77272

THE WORLD BANK ECONOMIC REVIEW, VOL. 12, NO. 3: 457-82

The Bolivian Social Investment Fund: An Analysis of Baseline Data for Impact Evaluation

Menno Pradhan, Laura Rawlings, and Geert Ridder

The Bolivian Social Investment Fund (SIF) is a financial institution that promotes sustainable investments in the social sectors, principally in the areas of health, education, and sanitation. This article shows how to use preintervention data collected for evaluating the SIF to improve the targeting of a program, to test the quality of the evaluation design, and to define corrective measures if necessary. It finds that among SIF interventions the benefits in education are distributed relatively equally over the population, while the investments in health and sanitation favor better-off communities.

The article contributes to the methods used to evaluate social investment funds and similar programs. It compares two types of evaluation designs to assess social investment fund interventions in the education sector. The authors demonstrate that a simple matched-comparison design introduces a bias in the estimate of the program effect, whereas an experimental design based on random assignment does not. With preintervention data, the analyst can select a quasi or indirect experiment, where the choice of the indirect experiment coincides with the selection of valid instrumental variables. The availability of preintervention data makes it possible to compare the two types of evaluation designs as well as to test the validity of the instruments and to determine the loss of efficiency due to the use of quasi-experimental techniques instead of random treatment assignment.

The Bolivian Social Investment Fund (SIF) is a financial institution designed to promote sustainable investments in the social sectors, notably in the areas of health, education, and sanitation. The SIF cofinances initiatives providing infrastructure, training, and equipment by making funds available to requesting agencies that then subcontract the implementation of the projects. The SIF is characterized by rapid disbursement, institutional efficiency, and a demand-driven approach, and it is among the first of its kind in the world. The SIF has been a

Menno Pradhan is with the Department of Economics at the Vrije Universiteit in Amsterdam, Laura Rawlings is with the Development Economics Research Group at the World Bank, and Geert Ridder is with the Department of Economics at The Johns Hopkins University. The analysis presented in this article was made possible by a grant from the Dutch Trust Fund at the World Bank. The collection of the baseline data was financed by the Social Investment Fund Project (1991–94). This study is part of the World Bank–financed research project Evaluation of Social Sector Investments, which has provided support to the Bolivian Social Investment Fund for the design, application, and analysis of the impact evaluation. The authors wish to thank John Newman and Ramiro Coa for their support and the referees for their comments.

key player in funding social sector projects in Bolivia. Since its inception in 1991, the SIF has provided $187 million worth of financing to more than 3,000 projects.[1] It has caught the attention of policymakers worldwide, and similar funds have been introduced in Africa, Asia, and other parts of Latin America. In Latin America and the Caribbean alone, 16 SIFs have been created (Glaessner and others 1994).

Despite the growing popularity of social funds, few have been subject to empirical investigation to assess the effects of these interventions. This question is of central concern to the World Bank, which has been a principal supporter of social funds, and to policymakers in the social sectors more broadly. The Bolivian SIF provides an opportunity to assess this impact, a task that will be completed in 1999 when the follow-up data will be available and analyzed. This article makes an initial contribution by exploring a special feature of the evaluation made possible by the presence of baseline data and the use of different evaluation methodologies. For the evaluation of the SIF, preintervention (baseline) data were collected in select rural provinces in 1993, and a follow-up survey is being completed in 1998.

This article is based on an analysis of the baseline data, which were collected in two stages. In June 1993, a survey was conducted in five provinces in the Chaco region.[2] In October and November 1993, this survey was extended to 17 other provinces in select rural areas throughout Bolivia, hereafter named the Resto Rural region.[3] The survey gathered information on the communities, facilities, and individuals expected to benefit from projects financed by the SIF. The survey also collected information for control groups not expected to receive SIF interventions. In the education sector, two methods were used to construct the control groups. In the Chaco region, an experimental randomized design was used, while in the Resto Rural region, the control group was constructed by matching on observed characteristics. A matched comparison was also used in the sanitation sector, whereas the evaluation in the health sector was based on a reflexive comparison before and after the SIF intervention.

Analyzing the baseline data before collecting the follow-up data is useful for two reasons. First, information on the facilities that will be upgraded and benefit incidence analysis allow for midcourse corrections in implementing the SIF projects, particularly with respect to targeting. Second, in the case of experimental designs or matched-comparison designs, the evaluation methodology can be tested by as-

1. As of 31 March 1998 (the Bolivian Social Investment Fund, Program and Policy Support Unit, La Paz, May 1998).

2. The provinces in the Chaco region are O'Connor and Gran Chaco in the department of Tarija; Cordillera in the department of Santa Cruz; and Luis Calvo and Hernando Siles in the department of Chuquisaca.

3. The provinces in the Resto Rural region are Bernando Saaverdra, Camacho, Muñeca, and Franz Tamayo in the department of La Paz; Capinota, Tapacarí, Quillacollo, and Arque in the department of Cochabamba; Saucarí, Cercado, Carangas, Sur Carangas, and Nor Carangas in the department of Oruro; and Linares, Nor Chicas, Saavedra, and Modesto Omiste in the department of Potosí.

sessing the comparability of the treatment and comparison or control group. Noncomparability may have implications for the statistical methods used to determine the impact and the required sample size of the follow-up survey.

This article focuses on two issues. First, we look at the target population of the SIF in education, health, and sanitation projects and compare it with the total population in the Chaco and Resto Rural regions. We investigate the effects of the institutional design of the SIF on its targeting. The demand-driven approach of the SIF does not necessarily guarantee that the poorest in society will benefit from the investments. Second, we take advantage of the two types of evaluation designs in the education component of the SIF to investigate the adequacy of each of the evaluation designs employed: randomization and matched comparison.

In the evaluation literature, randomized designs are widely regarded as the most methodologically robust evaluation approach. The process of randomization ensures that before the interventions take place the treatment and control groups are statistically equivalent, on average, with respect to all characteristics, observed and unobserved. (For a general discussion of experimental and quasi-experimental methods for causal inference, see Holland 1986.) Therefore, any differences observed after the intervention takes place can be causally attributed to the effect of the intervention itself (Grossman 1994).

The standard alternative, in the absence of randomized assignment, is matched comparison using a nonrandom process to select a comparison group that resembles the treatment group already assigned to receive an intervention. The matching is obviously restricted to (readily) observable characteristics, and matching does not guarantee that the comparability extends to unobserved or unobservable characteristics. Indeed, Fraker and Mayard (1984) and LaLonde (1986) have shown that matching on observables may induce a large bias in estimates of program effects. Researchers have made considerable progress in the improvement of matching techniques, in particular by concentrating on the selection probability (or propensity score) as the relevant quantity for assessing the quality of the match (Rosenbaum and Rubin 1983, Heckman and others 1998, and Dehejia and Wahba 1995). These methods require a good understanding of the selection process and high-quality data to identify and use the variables that are relevant in the matching, that is, the variables for which the matching makes the treatment and control groups comparable.

In evaluating the Bolivian SIF, we cannot benefit from experience with nonrandom selection for similar programs, nor do we have access to the type of high-quality data often available in other countries, such as for evaluating job training programs in the United States. Under these circumstances, an indirect experiment is the best alternative. In an indirect experiment, some feature of the selection process for the intervention is assumed to provide exogenous variation in the selection probability. If this feature can be measured by a variable, then this variable is a valid instrument and can be used as an instrumental variable (IV) estimator of the program effect (Moffitt 1991).

An instrumental variable satisfies two conditions: first, the variable affects the probability of selection, and second, it does not affect the outcome or response variable that is used to evaluate the program. The first condition implies that the instrumental variable must be included among the variables that are used to match treatment and control groups. An important reason for preferring the IV method over matching is that the set of matching variables, and in particular the degree of comparability achieved by the set of observed variables, is critically important for the quality of the matching estimator. For the IV approach, the instrument only needs to be exogenous (the second condition); it does not need to balance the treatment and control group. The amount of exogenous variation in the selection probability induced by the instrument is important. A "weak" instrument may result in an IV estimator with poor small-sample performance. With randomized assignment, all variation in selection is exogenous. IV and matched-comparison estimates may be similar (Friedlander and Robins 1995), but there is no reason to expect this to be true in general.

There is still considerable controversy over the interpretation of the IV estimator when the program effect varies in the population and, in particular, when that variation cannot be explained by observable characteristics of the population units (Imbens and Angrist 1994 and Heckman 1997). We are not concerned with this discussion but simply assume that all effects of heterogeneity can be captured by observable (and observed) variables.

In general, the IV method is based on untestable assumptions. In particular, the second condition usually cannot be verified empirically. If the instrumental variable and the response or outcome variable are uncorrelated, then we do not know whether this is due to no program effect or to lack of induced variation in the selection probability. If the correlation is nonzero, then we do not know whether there is a direct effect on the outcome variable. (Additional functional form assumptions as in the normal sample selection model help to make the distinction, but these assumptions are untestable.) Hence, the choice of an instrument has to be justified with ad hoc arguments, which may be more or less convincing. Sometimes the instrument is obtained by randomization, as in randomized experiments with noncompliance (see, for example, Angrist 1990).

With preintervention data, we can test the second condition under much weaker assumptions. If we assume that selection for the program does not change preintervention behavior—for example, if the preintervention response is measured before selection for the program is made public—then a zero (partial) correlation between preintervention response and instrument is evidence that the instrument is valid. We shall use this procedure to identify valid instruments for evaluating the Bolivian SIF. Of course, with only preintervention data, we cannot assess program impact. For this, we need the postintervention data that will be collected. However, as we shall argue, we can compare the efficiency of the different designs and predict the accuracy of the estimated program effects.

This article does not contain an evaluation of the SIF. Such an evaluation is not possible using only preintervention data. A full evaluation of the SIF investments in education, health, and sanitation will be carried out in 1998–99 once the follow-up data are available. This article contributes to the methods that can be used to evaluate the SIF and similar programs. We show that preintervention data contain valuable information for choosing an appropriate evaluation method.

Section I presents the data and sample design. Section II describes the SIF in greater detail and looks at the targeting mechanisms used by the SIF. Section III compares the evaluation designs for the education component. Section IV proposes the instrumental variable method that eliminates the bias in the matched-comparison design and studies the loss of efficiency due to this method. Section V concludes.

I. DATA

The SIF impact evaluation considers investment projects in health care, education, and sanitation. The data collected for the impact evaluation were based on surveys applied to both the institutions that receive funding (schools and health centers) and the households and communities that benefit from the investments. Similar data were also collected from comparison institutions and households.

The household survey gathered information on a range of household characteristics including consumption, access to basic services, and each household member's health and education status. The household survey consisted of three subsamples. The first was a random sample of all households in the Chaco and Resto Rural regions, the second was a sample of households that live near the schools in the treatment or control group for the education component, and the third consisted of households that will benefit from the sanitation component. The surveys can be merged easily. For example, each school has a unique code that is recorded in the household survey if a child attends that school. The surveys for the Chaco and Resto Rural regions differed slightly. The baseline data collected in the Resto Rural region are more extensive because shortcomings discovered in the surveys in the Chaco region, which were conducted first, were corrected. Sample sizes are given in table 1.

The health facility survey gathered information on the quality of infrastructure, staffing, and visits to the center. Because the SIF planned to intervene in all health centers in the Chaco and Resto Rural regions, all were included in the survey. The survey distinguished between health clinics at the sector, area, and district levels. Sector health clinics are typically very small, providing basic health care. Area health clinics provide more sophisticated care and serve a larger geographical region. District health clinics are hospitals, the largest type of facility. The larger the health clinic, the more detailed the questionnaire that was administered. The questionnaires are, however, comparable and collected similar types of information on infrastructure, equipment, the availability of medicines, staffing, and the services provided.

The school survey used two questionnaires, one for the director and one for each teacher separately. It gathered information on infrastructure, equipment,

Table 1. *Sample Sizes in the Baseline Survey, by Region in Bolivia, 1993*

Category	*Chaco*	*Resto Rural*
Health centers		
Sector	82	119
Area	16	24
District	3	4
Total	101	147
Schools		
Treatment group	35	37
Control group	37	33
Total	72	70
Households		
Random sample	2,029	2,138
Education component	995	902
Sanitation component	666	569
Total	3,670	3,609

Source: Authors' calculations based on survey data.

teaching methods, and dropout and repetition rates of students. For the Chaco region, the sampling frame consisted of all primary and secondary schools that qualified for SIF interventions and were subject to the random-selection process. As explained in section III, for equity purposes, the worst-off schools were all selected to receive active promotion for an SIF project and none of the best-off schools received active promotion. Only those schools in the middle of the quality distribution were subject to random selection because the financing did not allow for all schools to be reached. For the Resto Rural region, the sampling frame consisted of schools already designated to receive SIF interventions. The sample was augmented by a comparison group of schools not receiving an SIF intervention. Section III provides more detail on the requirements for eligibility for schools to receive SIF investments and the construction of comparison groups.

The community survey collected data from community leaders on a range of topics, including the quality of the infrastructure, the distance to facilities, and the presence of local organizations.

II. TARGETING OF SIF INTERVENTIONS

The SIF has traditionally funded, but not executed, project proposals received from the private, public, and not-for-profit sectors. The SIF is a demand-driven institution because it does not initiate projects but responds to outside initiatives by providing cofinancing for investments in infrastructure, equipment, and training. The cofinancing provided by the SIF generally accounts for approximately 80 percent of project costs, and the requesting institution provides the remaining 20 percent. Regional SIF offices assist communities in preparing proposals. The decision on whether to fund a project is made at the SIF central offices in La

Paz. The final outcome thus depends on both the preferences and capabilities of the local communities, in particular the local authorities and local nongovernmental organizations (NGOs), with respect to preparing and cofinancing projects and on centrally defined targets (see also Newman, Grosh, and Jorgensen 1992).

Central objectives, such as targeting the poor, may not always be reflected in the final outcome of the program. In particular, the project approval process has historically favored more well-organized groups that have access to counterpart financing and are rarely found in poorer areas. This is an inherent conflict of targeted, demand-driven projects. With the introduction of the Ley de Participación Popular in 1994, the preferences of the local population are becoming more influential. Under this law, a proportion of the government budget is allocated directly to communities, municipal elections ensure accountability of the leaders, and communities are given discretion over budget allocations.

An analysis of the SIF's targeting mechanisms using the available preintervention data cannot deal with behavioral responses that may result from SIF interventions. Baseline data can only be used to characterize households that are using the facilities in which the SIF is planning to invest. Of course, changes in household behavior as a result of changes in the supply of public services may be an important factor in determining the net impact of the project (Jimenez 1995). With the current data, however, we cannot deal with these effects.

Figure 1 shows the nonparametric estimate of the density function of log per capita consumption based on the random sample of all households.[4] The figure is included to enable a comparison with the other figures relative to the distribution of consumption in the population and can be used as a guide for assessing the levels and concentration of poverty in the population.

Figure 2 shows a nonparametric regression estimate of a function that is proportional to the conditional probability that a child in a household attends a school that will receive SIF funding, given the log per capita household consumption. The estimate is based on the equality

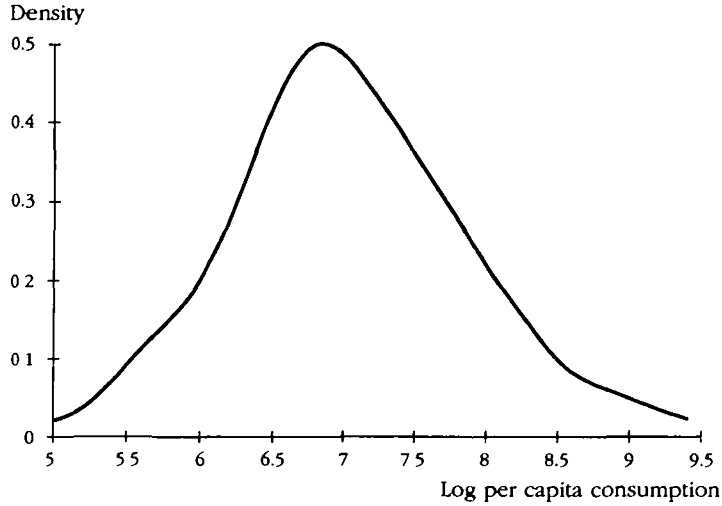
$$\Pr[\mathit{SIF\ school} \mid \log(\mathit{consumption})] = \frac{f[\log(\mathit{consumption}) \mid \mathit{SIF\ school}]}{f[\log(\mathit{consumption})]} \Pr(\mathit{SIF\ school})$$

where consumption is per capita household consumption. The probability that a child in the household attends a school that will receive support from the SIF is not known, but from the random sample of all households and the sample of households near schools that have been selected to receive SIF projects, we can estimate the densities in the numerator and denominator of the first term on the right-hand side of the equation. The estimated ratio is depicted in figure 2.

Figure 3 presents a nonparametric regression estimate of the probability that an individual visited a government health clinic in the month before the survey date as a function of log per capita household consumption. This estimate is

4. The nonparametric density and regression estimates are kernel smooth based on a Gaussian kernel. The bandwidth was chosen as suggested by Silverman (1986: 45).

Figure 1. *Density Function of the Log per Capita Consumption of Households in the Chaco and Resto Rural Regions of Bolivia, 1993*

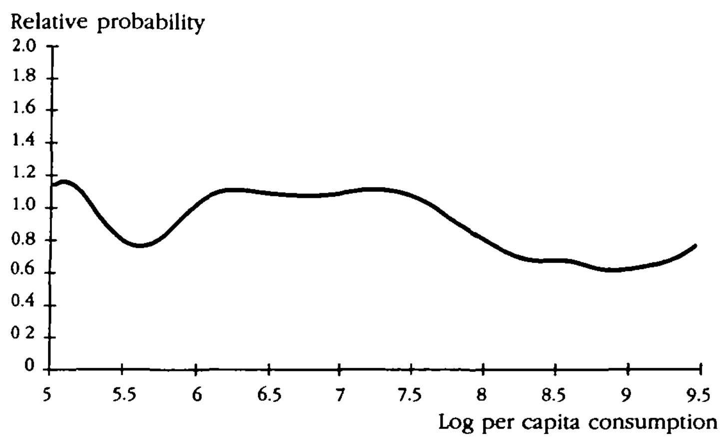


Source: Authors' calculations based on survey data.

based on the random sample of all households. The households that have a high probability of visiting a government health clinic are more likely to benefit from the SIF investments.

The nonparametric regressions estimate in figure 4 was obtained in the same way as for figure 2; it indicates the probability that the household lives in a community that is selected for an investment in sanitation.

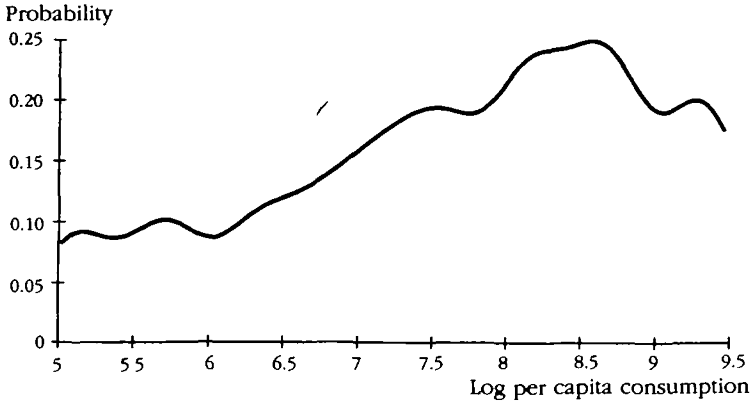
Figure 2. *Targeting of SIF Interventions in Education in Bolivia*



Note: Values are nonparametric estimates of the probability of benefit given household welfare. See the text for details on the calculation of the estimates.

Source: Authors' calculations based on survey data.

Figure 3. *Targeting of SIF Interventions in Health in Bolivia*

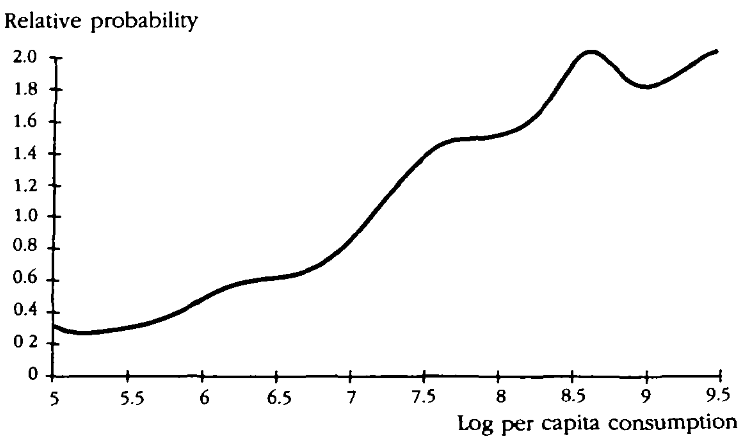


Note: Values are nonparametric estimates of the probability of benefit given household welfare. See the text for details on the calculation of the estimates.
Source: Authors' calculations based on survey data.

The figures reveal that there is no relation between household welfare, as measured by log consumption per head, and benefits from SIF interventions in education. However, SIF investments in health facilities and basic sanitation benefit households that are relatively better off.

District, area, and sector health clinics are not the only providers of medical care. Private doctors and particularly traditional healers are extensively consulted for medical care. Table 2 was constructed to determine who will benefit

Figure 4. *Targeting of SIF Interventions in Sanitation in Bolivia*



Note: Values are nonparametric estimates of the probability of benefit given household welfare. See the text for details on the calculation of the estimates.
Source: Authors' calculations based on survey data.

Table 2. *Health Status and Actual Health Care Consumption in the Month Preceding the Survey in Bolivia, 1993*
(percent)

Quartile level of per capita consumption[a]	*Reported good health*	*Sought medical care*	*If sought medical care, went to* Government clinic	Private doctor or midwife	Traditional care[b]	*If went to government clinic, type* Sector	Area	District
1-poor	84	44	44	10	47	48	33	19
2	83	55	58	10	32	42	35	23
3	80	58	59	20	21	38	34	28
4-rich	75	58	63	22	15	38	31	31
All	81	54	57	16	27	40	33	27

a. Quartiles correspond to the following levels of per capita yearly consumption: quartile 1, up to Bs746; quartile 2, Bs747–Bs1,224; quartile 3, Bs1,225–Bs2,158; quartile 4, above Bs2,158. Bolivia's currency is the boliviano.

b. Includes traditional healers, neighbors, family, and others without formal medical training.

Source: Authors' calculations based on survey data.

from the SIF investments in health centers. To this end, the population has been divided into four groups ranging from richer to poorer, depending on per capita consumption. The table presents results on the health status of future SIF beneficiaries and their use of health services prior to the SIF intervention. Health status is measured by asking the respondent whether he or she was in good health during the past month. Richer households tend to report a health condition that is worse than that of poorer households. However, this self-reported condition does not necessarily reflect the actual health status of the respondent, because richer people may be prepared to admit to being in bad health more readily than poorer people. Being in bad health is often associated with "not being able to work" or "seeking medical care," which is more readily affordable for the rich.

The data in table 2 show that, if ill, rich households seek health care more frequently and go more often to government health clinics. Poor households seek medical care less frequently and visit traditional healers more often. If the poor visit government health clinics, they mostly go to small (sector) health clinics. The results suggest that if the SIF wants to target investments in health facilities to poor communities, those investments should be concentrated in sector- and community-level health clinics. The results also show a need for information and outreach programs to encourage poor households to seek medical care when ill and to visit public health-care providers.

With respect to investments in sanitation, we find that households that will benefit from SIF investments in basic sanitation already have better sanitation facilities than most of the rural population. For example, 47 percent of the targeted households have access to piped water compared with only 26 percent of rural households in general. However, this is not necessarily inconsistent with SIF policy. Investments in basic sanitation are made most effectively in areas that already have access to a water system and are located in populated areas so that the project is able to take advantage of economies of scale. Constructing these facilities in remote rural areas may lead to better targeting but would be extremely costly relative to the per capita benefits achieved.

The results in this section show that evaluation of the impact of the SIF investments may be problematic. The selective targeting of investments in health care and sanitation biases a direct comparison of beneficiaries and nonbeneficiaries, because the latter were already worse off before the intervention. An impact evaluation should take this selectivity into account. Under the assumption that changes that occur in the time between the baseline survey and the date of measurement of the response variables affect all relevant units (facilities, regions, and communities) in the same way, we can estimate the impact of the SIF with a difference-in-differences estimator, as in Ashenfelter (1978). A direct comparison may be possible for the education component. In the next section, we investigate whether an unbiased estimate of the effect of SIF investments in education can be obtained.

III. RANDOMIZATION AND MATCHED COMPARISON IN EDUCATION

The German Institute for Reconstruction and Development had earmarked funding for education interventions in the Chaco region in 1991, yet the process for promoting SIF interventions in select communities had not been initiated. This situation provided an opportunity for assessing the need in schools for an SIF intervention and for applying a random-selection process. The school quality index for the Chaco region assigns each school a score from 0 to 9 based on the sum of five school infrastructure and equipment indicators: electric lights (1 if present, 0 if not); sewerage (2 if present, 0 if not); water source (4 if present, 0 if not); at least one desk per student (1 if so, 0 if not); and at least 1.05 square meters of space per student (1 if so, 0 if not).

Schools were ranked according to this index, with a higher value reflecting more resources. Only schools with an index below a particular cutoff value were eligible for an SIF intervention, and the worst-off schools were automatically designated to receive SIF-financed promotions and investments because of their extremely low quality. Both the worst-off and best-off schools were excluded from the randomization and sample, so that the restriction to eligible schools implies that the effects of the SIF cannot be generalized to all schools. The eligible schools in the middle of the quality distribution were allocated to the treatment or control group at random, creating the basis for an experimental evaluation design.

In the Resto Rural region schools had already been selected for SIF interventions, making randomization impossible. Therefore, treatment schools were sampled from the list of all schools designated for SIF interventions, and a comparison group was constructed by matching the sampled schools to non-SIF schools based on several observable characteristics.

The matching procedure used in the Resto Rural region consisted of two steps. First, using the 1992 census, cantons in which the treatment schools were located were matched to similar cantons with respect to population size and distribution by age, education level, gender, infant mortality rate, language, and literacy rate. Second, once the control cantons were identified, control schools were selected from the cantons to match the treatment schools with respect to the school quality index as developed for the Chaco region. No other data on schools, households, or communities were available for use in the matching exercise given the paucity of data in rural Bolivia.

Thus two distinct evaluation designs were used in the two regions: a classical experimental design with randomized assignment in the Chaco region and a matched-comparison design in the Resto Rural region. In recent years, there has been a controversy over the validity of various evaluation designs. As illustrated by Grossman (1994) in a review of the theory and practice of evaluation research, much of the controversy over evaluation techniques can be traced back to a seminal study by Fraker and Mayard (1984). They assess the impact of the National Supported Work (NSW) demonstration (a major employment program in the United States) using both a matched-comparison methodology and an experimental design made possible by the random assignment present in the NSW program. Fraker and Mayard calculate the impact estimates of NSW using both types of methodologies and find that the estimates based on the matched-comparison methodology do not come close to the impact estimated using the experimental design. In his own review of the NSW data, LaLonde (1986) supports this conclusion arguing that matched-comparison designs can be severely biased and that randomized assignment is the only design that can produce unbiased estimates of the effect of some intervention. In a reaction, Heckman and Hotz (1989) argue that careful modeling of the selection effect can remove most of this bias. However, because of the uncertainties in this approach, it seems that a more secure basis for identification of the intervention effect is needed.

The improved matching estimators of Dehejia and Wahba (1995) and Heckman and others (1998) cannot be used, because the selection for the SIF is not sufficiently understood to identify the relevant variables in the matching procedure and the current data are limited in their description of the project promotion and selection process. Hence, the most promising approach is to look for a variable that affects the selection into the treatment group, but not the relevant response variable. Such a variable is a valid instrument, and if such an instrument is available it can be used to obtain an unbiased, be it less efficient, estimate of the intervention effect (Angrist 1990 and Imbens and Angrist 1994). An instrument corresponds to an indirect experiment as opposed to a direct experiment due to randomization. Indirect experiments may be the only available evaluation design in many instances because, as in the Resto Rural region, random assignment often is not politically feasible and the information required for an unbiased matched comparison is rarely available.

In this section we check whether the assignment done in the Chaco region was indeed random, and we test whether the matching done in the Resto Rural region was selective and will give a biased estimate of the effect of the SIF with postintervention data. We test for random assignment because of the need to verify that the program administrators who were involved in selecting schools did not alter the planned evaluation design. Because the SIF cofinances investments in schools, it is natural to take the school as the unit in the evaluation. The goal of the investment is to improve the quality of education, and hence we need variables that measure this quality. An obvious choice would be the (average) score(s) on a standardized test of the achievement of pupils, but unfortunately such a test could not be administered during the baseline because of a major reform in the education sector. For that reason, we use two indirect measurements of education quality: repetition and dropout rates.

The repetition rate is defined as the fraction of pupils repeating a grade in the year of the survey; the dropout rate is defined as the fraction of students who dropped out of school in the same year. Both variables are indirect measures of school quality, and in particular a high repetition rate may result from either high standards or low-quality education. Although there is a lively debate over

Table 3. *Descriptive Statistics for Schools in the Chaco and Resto Rural Regions in Bolivia, 1993*

	Treatment schools		*Control schools*		*P-value of test of equal*[a]	
Region and variable	*Sample size*	*Mean value*	*Sample size*	*Mean value*	*Means*	*Distribution*
Chaco region						
Response variable						
Repetition rate	35	0.13	36	0.13	0.95	0.86
Dropout rate	35	0.13	36	0.09	0.20	0.51
School resources						
Blackboards per classroom	35	0.12	37	0.07	0.43	1.00
Desks per student	35	0.60	36	0.44	0.14	0.06
Students per classroom	35	24.3	36	22.1	0.36	0.12
Books per student	27	0.48	28	0.31	0.15	0.09
Students per teacher	35	19.9	36	20.0	0.97	0.90
Proportion of teachers with professional degrees	35	0.37	37	0.28	0.34	0.33
Characteristics of students						
Log per capita consumption of household[b]	31	7.05	34	7.03	0.81	0.98
Education of mother (years)	31	2.49	34	2.05	0.32	0.37
Education of father (years)	31	3.49	33	2.71	0.15	0.22
Community characteristics						
Number of nongovernmental organizations	28	0.54	28	0.14	0.08	0.44
Knowledge of the Social Investment Fund	28	0.46	28	0.46	1.00	1.00
Population	28	454.1	28	468.4	0.93	0.90
Distance to main road (kilometers)	28	17.8	28	13.3	0.40	0.44
Resto Rural region						
Response variable						
Repetition rate	24	0.08	21	0.08	0.92	0.98
Dropout rate	24	0.13	21	0.08	0.21	0.62
School resources						
Blackboards per classroom	37	0.53	33	0.22	0.00	0.02
Desks per student	37	0.72	32	0.45	0.03	0.04
Students per classroom	37	24.1	33	24.5	0.81	0.56
Books per student	37	0.38	33	0.28	0.27	0.66
Students per teacher	32	23.2	27	23.3	0.97	0.83
Proportion of teachers with professional degrees	32	0.55	27	0.57	0.86	0.88
Characteristics of students						
Log per capita consumption of household[b]	32	6.57	32	6.61	0.80	0.53
Education of mother (years)	32	1.35	32	1.43	0.80	0.94
Education of father (years)	32	3.85	31	3.75	0.86	0.95
Community characteristics						
Number of nongovernmental organizations	33	1.67	29	0.72	0.00	0.01
Knowledge of the Social Investment Fund	33	0.79	29	0.45	0.01	0.03
Population	33	293.4	29	344.7	0.52	0.10
Distance to main road (kilometers)	33	13.6	29	9.56	0.33	0.69

a. The test of equal means is the *t*-test; the test of equal distributions is the Kolmogorov-Smirnov test.
b. Household consumption is in bolivianos. In 1993 1 U.S. dollar = 4.31 bolivianos.
Source: Authors' calculations based on survey data.

the determinants of school effectiveness, most authors consider the resources available at the school to be an important determinant of the quality of education (Velez, Schiefelbein, and Valenzuela 1993). We are particularly interested in this dimension because the SIF aims to improve these resources. Hence, we also compare available resources at the schools, so that we use both the indirect measures and an indicator of the resources of the school as outcome variables. The results of the comparison are reported in table 3.

Table 3 shows compliance with the experimental design in the Chaco region. There are no significant differences between the treatment and control schools either in a comparison of the response variables or in a comparison of the school, student, or community characteristics. By contrast, in the Resto Rural region the matching of treatment schools with comparison schools on observable characteristics does not eliminate all differences between the schools. Although there are no significant differences for the response variables, the selected schools have significantly more resources in terms of the number of blackboards per classroom and desks per student. Moreover, they are located in communities with a larger number of NGOs and greater knowledge of the SIF.

Table 3 contains univariate comparisons. As a further check on treatment assignment, we estimate a probit that relates the probability of being a control or comparison school to the resources of the school and the characteristics of the student population.[5] The probit estimates in table 4 confirm the results of the univariate comparisons in table 3. There is weak evidence that even in the Chaco region schools with more resources have a higher probability of being selected for the SIF. However, the likelihood ratio test does not reject the hypothesis of random selection. That hypothesis is rejected for the Resto Rural region because better-off schools have a higher probability of selection. Note that the dummy for missing student data is not significantly different from 0, so that the selectivity is indeed due to nonrandom selection and not to nonrandomly missing observations.

Because assignment to the SIF seems to be related to the resources of the school and the resources affect the response variables, we estimate a linear regression in which the repetition rate and the dropout rate are related to school characteristics, student characteristics, and an indicator of being a comparison or control school. Besides the repetition and dropout rates, we also use an indicator of the resources of the school—the number of desks per student—as a response variable. Table 5 shows that for the Chaco region the control group indicator is not significantly different from 0 for all three response variables. Again this confirms compliance with the random assignment in the Chaco region.

For the Resto Rural region, we find that before the intervention, controlling for other observables, the comparison group had significantly lower dropout rates and fewer desks per student (table 5). As noted, comparison schools have significantly fewer resources but have repetition and dropout rates that do not

5. Control groups are randomly generated, whereas comparison groups are not.

Table 4. *Probit Estimate of the Probability of Selection for a Social Investment Fund Project in Bolivia, 1993*

Variable	*Chaco region*	*Resto Rural region*
School resources		
Blackboards per classroom	–0.36	–1.56
	(–0.53)	(–3.17)
Desks per student	–0.80	–1.17
	(–1.98)	(–2.27)
Students per classroom	–0.023	–0.042
	(–1.10)	(–1.43)
Books per student	–1.10	0.29
	(–2.30)	(0.45)
Students per teacher	–0.026	0.0011
	(–0.73)	(0.026)
Proportion of teachers with professional degrees	–0.62	0.12
	(–1.46)	(0.23)
Dummy for missing school data	0.18	–0.097
	(0.42)	(–0.082)
Characteristics of students		
Log per capita consumption of household	–0.29	0.32
	(–0.66)	(0.73)
Education of mother (years)	–0.081	0.29
	(–0.42)	(1.17)
Education of father (years)	–0.12	0.036
	(–0.78)	(0.30)
Dummy for missing student data	2.40	–2.27
	(0.78)	(–0.78)
Constant	2.05	1.72
	(1.96)	(1.45)
Likelihood ratio test of significance (p-value)	14.1	20.0
	[0.23]	[0.046]
Number of observations	71	69

Note: t-statistics are in parentheses; p-values are in square brackets.
Source: Authors' calculations.

differ significantly from treatment schools. For that reason, we find that, controlling for differences in resources, schools that perform poorly with respect to repetition and dropout rates are selected to receive SIF funding. The difference is significant for the dropout rate and the number of desks per student. None of the indicators for missing observations has a coefficient that is significantly different from 0. The results give an indication of the quality of the response variables. The repetition rate is weakly correlated with the resources of the school. That correlation is stronger for the dropout rate. This confirms our earlier doubts about the use of the repetition rate as a response variable.

We conclude that the matched-comparison design for the Resto Rural region does not yield directly comparable treatment and comparison groups. Comparison group schools have fewer resources but make better use of their resources, which results in lower repetition and dropout rates. The finding that there are

Table 5. *The Impact of School and Student Characteristics on Repetition Rate, Dropout Rate, and Desks per Student in the Chaco and Resto Rural Regions of Bolivia, 1993*

Region and variable	*Repetition rate*	*Dropout rate*	*Desks per student*
Chaco region			
School resources			
Blackboards per classroom	0.026	0.032	
	(0.43)	(0.61)	
Desks per student	-.055	0.035	
	(-1.53)	(1.14)	
Students per classroom	0.0022	0.0020	
	(1.23)	(1.28)	
Books per student	0.010	0.062	
	(0.24)	(1.70)	
Students per teacher	0.00075	-0.006	
	(0.26)	(-2.59)	
Proportion of teachers with professional degrees	0.013	-0.021	
	(0.33)	(-0.62)	
Dummy for missing school data	0.019	-0.019	
	(0.47)	(-0.54)	
Characteristics of students			
Log per capita consumption of household	0.046	0.040	-0.19
	(1.20)	(1.24)	(-1.42)
Education of mother (years)	-0.034	-0.013	0.0037
	(-1.89)	(-0.86)	(0.063)
Education of father (years)	0.022	0.0070	0.043
	(1.71)	(0.642)	(0.94)
Dummy for missing student data	-0.32	-0.29	0.99
	(-1.23)	(-1.28)	(1.08)
Control group	0.0039	-0.014	-0.13
	(0.123)	(-0.53)	(-1.24)
Constant	0.076	0.19	0.741
	(0.81)	(2.37)	(4.42)
R^2	0.21	0.22	0.10
Number of observations	71	71	71
Resto Rural region			
School resources			
Blackboards per classroom	-0.014	-0.056	
	(-0.38)	(-1.37)	
Desks per student	-0.067	-0.084	
	(-1.76)	(-2.02)	
Students per classroom	0.0014	-0.0010	
	(0.39)	(-0.28)	
Books per student	0.011	-0.060	
	(0.18)	(-0.91)	
Students per teacher	0.00052	-0.00046	
	(0.15)	(-0.12)	
Proportion of teachers with professional degrees	0.0025	0.084	
	(0.051)	(1.58)	
Dummy for missing school data	0.020	-0.20	
	(0.20)	(-1.84)	

(Table continues on following page.)

Table 5. *(continued)*

Region and variable	*Repetition rate*	*Dropout rate*	*Desks per student*
Characteristics of students			
Log per capita consumption of household	0.018	0.027	0.063
	(0.40)	(0.53)	(0.56)
Education of mother (years)	–0.008	–0.024	0.11
	(–0.49)	(–1.31)	(1.86)
Education of father (years)	–0.001	0.020	0.041
	(–0.12)	(1.88)	(1.09)
Dummy for missing student data	–0.15	–0.37	–0.87
	(–0.49)	(–1.12)	(–1.18)
Comparison group	–0.026	–0.094	–0.26
	(–0.78)	(–2.61)	(–2.30)
Constant	0.12	0.54	0.85
	(1.02)	(4.21)	(4.77)
R^2	0.28	0.58	0.26
Numbers of observations	44	44	69

Note: Results are from ordinary least squares. t-statistics are in parentheses.
Source: Authors' calculations.

no significant differences in the response variables between the comparison and treatment group schools ex ante does not imply in general that a direct comparison of the response variables between the two groups ex post yields an unbiased estimate of the true SIF effect. The resulting bias can be derived from a linear regression equation that relates the response variable y to the vector of school resources x (we omit student characteristics, which do not differ between treatment and comparison schools and are assumed to be the same before and after the intervention). The bias is derived for a linear regression, but this restriction is for ease of exposition only. The results can be generalized to arbitrary nonlinear relations. Consider

$$y^v_{stk} = \alpha^v_{st} + \beta' x^v_{stk} + \varepsilon^v_{stk} \quad t = 0,1; k = 1,\ldots,K_s; s,v = C,T$$

where k is the school, s indicates the assignment to the treatment ($s = T$) or comparison ($s = C$) group, v denotes the outcome with ($v = T$) or without ($v = C$) the SIF investment, and t is 0 if the data are preintervention or 1 if postintervention. We introduce the superscript v to stress that for a particular school the program effect is obtained by comparing the outcome variable for that school with ($v = T$) and without ($v = C$) the intervention. Of course, only one of these variables can be observed: the outcome where superscript v coincides with the assigned treatment s, that is, $s = v$.

The true SIF effect, that is, the difference in the expected outcomes with and without the SIF intervention, can be defined either for the treatment or the comparison schools. If we choose the first option, the average SIF effect is

$$(1) \qquad E(y^T_{T,1}) - E(y^C_{T,1}) = (\alpha^T_{T,1} - \alpha^C_{T,1}) + \beta'(\bar{x}^T_{T,1} - \bar{x}^C_{T,1})$$

where an overbar indicates the mean value over all schools. The counterfactual $E(y^C_{T,1})$, which is the expected outcome of the treatment group had they not received an SIF investment, is not observed. Under several assumptions, however, the average treatment effect can be derived directly from the observed preintervention and postintervention outcomes.

Using the (observed) postintervention difference between comparison and treatment groups, we obtain

$$E(y^T_{T,1}) - E(y^C_{C,1}) = (\alpha^T_{T,1} - \alpha^C_{C,1}) + \beta'(\bar{x}^T_{T,1} - \bar{x}^C_{C,1}). \tag{2}$$

Equation 2 is an estimate of the average SIF effect for the treatment schools, as defined in equation 1. The estimated effect in equation 2 results in a bias equal to (subtracting equation 1 from equation 2)

$$(\alpha^C_{T,1} - \alpha^C_{C,1}) + \beta'(\bar{x}^C_{T,1} - \bar{x}^C_{C,1}).$$

This bias is the average difference in outcomes between the treatment and comparison schools for the counterfactual case that the treatment schools do not receive the SIF support.

Under the strong assumption that the difference in the observed (x) and unobserved (α) characteristics between comparison and treatment groups in the absence of an intervention remains the same before and after the intervention,

$$\alpha^C_{T,1} - \alpha^C_{C,1} = \alpha^C_{T,0} - \alpha^C_{C,0} \text{ and } \bar{x}^C_{T,1} - \bar{x}^C_{C,1} = \bar{x}^C_{T,0} - \bar{x}^C_{C,0}$$

and under the weak assumption that the intervention has no effect for the treatment group before the start of the program (this assumption is also made with the IV estimator below),

$$\alpha^C_{T,0} = \alpha^T_{T,0} \text{ and } \bar{x}^C_{T,0} = \bar{x}^T_{T,0}$$

we find that the bias is equal to the difference between the preintervention average responses:

$$(\alpha^T_{T,0} - \alpha^C_{C,0}) + \beta'(\bar{x}^T_{T,0} - \bar{x}^C_{C,0}).$$

Thus the difference between the observed postintervention and preintervention averages gives an unbiased estimate of the SIF effect for the schools that are selected for the SIF. The estimator is the difference-in-differences estimator, as used by Ashenfelter (1978). Because of the long time interval between preintervention and postintervention surveys—five years—and the marked differences found between the two groups before the intervention, it seems very unlikely that the two groups would have evolved similarly in the absence of an intervention. Therefore, it is unlikely that the assumptions that support difference-in-differences estimation will hold.

From the estimates in tables 3–5, we have

$$\alpha^T_{T,0} < \alpha^C_{C,0}, \quad \beta' \bar{x}^T_{T,0} > \beta' \bar{x}^C_{C,0}.$$

The preintervention difference between the average responses is not significantly different from 0. This does not imply that the postintervention difference in the average response is an unbiased estimator of the SIF effect. The strong assumptions introduced above are needed to guarantee that this estimator is unbiased. As this example demonstrates, the fact that the difference between the preintervention average response of the treatment and comparison groups is 0 is neither a necessary nor a sufficient condition for using postintervention difference as an unbiased estimator of the SIF effect for the schools that receive SIF support. In the next section, we explore how to address the potential bias.

IV. Quasi-Experimental Evaluation

Besides matching and IV, economists have suggested other methods to deal with the bias due to selective treatment. These other methods are based on modeling the selection decision and are appealing if the assignment is the result of choice by an economic agent who has an interest in the outcome of the choice. The estimates in table 4, where the assignment is related to observable characteristics of the schools and the students, are a first step in this direction. However, in our case, because several actors are involved in selecting schools in the Resto Rural region to receive SIF projects, a detailed economic model is hard to identify. A model of the assignment process will be at best a reduced-form model. The results in table 5 show that the bias is not induced by the observable variables in the model for the treatment assignment, but by the unobservables. Hence, we should allow for a correlation between the error of the regression for the response variable and the error of the binary response model for treatment assignment (Heckman 1979). To apply this method, we must specify the joint distribution of the unobservable errors. It is well known that this method gives sensible results only if there are variables that affect the treatment assignment, but not the response variable. Such variables are also essential for the quasi-experimental IV approach, which does not require arbitrary distributional assumptions and for that reason is more robust.

The quasi-experimental IV approach starts from the observation that randomized assignment as used in a classical experiment induces exogenous variation in the intervention indicator, and this variation is not correlated with the response variable. If we can find a variable that affects the treatment assignment, but not the response variable, we have exogenous variation that mimics the type of variation induced by randomization. A variable with these properties is called an instrumental variable, and the corresponding experiment is referred to as a quasi or indirect experiment. The instrumental variable estimator of the intervention effect is consistent, but it is less accurate, that is, it has a larger variance, than the estimator of the intervention effect that could be used if the treatment assignment had been random.

In this section, we use preintervention data to show that some community characteristics provide valid instruments, because they affect the selection into the SIF, while they do not have an effect on the preintervention response variables. Hence, these instruments can be used in a quasi-experimental evaluation of the SIF. We also study the efficiency of this design. We show that the relative efficiency, that is, relative to a randomized design, is independent of the true treatment effect and hence can be estimated using baseline data only. This allows us to determine the sample size that will compensate for the loss of efficiency due to the quasi-experimental design.

The comparison in table 3 suggests that for the Resto Rural region the number of NGOs and the community leaders' knowledge of the SIF have a significantly positive effect on selection into the SIF. This is confirmed by a linear regression—the first step in a two-stage least squares estimation procedure for the SIF effect—of the indicator of selection into the program on the exogenous variables and the set of potential instruments in table 6. The number of NGOs and knowledge of the SIF have a significant negative effect (on the selection into the control group) in this linear probability model with coefficients –0.23 and –0.19, respectively (see table 6). This result is expected because NGOs often act as subcontractors for the implementation of projects and because of the role that local leaders have in the selection process. The variables are valid instruments if they have no effect on the response variables. The evidence for the Resto Rural region is in table 7. Because the regression coefficients for the potential instruments are not significantly different from 0, we conclude that they are valid instruments.

The instrumental variables allow us to obtain a consistent estimate of the effect of the SIF intervention from the postintervention data from the regression equation[6]

$$y_{s1k} = \alpha_{01} + \alpha_{11} d_{sk} + \beta' x_{s1k} + \varepsilon_{s1k} \quad k = 1, \ldots, K_s; s = C, T$$

where d is the vector of observations on the SIF indicator, and $d_{sk} = 1$ for the schools that receive SIF funding. X is the matrix with the observations on the independent variables including the constant but excluding the SIF indicator, and Z is the matrix with the observations on the instrumental variables. We define

$$M = I - X(X'X)^{-1}X'$$

with I as the identity matrix and M as the least-squares projection matrix. The

6. Because the denominator of an IV estimator is random, it is biased in small, but not in large, samples.

Table 6. *Linear Probability Model of Membership in the Control or Comparison Group in Bolivia, 1993*

Variable	*Chaco region*	*Resto Rural region*
School resources		
Blackboards per classroom	–0.2037	–0.3133
	(–0.93)	(–2.51)
Desks per student	–0.2553	–0.1541
	(–1.96)	(–1.44)
Students per classroom	–0.0054	–0.0086
	(–0.82)	(–1.15)
Books per student	–0.4108	–0.0512
	(–3.15)	(–0.35)
Students per teacher	–0.0130	0.0069
	(–1.29)	(0.74)
Proportion of teachers with professional degrees	–0.2747	0.1362
	(–1.82)	(1.13)
Dummy for missing school data	0.0788	–0.1149
	(0.48)	(–0.40)
Characteristics of students		
Log per capita consumption of household	–0.1009	0.0485
	(–0.84)	(0.38)
Education of mother (years)	–0.0368	0.1174
	(–0.53)	(2.15)
Education of father (years)	–0.0161	–0.0163
	(–0.33)	(–0.55)
Dummy for missing student data	0.8254	–0.1057
	(0.98)	(–0.13)
Potential instruments		
Knowledge of the Social Investment Fund	0.0455	–0.1877
	(0.38)	(–1.30)
Number of nongovernmental organizations	–0.1221	–0.2252
	(–1.37)	(–4.15)
Population (thousands)	0.0683	0.0359
	(0.52)	(0.26)
Distance to main road (kilometers)	–0.0043	0.0006
	(–1.11)	(0.17)
Dummy for missing instruments data	–0.1149	0.2934
	(–0.72)	(1.68)
Constant	1.3374	0.5865
	(4.17)	(1.77)
F-test instruments (*p*-value)	1.53	8.07
	(0.19)	(0.00)
R^2	0.24	0.43
Number of observations	71	69

Note: Values are from a linear probability model regressing membership in control group or school and student characteristics and potential instruments. Huber-corrected standard errors are in parentheses.

Source: Authors' calculations.

Table 7. *The Impact of Response Variables on Repetition Rate, Dropout Rate, and Desks per Student in the Resto Rural Region in Bolivia, 1993*

Variable	*Repetition*	*Dropout rate*	*Desks per student*
School resources			
Blackboards per classroom	–0.013	–0.053	
	(–0.32)	(–1.07)	
Desks per student	–0.036	–0.078	
	(–0.88)	(–1.59)	
Students per classroom	0.0015	–0.0012	
	(0.45)	(–0.30)	
Books per student	0.0037	–0.065	
	(0.060)	(–0.88)	
Students per teacher	0.00052	0.00006	
	(0.14)	(0.013)	
Proportion of teachers with professional degrees	–0.043	0.082	
	(–0.82)	(1.32)	
Dummy for missing school data	0.033	–0.21	
	(0.32)	(–1.68)	
Characteristics of students			
Log per capita consumption of household	0.040	0.030	0.045
	(0.81)	(0.51)	(0.37)
Education of mother (years)	–0.021	–0.022	0.14
	(–1.11)	(–1.00)	(1.91)
Education of father (years)	0.00072	0.019	0.032
	(0.065)	(1.44)	(0.80)
Dummy for missing student data	–0.27	–0.38	–0.81
	(–0.83)	(–1.00)	(–1.00)
Potential instruments			
Knowledge of the Social Investment Fund	–0.035	–0.0026	0.067
	(–0.84)	(–0.053)	(0.44)
Number of nongovernmental organizations	0.022	–0.011	0.0024
	(1.21)	(–0.49)	(0.031)
Population (thousands)	0.038	0.010	–0.26
	(0.70)	(0.16)	(–1.20)
Distance to main road (kilometers)	–0.0010	–0.00034	0.0017
	(–0.80)	(–0.23)	(0.41)
Dummy for missing instruments data	–0.097	–0.00064	0.20
	(1.44)	(–0.008)	(0.90)
Constant	0.19	0.55	0.71
	(1.52)	(3.77)	(3.14)
F-test instruments (p-value)	1.22	0.10	0.60
	[0.32]	[0.99]	[0.70]
R^2	0.42	0.59	0.29
Number of observations	44	44	69

Note: Values are ordinary least squares estimates. t-statistics are in parentheses; p-values are in square brackets.

Source: Authors' calculations.

variances of the ordinary least squares and IV estimators of α_{11} are given respectively by

$$\frac{\sigma^2}{d'Md}$$

and

$$\frac{\sigma^2}{d'MZ(Z'MZ)^{-1}Z'Md}.$$

The ratio of these variances does not depend on the variance of the disturbance of the postintervention regression, σ^2, and hence can be computed with preintervention data.

For the Resto Rural region, we find that the ratio of the standard errors of the estimates of the SIF effect is 3.97. Hence the standard error is four times larger than could have been obtained if the matched comparison had succeeded or if the SIF assignment had been random. As a consequence, 242 schools instead of 61 are needed to estimate the SIF effect with the same precision as with randomized assignment. Because the assignment was random for the Chaco region, we find that the ratio is 11.20 for that region. Using an instrumental-variable estimator with randomized assignment gives a very inaccurate estimate of the SIF effect.

V. Conclusions

We used preintervention data to study the targeting of Bolivian SIF interventions in the education, health, and sanitation sectors and to examine selection of SIF interventions. For the health and sanitation components, we found that households that are better off are more likely to be beneficiaries of SIF investment. For these components, the selectivity of the SIF complicates the impact evaluation.

For the education component, the random selection of a group of schools eligible for SIF interventions that then received active promotion for SIF education projects facilitated evaluation of the impact of the SIF. In the Resto Rural region, an attempt was made to mimic randomized assignment by matching treatment and comparison group schools on observable characteristics. We found that this attempt was not fully successful. The matched-comparison approach in the Resto Rural region yielded less comparable treatment and comparison groups than the random-selection process used in the Chaco region. We proposed using an alternative indirect procedure to evaluate the intervention in the Resto Rural region using an instrumental variable approach to control for nonrandom selection. The preintervention data allowed us to verify that our instrumental variable proposal will produce an unbiased estimate of the SIF effect but remained a less efficient approach compared with randomization. We computed the loss of accuracy due to the indirect experiment and estimated the number of schools that will be needed to obtain an estimate with a precision that is comparable to that obtained from randomized assignment.

Our analysis demonstrated that a simple matched-comparison design introduces a bias in the estimate of the program effect. We showed that with preintervention data we can select a quasi or indirect experiment that coincides with the selection of valid instrumental variables and that preintervention data allow us to test the validity of the instruments. We revealed that the loss of efficiency due to the use of quasi-experimental techniques instead of random treatment assignment can be determined from preintervention data. With post-intervention data, we will be able to compare estimates of the program effects obtained using the different methods.

REFERENCES

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

Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80(3, June):313–36.

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60(1, February):47–57.

Dehejia, R. H., and S. M. Wahba. 1995. "Causal Effects in Non-Experimental Studies." Working Paper. Department of Economics, Harvard University, Cambridge, Mass. Processed.

Fraker, Thomas, and Rebecca Mayard. 1984. *An Assessment of Alternative Comparison Group Methodologies for Evaluating Employment and Training Programs.* Princeton, N.J.: Mathematica Policy Research, Inc.

Friedlander, Daniel, and Philip K. Robins. 1995. "Evaluating Program Evaluations: New Evidence on Some Commonly Used Nonexperimental Methods." *American Economic Review* 85(September):923–37.

Glaessner, Phillip J., Kye Woo Lee, Anna Maria Sant'Anna, and Jean-Jacques de St. Antoine. 1994. *Poverty Alleviation and Social Investment Funds: The Latin American Experience.* World Bank Discussion Paper 261. Washington, D.C.: World Bank.

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

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1, January):153–61.

———. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32(3, summer):441–62.

Heckman, James J., and V. J. Hotz. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association* 84(408):862–80.

Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data: A Study of Adult Males in JTPA." *Econometrica* 66(5, September):1017–98.

Holland, P. W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(408):945–70.

Imbens, Guido, and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2, March):467–76.

Jimenez, Emmanuel. 1995. "Human and Physical Infrastructure: Public Investments and Pricing in Developing Countries." In Jere Berman and T. N. Srinivasan, eds., *Handbook of Development Economics,* vol. 3B, chap. 47. Amsterdam: North-Holland.

LaLonde, R. J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76(4, September):604–20.

Moffitt, Robert. 1991. "Program Evaluation with Nonexperimental Data." *Evaluation Review* 15(3):291–314.

Newman, J. L., Margaret Grosh, and Steen Jorgensen. 1992. "Demand-Driven Funds: Managing Their Conflicts." In Steen Jorgensen, Margaret Grosh, and Mark Schacter, eds., *Bolivia's Answer to Poverty, Economic Crisis, and Adjustment: The Emergency Social Fund,* chap. 7. A World Bank Regional and Sectoral Study. Washington, D.C.: World Bank.

Rosenbaum, P. R., and D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Velez, Eduardo, Ernesto Schiefelbein, and Jorge Valenzuela. 1993. *Factors Affecting Achievement in Primary Education: A Review of the Literature for Latin America and the Caribbean.* Washington, D.C.: Human Resources and Development Operations Policy, World Bank.