mu(y)} \int_0^p F^{-1}(\pi)d\pi$. In other words, it plots the share of income accruing to the bottom p percent of the population against p. For a Lorenz curve A to lie everywhere above another B means that in A the poorest p percent of the population receive a greater share of the income than in B, for every p. Atkinson (1970) has shown that if this condition holds, inequality in A is lower than in B according to any inequality measure that satisfies the Pigou-Dalton transfer axiom.

Table 4 presents the results of these three types of dominance comparisons among the four years for which we have CASEN data, using both income concepts. We first made the comparisons at the percentile level of aggregation, and then checked the completely disaggregated sample, testing its statistical significance according to the endogenous bounds method of Howes (1993). The results (including their significance tests) for both the household income per capita and the household income per equivalent adult distributions are identical in every case except one. The exception is the (statistically insignificant) Lorenz dominance of

Table 4. *Welfare and Inequality Stochastic Dominance Comparisons*

Year	*1987*	*1990*	*1992*	*1994*
1987				
1990				
1992	F^*, S^*	S^*		
1994	F^*, S^*	F^*, S^*, (L)		

Note: A letter F, S, or L in cell (i, j) indicates that year i respectively first-order, second-order, or Lorenz dominates year j. The letter is inserted when dominance is found at the percentile level. An asterisk indicates that the dominance is statistically significant at the 95 percent confidence level over a range greater than or equal to 99 percent in Howes's endogenous bounds test for the complete sample, and hence that the hypothesis of no population dominance can be rejected at that level. Parentheses indicate that dominance was statistically insignificant.

Source: Authors' calculations.

1994 over 1990, which is found only for the per capita income distribution. All Lorenz comparisons for equivalized income display crossings.

We can immediately make two observations. First, there are a number of significant welfare dominance results below the diagonal, indicating that welfare rose unambiguously from some earlier to some later years. Second, with the exception of the statistically insignificant case of 1994 over 1990 for per capita income, there are no instances of Lorenz dominance. This suggests that, with the possible exception of an improvement between 1990 and 1994, inequality comparisons between the years in this period are ambiguous and will depend on the specific measure used. This result confirms those presented in table 2.

Much more can be said in terms of the evolution of social welfare. Since our measure of welfare depends entirely on income, as discussed in the introduction, we would expect rapid economic growth to have a powerful impact. Nevertheless, the dominance results are interesting because they tell us something about the distribution of the gains from growth across households. Rapid growth between 1987 and 1990, for instance, was not sufficient to lead to unambiguous welfare gains, because the poorest 1 or 2 percent of the population were worse off in 1990. Gains above the second or third percentile indicate that there was still a rise in the mean income of the first decile, as reported earlier in this section. The highly disaggregated nature of dominance analysis allows us to capture finer changes. This loss to the poorest people in Chile at the end of the 1980s means that welfare functions very sensitive to their circumstances would not show an increase in social welfare since 1987, despite the substantial increase in incomes elsewhere in the distribution.

Growth from 1990 to 1992 did not seem to have this perverse effect at the bottom of the distribution. Rises in income across every percentile ensured that both 1992 and 1994 first-order dominate 1987. Both years also second-order dominate 1990.[19] However, 1994 and 1992 cannot be ranked by either the first- or second-order criterion. This is because a decline in income for those below the

19. Because of a crossing above the ninety-ninth percentile, 1992 does not first-order dominate 1990. Except for the very rich, everyone (in an "anonymous" sense) was better off in 1992 than in 1990.

eighth or ninth percentile (which was sufficient to lower the mean income of the first decile, as discussed earlier) co-existed with gains for all other social groups. This welfare loss to the very poor means that we cannot say unambiguously that social welfare rose over the last two years in our sample. The loss was not sufficient, however, to outweigh gains to those at the bottom of the distribution since 1990: 1994 does first-order dominate both 1987 and 1990.

These results add rigor to our earlier analysis and broadly support its findings. The first fundamental feature of the period is economic growth, which led to welfare dominance of the last two years over the first two. On two occasions, however, economic growth failed to improve measured living standards for the most vulnerable people in society: from 1987 to 1990 and, more famously, from 1992 to 1994. For these two periods we cannot make such an unambiguous welfare comparison.

The second fundamental feature—the relative stability in the dispersion of the distribution, but with a slight compression at the bottom and a stretching at the top—is also compatible with the absence of significant Lorenz dominance results. Such changes in the shape of the density function, reducing distances in one part of the distribution while increasing them elsewhere, is exactly what causes different inequality indexes to rank distributions in opposite ways. Although this section emphasizes that the evidence on inequality is not sufficiently clear to identify any real tendency in either direction, so that the most appropriate description is one of broad stability or inconclusive changes, the (insignificant) Lorenz dominance of 1994 over 1990 for the per capita income distribution suggests that, if one were pushed to indicate a tentative direction for change in inequality over the period, it would more likely be downward.

III. The Evolution of Poverty

The high GDP growth rates that Chile achieved over this period undeniably contributed to a considerable reduction in poverty from the relatively high levels of the mid-1980s. In this section we present detailed results on the changes in the incidence of poverty between 1987 and 1994, relying on our adjusted data set and comparing the numbers for household per capita income with those for household income per equivalent adult. Although the general downward trend confirms previous findings (see, for example, Larranaga 1994 and Contreras 1995), these new numbers reflect our adjustments, such as the incorporation of regional price differences, the adoption of a new equivalence scale, and the improved treatment of domestic servants (see appendix A). Before presenting the specific results, we briefly discuss the derivation of the poverty lines, with respect to which all of the measures must be understood.

We use three poverty lines in this study, all of them expressed in 1994 Santiago pesos per month: an indigence line set at P$15,050, a lower-bound poverty line (L) set at P$30,100, and an upper-bound poverty line (H) equal to P$34,164. The first two are the official indigence and poverty lines widely used in Chile.

The incomes against which we compare them, however, differ from most earlier studies in that we have "converted" them to 1994 Santiago pesos, using as an "exchange rate" the regional price index in table A-1 and the November consumer price index The derivation of the upper-bound line is explained below.

All three lines are absolute poverty lines, deriving from a standard food basket specified by CEPAL. The basket is chosen so as to provide 2,187 Kcal per person per day, the national average caloric requirement, which is obtained from the demographic characteristics of the population and from the Food and Agriculture Organization/World Health Organization recommended caloric intakes for different age and gender groups. The specific commodity composition of the basket is based on the actual consumption patterns of a reference group chosen by CEPAL. The reference group is the third quintile, measured according to consumption expenditures, in the Household Expenditure Survey of 1987–88. It is valued at average prices for November 1994 in Santiago. The monthly cost of this standard CEPAL food basket has traditionally been used in Chile as an indigence (or extreme poverty) line, separating the hard-core poor—those whose current monthly incomes are insufficient even to purchase a minimum diet—from the rest of society. In 1994 this amount was P$15,050, which we report below as the indigence line.

International practice, however, deems the indigence line too strict a criterion to identify the poor. There are, after all, other basic expenditures in addition to food that everyone must make, such as shelter, clothing, and public transport. We apply a standard methodology to arrive at a sensible poverty line: we multiply the cost of the food basket by the inverse of the share of food in total expenditures (the Engel coefficient) for some suitable reference group. Based on the estimates of the Engel coefficient for the lower quintiles of the Chilean expenditure distribution, (reported in table 5), we adopt the standard value of 0.5, which implies a doubling of the indigence line, to arrive at the poverty line.[20] However, although table 5 is reassuring in that the coefficients vary little from total household expenditures to expenditures per capita (suggesting robustness with respect to the equivalence scale adopted), they are substantially lower for the concept of per capita income including imputed rent. Since this is the welfare concept that we adopt for a large part of our analysis, we feel that we cannot ignore its implications for poverty measurement.[21] Weighing the relevant coefficients for the first and second quintiles by 0.8 and 0.2, respectively, we derive an Engel coefficient of approximately 0.44. Applying its inverse (2.27) to the cost of the food basket yields our upper-bound poverty line (H) of P$34,164.

We also use the same three lines discussed above when computing poverty measures using the vector of real household incomes per equivalent adult. Given our choice of two-adult households as the reference type, the per capita poverty

20. Doing so has the important advantage of allowing some comparability with the findings of earlier studies, most of which have used this line.

21. We also could not ignore it since the household income per equivalent adult concept also includes imputed rents.

Table 5. *Engel Coefficients*

Quintile	*Total expenditures*	*Expenditures per capita*	*Income per capita (including imputed rent)*
1	0.530	0.539	0.451
2	0.494	0.489	0.397

Source: CEPAL (1996, p. 31, table 13).

line is unchanged for that household type. As with any equivalence scale designed to take account of different relative costs of children and of economies of scale, ours implies different per capita poverty lines for household types other than the reference. This reflects the re-ranking of households that the scale inherently causes. Table 6 sets out what our equivalence scale implies in terms of household and per capita poverty lines.

Let us now turn to the poverty measures. For a given poverty line, z, we can define different poverty indexes, each aggregating information on the living standards of those below the poverty line in different ways. We work with three of the most common measures, all of which can be expressed as members of the following parametric class, proposed by Foster, Greer, and Thorbecke (1984):

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{n}\left[\max\left(\frac{z-y_i}{z},0\right)\right]^\alpha \tag{7}$$

where y_i is the income of the ith individual (of which there are n). The Foster-Greer-Thorbecke (FGT) measures are rather intuitive. As is well known, when $\alpha = 0$, P simplifies to p/n, the headcount index. When $\alpha = 1$, we get the normalized poverty deficit, and $\alpha = 2$ yields the FGT(2) measure, which incorporates some convexity to the distances between incomes and the poverty line, and is hence sensitive to inequality among the poor. Table 7 lists the values of each of these measures for the whole country, in each relevant year, for the distribution of household per capita income. Each index is listed for each of the three poverty

Table 6. *Implied Household and Per Capita Poverty Lines*

Household type	*Equivalence factor*	*Household poverty line*[a]	*Implied per capita poverty line*[b]
Single Adult	1.2	36,120	36,120
Couple	2.0	60,200	30,100
Couple + Child (age 0–4)	2.3	69,230	23,077
Couple + Child (age 5–10)	2.4	72,240	24,080
Couple + Child (age 11–15)	2.8	84,280	28,093
Couple + Child (age 0–4) + Child (age 11–15)	3.1	93,310	23,328

a. Equivalence factor x 30,100 = household poverty line.
b. Household poverty line / number of persons = implied per capita poverty line.
Source: Authors' calculations.

Table 7. *Poverty Measures for Household Income Per Capita*

Index	*1987*	*1990*	*1992*	*1994*
Indigence line, P$15,050				
Headcount	0.2209	0.1646	0.1403	0.0996
Poverty deficit	0.0756	0.0561	0.0325	0.0336
FGT (2)	0.0382	0.0295	0.0172	0.0184
Lower poverty line (L), P$30,100				
Headcount	0.5137	0.4427	0.3603	0.3386
Poverty deficit	0.2274	0.1838	0.1329	0.1269
FGT (2)	0.1299	0.1017	0.0681	0.0663
Upper poverty line (H), P$34,164				
Headcount	0.5679	0.5002	0.4206	0.3940
Poverty deficit	0.2647	0.2181	0.1637	0.1554
FGT (2)	0.1560	0.1240	0.0861	0.0831

Note: See text for definition of poverty indexes.
Source: Authors' calculations.

lines derived above. Table 8 is analogous to table 7, and lists the values of the same measures for the whole country, in each relevant year, for the distribution of household income per equivalent adult.

According to all three measures there has undoubtedly been a remarkable reduction in both poverty and extreme poverty from 1987 to 1994. Poverty in Chile was quite high in the mid-1980s, in the aftermath of the serious recession of 1982–84. Between 51 and 57 percent of the population lived in poverty in 1987, according to the per capita income concept, or 41–47 percent according to the income per equivalent adult concept.[22] By 1994 the figures were 34–39 percent for per capita income, or 23–29 percent for equivalized income.[23] The incidence of indigence fell from 22 to 10 percent according to the per capita concept, or 13 to 5 percent according to the equivalized concept.

22. These poverty ranges refer to the values with respect to the lower-bound and upper-bound poverty lines.

23. The poverty headcount figures for the per capita distribution are higher than those reported by several earlier studies (see Haindl 1996 for a survey). This is due primarily to the following methodological corrections that we make. First, some earlier studies report as a headcount the proportion of households below the poverty line. We report the proportion of individuals below the poverty line. Since poorer households tend to be larger, our method increases the figure. Second, incorporating regional price differences raises poverty since Santiago prices are the lowest in the country. Deflating incomes elsewhere to take this into account reduces their real incomes. Third, because we do not apply the arbitrary 0.66 factor to reduce prices in rural areas (see appendix A), we do not lower poverty lines for rural areas. We believe that there is no justification for that practice, particularly in light of the fact that reported figures for own consumption from household production among the rural poor do not appear to be unrealistic. The empirically baseless reduction of rural prices by an arbitrary factor might have been contributing to an underestimation of rural poverty. Finally, our inclusion of domestic servants as individuals with their own incomes—rather than those of their employers—might add some individuals to the ranks of the poor. See appendix A.

Table 8. *Poverty Measures for Household Income per Equivalent Adult*

Index	*1987*	*1990*	*1992*	*1994*
Indigence line, P$15,050				
Headcount	0.1268	0.0894	0.0474	0.0511
Poverty deficit	0.0412	0.0311	0.0174	0.0192
FGT (2)	0.0213	0.0176	0.0108	0.0118
Lower poverty line (L), P$30,100				
Headcount	0.4069	0.3306	0.2418	0.2308
Poverty deficit	0.1568	0.1196	0.0776	0.0762
FGT (2)	0.0822	0.0614	0.0376	0.0382
Upper poverty line (H), P$34,164				
Headcount	0.4726	0.3889	0.3000	0.2852
Poverty deficit	0.1905	0.1483	0.1006	0.0978
FGT (2)	0.1028	0.0777	0.0491	0.0492

Note: See text for definition of poverty indexes.
Source: Authors' calculations.

As others have pointed out before, the reductions were largest in the years of faster growth, from 1987 to 1992, and smallest in the more sluggish years of 1992–94. In fact, although all headcount indexes for the per capita income distribution record poverty reductions in 1992–94, as do the other two poverty measures, it is noteworthy that the poverty deficit and the FGT(2) measures for indigence actually record slight increases in extreme poverty between 1992 and 1994. The picture is even more severe for the equivalized income distribution, for which, in addition to increases in P_1 and P_2, the indigence headcount also rises, suggesting that the number of people living in extreme poverty increased from 1992 to 1994, despite continued economic growth. Furthermore, FGT(2) also rises (marginally) for the other two poverty lines as well.

This is a result of the decline in real incomes in the first decile of the income distribution (discussed in the previous section). Clearly, the closer a poverty line is to that decile, the likelier it is to record an increase in poverty, whereas more generous lines still record declines as a result of income gains to people in the second and third deciles. But even for those lines, poverty measures that are more sensitive to large distances between incomes and the poverty line (that is, those that place greater weight on greater destitution), such as FGT(2), are liable to have the losses at the very bottom outweigh gains closer to the upper poverty lines.

This picture of a considerable reduction in poverty throughout the period, albeit with some ambiguity between 1992 and 1994, is confirmed by stochastic dominance analysis. It has been shown that if a distribution A displays poverty mixed dominance [PMD(z^-, z^+)] over a distribution B, then any poverty measure that is decreasing in income, and satisfies the focus axiom and the transfer axiom (in situations where a crossing of the poverty line does not occur), will indicate that poverty is lower in A than in B for any poverty line between z^- and z^+.[24] This class of poverty measures includes all members of the FGT P_α class and

24. See Howes (1993) for a discussion, and Ferreira and Litchfield (1996) for an application to Brazil.

Table 9. *Poverty Mixed Stochastic Dominance Comparisons*

Year	*1987*	*1990*	*1992*	*1994*
1987				
1990				
1992	*P*	*P*		
1994	*P*	*P*		

Note: The letter *P* in cell (*i*, *j*) indicates that year *i* dominates year *j*.
Source: Authors' calculations.

is therefore particularly appropriate for this study. Poverty mixed dominance essentially requires that distribution *A* display second-order dominance over *B* from zero to the lower-bound poverty line (z^-), and first-order dominance from z^- to z^+. Table 9, which is analogous to table 4, presents the results for Chile, with z^- set at the indigence line and z^+ set at the upper-bound poverty line (*H*). As before, we checked dominance for both per capita income and income per equivalent adult; both concepts yield exactly the same results.

These results reveal that there was unambiguously less poverty in 1992 and in 1994 than in either 1987 or 1990, whether poverty is measured by the headcount, the poverty deficit, or indeed any of a host of other sensible poverty measures, and for any poverty line set between P$15,050 and P$34,164 per month in 1994 Santiago pesos. This sort of unambiguous poverty reduction, independent of the specific measure used and valid for such a large range of poverty lines, is not common. Its achievement confirms the widely held view that Chile has made substantial strides in the fight against poverty during the past decade.[25]

Yet there is also confirmation that growth did not lead to unambiguous poverty reductions in two subperiods: from 1987 to 1990 and from 1992 to 1994. On both occasions, although the headcount for the headline poverty line indicates a reduction in the number of poor people, there were income losses in the lowest percentiles of the distribution. These losses imply that some poverty measures in the wide class defined above would indicate increases in poverty for at least some of the poverty lines in the covered range. Indeed, in 1992–94 this was the case for all three indigence measures reported in table 8, for the equivalized income distribution. Still, the dominance of 1994 over both 1987 and 1990 indicates that the losses to some of the poor in the last two years were at least not sufficient to outweigh the gains made between 1990 and 1992.

Overall, there is no question that Chile's growth and social policies were tremendously successful in reducing the incidence, intensity, and inequality of poverty between 1987 and 1992, with the poverty deficits being roughly halved across all three poverty lines (a little less for per capita income, a little more for equivalized

25. All but one of the positive dominance results reported in table 9 follow directly from the first-order dominances reported in table 4. Table 9 contains new information only for the cases in which there was no dominance, as well as for the dominance of 1992 over 1990, the distribution functions of which clearly do not cross between z^- and z^+.

income).[26] The recent reversal in the performance of the poorest of the poor, between 1992 and 1994, does, however, provide a cautionary signal against complacency.[27]

IV. Conclusions

In presenting an overview of welfare, inequality, and poverty trends in Chile during 1987–94, this article has sought to establish a number of empirical conclusions that can be drawn with confidence from a careful treatment of the data. These can be summarized as follows. Chilean inequality is high by international standards, and it remained largely unchanged between 1987 and 1994. Scalar inequality measures—such as the Gini coefficient and the Theil index—varied little over the period, and there were no statistically significant Lorenz dominances between any of the years in the sample.

Within this broad picture of stability in inequality, however, there is some evidence that the shape of the density function may have altered slightly, with a compression at the lower tail (reducing bottom-sensitive inequality measures such as the mean log deviation), and an increase in dispersion at the upper tail (leading to rises in top-sensitive measures such as the coefficient of variation).

Economic growth has had a substantial and clearly beneficial impact, helping shift the distribution function to the right. Welfare, as measured by any reasonable social welfare function, was unambiguously higher in both 1992 and 1994 than it had been in either 1987 or 1990. Similarly, poverty was incontrovertibly lower. The bulk of these welfare improvements took place between 1987 and 1992, however, and the slowdown in growth and increase in unemployment after 1992 were associated with losses to the poorest 8 or 9 percent of the population. Although other social groups—including many of the poor—continued to gain, the losses at the very bottom of the distribution were associated with an increase in indigence between 1992 and 1994. Over the whole period, however, growth and other factors led to a remarkable decline in poverty, however it is measured, and for a wide range of plausible poverty lines. For some of these lines, headcounts and poverty deficits were halved from 1987 to 1994.

Although we have found that some assumptions underlying previous studies contributed to underestimating both poverty and inequality—whether by arbitrarily reducing prices when calculating poverty lines for rural areas, taking households rather than individuals as the unit of analysis, or ignoring live-in domestic

26. Larranaga (1994) decomposes the changes in poverty in Chile between 1987 and 1992 into a growth and a redistribution component, using a methodology of Datt and Ravallion (1992). Although he finds that some 80 percent of the reduction could be explained by the effects of growth, some of the changes were also due to a redistribution effect, which may have followed—at least in part—from the government's social policies and expenditures.

27. Although the unit record data set of the CASEN 1996 was not made available for this study, the tabulated findings reported by MIDEPLAN (1997) suggest that faster growth after 1994 led to a resumption of previous rates of poverty reduction and that inequality remained broadly stable.

servants—another important adjustment made in this study goes in the opposite direction, suggesting that previous reports may have overestimated poverty and inequality. We have found that both poverty and inequality measures for Chile are considerably reduced when household income is adjusted to take into account differences in needs between children and adults, and to take into account economies of scale inherent in the sharing of fixed costs within the household.

Space constraints prevented us from investigating the determinants of the structure of and the changes in inequality in Chile in this article. However, our previous work suggests that education is by far the most important candidate variable. Ferreira and Litchfield (1998) find that differences between groups partitioned by educational attainment explain a much greater share of overall inequality than do differences in any other household attribute, and that changes in returns to education appear to lie behind the reduction in bottom-sensitive inequality measures over the period (see also World Bank 1997).

In that paper we argue that the importance of increases in education and the stability of overall inequality, taken together, also suggest that Chile may have reached a dynamic equilibrium between rising demand for and supply of skills. This would explain the recent reversal (or flattening) of the upward trend in both earnings and personal income dispersion that other authors had identified for periods prior to 1990. With the profound political and economic changes of the 1970s and 1980s behind it, and regardless of its trade patterns, Chile's income distribution is, for the moment, calm.

Appendix A. The Data Set

The CASEN sampling methodology can be described as multi-stage random sampling with geographical stratification and clustering. The country is first divided into strata comprising the rural and urban sectors of each of the 13 regions.[28] The rural sectors are final-level strata. The urban sectors are further subdivided into three categories of towns, according to population: towns with between 2,000 and 9,999 inhabitants; towns with between 10,000 and 39,999 inhabitants; and towns with 40,000 or more inhabitants. All of the latter are sampled (that is, they are final-level strata). For other towns there is a level of clustering in the selection of towns for sampling. At this level, with selected small towns, all large towns, and the rural sectors, a first stage samples primary units (*zonas de empadronamiento*), with probabilities proportional to the population. A second stage samples households. This process is described in more detail in annex III to MIDEPLAN (1992). The sample sizes for our analysis are as follows: the 1987 sample includes 23,403 households; the 1990 sample consists of 26,248 households; the 1992 sample numbers 36,587 households; and the 1994 sample covers 45,993 households.[29]

28. For this purpose an urban area is any grouping of dwellings with a population in excess of 2,000.

29. These sample sizes are slightly larger than those reported in the official MIDEPLAN records of the surveys, reflecting our treatment of live-in domestic servants as separate households. See below.

Once each survey is completed, the data are entrusted to CEPAL, which conducts two types of adjustments to the raw figures. The first type corrects for nonresponses, which are made in three instances: when people who declare themselves employed report no income from their main occupation, when people who state that they receive an occupational or widow's pension do not report a value for this benefit, or when owner-occupiers of their domiciles do not report a value for imputed rent. In all three cases missing income values are replaced by the average value of the specific income variable in the group to which the household belongs, where the group is defined by a partition according to a number of variables, including region, gender of household head, educational attainment of household head, occupational sector, and category. See appendix 1 to CEPAL (1995) for details.

The second type of adjustment seeks to correct for underreporting or overreporting of different income categories, a common problem with household income surveys everywhere. For this purpose CEPAL uses as the reference point for aggregate income flows the Household Incomes and Expenditures Account of the National Accounts System (SCN) of the Central Bank of Chile. First, by a careful process the information in the original Central Bank accounts is converted to the income concepts surveyed by CASEN. Once that is achieved, totals by specific income category are compared for CASEN (with recourse to the appropriate expansion weights) and the national accounts. Finally, the proportional differences for each income category between the two sources are imputed uniformly to each income recipient in CASEN, with two notable exceptions: the adjustment in capital incomes is applied only to the top quintile (of households), proportionally to the primary incomes (*ingresos autonomos*) of all recipients there; and incomes from entitlement transfers and gifts are not adjusted.[30] The underlying assumption justifying this procedure is that misreporting differs fundamentally across income categories, rather than income levels.[31] In fact, the imputation would be strictly correct only if the income elasticity of misreporting within each income category was unitary. The only exception to this assumption, as noted, is in the treatment of capital incomes, which are imputed proportionally, but exclusively within the richest 20 percent of households. A detailed account of CEPAL's adjustment methodology, complete with the numbers used in each of the four years, is available in CEPAL (1995).

We also make three adjustments to the data set after it is processed by CEPAL. The income variable from the CASEN records on which our analysis is primarily based is total adjusted household income (YTOTHAJ), which includes all primary

30. It is suggested that the main reason for not adjusting entitlement transfers and gifts is that underreporting of transfers consists mostly of complete omissions of benefits by some households, rather than proportional underreporting of values by all recipients. There being no way to identify which households are omitting this information, no adjustment was found that would have improved the picture obtained from the survey.

31. It may be interesting to note that the proportional adjustments do vary substantially across income categories. In fact, imputed rents are consistently found to have been overreported and are adjusted downward in every survey.

income, monetary transfers,[32] and gifts, as well as imputed rent, after the CEPAL adjustments.[33] It is from this variable that we construct both of the income concepts listed above—household income per capita and household income per equivalent adult—by appropriate choice of denominator. Our first adjustment is in the treatment of live-in domestic servants. It is unclear how previous studies treated them, since YTOTHAJ is defined to exclude their incomes. Household-based studies are likely to have unwittingly excluded them from the sample altogether, by simply imputing YTOTHAJ to the household. For this article, in households with live-in domestic servants, all other members received YTOTHAJ divided by the appropriate denominator (their number in the per capita income case, or the equivalence scale defined over them), while the servants were treated as a separate household whose income was the sum of total adjusted individual incomes (YTOTHAJ) over them.

The second adjustment is to exclude from the analysis the three richest households in the 1994 sample. This decision was carefully considered, and was based on the impression that these households reported sufficiently disproportionate incomes to be regarded as genuine outliers.[34] This impression was reinforced by the fact that two of these households were identical in every respect, having clearly been double-sampled,[35] and by the position of the three households as outliers in a plot of the Pareto distribution of the top 1 percent of the sample.

The third adjustment, as described in section I, is to deflate all incomes by a regional price index. The values of the regional deflators are given in table A-1.

32. We believe that the questionnaire coverage of monetary transfers from the state, whether at the federal or municipal level, is exhaustive. Questions are asked and amounts are registered for the following benefits: *Asignaciones Familiares*, PASIS, SUF, *Subsidio al Consumo de Agua Potable*, and *Subsidio de Cesantia*. See MIDEPLAN (1996) for a description of each of these entitlement benefits. Other benefits, associated with formal employment, such as *jubilaciones*, *pensiones de invalidez*, and *montepios* are also included, although they are aggregated as part of primary, rather than secondary, income.

33. We would have liked to use an even broader income concept, which takes into account the values of the transfers in-kind that the government makes to many households through programs in the areas of education, health, and housing. Monitoring these expenditures is in fact one of the objectives of the CASEN, and an exercise of valuation of these benefits is carried out by MIDEPLAN, relying on answers to survey questions about the usage of services and on cost data provided by the relevant ministries. Although the methodology for these valuations is discussed in MIDEPLAN (1993) and a tabulation of the imputations is available for quintiles of households in MIDEPLAN (1994), the disaggregated imputations at the household level—which would have been necessary for our analysis—are not made available with the other CASEN variables. This is due to alleged problems of methodology and reliability. The results are therefore only available at the quintile level, for distributions of households, and are of very limited comparability with the distributions used in this study. One should bear in mind that the—often substantial—value of these services, many of which are targeted, is omitted from the income data, when interpreting the distributional results that follow, or attempting to draw any conclusions about the social policies followed during the period.

34. The value of $E(2)$—see below—when the outliers are included, is 6.58 in 1994. Its range over the other years is from 1.39 to 1.74.

35. The practice of imputing all values from one household (randomly selected within the cluster) to another when the latter has failed to respond, is often adopted as a way of maintaining representativeness within a small cluster. When the doubled household is the Rockefellers, the practice requires revision.

Table A-1. *Regional Price Indexes*

Region	*Price index*
I	1.2163
II	1.1668
III	1.1112
IV	1.0859
V	1.0420
VI	1.0269
VII	1.0136
VIII	1.0441
IX	1.0500
X	1.0383
XI	1.0856
XII	1.2208
XIII (RM)	1.0000

Appendix B. Inequality and Welfare Analysis for Per Capita Incomes Without Regional Price Deflation

Table B-1. *Descriptive Statistics for Household Income Per Capita with No Regional Price Adjustment*

Statistic	*1987*	*1990*	*1992*	*1994*
Mean income[a]	56,558	66,963	77,548	80,944
Median income[a]	29,803	36,520	41,793	45,116
Gini coefficient	0.5596	0.5523	0.5512	0.5429
$E(0)$	0.5624	0.5444	0.5310	0.5219
$E(1)$	0.6300	0.6401	0.6432	0.6079
$E(2)$	1.3726	1.7297	1.6227	1.6468

a. In 1994 Santiago pesos.
Note: See text for definition of statistics.
Source: Authors' calculations.

Table B-2. *Decile Income Shares for Household Income Per Capita with No Regional Price Adjustment*
(percent)

Income category	*1987*	*1990*	*1992*	*1994*
Mean income (1994 Santiago pesos)	56,558	66,963	77,548	80,944
Decile 1	1.24	1.30	1.44	1.35
Decile 2	2.20	2.33	2.39	2.36
Decile 3	2.96	3.09	3.16	3.15
Decile 4	3.78	3.93	3.94	4.00
Decile 5	4.73	4.88	4.86	5.00
Decile 6	5.95	6.12	6.09	6.26
Decile 7	7.67	7.74	7.72	7.94
Decile 8	10.41	10.25	10.19	10.65
Decile 9	16.02	15.50	15.05	15.92
Decile 10	45.04	44.85	45.16	43.36
Top percentile	12.44	13.52	14.08	12.82

Source: Authors' calculations.

Table B-3. *Poverty Measures of Household Income per Capita with No Regional Price Adjustment*

Index	*1987*	*1990*	*1992*	*1994*
Indigence line, P$15,050				
Headcount	0.2132	0.1417	0.0968	0.0928
Poverty deficit	0.0726	0.0479	0.0303	0.0313
FGT(2)	0.0367	0.0255	0.0163	0.0173
Lower poverty line (L), P$30,100				
Headcount	0.5053	0.4151	0.3456	0.3212
Poverty deficit	0.2214	0.1657	0.1259	0.1191
FGT(2)	0.1258	0.0898	0.0641	0.0620
Upper poverty line (H), P$34,164				
Headcount	0.5584	0.4725	0.4068	0.3774
Poverty deficit	0.2584	0.1989	0.1558	0.1465
FGT(2)	0.1515	0.1106	0.0814	0.0779

Note: See text for definition of poverty indexes.
Source: Authors' calculations.

Appendix C. Inequality and Welfare Analysis for Per Capita Income

For comparability purposes this appendix replicates some of the analysis described in section II for household income per equivalent adult, using instead the distributions of household income per capita. This is intended not only to show that the main results of the paper are remarkably robust to the choice of unit, but also to enable comparisons with other studies found in the earlier Chilean literature.

Table C-1 is analogous to table 2, and reports mean and median incomes, as well as the same four inequality measures, for the per capita distributions. Two changes from table 2 are noteworthy. First, there is a decrease in the absolute values of the mean and median incomes for each year, which follows from the fact that there are many more households for which the denominators are reduced by the application of the equivalence scale used in section II, than there are households for which they are increased.[36] Second, all inequality measures are higher for this distribution than for the distribution of income per equivalent adult. This is in keeping with international experience, where it has been repeatedly found that the per capita income distributions generate upper-bound values for inequality, when compared to other assumptions about differences in needs and economies of scale within the household (see Coulter, Cowell, and Jenkins, 1992, and Ferreira and Litchfield 1996). The reason is essentially that large households, or those with many small children, are re-ranked upward from the per capita to the equivalized distribution, with the usual impact of reducing overall disparities.

36. Only single-person households would have an increased denominator and would have lower entries in the distribution of income per equivalent adult than in the per capita distribution.

Table C-1. *Descriptive Statistics for Household Income Per Capita*

Indicator	*1987*	*1990*	*1992*	*1994*
Mean income[a]	55,367	63,293	75,371	78,281
Median income[a]	29,148	34,153	40,378	43,277
Gini coefficient	0.5603	0.5563	0.5534	0.5454
$E(0)$	0.5611	0.5495	0.5287	0.5212
$E(1)$	0.6349	0.6509	0.6551	0.6194
$E(2)$	1.3903	1.7447	1.6680	1.7121

a. In 1994 Santiago pesos.
Note: See text for definition of the statistics.
Source: Authors' calculations.

Despite those changes in level, there are no modifications to the perceived trends in inequality: the Gini still suggests a slight downward trend, but it is basically stable. The Theil index, too, is trendless. The mean log deviation continues to indicate a reduction in bottom-sensitive inequality, and the $E(2)$ continues to point to a gradual increase in top-sensitive inequality. This picture is broadly confirmed by the shares of each decile in the distribution of household income per capita, as revealed by table C-2.

Table C-2. *Decile Income Shares for Household Income Per Capita*
(percent)

Income category	*1987*	*1990*	*1992*	*1994*
Mean income	55,367	63,293	75,371	78,281
Decile 1	1.21	1.21	1.35	1.28
Decile 2	2.19	2.26	2.35	2.33
Decile 3	2.95	3.03	3.13	3.11
Decile 4	3.77	3.87	3.91	3.96
Decile 5	4.72	4.84	4.83	4.96
Decile 6	5.94	6.07	6.05	6.22
Decile 7	7.66	7.71	7.70	7.91
Decile 8	10.37	10.25	10.18	10.60
Decile 9	15.89	15.49	14.97	15.95
Decile 10	45.30	45.26	45.63	43.66
Top percentile	12.54	13.60	14.31	12.94

Source: Authors' calculations.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

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

Becker, Gary. S. 1965. "A Theory of the Allocation of Time." *Economic Journal* 75: 493–517.

Beyer, H. 1995. "Logros en la Pobreza; Frustracion en la Igualdad?" *Estudios Publicos* 60(1):15–33.

CEPAL (United Nations Economic Commission for Latin America and the Caribbean). 1995. "La Medicion de los Ingresos en la Perspectiva de los Estudios de Pobreza: El caso de la Encuesta CASEN de Chile: 1987–1994." CEPAL Working Paper LC/R1604. Santiago. Processed.

_____. 1996. "La Sensibilidad del Indicador de Pobreza." CEPAL Working Paper LC/R1657. Santiago. Processed.

Chauduri, Shubham, and Martin Ravallion. 1994. "How Well do Static Indicators Identify the Chronically Poor?" *Journal of Public Economics* 53(3):367–94.

Comision Nacional de Justicia y Paz. 1996. "Es Chile un Pais Equitativo?" Area Pastoral Social de la Conferencia Episcopal de Chile. Santiago. Processed.

Contreras, Dante. 1995. "Poverty, Inequality and Welfare in Chile: 1987–1992." Department of Economics, Universidad de Chile. Processed.

Coulter, F. A. E, F. A. Cowell, and S. P. Jenkins. 1992. "Equivalence Scale Relativities and the Extent of Inequality and Poverty." *Economic Journal* 102(414):1067–82.

Cowan, K., and J. De Gregorio. 1996. "Distribucion y Pobreza en Chile: Estamos mal? Ha habido progresos? Hemos retrocedido?" Ministerio de Hacienda de Chile. Santiago. Processed.

Cowell, F.A. 1995. *Measuring Inequality*. Second Edition. Hemel Hempstead: Harvester Wheatsheaf.

Datt, Gustav, and Martin Ravallion. 1992. "Growth and Redistribution Components of Changes in Poverty Measures." *Journal of Development Economics* 38(2):275–95.

Deaton, Angus. 1997. *The Analysis of Household Surveys: Microeconometric Analysis for Development Policy*. Baltimore, Md.: The Johns Hopkins University Press.

Deininger, Klaus, and Lyn Squire. 1996. "A New Data Set Measuring Income Inequality." *The World Bank Economic Review* 10(3):565–91.

Edwards, Sebastian, and A. C. Edwards. 1987. *Monetarism and Liberalization: The Chilean Experiment*. Chicago: University of Chicago Press.

Ferreira, Francisco H. G., and Julie A. Litchfield. 1996. "Growing Apart: Inequality and Poverty Trends in Brazil in the 1980s." LSE STICERD DARP Discussion Paper 23, Suntory and Toyota International Centres for Economics and Related Disciplines, London School of Economics, London.

_____. 1998. "Calm After the Storms: Income Distribution in Chile, 1987–1994." Policy Research Working Paper 1960, Development Research Group, World Bank, Washington, D.C.

Foster, James. E., Joel Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52(3):761–66.

Haindl, E. 1996. "Poverty and Income Distribution in Chile: A Survey." World Bank, Washington, D.C. Processed.

Howes, S. R. 1993. "Income Distribution: Measurement, Transition and Analysis of Urban China, 1981–1990." Ph.D. dissertation, London School of Economics.

Larranaga, Osvaldo. 1994. "Pobreza, Crecimiento y Desigualdad: Chile, 1987–1992." *Revista de Analisis Economico* 9(2):69–92.

Londono, J. L., and M. Szekely. 1997. "Persistent Poverty and Excess Inequality: Latin America during 1970–1995." Office of the Chief Economist, Inter-American Development Bank, Washington, D.C. Processed.

Marcel, M., and Andres Solimano. 1994. "The Distribution of Income and Economic Adjustment." In B. P. Bosworth, Rudiger Dornbusch, and Raul Laban, eds., *The Chilean Economy: Policy Lessons and Challenges*. Washington, D.C.: The Brookings Institution.

Meller, P. 1996. "Trade Liberalization and Wage Differentials in Chile." Paper presented at the XIV Latin American Econometric Society Meetings, Rio de Janeiro, August. Processed.

MIDEPLAN. 1992. *Poblacion, Educacion, Vivienda, Salud, Empleo y Pobreza:* CASEN *1990*. Santiago: Ministerio de Planificacion y Cooperacion.

_____. 1993. *Programas Sociales: Su impacto en los hogares chilenos,* CASEN *1990*. Santiago: Ministerio de Planificacion y Cooperacion.

_____. 1994. *Integracion al Desarrollo: Balance de la politica social, 1990–1993*. Santiago: Ministerio de Planificacion y Cooperacion.

_____. 1996. *Balance de Seis Anos de las Politicas Sociales: 1990–1996*. Santiago: Ministerio de Planificacion y Cooperacion.

_____. 1997. "Pobreza y Distribucion del Ingreso en Chile, 1996: Resultados de la CASEN." Santiago. Processed.

Montenegro, C. 1996. "The Structure of Wages in Chile, 1960–1993: An Application of Quantile Regressions." Inter-American Development Bank, Washington, D.C. Processed.

OECD (Organisation for Economic Co-operation and Development). 1982. *The* OECD *List of Social Indicators*. Paris.

Ravallion, Martin, and Benu Bidani. 1994. "How Robust is a Poverty Profile?" *The World Bank Economic Review* 8(1):75–102.

Riveros, L. 1983. "Distribucion del Ingreso en el Gran Santiago." *Revista de Economia*, 9, Universidad de Chile.

Robbins, D. 1994. "Worsening Relative Wage Dispersion in Chile During Trade Liberalization and its Causes: Is Supply at Fault?" Harvard Institute for International Development, Cambridge, Mass. Processed.

_____. 1995. "Relative Wage Structure in Chile, 1957–1992: Changes in the Structure of Demand for Schooling." *Estudios de Economia* 21(1):51–78.

_____. 1996. "HOS Hits Facts: Facts Win. Evidence on Trade and Wages in the Developing World." Harvard Institute for International Development, Cambridge, Mass. Processed.

Saposnik, R. 1981. "Rank-Dominance in Income Distribution." *Public Choice* 36(1):147–51.

Scott, C. 1996. "The Distributive Impact of the New Economic Model in Chile." In V. Bulmer-Thomas, ed., *The New Economic Model in Latin America and its Impact on Income Distribution and Poverty*. London: Macmillan.

Sen, Amartya K. 1981. *Poverty and Famines: An Essay on Entitlement and Deprivation.* Oxford: Clarendon Press.

Shorrocks, A. F. 1983. "Ranking Income Distributions." *Economica* 50(1):3–17.

UNDP (United Nations Development Programme). 1996. *Desarrollo Humano en Chile, 1996.* Santiago.

Wood, Adrian. 1997. "Openness and Wage Inequality in Developing Countries: The Latin American Challenge to East Asian Conventional Wisdom." *The World Bank Economic Review* 11(1):33–57.

World Bank. 1997. *Chile—Poverty and Income Distribution in a High-Growth Economy: 1987–1995.* Report 16377-CH, Latin America and the Caribbean Region, Washington, D.C.

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 539-59

Changes in the Perception of the Poverty Line During the Depression in Russia, 1993–96

Branko Milanovic and Branko Jovanovic

Economic transition in Russia was accompanied by a precipitous decline in real income for most of the population. This article analyzes how the decline affected people's perception of the minimum level of income needed to make ends meet. Individual-level data collected from repeated surveys between March 1993 and September 1996 reveal that the elasticity of subjective minimum income with respect to actual median income was 1.5 or that people's subjective estimate of the minimum income for an adult Russian fell about 1.7 percent each month. This sharp reduction in the face of a decrease in real income meant that the percentage of the population who felt that they were poor declined, even though poverty remained at a very high level (more than 60 percent of the population) throughout the period. This self-perception is in marked contrast to an "objective" measure of poverty: the percentage of the population whose income was less than a given real poverty line rose.

In the course of its transition to a market system, the Russian economy experienced a series of shocks. Its output fell sharply: gross domestic product (GDP) in 1997 was almost one-third less than it was in 1987. It suffered from rapid and continuing inflation: during the period under study here (March 1993 to September 1996) the price level increased 46 times. Open unemployment appeared, affecting about 10 percent of the labor force by 1997. Real wages and pensions declined to half their level before the transition, and delays in their payment became endemic. A few individuals who were politically well connected, enterprising, or lucky were able to amass considerable wealth. As a consequence, since the transition began, income inequality has risen by an unprecedented amount (the Gini coefficient increased four to five times faster in Russia than in the United States during the 1980s; see Milanovic 1998). The number of families living in poverty also increased rapidly.[1]

Given these developments, the population's views about what constitutes poverty and the minimum income needed to "make ends meet" likely evolved as well. Because the decline in income was so sharp, it reveals, within a very com-

1. See, for example, Braithwaite (1997); Glinskaya and Braithwaite (1998); Milanovic (1998); Lokshin and Popkin (1998); and Ovcharova, Turuntsev, and Korchagina (1998).

Branko Milanovic is with the Development Research Group at the World Bank, and Branko Jovanovic is with the Economics Department at Texas A&M University and the Development Research Group at the World Bank. Their e-mail addresses are bmilanovic@worldbank.org and bjovanovic@worldbank.org. The authors gratefully acknowledge comments from Jeanine Braithwaite, Christine Jones, Misha Lokshin, and three anonymous referees. The research was financed by World Bank research grant RPO 681-42.

pressed time period, how conceptions of well-being and deprivation respond to abrupt changes in income (see, for example, Kahneman and Thaler 1991). For most people in most countries these concepts remain relatively constant over considerable periods of time. It is therefore difficult to observe the impact of changes in external circumstances on the formation of attitudes or expectations. But the Russian experience allows us to do so. In addition, the question of what the population believes to be the minimum acceptable income has obvious political implications: if most people feel poor, they are unlikely to support reform. This article explores how the perception of the poverty line, among the population as a whole, changed in Russia during 1993–96.

I. The Model

In the literature on subjective welfare estimation the usual specification defines the minimum income necessary for a family to make ends meet (MY_f) as a dependent variable. MY_f may be considered a point on a household cost function related to a specific level of welfare u_{min}. In the most parsimonious formulation total household income (Y_f) and family size (n) are explanatory variables (Goedhart and others 1977; Hagenaars and van Praag 1985; van Praag and Van der Saar 1988):

$$\ln MY_f = fct(\ln Y_f, \ln n). \tag{1}$$

The minimum income necessary for a family to make ends meet is obtained from a so-called minimum-income question, such as "What do you consider as an absolute minimum net income (per period of time) for a household such as yours?" (see Flik and van Praag 1991). Obviously, family size positively influences the minimum income or subjective poverty line—we use the terms interchangeably. In addition, the actual level of family income, which may be regarded as a proxy for the family's permanent income, positively influences the subjective poverty line. The rationale is that families accustomed to a higher standard of living will, everything else being the same, have higher aspirations and hence a higher estimate of their minimum income. van Praag (1971) calls this "the preference drift." Its value, in a double-log formulation such as equation 1, lies between 0 and 1. If the preference drift equals 0, then the subjective poverty line becomes an absolute poverty line. At the other extreme, when the preference drift equals 1, every increase in real income exacts the same percentage increase in what is perceived to be the poverty line. The poverty line then becomes fully relative. Not surprisingly, most research has yielded values of the preference drift between 0.4 and 0.7 (see, for example, Flik and van Praag 1991 and van Praag and Flik 1992). These values fit well with our intuitive perception that, as people get richer, they set the necessary minimum higher, but do not raise it (in percentage terms) as much as the rise in their income.

Answers to the minimum-income question yield several observations. We fit the regression based on these observations; the intersection of the regression and

45-degree lines is defined as the social subjective poverty line (see point *A* in figure 1). Households to the left of *A* have income below the regression line (that is, less than society deems needed). They are considered poor. Households to the right of *A* are not considered to be socially poor because their actual income is above the regression line—even if they may consider themselves to be poor (for example, their required minimum income may lie at C, much above their own income).

Writing equation 1 in log-linear form, we derive equation 2:

$$\ln MY_f = \beta_0 + \beta_1 \ln Y_f + \beta_2 \ln n. \tag{2}$$

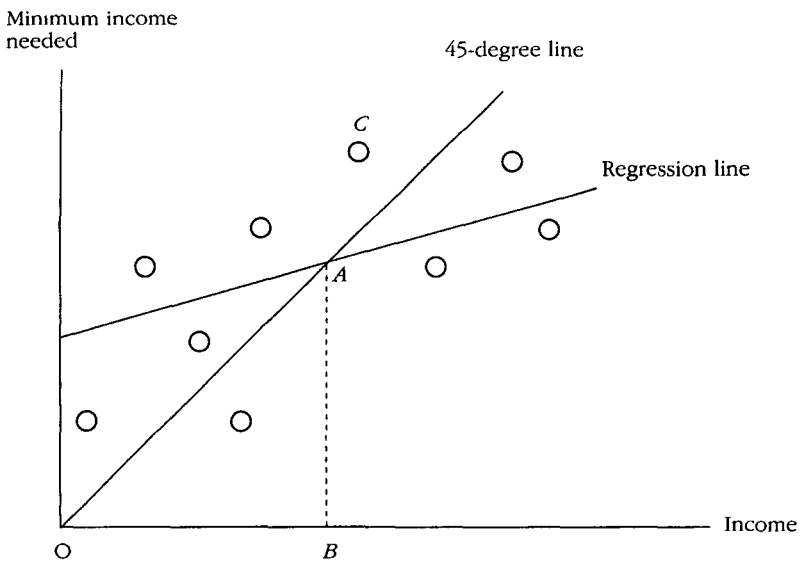
And letting $MY_f = Y_f$, we derive equation 3:

$$(1 - \beta_1) \ln Y_f = \beta_0 + \beta_2 \ln n. \tag{3}$$

The elasticity of family size with respect to the subjective poverty line (that is, the parameter θ in the expression of the equivalent income, Y/n^{θ}, which is defined below) becomes $\frac{\beta_2}{1-\beta_1}$.

Unfortunately, the data set that was available to us does not contain the minimum-income question as we have explained it (see section II). Instead of asking the household head for his or her opinion on the minimum income for the entire family, the enumerator asked, "What income, in your opinion, constitutes the subsistence minimum per person at the present time?" This is a very general minimum-income question, in effect asking for the respondent's view of the minimum income for an adult (since "person" is likely to be interpreted as "adult").

Figure 1. *The Social Poverty Line*



Note: The social subjective poverty line is defined by point *A*.

It does not ask what the minimum income per person would be for that family.[2] Therefore, we cannot apply our theory in a straightforward fashion; we must use an alternative approach.

Equation 4 shows the effective formulation based on the question as asked and introduces other control variables that may be relevant:

$$(4) \qquad \ln AMY = fct(\ln Y^*, age, age^2, SETTLEMENT, REGION, time).$$

AMY represents answers to the question about the minimum income for an adult. The other variables are Y^*, the true income level of the household (income per equivalent adult in the household), age of the respondent, size of the settlement (a dummy variable), region where the family lives (a dummy variable), and time. The crucial variable is true income. In trying to determine what people believe to be the minimum income for an adult in Russia based on their own income, we must define Y^* so that it accurately reflects the household's economic welfare. Clearly, Y^* is unlikely to be total family income because total family income does not take into account the number of people in the family. But it could be per capita income or income per equivalent adult, which accounts for economies of size. Therefore, we define Y^* as Y/n^θ, where θ is a parameter for economies of size ranging from 0 (full economies of size) to 1 (no economies of size or per capita measurement).

Determining the right θ is problematic. We argue that the right θ (θ^*) will make the sign of the variable for household size (n) not significantly different from 0. The rationale is as follows. Once we identify true household income, there is no reason why a household's size or composition should affect what people regard as the minimum income for an adult. Therefore, we try different values of $Y(\theta^*)$ and choose the $\theta = \theta^*$ that makes the coefficient on $\ln n$ equal to 0 in equation 3. For values of $\theta < \theta^*$ we expect the coefficient on $\ln n$ to be negative because the economic welfare of large households is overestimated (they are not as rich as they seem). Their estimate of minimum income (*AMY*) is systematically biased downward, which in turn leads to a negative correlation between *AMY* and $\ln n$ and to a negative regression coefficient. For values of $\theta > \theta^*$ the opposite is true, and we expect the regression coefficient to be positive.

Including the *age* and age^2 variables accounts for the (parabolic) life-cycle effect whereby perceived needs increase until they reach a peak and decrease thereafter. We have to be careful with the interpretation of these variables because they record the age of the respondent and not necessarily the age of the household head.

We capture the importance of the environment on the perception of the poverty line by introducing a dummy variable for the size of the settlement and a dummy variable for the region where the settlement is located. People living in big cities or richer regions (for example, Moscow or St. Petersburg) face higher

2. If respondents are rational, there is no difference between asking them the minimum total income for their family and the minimum per capita income for their family. The answers to the latter can simply be multiplied by the number of family members to obtain the minimum family income.

prices and would be expected to pitch their poverty line higher.[3] The social reference (demonstration) effect may also be important in larger cities, as people seeing the wealth of others come to expect more. Living in a harsh climate might also increase people's perception of the necessary minimum income.

Finally, we introduce a variable for time in our model in order to capture changes in the perception of the poverty line over time. Our hypothesis is that the subjective poverty line will fall as time passes and people adapt to worsening conditions. Our data span March 1993 to September 1996, during which time the real income of the Russian population declined severely. The decline is estimated at 14 percent based on our survey results or at almost 20 percent based on official (*Goskomstat*) monthly estimates of population income over the same period (see figure 2). Note from figure 2 that our data underestimate income by about 40 percent compared with the official data, but also that the underestimation diminishes with time. Most of the difference is due to the omission of income in kind from the survey data. Both official and survey data almost certainly underestimate the gray or black economy.

We ask whether, in addition to the income effect, the passage of time and the realization of what seem to be ever-worsening circumstances lead the public to scale down its expectation of the minimum tolerable income. As adaptation to less-fortunate circumstances proceeds, we expect the *time* variable to enter negatively in equation 3.

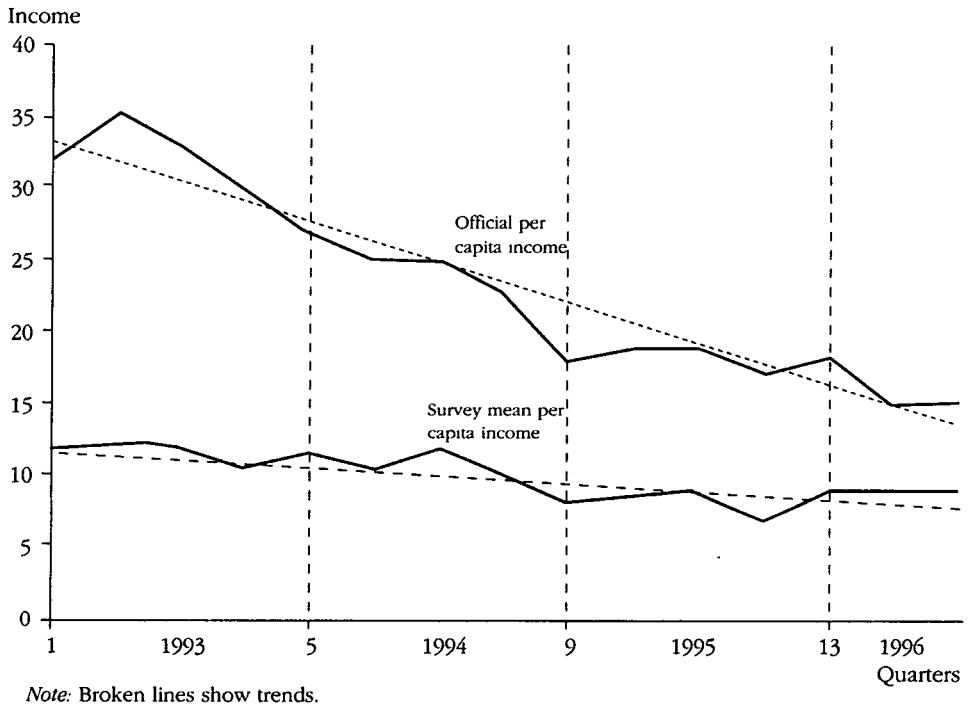
II. The Data

We use the 29 cross-sectional data sets from the All-Russian Centre for Public Opinion Research (VCIOM by its Russian abbreviation), covering the period from March 1993 to September 1996. The survey is a representative sample of Russian households. It was conducted monthly between March 1993 and January 1994, and approximately every second month since then. Although most of the survey questions are concerned with the household (family), some questions target individuals. These variables include, among others, gender, age, and education. In most surveys such questions are targeted specifically to the head of the household. Here, however, the respondent is not necessarily the household head. The fact that the respondent need not be the household head might jeopardize the accuracy of some data (for example, the respondent may not be fully aware of all the components of household income).

The original data set consisted of 91,090 observations spread over 29 cross sections. We reduced the number of observations to 80,826 after omitting those that did not contain information on family income (total or by components). Individual cross sections contained between 3,626 (January 1994) and 2,034 (September 1996) observations. The reduction of the sample size over time was

3. Our database is not deflated for regional price differences because regional consumer price indexes are not available.

Figure 2. *Real Population Income, 1993–96*
(thousands of constant March 1993 rubles per capita per month)



Note: Broken lines show trends.
Source: Survey per capita income calculated from VCIOM surveys. Official per capita income from monthly *Goskomstat* statistics.

considerable; however, according to the VCIOM staff, it did not make the sample less representative because the sampling procedures were improved.[4]

Table 1 gives summary statistics for the basic characteristics of the households and respondents surveyed.[5] Total family income is computed as the sum of income components: main income and income from a second job, private sector activities, pensions, other social transfers (including family, unemployment, and disability benefits), stipends, alimony, income from financial papers (stocks and bonds, vouchers, interest from savings accounts), sale of self-produced goods, and other monetary income (appendix A provides details). We use the all-Russia monthly consumer price index (CPI), with March 1993 as the base, to deflate all the monetary variables. We assume that inflation affects all regions equally (regional CPIs are not available).

In real terms the subjective minimum income for an adult (*AMY*)—calculated as an individual-weighted average of *AMY*s over all households—decreased dramatically between March 1993 and September 1996. It started out higher than

4. We owe this information to Jeanine Braithwaite.
5. More detailed statistics are available from the authors.

Table 1. *Summary Descriptive Statistics, 1993–96*

Variable	*1993*	*1994*	*1995*	*1996*
Women (percent)	60.06	60.12	57.04	56.65
Age of respondent (years)	42.71	42.40	42.58	43.29
Education attained by respondent (percent)				
Primary or less	6.10	5.85	5.82	5.51
Secondary, incomplete	8.08	8.15	8.75	8.56
Secondary, completed with no diploma	2.69	2.58	2.31	2.53
Technical school, less than secondary	4.44	4.07	3.99	4.45
Secondary, completed with diploma	15.34	14.88	15.16	13.79
Technical school and secondary	9.80	10.22	9.87	10.29
Vocational school	27.57	28.33	27.34	27.19
University, 3–4 years	3.20	3.44	3.81	3.90
University, completed	22.77	22.49	22.96	23.78
Number of observations	34,759	23,040	12,189	10,298

Note: Values are averages of the monthly averages, with 10 months for 1993, 8 months for 1994, 6 months for 1995, and 5 months for 1996.

Source: Authors' calculations from the VCIOM data.

Rs35,000 (rubles) in the early surveys and ended at Rs15,000 (see figure 3). The Ministry of Labor's official minimum income for an adult (*prozhitochnyi minimum*) remained constant in real terms at around Rs10,000.[6] The gap between the subjective and the official minimum diminished steadily, as the public perception of the minimum income for an adult Russian gradually approached the official minimum.

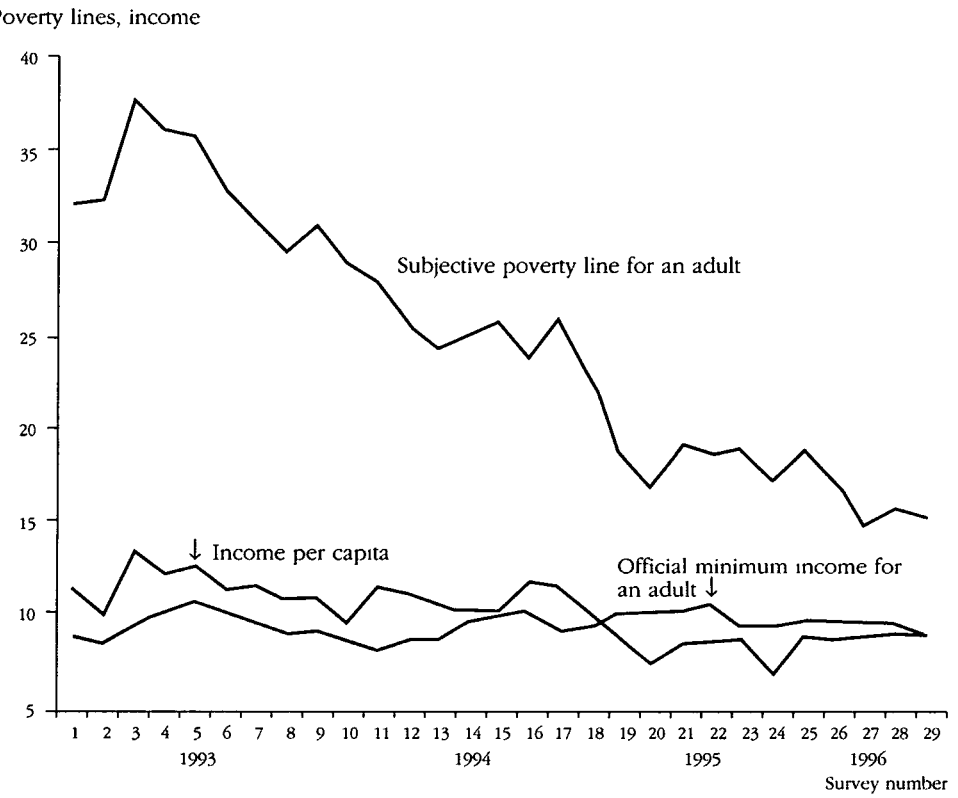
The composition of households, as well as the demographic characteristics of the respondents, stayed roughly the same over time (see tables 1 and 2). Over the entire survey period the average household consisted of 3.1 members, with 0.7 children (table 2). For comparison, according to the all-Russia official statistics for 1994, the average household size was 2.84 members (Goskomstat Rossii 1995). The average survey respondent was 42.7 years old and spent 11.2 years in school (table 1). The average duration of schooling of the population over 15 years of age calculated from the 1993 Russian Longitudinal Monitoring Survey was a little over nine years. Of all the respondents, 59.2 percent were women; according to official statistics, women made up 53 percent of the Russian population in 1995.

Most of the respondents (76.5 percent) lived in urban areas, a percentage close to the official 1995 statistics (73 percent). A plurality of respondents (46 percent) lived in cities with populations less than 100,000, followed by 25.1 percent who lived in cities with populations over 1 million.

In order to check for possible outliers in the data, especially in Y_f and AMY, we create a flag variable. We compute the variable for each cross section sepa-

6. The official minimum is composed of a given bundle of food and nonfood goods. Its slight oscillations around Rs10,000 in March 1993 prices are due to the fact that the CPI we use to deflate the nominal monthly values of the official minimum at times might have increased faster or slower than the cost of the minimum bundle of goods.

Figure 3. *The Average Subjective Poverty Line for an Adult, the Official Poverty Minimum for an Adult, and Average Per Capita Income, 1993–96*
(thousands of March 1993 rubles per month)



Note: The average subjective poverty line for an adult is the simple individual-weighted average of poverty lines (*AMY*) in the surveys; the official poverty line for an adult is the all-Russia official poverty line (*MinTruda Rossii*); the average per capita income is the average income from VCIOM surveys (as in figure 2).
Source: Authors' calculations.

rately according to the method developed by Hadi (1992, 1994) using the *hadimvo* procedure in STATA; 2,634 observations are identified as possible outliers, representing 3.28 percent of the total sample.

III. Estimating the Subjective Poverty Line

We first try to estimate true household income using different values of θ. To do so, we run the basic model, equation 3, including in addition ln n as the control variable for household size. Figure 4 shows how the coefficient on ln n changes as θ in $Y(\theta^*)$ varies from 0 to 1. For $\theta = 0.62$, the coefficient equals 0. (It is insignificantly different from 0 for a few other values around 0.6, but takes its lowest value for $\theta = 0.62$.)

Table 2. *Poverty Lines and Household Characteristics, 1993–96*
(thousands of rubles per month at March 1993 prices)

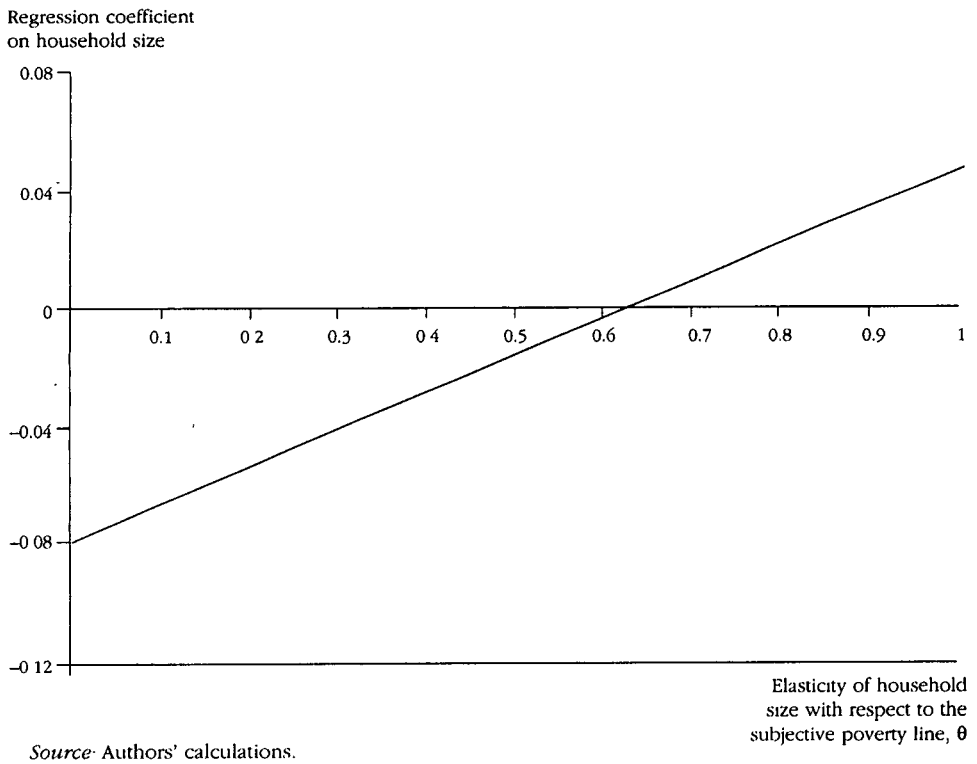
Variable	*1993*	*1994*	*1995*	*1996*
Total real family income				
Average	35.29	34.14	25.11	27.22
Bottom quartile	8.42	7.44	6.16	6.61
Top quartile	90.45	88.93	61.38	68.44
Per capita real family income[a]				
Average	12.51	11.97	8.89	9.75
Bottom quartile	3.42	3.02	2.48	2.55
Top quartile	32.38	31.48	21.92	24.52
Subjective per adult poverty line	32.63	25.30	18.24	16.27
Size of household	3.08	3.13	3.07	3.07
Children per household	0.69	0.72	0.68	0.65
Share of pensions in household income (percent)	15.8	20.6	22.9	22.7
Gini coefficient (all Russia)	44.6	45.4	41.4	45.2
Number of observations	34,759	23,040	12,189	10,258

Note: Values are averages of the monthly averages, with 10 months for 1993, 8 months for 1994, 6 months for 1995, and 5 months for 1996.

a. Per capita real family income is calculated using number of people per household as weights.

Source: Authors' calculations from VCIOM data.

Figure 4. *Household Size and the Elasticity of Household Size with Respect to the Subjective Poverty Line*



Source: Authors' calculations.

We then directly estimate equation 3 using $Y^* = \frac{Y}{N^{0.62}}$ (table 3).[7] All the regressions are run with Huber (robust) variances to adjust for the fact that the observations are drawn from different time clusters (that is, from the 29 surveys). The Huber correction also adjusts for the fact that the variability of the observations within each survey is less than it would be if all observations were randomly drawn from the population at large. That is, the variability of the observations from the pooled cross sections is less than what it would be if all 80,000 of our observations were drawn from one cross section. Nevertheless, the *t*-values in table 3 show that most of the coefficients are significant at a probability greater than 99 percent.

The elasticity of the subjective poverty line for an adult with respect to income is 0.144 for the overall sample and 0.132 in a regression that excludes Hadi outliers (see appendix B on the *hadimvo* procedure). These values are significantly lower than those reported by Frijters and van Praag (1994) in their study of the former Soviet Union and Russia. Frijters and van Praag report preference drift values of 0.62 and 0.64 for Russia in 1993 and 1994, respectively. For the Soviet Union in 1991, they report a value of 0.41. However, our results and their results are not entirely comparable. Frijters and van Praag use a variant of the so-called income evaluation question in order to obtain the left-side variable (the Leyden poverty line).[8] We use the income level that people consider the minimum value for an adult.

We calculate a value for preference drift that is significantly lower than the value found in some Western countries. Flik and van Praag (1991), for example, report a value of 0.59 for the Netherlands. Hagenaars and van Praag (1985) report a value of 0.54 for a collection of West European countries. Part of the difference may be due to the richer choice of control variables included here (regional and size of settlement dummies), as well as to the introduction of the *time* variable. In effect, if we run a very parsimonious formulation, such as equation 1, which is basically what Hagenaars and van Praag (1985) do (without a control variable for household size), the preference drift increases from 0.14 to 0.23.

The latter value (0.23) is almost identical to the preference drift of 0.223 obtained by Ravallion and Lokshin (1999). Ravallion and Lokshin use what they dub the economic ladder question, whereby individuals rank their own subjective level of living from 1 (the poorest) to 9 (the richest). Similar to the rest of the subjective poverty literature, underlying differences in real income explain the rankings. Ravallion and Lokshin use a representative sample of the Russian population in 1996 from the Russian Longitudinal Monitoring Survey.

7. We also estimate equation 3 for each individual year using θ calculated for that year. The results are available on request. The coefficients of equation 3 are stable regardless of whether we use yearly cross sections or pooled data.

8. Under the income evaluation question methodology a respondent is asked to write the level of income that the respondent's family would consider to be very bad, bad, middling, good, and very good. The mean of the five answers is defined as the Leyden poverty line. For more on the methodology see Hagenaars and van Praag (1985) and Flik and van Praag (1991).

Table 3. *Regression Results for the Subjective Minimum Income for an Adult*

Variable	*(1) Basic equation with Huber (robust) variances*	*(2) Regression without Hadi outliers*	*(3) Regression with Gini coefficient*	*(4) Regression with median survey income*
Ln equivalent income (Y^*)[a]	0.144	0.132	0.132	0.126
	(24.0)	(22.3)	(22.3)	(21.3)
Age	0.016	0.017	0.017	0.017
	(14.0)	(17.1)	(17.1)	(16.6)
Age^2	−0.0002	−0.0002	−0.0002	−0.0002
	(−18.9)	(−20.5)	(−20.6)	(−20.8)
Small towns and villages (less than 100,000 people)	−0.062	−0.065	−0.066	−0.073
	(−4.2)	(−4.3)	(−4.4)	(−4.7)
Towns (100,000–500,000 people)	0.064	0.056	0.056	0.054
	(4.0)	(3.4)	(3.4)	(3.2)
Medium-size cities (500,000–1 million people)	0.058	0.059	0.058	0.059
	(4.7)	(4.8)	(4.9)	(5.0)
Northern region	−0.243	−0.220	−0.223	−0.226
	(−12.0)	(−11.7)	(−10.7)	(−11.8)
Central and Central Black Earth	−0.330	−0.307	−0.311	−0.328
	(−13.9)	(−13.0)	(−10.9)	(−12.1)
North Caucasus	−0.225	−0.210	−0.210	−0.209
	(−5.7)	(−5.6)	(−5.5)	(−5.6)
Volga-Vyatka	−0.324	−0.291	−0.295	−0.315
	(−13.1)	(−12.3)	(−10.6)	(−11.5)
Volga	−0.256	−0.236	−0.239	−0.253
	(−6.9)	(−6.9)	(−7.0)	(−7.7)
Urals	−0.194	−0.179	−0.182	−0.201
	(−7.8)	(−7.5)	(−6.7)	(−7.6)
West Siberia	−0.150	−0.131	−0.132	−0.139
	(−6.4)	(−6.0)	(−5.6)	(−6.2)
East Siberia and Far East	0.035	0.030**	0.028**	0.199**
	(1.6)	(1.4)	(1.3)	(1.0)
Time	−0.017	−0.017	−0.017	
	(−17.7)	(−18.0)	(−17.3)	
Ln median income (by survey)				1.50
				(11.3)
Regional Gini coefficient (by survey)			−0.054**	−0.284**
			(−0.4)	(−1.9)
Constant	2.889	2.822	2.847	−1.93
	(82.7)	(97.8)	(40.0)	(−4.8)
Sample size	79,595	76,965	76,965	76,965
R^2	0.189	0.191	0.191	0.178
F-value	210.6	246.9	243.4	408.8

** Not significant.

Note: *t*-values are in parentheses. All coefficients are significant at the 1 percent level, unless noted. For size of settlement, the omitted category is larger cities (population over 1 million). For region, the omitted variable is the city of Moscow.

a. Defined as $Y/N^{0.62}$.

Source: Authors' calculations.

The fact that two independent studies using different surveys both derive very low values for preference drift in Russia requires explanation. There are, we believe, two possible reasons. First, people's views may vary less with income when they are asked what they consider to be a minimum amount for an adult in general (as in the VCIOM survey) than when they are asked the minimum for their own family. When referring to their own family, poor people may pitch their minimum fairly low, while the rich may find it hard to imagine living without a relatively high income. But the opinions of the poor and rich may not be so far apart in reference to an abstract (adult) individual.

Second, low preference drift may suggest relative homogeneity in people's perceptions. People at the top of the income scale may not evaluate the minimum income needed to make ends meet much differently than poor people. The relatively recent explosion of income inequality may explain this homogeneity. People who recently had similar incomes will not suddenly diverge very much in their perception of the poverty line. In countries such as those in Western Europe, income differences historically have been greater and income mobility lower (in the sense that people who currently have high incomes probably had high incomes five or ten years ago). In those countries rich and poor people's perceptions of the poverty line may differ significantly. By contrast, Russia, until recently, was very egalitarian and was then subjected to an almost random and huge income shock. Some people's incomes increased manifold, and other people's incomes dropped significantly, but both groups' perception of the minimum income may have remained similar.

The economies of scale parameter (θ) is 0.62. This result is close to the value of 0.5 reported by Frijters and van Praag (1994) for Russia in 1994 and 0.42 reported by Ravallion and Lokshin (1999) for Russia in 1996.[9] Compared with other methods, subjective methods yield a relatively low value for the equivalence scale (see Atkinson, Rainwater, and Smeeding 1995).

The parabolic age effect implies that the subjective poverty line rises with age until a certain point, after which needs decrease. The peak occurs at around 40 years, some four and a half years later than reported by Frijters and van Praag (1994). However, because the variable captures the respondent's age, it may not represent the age composition of the household.

The dummy variables adjust for the size of the settlement where the family lives and for the region. For the size of the settlement the omitted category is larger cities (with populations over 1 million). The subjective poverty line is lower in small towns and villages. Surprisingly, the perceived minimum income for an adult is higher in towns and medium-size cities than in very large metropolitan areas. We would expect that needs would increase monotonically with the size of the settlement, perhaps because the cost of living is higher or the demonstration effect is greater. The absence of this regularity for large metropolitan areas may be due to the fact that the regional variables are picking up some of the effect.

9. The standard error for the Ravallion-Lokshin estimate is 0.148.

For the regional variables the city of Moscow is the omitted category. Of course, subjective needs in all other regions except East Siberia and Far East are less than those in the city of Moscow. (The city of Moscow does not include the Moscow region, which is part of the Central and Central Black Earth region.) Compared with Moscow, the subjective poverty line is lower, ceteris paribus, by between 13 percent in West Siberia and 30 percent in the Central and Central Black Earth and Volga-Vyatka regions. In East Siberia and Far East the subjective needs are about the same as in Moscow. The high poverty line in East Siberia and Far East is explained by the harshness of the climate (which raises housing and energy expenditures) and the regions' remoteness (which means that prices of consumption goods are higher). We discuss the difference between the regional subjective and official poverty lines in section IV.

The variable *time*, measured in months with March 1993 as the starting point, shows how the subjective poverty line for an adult changes as people downscale their expectations.[10] In principle, we would expect this effect to operate through the income variable—lower income would, through the preference drift, reduce the subjective poverty line. But given the rapid decline in real income in Russia during 1993–96, people downscaled their expectations even faster. Thus the passage of each month (after March 1993) reduced the subjective poverty line 1.7 percent. After more than three years of depression (by the fall of 1996), the public's perception of the minimum per capita income was about half of what it would have been had people maintained the same real income that they held at the beginning of the period (spring of 1993). We introduce the $time^2$ variable to determine whether the time effect subsided as the surveys progressed. It is not significantly different from 0.

In variant 3 in table 3 we introduce a measure of income inequality (the regional Gini coefficient) to account for a possible increase in the subjective poverty line due to higher inequality. Hagenaars and van Praag (1985) find that such an influence is explained by the demonstration effect (greater inequality and therefore the presence of higher incomes invite people to pitch their poverty lines higher). We calculate the Gini coefficient for per capita income for each region and for each survey and include it in the regression. However, we find no evidence that inequality influences the subjective poverty line.

Finally, we replace *time* by median per capita income (see variant 4 in table 3). We expect that the subjective poverty line will fall with people's real income. Indeed, figure 2 illustrates the decrease in mean per capita real income, and figure 3 illustrates the decrease in the subjective poverty line. We find the elasticity of the subjective poverty line with respect to survey median income to be very high: 1.5. Section V shows how this elasticity affects the calculation of the number of people who are subjectively poor. If the elasticity of the subjective poverty line with respect to mean or median income is greater than unity, the percentage

10. An alternative formulation would be to use survey dummies. The results (available on request) suggest a decreasing subjective poverty line beginning at the end of 1993.

of people who feel poor will tend to decrease as income goes down. Paradoxically, (subjective) poverty would become less widespread as people's (objective) circumstances deteriorate.

The other coefficients in variant 4 in table 3 are stable, with one exception. The inequality variable (the regional Gini) increases, but is still not significant at the 5 percent level.

IV. Comparing Regional Subjective and Official Poverty Lines

We have already seen that the subjective poverty line for an adult is several times higher than the official poverty line (*prozhitochnyi minimum*) for an adult, although the gap between the two diminishes over time. Here we look at differences in the structure (rankings) of the regional official and subjective poverty lines. Table 4 shows the ruble amounts for the official and subjective regional poverty lines in 1996. As we would expect, subjective poverty lines are always higher, but the extent of the gap differs among regions: the official poverty line is less than half of the subjective line in North Caucasus, but is almost two-thirds of the subjective line in the North.

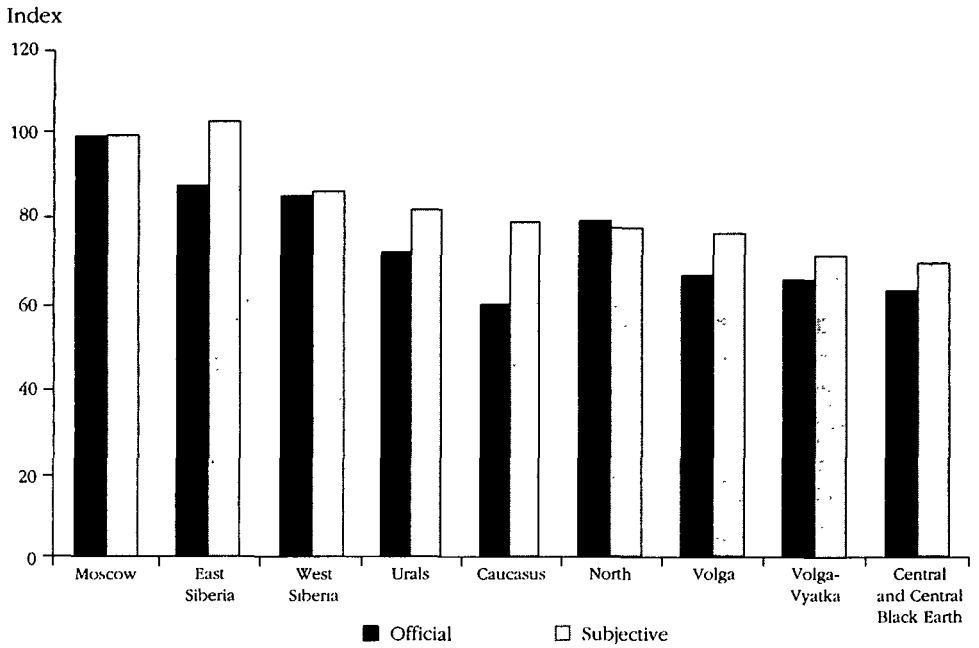
The regional differences imply that the official poverty lines do not accurately reflect the population's perception of the differences in subjective needs among the regions, even though the correlation coefficient between the official and subjective poverty lines is 0.85. Figure 5 shows that if the subjective and official poverty lines for the city of Moscow are set at 100, the relative subjective poverty lines for all but one region (North) are higher than official lines. This finding suggests a pro-Moscow bias in setting official poverty lines. For example, the official poverty line for an adult in the Caucasus is 40 percent lower than that for an adult in Moscow; but the public perception is that it should be only 20 percent less (figure 5).

Table 4. *Official and Subjective Regional Poverty Lines, 1996*
(thousands of March 1993 rubles)

Region	*Official poverty line*	*Subjective poverty line*	*Ratio of official to subjective poverty line*
North	8.30	12.7	0.65
Central and Central Black Earth	6.55	11.3	0.58
North Caucasus	6.17	12.9	0.48
Volga-Vyatka	6.82	11.5	0.59
Volga	6.84	12.4	0.55
Urals	7.47	13.4	0.56
West Siberia	8.79	14.1	0.62
East Siberia and Far East	9.06	16.8	0.54
Moscow	10.30	16.3	0.63

Note: The subjective poverty lines cover the period January–September 1996.
Source: Authors' calculations. Official poverty lines calculated from Goskomstat Rossii (1998, tables 2.7 and 4.20). Subjective poverty lines calculated from regression 2 (table 3).

Figure 5. *Official and Subjective Regional Poverty Lines Indexed to Moscow, 1996*
(Moscow = 100)



Source: Authors' calculations.

V. How Many People Are Poor?

In this section we look at the proportion of people who are poor, defined according to three criteria. The first criterion labels "subjectively poor" those households that assess themselves as poor, that is, households whose view of the minimum income for an adult is greater than their actual adult equivalent income ($AMY_f > Y_f^*$ for a given family). A problem with this criterion is that two identical households with the same income may be classified as poor and nonpoor, depending on how they perceive their own well-being.

For the second criterion, the "socially subjectively poor," we impose a social equivalence scale ($\theta = 0.62$) that may not correspond to a household's own equivalence scale. This criterion defines as poor those households whose current income per equivalent adult (Y_f^* using $\theta = 0.62$) is less than the social subjective minimum income (per adult) for such a household. Regression 3 (the variant with Huber-robust variances and excluding Hadi outliers) predicts this income measure.

According to the third criterion, the poor are those whose current income per equivalent adult (Y_f^* using $\theta = 0.62$) is less than the official all-Russia poverty line (per working adult).

Figure 6 shows the share of individuals who are poor according to the three criteria. We can draw several conclusions. First, an extremely high percentage of the population (almost always greater than 60 percent) is subjectively poor, whatever (subjective) criterion is used. This percentage is consistently higher than the percentage judged to be poor by the official poverty line.

Second, there is a clear tendency for the subjective poverty headcounts to decrease with time. The subjective poverty line fell faster than real income, so that fewer people assessed themselves as poor. The percentage of the socially subjective poor dropped from 90 percent in March 1993 to less than 60 percent in September 1996. At the same time, real average per capita income decreased 14 percent.

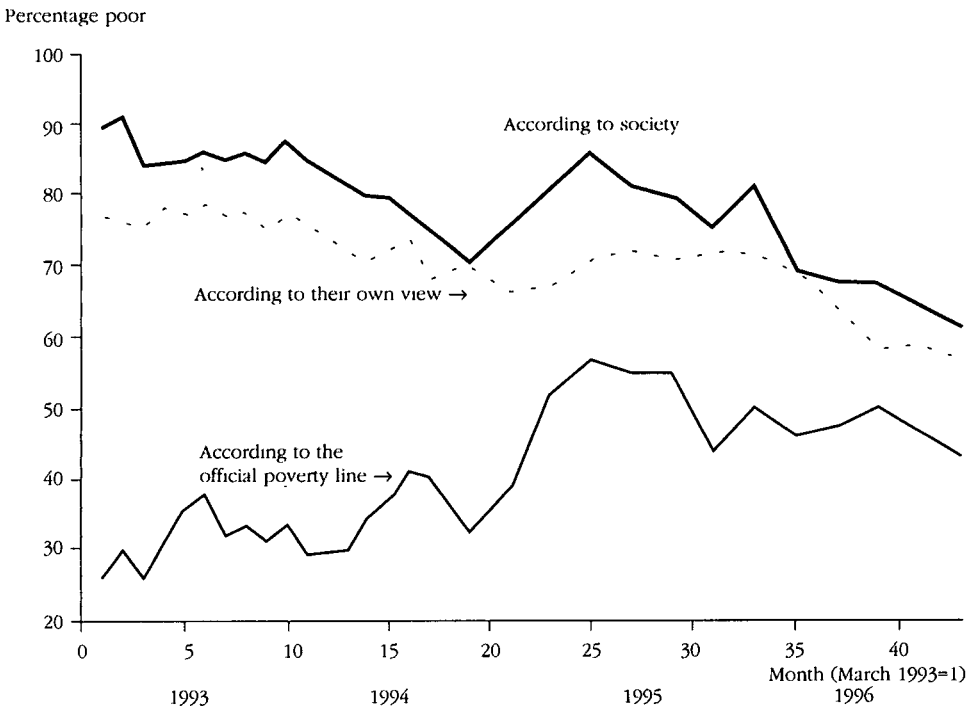
Third, because real income fell while the official poverty line remained the same, the proportion of people whose income was less than the official poverty line increased from a third of the population in 1993 to more than half in 1995, before dropping to near 40 percent in late 1996.[11] Thus the decrease in real income has made more people poor according to an objective and fixed yardstick. Ironically, the same reduction in real income has reduced people's perception of the minimum income they need to survive and has made fewer of them feel poor. Figure 7 shows that the decrease in the percentage of self-assessed poor coincided with the decline in real income. The decline in the percentage of self-assessed poor decelerated only between mid-1994 and mid-1995, when current real income took another sharp dip: a larger than usual decrease in real income was required to keep the percentage of the self-assessed poor constant. Thus there are two circumstances under which fewer people feel poor: when their real incomes grow quickly or when their real incomes fall (equally) quickly. In Russia, unfortunately, it was the second alternative that occurred.

Fourth, in all but two surveys the social subjective poverty line yields higher poverty headcounts than do people's own assessments (figure 6). This means that some households that are socially considered poor do not view themselves as such.[12] These households fall in the triangle *OAB* in figure 1, to the left of the social poverty line *AB*. These households' assessment of their well-being is better than the "social" assessment, possibly indicating the presence of the much-discussed pockets of social resilience and patience often associated with the Russian population.

11. We cannot compare this figure with the percentage of poor from the official Goskomstat statistics, which ranged between 22 and 31 percent over the same period (Goskomstat Rossii 1998: 79), or the percentage obtained from the Russian Longitudinal Monitoring Survey (see World Bank 1998: 5). Income in these surveys is defined to include noncash sources, while in the VCIOM income includes only cash sources.

12. A caveat is in order here. Since we assume that all households have the social equivalence scale reflected in $\theta = 0.62$, some households that classify themselves as poor may in fact have a lower θ and thus not regard themselves as poor. The opposite classification mistake is possible for nonpoor households whose θ is greater than 0.62.

Figure 6. *Share of Poor Individuals in the Total Population According to Three Concepts of Poverty, 1993–96*



Note: All shares are individual-based
Source: Authors' calculations.

VI. Conclusions

In the three and a half years (March 1993 to September 1996) covered by the VCIOM surveys of the Russian population, real per capita income fell 15–20 percent. This decrease came on top of severe income contractions in 1991 and 1992. Thus the Russian population experienced one of the worst peacetime depressions in the twentieth century. At the same time, income inequality increased substantially. In this article we have analyzed what happened to the public's perception of the minimum income needed to make ends meet under these exceptional conditions.

We would expect the subjective poverty line to decrease as well. Indeed, the elasticity of the subjective poverty line with respect to median population income is above unity—a very high 1.5. The *time* variable is significant, as the subjective poverty line fell 1.7 percent each month. Thus after more than three years of depression, the public's perception of the minimum income an adult would need to survive was about half of what it would have been had real incomes remained what they were at the beginning of the period.

Figure 7. *Real per Capita Income and the Self-Assessed Poor, 1993–96*

Note: Real per capita income is calculated from VCIOM surveys.
Source: Authors' calculations.

At the same time, the cross-sectional preference (income) drift parameter is relatively low, at slightly less than 0.15: each percentage point fall in real income, on average, reduces the public's perception of the poverty line by 0.15 percent. Even after dropping the *time* variable, income drift remains low (0.23) compared to West European countries, where it ranges from 0.4 to 0.7. This seems to suggest that people's perception of the subjective poverty line in Russia was relatively homogeneous. Those at the top of the income scale did not evaluate the minimum income needed to survive much differently than did the poor. We explain this result in that the poverty line question may have been formulated so as to address implicitly the needs of an adult or that income inequality had exploded relatively recently. The poverty line question might have influenced the answers in the sense that the rich and the poor might differ less in their responses when asked to assess how much the general person needs to survive than when asked how much they themselves need. The recent increase in inequality might mean that people who had similar incomes until only recently will not suddenly diverge very much in their perception of the poverty line.

We also find that subjective needs vary across regions. The poverty line is highest in East Siberia, Far East, and the city of Moscow. The poverty line in other regions is between 13 percent (West Siberia) and 30 percent (Central and

Central Black Earth, and Volga-Vyatka) less than that in the city of Moscow. These differences are smaller than the differences in the official regional poverty lines, suggesting the existence of a pro-Moscow bias in setting the official lines.

During the time period studied, a very high percentage of the population (always more than 60 percent) considered itself poor according to the social subjective poverty line. The percentage of the subjectively poor tended to decline more than proportionately with the decline in real income. We thus face a somewhat unusual situation in that the percentage of the subjectively poor decreased more or less in step with the reduction in people's real income. Only larger than usual declines in income kept the percentage of the poor unchanged.

It is also noteworthy that the percentage of the self-assessed poor was always lower than the percentage of the socially subjective poor. Thus part of the population regarded their own income as adequate, although they were deemed poor according to the public's perception of minimum income. These last two findings—the decline in the percentage of the subjectively poor as real income fell and the lower percentage of self-assessed poor than socially subjective poor—suggest that to adapt to the worsening circumstances, people sharply reduced their perception of the minimum income needed for survival.

Appendix A. Construction of the Income Variables

The VCIOM data set contains several income variables that are measured on both the individual and household levels. Two reported income variables are individual main income and individual income from a second job. Household (family) income components include family income from a main job, family income from a second job, income from private sector activities, pensions, other social transfers, stipends, alimony, income from financial assets (stocks, bonds, vouchers, interest income), income from sale of self-produced goods, and other monetary income.

The total family income variable is also included in the data set, and it is supposed to equal the sum of the family income components, but rarely does. In some cases total family income was reported missing, although the income components were available. Also, in some cases, even though all the income components were missing, total family income took a positive value. Furthermore, there were inconsistencies in reported individual and family main income, as well as between individual and family income from a second job.

We computed total family income as the sum of the family income components, that is, as the sum of main job income, income from a second job, and income from private sector activities, pensions, benefits and subsidies, stipends, alimony, income from financial assets, income from sale of self-produced goods, and other monetary income. This was done as follows:

- The individual main income variable before April 1994 corresponds to the variable *main_inc*, and to the variable *main_in2* thereafter (the two variables

have the same definition, only the name has been changed). Thus for any survey after April 1994, we replace the value of *main_inc* with *main_in2*. However, no data are available for November 1995 because that survey did not ask a question concerning the main income; therefore we do not have observations on the main individual income for that survey.

- The variables family income from a main job and from a second job are at least equal to the corresponding variables for the individual. Therefore, where the data on family income are missing or less than the individual income, we replace the value of family income with the observation on individual income.

In cases where all the income components are missing, we replace our total income with VCIOM-computed total income. Also, where our total income is less than the VCIOM-computed total income, we take the VCIOM value. Although the code book reports that a value of zero should be treated as a missing variable, both true zero and missing responses seem to be coded as zero. We do not attempt to distinguish between the two because efforts in that direction are unlikely to be fruitful.

Appendix B. The Hadi Procedure

The basic outline of the Hadi (1992, 1994) procedure is as follows. We define a measure of distance from an observation to a set (cluster) of points. The points are scatter plots, with one variable on the x-axis and one on the y-axis. Initially, we have only three points because we operate with only two variables (we assume explicitly that outliers are in the income variables and not in human capital or other categorical variables). We introduce an additional point (observation of total income and *pov_line*) and measure the distance between that point and the initial group of points. Once the base cluster is established, a more standard mean-based center of the r-observations cluster is defined, and the $r + 1$ observations closest in the covariance-matrix sense are chosen as a new base cluster. This is repeated until the base cluster has a certain number of points, and when it reaches the critical size, the distance rule changes. A base cluster of r points is selected (r is defined as $k + 1$, where k is the number of variables), and then that cluster is continually redefined by taking the $r + 1$ point "closest" to the cluster as the new base cluster. The distance rule used in the STATA *hadimvo* procedure is based on a matrix of second moments, with the median of variables removed). In this article we use variables *tot* (total income) and *pov_line* to flag the Hadi outlier.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Atkinson, Anthony B., Lee Rainwater, and Timothy M. Smeeding. 1995. *Income Distribution in* OECD *Countries: Evidence from the Luxembourg Income Study.* OECD Social Policy Studies 18. Paris: Organisation for Economic Co-operation and Development.

Braithwaite, Jeanine. 1997. "The Old and the New Poor in Russia." In Jeni Klugman, ed., *Poverty in Russia.* Washington, D.C.: World Bank.

Flik, Robert J., and Bernard M. S. van Praag. 1991 "Subjective Poverty Line Definitions." *De Economist* 139(3):311–31.

Frijters, Paul, and Bernard M. S. van Praag. 1994. "Estimates of Poverty Ratios and Equivalence Scales for Russia and Parts of the Former USSR." Rotterdam. Processed.

Glinskaya, Elena, and Jeanine Braithwaite. 1998. "Poverty, Inequality, and Income Mobility in Russia, 1994–96." Paper prepared for the American Population Association, Chicago, April 2–4. Processed.

Goedhart, Theo, Victor Halberstadt, Arie Kapteyn, and Bernard M. S. van Praag. 1977. "The Poverty Line: Concept and Measurement." *The Journal of Human Resources* 12(4): 503–20.

Goskomstat Rossii. 1995. *Rossiiski Statisticheskiy Ezhegodnik 1995.* Moscow.

———. 1998. *Rossiya v tsifrah 1998.* Moscow.

Hadi, A. S. 1992. "Identifying Multiple Outliers in Multivariate Data." *Journal of the Royal Statistical Society,* Series B, 54(3):761–71.

———. 1994. "A Modification of a Method for the Detection of Outliers in Multivariate Samples." *Journal of the Royal Statistical Society,* Series B, 56(2):393–96.

Hagenaars, Alvi J. M., and Bernard M. S. van Praag. 1985. "A Synthesis of Poverty Line Definitions." *Review of Income and Wealth* 31(2):139–54.

Kahneman, Daniel, and Richard Thaler. 1991. "Economic Analysis and the Psychology of Utility: Applications to Compensation Policy." *American Economic Review Papers and Proceedings* 81(2):341–46.

Lokshin, Michael, and Barry M. Popkin. 1999. "The Emerging Underclass in the Russian Federation: Income Dynamics, 1992–96." *Economic Development and Cultural Change* 47(4):803–29.

Milanovic, Branko. 1998. *Income, Inequality, and Poverty during the Transition from Planned to Market Economy.* Washington, D.C.: World Bank.

Ovcharova, Liliana, E. Turuntsev, and Irena Korchagina. 1998. "Bednost': gde porog." *Voprosi ekonomiki* 7(2):61–73.

Ravallion, Martin, and Michael Lokshin. 1999. "Subjective Economic Welfare." Policy Research Working Paper 2106. World Bank, Washington, D.C. Processed.

van Praag, Bernard M. S. 1971. "The Individual Welfare Function in Belgium: An Empirical Investigation." *European Economic Review* 2(3):227–69.

van Praag, Bernard M. S., and Robert J. Flik. 1992. "Subjective Poverty." Final report in the framework of the European Eurostat Project, Enhancement of Family Budget Surveys to Derive Statistical Data on Least Privileged Groups. Foundation for Economic Research and Research Institute for Population Economics, Rotterdam. Processed.

van Praag, Bernard M. S., and Nico Van der Saar. 1988. "Empirical Uses of Subjective Measures of Well-Being: Household Cost Functions and Equivalence Scales." *Journal of Human Resources* 23(2):193–210.

World Bank. 1998. "Poverty Policy in Russia: Targeting and the Longer-Term Poor." Eastern Europe and Central Asia Region. Poverty Reduction and Economic Management (PREM), Washington, D.C. Processed.

THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 561-78

Labor Market Analysis and Public Policy: The Case of Morocco

Julia Lane, Guillermo Hakim, and Javier Miranda

This article uses detailed industry and household data to understand why Morocco's labor market performed poorly in 1985–95. The data indicate that marked structural changes and weak demand in the product market were responsible. This article makes two contributions to the literature. The first is specific: it underscores that the demand for labor is a derived demand and that the performance of the product market is an important determinant of the performance of the labor market. The second is more general: it demonstrates that this kind of microeconomic analysis, using data sets that are often available in developing countries, can inform policy design.

Since 1985 Morocco has experienced increasing unemployment in urban areas and stagnant wages in the manufacturing sector (World Bank 1995). We must understand the reasons for the labor market's poor performance in order to design policies that will help to generate adequate jobs and earnings. In this article we use two data sets for Morocco to demonstrate how standard econometric techniques can disentangle the causes of such poor performance. These types of data sets are often available in developing countries.

We look at demand- and supply-side factors to explain Morocco's stagnant earnings growth. On the supply side, this stagnation could reflect a deterioration in the quality of the labor force. If so, policy reform should concentrate on improving the external and internal efficiency of expenditures on general education and vocational training. On the demand side, weak job and earnings growth may be a result of inappropriate government intervention: a stringent labor code that impedes firms' ability to hire and fire workers or a minimum wage policy that reduces firms' incentives to create jobs. If so, policy reform should include legislation to liberalize labor market regulations.

In this article we argue that the labor market's poor performance during 1986–95 can be traced to, first, a shift in the employment composition of manufacturing to low-paying industries linked to the export sector, which depressed average wages, and, second, stagnant product demand, which weakened the derived demand for labor. Stagnant product demand also explains Morocco's high unem-

Julia Lane is with the Urban Institute, Guillermo Hakim is with the Human Development Group of the Middle East and North Africa Region at the World Bank, and Javier Miranda is with the Department of Economics at American University. Their e-mail addresses are jlane@ui.urban.org, lhakim@worldbank.org, and jmiranda@ces.census.gov.

ployment: although there has been some employment growth in the formal manufacturing sector, that growth has been too weak to offset the expansion of the urban labor force.

Supply-side changes over the past decade have put upward, not downward, pressure on earnings. A much higher proportion of workers were in their prime earning years in 1993 than in 1986. The average education level of the labor force was also higher in 1993 than in 1986. As for demand, standard measures of labor market flexibility show that labor market rigidity is not the prime reason for poor performance: Morocco's recourse to temporary contracts has allowed firms to adjust their employment levels as freely as firms in countries that are part of the Organisation for Economic Co-operation and Development (OECD). In particular, the elasticity of labor demand with respect to output, rates of job creation and destruction, and speed of adjustment to demand shocks in Morocco are similar to those in other OECD countries.

One reason for the stagnant wages in Morocco is that low-paying industries have been the major source of employment growth. The three industries with the fastest employment growth over the past decade—processed food, textiles, and clothing—are also the three with the lowest average earnings per worker. An obvious policy measure for stimulating the labor market would be one that seeks to increase sales as well as the earnings and productivity of workers in such firms. Although the average performance of firms has been weak (in terms of generating revenue, value added, or sales), key industries, such as textiles and clothing, have outperformed other industries, with clothing increasing its employment more than 300 percent over the decade.

I. DATA

We use two data sets in the analysis. Data describing the supply side come from the World Bank's 1990–91 Moroccan Living Standards Measurement Survey (LSMS). This household survey provides detailed socioeconomic information on about 19,000 individuals. Of these, 9,685 are between the ages of 18 and 65, and 1,946 report wage and salary earnings (40 percent of men and 8 percent of women). This last group is the basis of our analysis.

The fact that the analysis focuses on a subset of the working population highlights an important point: because of the difficulty of collecting data on the informal sector, our analysis applies only to the formal sector. Although this limitation makes the results less general, stimulating the growth of the formal sector is a key policy focus of both the Moroccan government and the World Bank. Indeed, a healthy formal sector is critical to economic development.

Most of the key variables used are standard for this type of study—age, education, industry and occupation, and region of residence. However, the earnings measure is monthly earnings, which does not include benefits, such as pensions, holidays, and in-kind transfers. We measure education in years according to the level of qualification achieved. Since the survey does not capture actual experi-

ence, we construct a measure of potential experience as the age of the worker minus years of education minus seven.

Data on the demand side come from an annual (1984–95) census of all manufacturing firms in the formal sector with more than 10 workers and a sample of those with 10 or fewer. There are a total of 61,897 observations (starting with 3,692 firms in 1984 and ending with 6,036 firms in 1995), dropping firms that report having no employees. Of the full sample of firms, 59,398 report in three or more years. But the usual problems with nonreporting arise: only 51,858 observations have value added measures, for example. There is also the possibility that firms spuriously drop into and out of the data set. But since more than 68 percent of the firms report positive employment in at least 9 of the 12 years, and these firms account for 80 percent of employment, the qualitative impact on the data is minimal.

Although the key variables we use are standard, it is worth discussing the measurement of some of them. Firms report both permanent employment and total days worked by temporary employees. We transform this measure into total employment by summing permanent employment and the annual equivalent of temporary employment (total days worked divided by 220). We calculate average wages by dividing total payroll by total employment—so lower wages may reflect either a change in the composition of workers or a true decrease in wages for a particular type of worker.

The measure of capital stock is difficult to construct, and we follow the literature in using the gross book value of fixed assets, which is reported in 1990, and the perpetual inventory method of depreciating investment, which is reported every year. There are obvious problems with this approach. First, and most critical, the key base variable—book value—does not represent the economic value of capital. Second, firms do not report the proportion of their capital that is plant and the proportion that is equipment. Third, many firms do not report the book value at all. Finally, the estimate of the depreciation rate (chosen to be 5 percent) is arbitrary and does not reflect industry differences or different proportions of plant and equipment. We thus follow Currie and Harrison (1997) and use a variation of firm fixed effects to verify the robustness of the results when capital stock is used.

II. Wages and Employment in Morocco

Data from the annual employment survey conducted by the government's Statistical Institute show that in the past decade Morocco's population growth has averaged 2.2 percent a year, and its labor force growth has been slightly lower. But the growth rate of the urban labor force has been much higher because of strong migration from rural areas to the cities. Between 1986 and 1995 the urban labor force grew at 5 percent a year, while the rural labor force declined in absolute terms. However, net job growth in urban areas was only 4 percent a year. This differential, coupled with increasing participation rates by

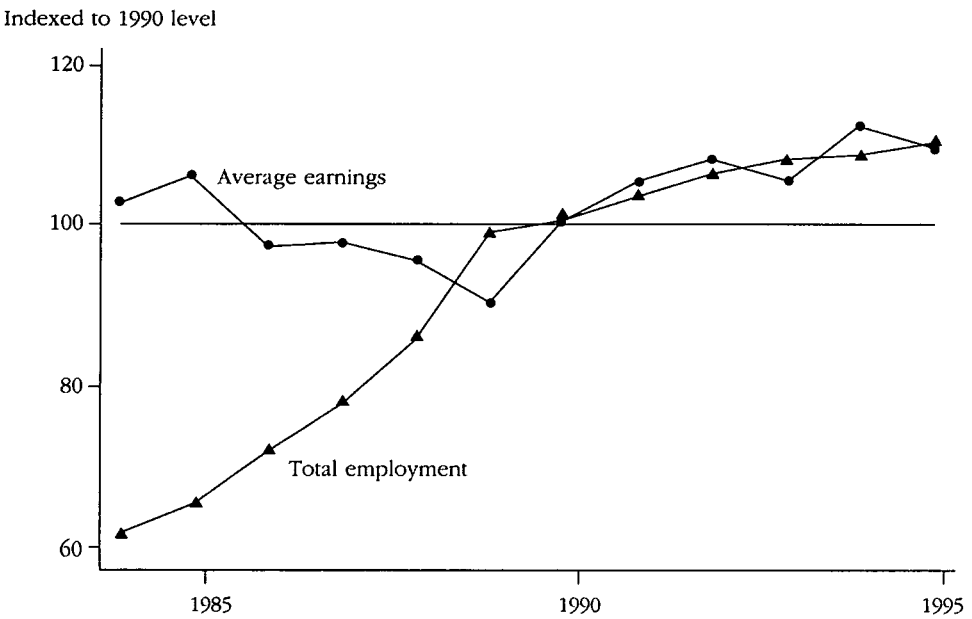
women, boosted unemployment rates and shifted jobs to the informal sector. The official urban unemployment rate hovers around 20 percent, affecting about 1.2 million people.

Between 1985 and 1995 employment growth in the formal urban sector was strongly correlated with the performance of manufactured exports. Employment grew with manufactured exports in the late 1980s, but slowed considerably in the 1990s (figure 1).

The period of growth in the 1980s was the result of several factors. First, Morocco benefited from its privileged access to European markets, which it secured through various trade agreements with the European Union. Second, manufactured exports became more competitive when the real effective exchange rate depreciated because of falling unit labor costs (measured in local currency). During the 1980s manufactured exports grew at an annual rate of 23 percent, and employment grew at 11 percent. Although this combination of factors produced strong employment expansion, it was short-lived.

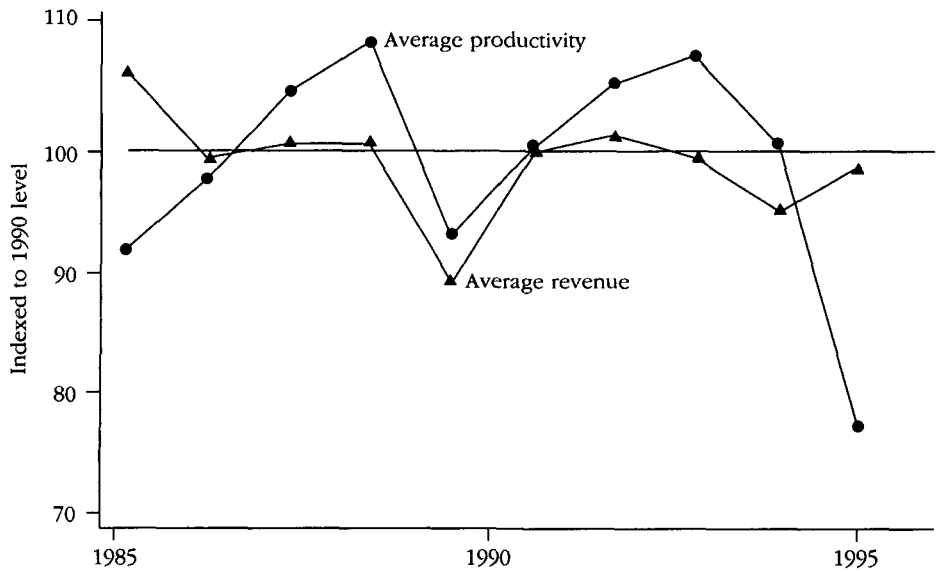
In the 1990s unit labor costs reversed their downward trend as real wages rose slightly and productivity declined (figures 1 and 2). At the same time, the real effective exchange rate appreciated, making Moroccan exports more expensive abroad. All of these elements came together at a moment when Eastern European and Asian competitors were also granted access to European markets, pushing out Moroccan goods. Manufactured exports grew at only 4 percent a year during 1990–95, and employment grew at less than 2 percent a year.

Figure 1. *Indicators of Manufacturing Performance, 1985–95*



Source: Authors' calculations.

Figure 2. *Indicators in Firm Performance, 1985–95*



III. Conceptual Framework

Our formal analysis follows a simple supply and demand approach of the type proposed by Bound and Johnson (1992). The aggregate manufacturing workforce is composed of i demographic groups defined by age and education, working in j different industries. The wage rate of each group of workers is W_i, and the relative rent of each worker in industry j is μ_j. The average wage in manufacturing at time t, w_t, is a simple weighted average of the number of workers in each skill group (p_{it}) at time t and their relative rents:

$$w_t = \Sigma_i \Sigma_j (p_{it} W_i + p_{jt} \mu_j). \tag{1}$$

Wage changes in this model can be written as

$$\begin{aligned} w_t - w_{t-1} = {} & \Sigma_i [(W_{it} - W_{it-1}) p_{it} + W_{it-1} (p_{it} - p_{it-1})] \\ & + \Sigma_j [(\mu_{jt} - \mu_{jt-1}) p_{jt} + \mu_{jt-1} (p_{jt} - p_{jt-1})]. \end{aligned} \tag{2}$$

If we assume that the wage of each demographic group is exogenous and not a function of the group's relative supply, then wage changes are simply driven either by changes in the number of workers in each demographic group or by the allocation of the same group of workers across different industries. This is clearly a very strong assumption, since it reduces the importance of wages as a mechanism for allocating resources. However, since neither linked worker-firm data nor repeated, relatively large-scale household surveys exist, we are forced to focus on quantity rather than price adjustments.

In the next two sections we apply this model to examine explanations that focus on changes in the relative supply of different groups of labor and those that focus on shifts in product demand.

IV. Labor Supply

Earnings may become stagnant because of a deterioration in the quality of the labor force. It may be that the average educational level has deteriorated or that the workforce has fewer experienced prime-age workers. In this section we examine each of these possibilities by simulating how a shift in the composition of the labor force would have affected average earnings had wages remained constant during 1986–95.

The Distribution of Education in the Urban Labor Force

The effect that the distribution of education will have on average earnings depends on the return to education and the distribution of workers with different types of education. We estimate returns to education with standard Mincerian earnings models using LSMS data and examine the distribution of workers using published national statistics.

In most countries age-earnings profiles vary with the level of education: workers with relatively little education have flatter earnings profiles than workers who are highly educated. This pattern holds in Morocco. The results of our Mincerian regressions indicate that the return to investment in education is about 10 percent a year, using hourly earnings as the dependent variable (table 1). At least part of this return comes from an economic rent associated with working in certain industries and occupations. Once industry and occupational dummies are included in the regression, the return to education drops to about 4–5 percent.

Given that the return to education is positive and reasonably large, we then ask whether stagnant real earnings result from a deterioration in the quality of the labor force. In general, this is unlikely, since the education level of the *potential* labor force does not change rapidly over time. However, the education level of the *employed* labor force may vary markedly if workers at one end of the education spectrum drop out or if large cohorts of uneducated workers enter. In Morocco the change has been an increase in the quality of labor: the proportion of workers with no primary school qualifications fell between 1986 and 1990 and again, more sharply, between 1990 and 1993.

Thus the net effect of the changing educational structure of the employed labor force was to raise earnings over what they would have been otherwise. In particular, because of the increase in the education level of the employed labor force during 1986–93, average earnings actually rose by DH240 (dirhan) relative to what they would have been had education remained at the 1990 level. These results suggest that the immediate source of stagnant earnings and productivity growth was not a deterioration in the educational quality of the employed formal labor force. The proportion of highly educated workers rose. And since the wages of

Table 1. *Estimates of the Returns to Education, 1990–91*

Independent variable	*1*	*2*	*3*[a]	*4*[b]
Experience	0.003	0.002	0.002	0.003
	(12.18)	(0.24)	(8.35)	(10.10)
Experience^2	−3.31e−06	−1.95e−06	−2.11e−06	−2.53e−06
	(−7.52)	(−6.56)	(−6.61)	(8.50)
Years of schooling	0.099	0.053	0.051	0.065
	(32.85)	(12.99)	(10.16)	(14.13)
Employee is female		−0.135	−0.089	−0.034
		(−2.41)	(−1.46)	(−0.59)
Public employer		−0.048	−0.020	−0.120
		(−0.62)	(−0.03)	(−1.36)
Private employer		−0.158	−0.302	0.285
		(−2.14)	(−4.83)	(−3.19)
Agriculture and other		−0.368	−0.495	−0.482
		(−4.08)	(−5.03)	(−4.50)
Household worker		−0.319	−0.311	−0.227
		(−2.68)	(−2.47)	(−1.54)
Employee lives in urban area		0.118	0.111	0.128
		(2.82)	(2.40)	(2.80)
Inverse Mills ratio		−0.137	−0.202	−0.255
		(−4.61)	(−2.72)	(3.44)
Constant	2.223	7.041	3.183	3.074
	(51.60)	(61.96)	(25.12)	(24.01)
R^2	0.33	0.39	0.46	0.46

Note: The log of hourly wages is the dependent variable. All regressions have regional fixed effects. *t*-statistics are in parentheses.

a. Regression controls for industry.

b. Regression controls for occupation.

Source: Authors' calculations from LSMS data.

highly educated workers are higher than those of less educated workers, this change in proportion should have put upward pressure on earnings.

The Age Structure of the Urban Labor Force

Another potential source of change in labor quality is the age structure of the labor force. Part of the stagnation in average earnings in the United States in the 1970s came from an increase in the number of youth and women in the labor force, both of which have historically had fewer skills than prime-age males. This was not the case in Morocco, despite the fact that the labor force grew markedly over this period.

Much of this growth came from the disproportionately large increase in the number of workers ages 35–44, that is, prime-age workers, presumably those who moved to urban areas from the countryside. According to published national statistics, the number of workers in this age group grew 39 percent in 1986–93, increasing their proportional representation in the labor force 3.56 percent. Thus we are led to the interesting question of how this pure age-related change affected average earnings (the effect can be calculated in the same fashion

as above). The change in the age distribution actually increased mean earnings in Morocco by DH52 in excess of what they would have been had the age distribution remained the same.

In sum, then, changes in the age and education composition of the workforce raised the level of earnings and should have increased labor productivity in urban areas. Other explanations for the stagnation must be found.

V. Labor Demand

Since stagnant average earnings are clearly not the result of changes in labor supply, we turn to the demand side of the market for an explanation. Earnings could stagnate if the demand for labor does not respond normally to demand shocks, if the labor market is unable to reallocate workers from less productive to more productive activities, or if there has been a structural shift in employment from high-paying to low-paying industries. We examine each of these in turn.

The Derived Demand for Labor

Since the demand for labor is a derived demand, it is extremely important to examine the health of the product market. In Morocco average firm performance did not change markedly over 1985–95 (figure 2). Revenue generated per worker actually fell below the 1985 level, as did productivity. It is worth noting in this context that although exports per worker grew dramatically in the late 1980s, the 1990s have seen a substantial drop in the growth rate, with investment actually falling in 1994 and 1995.

We can examine the effect of this change on labor demand by estimating fairly standard labor demand equations of the form:

$$\ln L_{ft} = \beta_0 + \beta_1 \ln w_{ft} + \beta_2 \ln VA_{ft} + \beta_3 t + \epsilon_{ft}. \tag{3}$$

Thus the demand for labor, L, in firm f at time t is determined by the average wage of the workers, w, and the value added, VA, in firm f at time t. We face two estimation problems. First, the Moroccan data are very similar to U.S. data, which are characterized by vast heterogeneity across firms because of differences in the quality of inputs, outputs, and managers (Dunne and Roberts 1993; Haltiwanger, Lane, and Spletzer 1999). Fortunately, however, our data set is quite rich: it comprises longitudinal data spanning 12 years and about 5,000 firms. Thus we are able to follow Dunne and Roberts (1993) in using fixed effects both to account for unobservable heterogeneity and to minimize simultaneity bias. The second problem is that both wages and value added are likely to be contemporaneously correlated with the error term. We address this issue by also reporting the results using lagged values as instruments.

The first column in table 2 reports the results of estimating equation 3 by ordinary least squares (controlling for fixed effects). The elasticity of labor demand with respect to value added is 0.368, indicating that employment is indeed

Table 2. *Determinants of Manufacturing Labor Demand, 1985–95*

Independent variable	*Ordinary least squares*	*Instrumental variables*
Value added	0.368	0.181
	(74.80)	(37.70)
Average wage	–0.482	–0.143
	(59.07)	(18.98)
Year	0.036	0.013
	(46.51)	(14.91)
Constant	–1.209	1.14
	(18.22)	(14.15)
R^2	0.95	0.93
Number of observations	62,179	55,990
Number of firms	8,161	7,531

Note: *t*-statistics in parentheses.
Source: Authors' calculations.

responsive to product demand shocks. Both this elasticity and the own-wage elasticity are well within the range of empirical estimates for other countries (Hamermesh 1993), regardless of whether we use the least squares estimates or the instrumental variable estimates. Hence firms in Morocco respond to market stimuli in similar ways as firms in other countries. However, the growth rate of employment in each firm is extremely sluggish: firms expand about 1 percent a year.

The results are stable when we examine firms at a more detailed level (table 3). The own-wage elasticity always falls within reasonable bounds and, in fact, is remarkably stable for firms of different sizes. Export-oriented firms (defined as firms whose sales are more than 50 percent exports) are more responsive to wage increases than are their domestically oriented counterparts. State-owned firms (defined as firms with more than 50 percent state ownership) are more wage sensitive than private firms.

These results suggest that Moroccan labor demand is not much different from that in other countries, at least with respect to observable responses to market stimuli. It may be true that regulation lowers labor demand in all countries, but the evidence does not suggest that the effect in Morocco is unusually strong. We now examine labor market flexibility by comparing job creation and job destruction rates in Moroccan firms.

Labor Flexibility

Foster, Haltiwanger, and Krizan (1998) establish that in the United States more than half of productivity growth is due to the reallocation of resources from less productive to more productive firms. Thus the movement of workers across firms is likely to be an important source of productivity (and hence wage) growth. In this section we examine the two components of job reallocation: job creation (the positive change in employment at the firm level, divided by average firm employment) and job destruction (the negative change in employment at the

Table 3. *Labor Demand by Type of Firm, 1985–95*

Indicator	*Microenterprise (fewer than 20 employees)*	*Small (20–49 employees)*	*Medium-size (50–249 employees)*	*Large (250+ employees)*	*Exporting*	*Not exporting*	*State-owned*	*Not state-owned*
Own-wage elasticity	0.24	0.29	0.34	0.35	0.32	0.22	0.36	0.25
Elasticity with respect to value added	0.25	0.20	0.25	0.29	0.32	0.21	0.23	0.23
Job creation rate	0.13	0.14	0.15	0.14	0.18	0.12	0.12	0.13
Job destruction rate	0.12	0.11	0.11	0.11	0.12	0.10	0.10	0.10

Source: Authors' calculations.

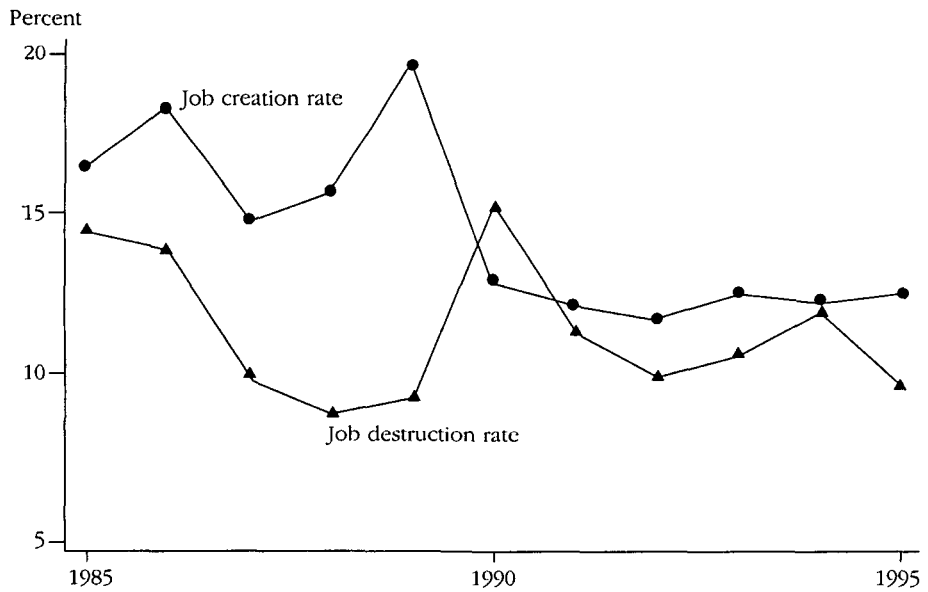
firm level, divided by average firm employment), which have often been used as measures of labor market flexibility (see Davis, Haltiwanger, and Schuh 1996). Indeed, OECD (1996) finds a negative relationship between the job destruction rate and the index of labor market regulation.

With average job creation and destruction rates hovering around 12 percent, Morocco shows a degree of labor market flexibility similar to that of OECD countries (figure 3). These figures confirm earlier work by Currie and Harrison (1997), who offer anecdotal evidence that labor market adjustment is quite fluid and demonstrate that another measure of labor market flexibility—firms' speed of adjustment to demand shocks—falls within international norms.

The difference between the volatility in permanent and temporary employment is also marked, suggesting that Moroccan firms are using temporary employment as a way of adjusting to demand shocks (figures 4 and 5). Using temporary employment in this way, while consistent with employment practices in other countries, does pose some long-run concerns for the development of a skilled and productive workforce and for the flexibility of permanent workers.

Again, a closer examination of these trends suggests that all classes of firms, defined by size, export orientation, and state ownership, are roughly equivalent to international standards of flexibility (see the last two rows of table 3). The result for state-owned firms is particularly surprising and suggests that there is widespread avoidance of the labor code, even in the state-owned sector.

Figure 3. *Job Creation and Destruction Rates for All Manufacturing Employment, 1985–95*



Source: Authors' calculations.

Figure 4. *Job Creation and Destruction Rates for Permanent Employment, 1985–95*

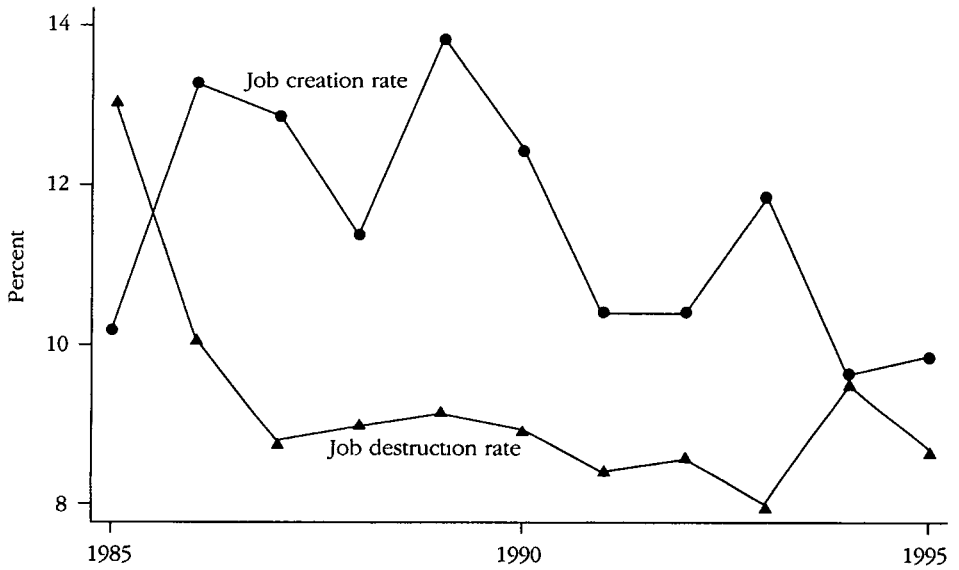
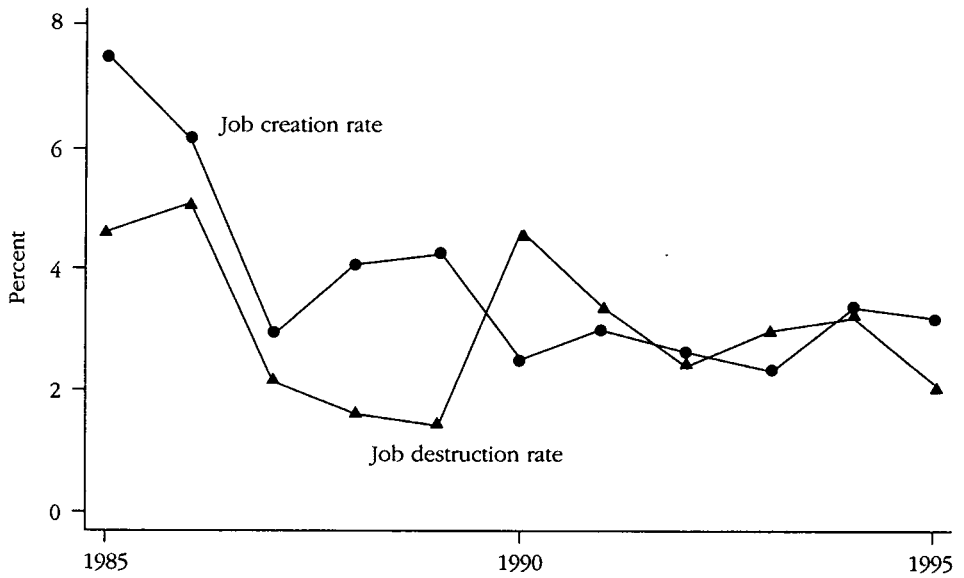


Figure 5. *Job Creation and Destruction Rates for Temporary Employment, 1985–95*



Sectoral Shifts

The last possible explanation for stagnant wage growth is that the structure of the economy has changed from high-productivity and high-wage industries to low-productivity and low-wage industries. That this is a possibility is immediately evident from an examination of individual industries (table 4). The three largest industries that also display some of the highest employment growth—processed food, textiles, and clothing—are the three with some of the lowest earnings (they are ranked sixteenth, fourteenth, and eighteenth, respectively, in terms of average payroll). Although earnings are much higher in beverages and tobacco, electrical and electronic equipment, and base metal products, job creation in these industries has been relatively flat or declining. The only industry to combine both high earnings and high employment growth is the chemicals industry.

Another way of looking at sectoral changes is through the job creation and destruction rates of high-wage and low-wage firms. We call firms "high-wage" if their average payroll exceeds the average payroll for all firms in a given a year. We classify "low-wage" firms similarly. We then track their patterns of job creation and job destruction over time (figure 6). It is clear that job creation and job destruction are higher in low-wage than in high-wage firms and that net job creation is positive for low-wage firms and negative for high-wage firms. In other words, employment growth is occurring at the low end of the wage spectrum.

Finally, we decompose the growth in average wages between 1984 and 1995 by simplifying equation 2 to focus solely on changes in industry composition, rather than on changes in workforce composition. We thus note that $w_t = w_{jt} p_{jt}$, where w_t is the average wage at time t, w_{jt} is the average wage in industry j at time t, and p_{jt} is industry j's employment share. Thus:

$$w_t - w_{t-1} = \Sigma_j \, [p_{jt}(w_{jt} - w_{jt-1}) + w_{jt-1} \, (p_{jt} - p_{jt-1})] \tag{4}$$

or, the change in the average wage is the sum of the change in the wage of each industry weighted by that industry's share of employment (the first term on the right side) and the change in the industry's share of employment, weighted by average wages at time $t - 1$ (the second term on the right side). Applying this simple decomposition reveals that had the sectoral distribution of employment remained the same between 1984 and 1995, average wages would have been 9.8 percent higher. The same holds true for a similar decomposition of productivity: because employment has grown in the least productive sector of the economy, if employment shares had stayed the same between 1985 and 1995, average productivity would have been 20.1 percent higher.

VI. Policy Implications

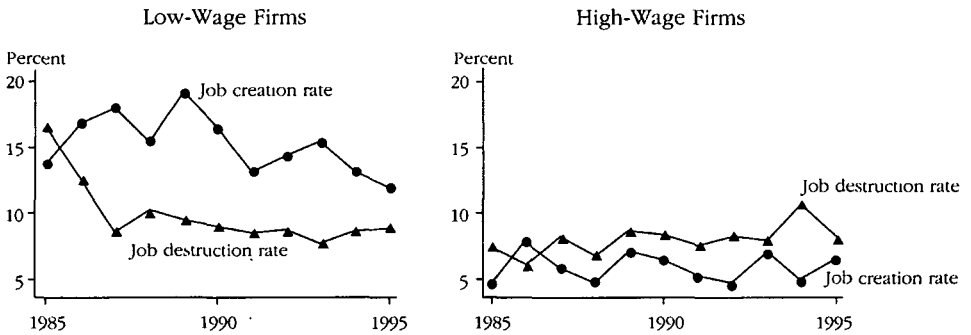
Our initial concern was to explain poor employment and earnings growth in the Moroccan economy. The evidence presented above strongly suggests that the poor performance in the past 12 years is not a result of inadequate education or

Table 4. *Labor Market Indicators by Industry*

Industry	*Average earnings (1995)*	*Percentage growth in average earnings (1984–95)*	*Rank (earnings level)*	*Number of employees (1995)*	*Percentage employment growth (1984–95)*	*Rank (employment growth)*
Food products	33.15	9.03	9	27,197	26.81	14
Processed food products	23.02	28.45	16	60,748	64.41	6
Beverages and tobacco	56.64	47.58	1	10,201	2.73	17
Textiles and hosiery	25.79	4.76	14	70,213	48.89	10
Clothing excluding footwear	16.66	30.10	18	101,771	352.71	1
Leather products including shoes	19.03	-8.42	17	15,714	35.19	12
Lumber and wood products	24.61	24.94	15	12,850	18.76	16
Paper, cardboard, and printing	33.63	3.31	8	16,825	66.24	5
Glass, cement, and ceramic products	31.45	13.18	12	33,266	61.53	7
Base metal products	49.03	20.95	3	1,932	–33.87	18
Metalwork, excluding machinery, and transport equipment	34.40	10.30	7	23,948	46.55	11
Machinery and equipment	35.79	6.90	6	7,137	27.58	13
Transportation equipment	44.30	6.73	5	12,498	122.02	2
Electrical and electronic equipment	45.49	17.17	4	10,703	24.57	15
Office machines, measuring	29.27	–22.79	13	1,013	55.70	9
Chemicals and related products	51.13	10.64	2	31,342	89.96	4
Rubber and plastic products	32.87	3.09	10	12,362	55.94	8
Miscellaneous manufactures	32.01	36.63	11	651	103.53	3

Source: Authors' calculations.

Figure 6. *Job Creation and Destruction Rates by Wage Category, 1985–95*



Source: Authors' calculations

labor market rigidities. Rather, structural changes in Morocco were behind the labor market's poor performance. If the strong growth of the clothing industry, for example, had been seen in the rest of the manufacturing sector, overall employment would have grown by an astounding 352 percent and presumably would have reduced the unemployment problem commensurately. However, the stagnation of earnings is also directly attributable to the growth of this relatively low-paying sector of the economy.

In order to examine the structural changes in more detail, we estimate (but do not report) industry-specific labor demand regressions and find wide variation in the elasticity of responsiveness to product demand (that is, value added). In particular, the industry with the greatest employment response to increases in product demand is also one of the lowest paid: the textiles and hosiery industry. An increase of 10 percent in value added would increase employment anywhere from 2.4 percent to 4 percent—about 1,700 workers. We see the same responsiveness in another important, but low-wage, Moroccan industry, the clothing industry: a 10 percent increase in product demand would increase employment by almost 5 percent, or more than 5,000 workers. A rise in demand in other industries would have much less impact on employment either because employment levels are lower or employment is less responsive to demand shocks, or both. For example, the employment responsiveness of the relatively highly paid beverage industry to demand shocks is about one-seventh that of textiles.

This analysis suggests that the twin targets of increasing employment and increasing average earnings may not be feasible given the current structure of Moroccan industry. A more detailed study of the two strongest industries, clothing and textiles, indicates that even within these growing industries, the most successful firms are competing on the basis of relatively low wages. In the clothing industry the firms that are in the top quartile of employment growth in any given year have a higher ratio of exports to revenue than firms in the bottom quartile of employment growth (73 percent compared with 65 percent). However, these fast-growing firms are also less productive and offer lower wages (productivity

per worker is 20 percent less, but wages are equally discounted). They are also younger—the average age of a firm in the top quartile is 8.9 years, whereas that of a firm in the bottom quartile is 11.13 years. The textile industry has a similar structure. The firms in the top quartile are 20 percent less productive (mirrored by equally discounted wages) and are younger, averaging 17 years compared with 20 years for firms in the bottom quartile.

The role of exports in the labor market is important since job growth is highly correlated with export growth. Also, Morocco substantially reduced both tariffs and import restrictions in 1983, which had substantial effects on employment, probably bringing about the subsequent structural changes.

The view that the future of employment growth lies in the relatively low-wage sector is reinforced by a recent World Bank report (World Bank 1993) showing that the clothing industry is the most export-oriented of all Moroccan industries. It exported 90 percent of its output in 1995. Although less visible, the textile industry surged in its percentage of exports, moving from 15 percent of its output in 1984 to 41 percent in 1995. The only other industries to display similar export competitiveness in 1995 are processed food (28 percent), leather products (60 percent), and chemical products (38 percent). It is possible that wages in declining sectors were artificially high and that competition is eliminating these jobs. If so, structural adjustment assistance might be justified to compensate the affected workers. The main policy change would then be to encourage export competition.

Clerides, Lach, and Tybout (1998) find that the most efficient firms become exporters and that learning-by-exporting may have characterized Moroccan firms in the apparel and leather industries. So, clearly, modernizing the capital stock and introducing new technology to firms in exporting industries would make sense. Doing so could also increase average earnings in the low-paid clothing and textile industries, especially if accompanied by on-the-job-training. A recent World Bank study on training in Morocco (World Bank 1996) reveals that only 4 percent of Moroccan firms engage in any formal training and that average earnings and productivity in these firms are much higher than in others.

VII. Conclusions

In this article we explain Morocco's poor labor market performance in the 1990s. We use detailed industry and household data to determine that this poor performance is primarily the result of marked structural changes and reflects the country's weak product market. This article underscores the fact that the demand for labor is a derived demand and that the performance of the product market is an important determinant of the performance of the labor market. It also demonstrates that this kind of microeconomic analysis, using data sets that are often available in developing countries, can inform policy design.

But the article leaves unanswered a more fundamental question—the role of wages and worker mobility in the adjustment process. This analysis does not

track workers over time and across industries, and hence cannot determine whether wages changed for the same set of workers over time or whether one set of workers was replaced by another set. Both may have occurred. In the first case the notion that there are interindustry and even intraindustry wage differentials for similar workers has a long tradition in labor economics (Dunlop 1957). Recent, compelling empirical evidence has reinforced this conclusion for industrial countries (Krueger and Summers 1988; Groshen 1991; Abowd, Kramarz, and Margolis 1999). The latter case is also possible, given the different skill needs of expanding and contracting industries. The Moroccan data set that we use does not permit us to investigate this question—although knowing whether workers in the low-wage clothing and textile industries in 1995 are the same workers who were in the high-wage base metals industry in 1985 would have obvious policy implications. Developing a linked employer-employee data set that enables us to answer this intriguing question is the next part of our research agenda.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Abowd, John, Francis Kramarz, and David Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67(2):251–333.

Bound, John, and George Johnson. 1992. "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations." *American Economic Review* 82(June):371–92.

Clerides, Sofronis, Saul Lach, and James Tybout. 1998. "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco." *Quarterly Journal of Economics* 113(3):903–47.

Currie, Janet, and Anne Harrison. 1997. "Sharing the Costs: The Impact of Trade Reform on Capital and Labor on Morocco." *Journal of Labor Economics* 15(July):S44–S71.

Davis, Steven, John Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction.* Cambridge, Mass.: MIT Press.

Dunlop, John T. 1957. "The Task of Contemporary Wage Theory." In George W. Taylor and Frank C. Pierson, eds., *New Concepts in Wage Determination.* New York: McGraw-Hill.

Dunne, Timothy, and Mark Roberts. 1993. "The Long-Run Demand for Labor: Estimates from Census Establishment Data." CES Working Paper 93-13. U.S. Bureau of the Census, Washington, D.C. Processed.

Foster, Lucia, John Haltiwanger, and C. J. Krizan. 1998. "Aggregate Productivity Growth: Lessons from Macroeconomic Evidence." NBER Working Paper 6803. National Bureau of Economic Research, Cambridge, Mass. Processed.

Groshen, Erica. 1991. "Sources of Intra-industry Wage Dispersion: How Much Do Employers Matter?" *The Quarterly Journal of Economics* 106(August):869–84.

Haltiwanger, John, Julia Lane, and James Spletzer. 1999. "Productivity Differences across Employers: The Role of Employer Size, Age, and Human Capital." *American Economic Review* 89(May):94–98.

Hamermesh, Daniel. 1993. *Labor Demand*. Princeton, N.J.: Princeton University Press.

Krueger, Alan, and Lawrence Summers. 1988. "Efficiency Wages and the Inter-industry Wage Structure." *Econometrica* 56(March):259–93.

OECD (Organisation for Economic Co-operation and Development). 1996. *Employment Outlook*. Paris.

World Bank. 1993. "Developing Private Industry in Morocco." MENA, Washington, D.C. Processed.

———. 1995. "Kingdom of Morocco Country Economic Memorandum: Towards Higher Growth and Employment." Middle East and North Africa Region, Country Department I, Washington, D.C. Processed.

———. 1996. "Staff Appraisal Report: Kingdom of Morocco, Third Private Sector Development Project (In Service Training)." Middle East and North Africa Region, Washington, D.C. Processed.

Coming in the next issue of

THE WORLD BANK ECONOMIC REVIEW

January 2000
Volume 14, Number 1

A symposium on trade and technology, including . . .

- International Knowledge Flows and Economic Performance: An Introductory Survey of the Evidence
  Giorgio Barba Navaretti and David Tarr
- How Trade Patterns and Technology Flows Affect Productivity Growth
  Wolfgang Keller
- Avenues of Technology Transfer: Foreign Investment and Productivity Change in the Czech Republic
  Simeon Djankov and Bernard M. Hoekman
- Productivity and Turnover in the Export Market: Micro-level Evidence from Taiwan (China) and the Republic of Korea
  Bee Yan Aw, Sukkyun Chung, and Mark J. Roberts
- Vintage Technology and Skill Constraints: Evidence from U.S. Exports of New and Used Machines
  Giorgio Barba Navaretti, Isidro Soloaga, and Wendy Takacs
- Why Agricultural Technology Transfers to Developing Countries Should Be Deregulated
  David P. Gisselquist and Jean-Marie Grether

A symposium on geographical targeting including . . .

- Geographical Targeting for Poverty Alleviation: An Introduction to the Special Issue
  David Bigman and Hippolyte L. Fofack
- Combining Census and Survey Data to Study Spatial Dimensions of Poverty: A Case Study of Ecuador
  Jesko S. Hentschel, Jean Lanjouw, Peter F. Lanjouw, and Javier Poggi
- Community Targeting for Poverty Reduction in Burkina Faso
  David Bigman, Stefan Dercon, Dominique Guillaume, and Michel Lambotte
- Using Light Monitoring Surveys in Conjunction with Integrated Surveys to Improve Targeting for Poverty Reduction: The Case of Ghana
  Hippolyte L. Fofack

The World Bank

1818 H Street, N.W.
Washington, D.C. 20433, U.S.A.
Telephone: (202) 477-1234
Facsimile: (202) 477-6391
Telex: MCI 64145 WORLDBANK or MCI 248423 WORLDBANK
World Wide Web: http://www.worldbank.org/
E-mail: books@worldbank.org

14363
9 780821 343630

ISBN 0-8213-4363-7