2013

The Effectiveness of Public Development Banks: Designing Good Impact Evaluations

Alessandro Maffioli
Inter-American Development Bank

Cesar M. Rodriguez
Portland State University, cesar.rodriguez@pdx.edu

Citation Details

Maffioli, Alessandro and Rodriguez, Cesar. (2013). "The Effectiveness of Public Development Banks: Designing Good Impact Evaluations," in Public Development Banks: Toward a New Paradigm? IDB Publications, Chapter 3, pages 39-70. ISBN 978-1-59782-172-8

CHAPTER 3

The Effectiveness of Public Development Banks: Designing Good Impact Evaluations

Alessandro Maffioli and César M. Rodríguez

- Although still in the early stages, evidence from Latin America shows that there have been significant efforts to measure the effectiveness of programs closely linked to public development banks (PDBs) and of PDBs themselves.
- Sound impact evaluations require a clear definition of the impact indicators to measure, access to a full and reliable database, selecting the best methodological approach in accordance with the questions and the available data, and having the necessary financial and human resources.

WHICH EVALUATIONS ARE PERTINENT?

PDB programs have become a fundamental ingredient of productive development policy strategies in most emerging economies. Although the overall need for these interventions is rarely questioned, academics and policymakers often debate their effectiveness, as well as the optimal approaches and instruments necessary to implement them. Therefore, the need to produce rigorous evaluations of PDBs has become increasingly relevant for both government and civil society (see Chapter 2).

This chapter presents the main concepts and operational arguments regarding the execution of in-depth impact evaluations of PDB initiatives and instruments. For a more practical approach, these arguments are presented with examples of such evaluations, which have either ended or are ongoing, as well as of other programs that relate to their activities. This, however, limits the scope of this chapter.

First, only one key aspect of the evaluation process is included: the attribution of effects. This suggests that all the methods and techniques covered address the fundamental problem of identifying the causal relationship between public policy intervention and the observed changes in the study's target population. Other important elements relating to a comprehensive evaluation process—such as efficiency, relevance, and institutional coherence—fall beyond the scope of this analysis.

Second, only quantitative approaches are included, in order to solve the problem of attribution. This does not, in any way, imply that the contribution made by the qualitative approaches to the study of PDBs is not appreciated. On the contrary, quantitative and qualitative approaches are complementary, but much more exhaustive studies are required to include both. This chapter focuses mainly on the methodological literature based on counterfactual analysis, which stems from applying experimental and quasi-experimental methods to the evaluation of public policy.

Interventions that are applicable to a PDB can cover a wider range of sectors than can be dealt with in one chapter alone. Therefore, the analysis and discussion in this chapter is restricted to those PDB initiatives that improve access to credit for the productive sector (business and agriculture). As such, it is possible to discuss more specific ideas and suggestions, while acknowledging that the complexity and characteristics of PDB programs call for more specific studies.

The rest of this chapter is structured as follows. It begins with a section posing the most important questions that are (or should be) included in any study of PDB effectiveness. The following section will identify the most commonly used indicators in these studies and the potential sources of information needed to establish such indicators. Later, the chapter analyzes the methods to respond to important questions relating to an evaluation, thus ensuring that the effects are correctly attributed. The final section explores the resources required to carry out a rigorous PDB evaluation.

AN EFFICIENT EVALUATION: ESSENTIAL QUESTIONS

One of the first issues to determine prior to conducting a PDB impact evaluation relates to the evaluation's principal objectives. These can be divided into two basic groups: (i) those that relate to the Average Treatment Effect on the Treated (ATT) and (ii) those that relate to an analysis of the program's secondary effects. The majority of analyses of the effectiveness of PDB programs seeks answers in terms of the ATT; for example, an analysis of the impact of the creation of a credit line on the quality of access to credit—or on the performance of the beneficiary businesses—should focus on the ATT.

Once a careful selection of the apparent outcomes and their indicators has been made, an evaluation of the impact of PDB programs is not a trivial task, especially in terms of measuring the causal impact that these programs have on expected outcomes. The definition of causality in any impact evaluation is based on counterfactual analysis; in other words, what would have happened if the program had not existed? For example, if a business receives a subsidized loan or a specific line of credit, and the value of a certain outcome is observed (credit quality, performance, etc.), the public subsidy will have a causal impact when it can be demonstrated that, in its absence (all other factors being equal), the outcome would have been different.

Although this is a relatively simple and inherent definition of causality, it does present an important empirical complication: by definition, the counterfactual result can never be observed. In other words, if a firm receives a subsidy, it becomes impossible to determine what outcomes that firm might have achieved without the subsidy, or vice versa. Holland (1986) refers to the impossibility of observing a determined unit concurrently with and without treatment as the "fundamental problem of causal inference."

The challenge of defining an adequate counterfactual cannot be resolved based solely on an individual observation (in other words, it is impossible to generate a counterfactual for a specific beneficiary of a public intervention). However, it can be resolved efficiently in terms of the average values of a combination of beneficiaries.[1] Impact evaluations, therefore, focus on calculating the average, rather than the individual, effect of the treatment.

It is possible to estimate this average effect in various ways. The parameter for the widest scope is the average impact of the treatment on the population, as a whole: Average Treatment Effect (ATE). Calculating the ATE involves constructing two counterfactuals (and, therefore, measuring two parameters): first, the counterfactual of what would have been the outcome for beneficiaries if they had not been beneficiaries (the ATT) and, second, the counterfactual of what would have been the outcome for the nonbeneficiaries if they had, in fact, been beneficiaries (also known as the Average Treatment Effect on the Untreated, or ATU).

These parameters will be biased in any study that does not incorporate the random assignment of beneficiaries (see Appendix 3.1). In all other cases, econometric techniques should be applied to eliminate biases and accurately calculate the program's average impact.

Although both ATE and ATT are extremely important for evaluating the effectiveness of an intervention, a well-designed evaluation can provide additional information to aid in the analysis of this effectiveness, and derive adequate conclusions that contribute toward a successful policy design. One should consider the following aspects in the design of an evaluation.

EXTERNALITIES

When a program is implemented, a producer or a business can experience varying types of externalities or indirect effects. For example, the fact that a business receives a loan according to a PDB policy could mean that it will undergo changes in its production chain that will augment its productivity, which could, in turn, affect other neighboring firms, or those with which it is linked. These other firms may be either geographically close or linked through the production process, in which case they can be considered as indirect beneficiaries. In principle, a distinction can be drawn between the monetary and nonmonetary effects of externalities. For example, monetary effects could be those related to cost reductions in the production chain, whereas nonmonetary effects could be changes in actual production technique.

[1] See Appendix 3.1 for an analysis of this subject.

Economies of agglomeration may even arise, which can result from a combination of positive externalities occurring simultaneously, specific to an industry or location. For example, in a PDB program, beneficiary enterprises could take advantage of economies of agglomeration and enhance their performance through information flow and new technologies, generated by both formal and informal links between enterprises and organizations. In turn, these effects could generate negative and positive externalities and/or general equilibrium effects. Therefore, these aspects should be considered within the evaluation; otherwise the overall impact of the PDB programs could be misinterpreted.

Distribution of Effects over Time

It is possible that the effects of certain PDB programs, such as economic performance, take time to reveal themselves. In fact, the process of incorporating new credit, recruiting adequate staff, and organizing the business will delay the effect on economic performance.

These time lags can vary, according to the economic performance indicator selected. For example, an intervention could generate a temporary increase in results, or it could have significant impacts that would dissipate progressively over time. Alternatively, the program's impact may become apparent only after a determined period, or there could be an initial decrease in results, but later an improvement.

Therefore, a PDB program impact evaluation should contain an adequate idea of the distribution of the effects on beneficiary enterprises over time. A distinction should be clearly made between the short-, medium-, and long-term effects, in order to adequately evaluate the costs and benefits of a public program. In fact, focusing only on a brief period after an intervention could lead to an underestimation of the impacts in the event that the program's effects take several years to be recognized. On the other hand, evaluations that only take into account periods following implementation of the intervention could result in an underestimation of costs, should an adjustment process take place during the initial few years.

Intensity of Treatment and Dosage Effects

Literature relating to impact evaluations generally analyzes the binary case of participation versus the lack of participation in a determined program. In practice, units may often differ, not just according to their binary treatment status (participants versus nonparticipants), but also according to treatment intensity. For example, enterprises may receive different amounts of financing from a PDB loan program, or they may participate by taking out loans at varying times. This highlights the important aspects that need to be kept in mind during the evaluation design. It is useful to recognize whether participants perform better than nonparticipants, as well as how different degrees of treatment intensity influence performance, and whether it is possible to locate an "optimal level" for intervention (e.g., the amount of financing that maximizes the effects on corporate performance).

Multiple Treatments

In contexts of multiple treatments, the evaluator may be interested not only in the individual effects of each treatment, but also in the effects generated by their interactions. It is far from clear whether the effect of multiple programs is always cumulative. However, research indicates that combining different interventions can produce multiplicative effects, but also that the effects of one treatment can sometimes cancel the effects of another (e.g., when enterprises take investment loans from a public bank and, at the same time, a subsidy for innovation is granted by a funding program). Investigating the combined effect of different types of interventions is crucial to effective PDB program design.

Heterogeneity of the Impact

In most contexts, it is hard to assume that a certain intervention will have a constant effect on all of the units reviewed. In particular, two main types of impact heterogeneity can emerge. The first occurs when the interventions have varying effects on different groups (e.g., when the effect of a PDB credit is stronger in those businesses that would otherwise be experiencing liquidity constraints). The second relates to distribution of the effects of the program throughout the population; for example, two programs might have the same average effect, but the effects of one could be concentrated in the lower half of the distribution (Frölich and Melly, 2009).[2]

INDICATORS AND DATA

Indicators

Various indicators can be used to evaluate the impact of PDB programs on business performance.[3] These include productivity, innovation, and employment, as well as others that relate to exports.

Productivity

There are various ways to measure productivity in an enterprise. The term may refer to the productivity of an input (e.g., labor productivity) or to the productivity of all inputs (i.e., the total factor productivity,

[2] Heckman, Smith, and Clements (1997) list other parameters that might be of interest to the evaluator: (i) the percentage of persons that accept the program and benefit from it, (ii) the percentage of the total population that benefits from the program, (iii) certain impact distribution quantiles, and (iv) the distribution of gains to certain base state values. In these contexts, restricting the analysis to the average impact on the overall population (or on the treated population) might lead to an inaccurate, or at least incomplete, evaluation of the program's effects.

[3] As Gertler et al. (2011) reveal, the foremost indicators should be Specific, Measurable, Attributable, Realistic, and Targeted (SMART).

or TFP). Special caution must be taken when measuring labor productivity, which is expressed as the ratio between total production and the work factor.[4] In practice, enterprises produce diverse goods, and these have to be aggregated in a single measurement of production (e.g., sales or added value). Generally, there is information available regarding the number of employees and the labor costs, despite the fact that nominal variables should be qualified to obtain true variables.

With regard to TFP, the various methods of calculation make assumptions about the production process and market competitiveness. Each method, therefore, has its strengths and weaknesses.[5]

Given the difficulty of observing this variable, many PDB programs are designed to directly address the improvement of diverse related variables that are easier to observe, such as, for example, the value of exports, research and development (R&D) costs, innovation, total sales, and employment levels.

Export-related indicators

In some cases, PDB programs can focus on promoting the exports of beneficiary firms. In order to measure the effects of these kinds of programs, different indicators can be used, including the value of exports, the probability that a firm becomes an exporter, the number of goods exported, and the number of export markets.

Some impact evaluations in the Latin American and Caribbean (LAC) region use these indicators. For example, using the database that includes enterprises in Peru, throughout the period 2001–05, Volpe and Carballo (2008) find a relationship between the initiatives for promoting and for increasing exports, both in terms of markets and products. Likewise, according to a combination of corporate data covering the period 1996–2008, Castillo et al. (2011) observe that Argentina's Business Restructuring Program (Programa de Apoyo a la Reestructuración Empresarial, or PRE), which aims to strengthen the small- and medium-sized enterprises (SMEs), does enhance the chances of a firm becoming an exporter.

Innovation-related indicators

PDB programs can aim to correct market gaps by promoting investment by enterprises into R&D.[6] Instruments used to tackle this problem—as well as financial limitations on innovation—include public subsidies (through support and nonreturnable grants), specific credits, tax incentives, and tools related to

[4] Ideally, these should be measured as the quantity of goods produced and the number of hours worked to produce those goods, respectively.
[5] For studies relating to the estimation of productivity at the enterprise or establishment level, see Hulten (2001), Bartelsman and Doms (2000), and Van Biesebroeck (2009).
[6] With regard to the innovative initiatives, market gaps can arise due to the difficulty of the private sector to appropriate the social returns arising from such initiatives.

intellectual property. According to the nature of the impact to be measured, examples of these indicators include total spending on R&D, performance, and the number of patents granted.

Employment-related indicators

Finally, a PDB program can increase employment in participating enterprises. This combination of indicators could, therefore, include the number of employees, the type of employee in terms of qualification level, and the level of staff remuneration. For example, Castillo et al. (2011) presented evidence that the PRE program actually increased both the number of employees and the salaries.

Data

When evaluating the effectiveness of PDB programs, having access to high quality data can make all the difference. The data employed should be readily available, accurate, and reliable. One challenge that faces PDB program evaluations in the LAC region is that secondary data—in other words, data gathered for objectives other than evaluating a certain policy—is not usually available. Although surveys and censuses do exist that could well provide ample information for evaluating and monitoring PDB programs, they are not always available for these purposes. This lack of availability also hinders primary data gathering.[7]

Secondary data

There are three sources of secondary data: surveys, censuses, and administrative registers. Each one of these sources has its advantages and disadvantages, which should be considered during an evaluation.

Surveys have the advantage of enabling a group of businesses to be established with annual information. Furthermore, they provide information about different variables, enabling the evaluator to use matching methods to locate nonbeneficiaries with similar characteristics to the beneficiaries. However, the principal disadvantage of these surveys is that they include only samples of the population, and in many cases, these samples include only a small percentage of the beneficiaries.

Censuses, on the other hand, collect data concerning the total beneficiary population. Therefore, if the beneficiaries are active when the census is conducted, they will be included. Censuses tend to gather more information than surveys, which becomes significant when applying the statistical technique of propensity score matching (PSM). The main disadvantage of censuses is that they are not conducted every year.

[7] For example, as will be further examined below, in order to define a sample of businesses, a list of all the firms in the region or the country should be available. If there is census data, the task becomes more difficult.

Finally, administrative registers refer to a wide range of information about businesses, collected by various institutions for purposes other than evaluation. As with censuses and surveys, these databases can only be used within the institution that administers them, and only under a confidentiality agreement. The main advantage of these administrative databases, compared to surveys and censuses, is that in most cases, they provide annual information regarding each and every business. However, the information is limited and indicators, such as TFP, cannot always be expanded upon.

The administrative databases employed in the evaluations of productive development policies and PDB programs in the LAC region are the following: the Dynamic Analysis of Employment Database (Base de Datos para el Análisis Dinámico del Empleo, or BADE) in Argentina; the Annual Social Information Report (Relação Anual de Informações Sociais, or RAIS) in Brazil; the Internal Revenue Service (Servicio de Impuestos Internos, or SII) in Chile; and Superfinanciera in Colombia.[8] Castillo et al. (2011) consulted BADE in their evaluation of Argentina's PRE program; and both Ribeiro and De Negri (2009) and De Negri et al. (2011) gathered data from RAIS for their evaluations of the loan policies of the National Bank of Social and Economic Development (Banco Nacional de Desenvolvimiento Econômico y Social, or BNDES) and the PDB programs in Brazil, respectively. Finally, Arráiz, Henríquez, and Stucchi (2011) referred to the SII database for the evaluation of Chile's Supplier Development Program (Programa de Desarrollo de Proveedores).

Primary data

When there is no secondary data available, primary data should be collected. The main advantage of being able to collect primary data is that the questionnaire can be made to measure. The disadvantage is cost, as well as the fact that these data tend to cover only a short period of time.[9] A sample questionnaire should be designed, together with an established plan of activities for the PDB program evaluations that require primary data.[10]

---

[8] The Employment and Business Dynamics Observatory (Observatorio de Empleo y Dinámica Empresarial, or OEDE), part of Argentina's Ministry of Work, Employment, and Social Security, created and administers BADE. Brazil's Ministry of Work and Business (Ministério do Trabalho e Emprego, or MTE) administers RAIS. Colombia's Financial Superintendance (Superintendencia Financiera) maintains Superfinanciera's database. Both access to, and use of, these databases are limited and regulated, according to regulations relating to statistical confidentiality, which are applied by respective administrative authorities.

[9] While it is possible to obtain accounting data for businesses from previous years, this strategy may not be entirely effective since, sometimes, part of the information is not taken from accounting registers, or it is difficult to extract from old registers. Moreover, not all firms have adequate accounting systems, especially micro- and small enterprises.

[10] Sample design is one of the most important activities in any study. In particular, the following elements (at least) should be addressed: (i) the unit of analysis and the strategy to define it, (ii) the selection strategy and the sample size, and (iii) the data-gathering plan. As in most cases, a pilot test should be implemented before a base and monitoring survey can be initiated. Finally, all data-gathering exercises should have a schedule of activities (agreed by all the stakeholders involved), within which the dates for each activity, and the stakeholders responsible, are stipulated.

METHODOLOGICAL STANDARDS

The key element in any evaluation is to construct a credible counterfactual that accurately attributes the results of the policy intervention under evaluation. In particular, there are experimental and quasi-experimental methods for evaluating PDB programs.

This section presents a general description of some of the most commonly applied methodologies. The first subsection examines the experimental design method, currently considered to be the "gold standard" for impact evaluations. Even in those cases in which a complete experimental design is not viable, this often becomes the benchmark for comparison with other methods. In the second subsection, different quasi- and nonexperimental methods are discussed, which can be applied whenever an experimental design is not viable.

The "Gold Standard": Experimental Design

Impact evaluation literature describes the experimental design as being accorded a special status. This type of design is based on randomly dividing a representative sample into a treatment group and a control group. This ensures that there is equilibrium between the treated and untreated units with regard to the average observable and unobservable characteristics. The groups thus become comparable and the selection bias can be eliminated.[11]

Apart from their proven efficacy in solving the problem of the missing counterfactual, experimental designs have other practical advantages. First, randomization allows the average impact of a program to be calculated as a simple difference in means between the treated and control groups, without recourse to the sophisticated econometric techniques necessary in nonexperimental contexts to allow for different types of bias.

Second, randomization can reduce data requirements vis-à-vis other nonexperimental techniques, due to the estimation of the average program impact. This random assignment only requires the post-treatment outcomes for each group, as well as a handful of ex-ante characteristics, to verify that randomization has been successful.

Of course, this does not imply that an efficient database is not an essential requirement for experimental evaluations; the more data available, the more accurate and encompassing the evaluation. For example, gathering data for many years after treatment can help establish a program's long-term effects. Likewise, a good supply of pretreatment outcome data, variables, and other observable factors can significantly improve the accuracy of the estimated impacts, which is of key concern in studies with small sample sizes.

[11] Consequently, it is possible to solve the fundamental problem of casual inference by using a randomly selected control group to calculate the counterfactual result of the treatment group.

Although randomization is becoming the widespread approach for evaluating the impact of public policy in sectors, such as development and labor economics (see, for example, Banerjee and Duflo, 2009), it has yet to be applied to the evaluation of PDB programs. One possible explanation is that it is unlikely that these programs fulfill the criteria (i.e., excess demand) that make a random assignment possible.[12] In general, randomized experiments for evaluating public intervention take advantage of high demand for these services and of supply-side limitations. Under these circumstances, an arbitrary selection of beneficiaries from a pool of possible candidates is a clear and transparent method of guaranteeing that all units (individuals, businesses, etc.) have the same opportunity to participate.

Banerjee and Duflo (2004) present an experimental design for PDB programs. The authors make use of an exogenous variation, generated by a policy change in India, to establish whether or not the enterprises that received direct credits increased their production. The results showed a significant acceleration in the growth rate of sales and profits among the beneficiary enterprises. Another example by Cotler and Woodruff (2008), applies the differences established in the introduction of a new loans program, designed to serve the clients of the largest fast-food company in Mexico (Bimbo). This is done to identify the impact of credit on the results of small retailing firms in Mexico City. The authors discovered that the loans positively impacted the smaller firms, and negatively the larger ones. They claim that these outcomes are consistent with their hypothesis—that smaller enterprises experience greater capital returns and face greater credit constraints.

Quasi- and Nonexperimental Methods

In the absence of a random assignment, the preexisting differences between program participants and nonparticipants can generate biases that severely hamper the estimation of the programs impact. Selection bias is of significant concern, due to two possible sources. First, there could be an administrative bias (or program placement bias), which occurs when program administrators select participants on the basis of specific criteria that differentiate them from nonparticipants. Second, there could be a case of self-selection, which occurs when individuals have agreed whether or not to participate, according to a type of cost–benefit analysis that, again, could lead to significant differences between the pool of participants and nonparticipants.

[12] It is worth highlighting that excess demand is not a necessary condition for applying an experimental design. In effect, randomization is compatible with treating the entire eligible population. For example, randomization is normally used to divide the eligible individuals into different groups, and to arbitrarily assign the order in which they receive treatment, instead of whether or not they actually receive it. This will enable the aggregations, which are treated later, to be used as control groups for those aggregations that were treated earlier. However, certain program characteristics, such as the type of project and the number of applicants, might mean that this type of randomization might not be politically or ethically feasible in some cases, while there remains the need to carry out an impact evaluation. Fortunately, there are numerous nonexperimental techniques that have been created to rep e random assignation as a way of estimating the impact of public programs.

In practice, it is highly likely that there will be a combination of both selection biases: in general, all public interventions have a target population, such as SMEs, young researchers willing to study abroad, or farmers willing to introduce new technologies. Within this target population, individuals or enterprises can decide whether to participate or not. Consequently, a simple preexisting difference in the values between the treated and untreated groups can affect the estimation of program impact and make it inaccurate.

To address this problem, an initial attempt to control the factors that generate selection bias should be made. A few adopted techniques include the following: regression methods, propensity score matching (PSM), difference-in-differences (DD) methods, and fixed effects (FE) methods. A second approach, represented by Instrumental Variables (IV) and the Regression Discontinuity (RD) design, consists of analyzing the specific characteristics of the assignment principles, in order to reproduce the experimental setting.

Regression methods and propensity score matching

As previously mentioned, impact can be calculated as the difference in value between the treated and the untreated groups, within an experimental design program. In turn, this can be equivalent to running a linear regression of the outcome of interest against a constant and a binary variable that indicates the treatment status (treated/untreated). In nonexperimental settings, this regression becomes inadequate, due to the biases previously referred to. However, if all the variables affecting both the treatment status and the outcomes are obvious, it becomes possible to implement control by adding these variables to the linear regression.[13]

To understand how Propensity Score Matching (PSM) works, suppose that treated and untreated individuals only differ by a single variable, X. Thereafter, the matching estimator assigns a unit of comparison to each treated individual with an untreated individual that has the most similar value to X. In this case, the effect of the treatment can be calculated as an average of the differences between the treated units and these units' nearest untreated neighbors in terms of their values of X.[14]

Currently, PSM appears to be the preferred approach in the evaluation of PDBs. For instance, Aivazian, Mazumdar, and Santor (2003) conclude that the World Bank's Small and Medium Scale Industry Program in Sri Lanka has contributed to reducing credit constraints and increasing investment levels in the enterprises that have received subsidies. However, this effect has been rather limited due, to a large

[13] The key assumption here is one that can control, explicitly, for all relevant variables, usually referred to as Conditional Independence Assumption (CIA) or Selection on Observables.

[14] However, when various factors differ between the groups, the idea of closeness is not clearly defined, given that individuals might be similar in some aspects and different in others. To overcome this dimensionality problem, Rosenbaum and Rubin (1983) show that, if all the relevant factors that determine program participation are known, the matching approach between treated and untreated individuals can be conducted. This is based on the conditional probability of participation or the propensity score, which represents the probability of participating in the program for a given value of the vector of characteristics X.

extent, to the relatively small quantity of resources employed for this purpose. Another conclusion arising from the mentioned analysis is that a public guarantee considerably lowers the cost of loans for SMEs.

Difference-in-differences and fixed effects models

DD models arise in the context of "natural experiments." In other words, they arise in situations in which treatment and control groups exist, but a researcher does not design them—rather they appear naturally. Studies often use these models to evaluate the impact of aggregate-level policy changes.[15]

The DD model consists of the application of a double difference. It compares the changes over time in the variable of interest (such as sales or productivity) between a beneficiary population of a program (treated group) and a nonbeneficiary population (comparison group).[16]

The identification assumption, which determines the DD and FE models, is that there are no unobserved factors that vary over time, nor are there any that affect the status of the treatment or outcome. In other words, relevant unobserved factors remain constant over time. Therefore, DD and FE models require that, in the absence of a treatment, it is assumed that the two groups (treated and control) would have the same trends.

The most commonly used approach is to apply the DD method to the databases, combined with PSM, in order to ensure a similarity between participants and nonparticipants. This approach works as follows: when there is available data for the years prior to the program, it is possible to apply PSM to establish nonbeneficiaries with the same ex-ante trends as beneficiaries in the outcome variables. When beneficiaries and nonbeneficiaries with the same characteristics are compared prior to program implementation, including the trends in the outcome variables, it is easier to assume the equilibrium of trends in the absence of the program. Thus, the combination of DD and PSM is a powerful procedure for obtaining effective impact estimations of PDB programs.[17]

For example, Zecchini and Ventura (2006), apply a DD approach to show that public guarantee funds for SMEs, in Italy, increased the credit that these enterprises received from the banking system. Based on

[15] For example, one of the most cited papers that applied this technique is that by Card and Krueger (1994), in which a change of legislation in New Jersey was analyzed to assess the impact of minimum salaries on employment, using Pennsylvania as a comparison group.

[16] This method can easily be extended to multiple groups and time periods, as well as to include control variables (e.g., Imbens and Wooldridge, 2009). Furthermore, the DD estimator can be adapted to cases in which the treatment is assigned at the individual level; this will overcome one of the most significant drawbacks of regressive and matching estimators, given that it allows for the control of selecting the unobserved factors, as long as they are constant over time. Thus, the DD method is an example of a fixed-effect (FE) estimator, which assumes that any unobserved heterogeneity that influences program participation and outcomes is fixed throughout the recorded period.

[17] This procedure comprises three stages: (i) calculate the pretreatment propensity score, (ii) define a common base for businesses through matching, and (iii) utilize a fixed effects model on this base. Heinrich, Maffioli, and Vázquez (2010) present directives for the application of this method, and various authors have carried out evaluations, based on its application.

this result, the authors conclude that, due to the relatively low cost and the State's high capacity to mobilize private capital, guarantee schemes are an effective instruments for promoting SME financing, as long as the focus is placed on those enterprises with the most significant financial constraints.

Also applying a DD methodology, Paravisini (2008) analyzes the effect of a loan program, using World Bank funding, for small Argentine enterprises. He observed that 93 cents out of every dollar invested would have reached the businesses in any case. This outcome suggests that banks implement programs targeting defined beneficiaries to reduce the cost of loans, without substantially increasing the amount of loans approved.

Finally, Bach (2011) demonstrates that the French loan program, the Industrial Development Savings Account (Compte pour le Développement Industriel, or COVEDI), does improve credit flow to small enterprises in France. The findings reveal that access to subsidies considerably augment the financing of loans to businesses. However, Bach concludes that this does not lead to a significant substitution between the subsidized and unsubsidized financing channels, which could be interpreted as financial constraints.

Furthermore, Hall and Maffioli (2008) present a summary of the empirical evaluations in Latin America. Their study reveals that credit programs usually have positive effects on intermediate outcomes, such as when allocating funds for R&D, vocational training, and the introduction of new quality control processes and procedures, especially in developing countries (López Acevedo and Tan, 2010). However, evidence of any impact on performance outcomes over the longer term, such as on sales, exports, employment, labor productivity, or PTF, varies.

As an example, Chudnovsky et al. (2010) analyze Argentina's Technological Fund (Fondo Tecnológico Argentino, or FONTAR), a program designed to improve R&D and technological development through nonreturnable payments. Although the authors establish positive effects that range from a 57 percent to a 79 percent increase in investment in innovation, they find no relevant impact on labor productivity or in sales of new products. Similarly, with regard to Brazil's National Technological Enterprise Development Support (Apoio ao Desenvolvimento Tecnológico da Empresa, or ADTE), a program of subsidies for R&D and technological development, Ribeiro and De Negri (2009) observe an increase of between 50 percent and 90 percent in R&D expenditure, but they find little impact on sales, employment, or labor productivity. Benavente, Crespi, and Maffioli (2007) examine Chile's National Fund for Technological and Productive Development (Fondo Nacional de Desarrollo Tecnológico y Productivo, or FONTEC), which is designed to promote the transfer and development of technologies and to support R&D. The authors calculate an estimated 40 percent increase in sales and a 3 percent increase in export concentration, but they do not find an impact on labor productivity.

Other examples, which are closely related to PDB programs targeting Latin America, are those examined by Ribeiro and De Negri (2009), De Negri et al. (2011), and Eslava, Maffioli, and Meléndez Arjona (2012a and 2012b). For example, De Negri et al. (2011) analyze the effectiveness of public credit lines to boost performance in Brazilian enterprises. The authors focus on the impact of credit lines, administered by BNDES and by the Brazilian Innovation Agency (Agencia Brasileña de Innovación), on the growth in employment, labor productivity, and exports. They apply a combination of panel data, developed by the Institute of Applied Economic Research (Instituto de Pesquisa Econômica, or IPEA), which gathers information about performance at the firm level and access to credit lines. This particular data setting allows them to apply quasi-experimental techniques to control selection biases when calculating the impact of access to public credit. The basis of their calculation includes a DD strategy, which they complement with matching methods in order to verify impact robustness. The results consistently demonstrate that access to public credit lines does have a significantly positive impact on growth in employment and exports. Additionally, they do not detect significant effects on productivity. It is interesting to note from the conclusions that the impact on exports is owed, primarily, to an increase in the volume of exports by exporting firms; however, there is no significant effect detected regarding the firms' themselves becoming possible exporters.

Eslava, Maffioli, and Meléndez Arjona (2012a) analyze the impact of lending activity on business performance in Colombia's Business Development Bank (Banco de Desarrollo Empresarial, or BANCOLDEX). The use data, gathered over several years, to evaluate the loans made by BANCOLDEX and the performance of manufacturing establishments with 10 or more employees. According to a combination of matching techniques and fixed-effect panel regressions to address the selection biases, they find significant increases in production (24 percent), employment (11 percent), investments (70 percent), and productivity (approximately 10 percent) over the four years following the first BANCOLDEX loan. However, the impact on investments, production, and productivity is derived, primarily, from long-term loans made by BANCOLDEX.

Similarly, Eslava, Maffioli, and Meléndez Arjona (2012b) examine the impact of BANCOLDEX on access to credit. For this purpose, they use a database containing key characteristics of all the loans administered to enterprises in Colombia, including data relating to the financial intermediary, through which the loan was arranged, and whether or not it was financed by BANCOLDEX. The authors compare BANCOLDEX loans with loans from other sources, and they study the impact of receiving a BANCOLDEX loan, based on the prior credit history of an enterprise. To address the problem of selection bias, they apply a combination of controlled models using fixed effects and matching techniques. The conclusions demonstrate that the credit terms relating to BANCOLDEX loans are characterized by lower-than-average interest rates, larger-than-average amounts loaned, and longer-than-average payment terms. However, the effect of the longer-than-average payment term could take up to two years before it can be observed. Finally, the conclusions present a demonstration effect: businesses with access to BANCOLDEX credit are capable of significantly expanding the number of intermediaries with which they share credit relations.

The instrumental variables approach

The institutional variables (IV) approach consists of exploiting certain features of the design and institutional setting of a program in order to find the source of an exogenous variation that best reproduces the conditions of a random trial. Although the theoretical aspects of the IV method may be complex, the perception is simple: it relates to establishing a variable that can influence the probability of participation, but that is not related to other variables that influence the outcome in any way. In other words, an instrumental variable (or, simply, an instrument) is a variable that influences the treatment status, but can also be considered to be "as good as random."

To illustrate how this method works, suppose a PDB program seeks to increase the sales of beneficiary firms in order to adopt new technologies that would enable them to access international markets. In this case, it can be anticipated that some unobserved factors that determine participation by businesses in the program (e.g., entrepreneurs' capacity and motivation) could also have some influence on sales capacity. In this context, a comparison between beneficiaries and nonbeneficiaries would not only reflect the project's impact, but also the intrinsic characteristics of the participating firms.[18]

Although the IV method is an effective tool for evaluating the impact of PDB programs, it is not always easy to find an instrument once a project has been designed. In this case, an effective approach is to implement the project with a so-called "random stimulus," an incentive that arbitrarily persuades firms to participate in the credit program through various mechanisms. For instance, flyers that are distributed to some firms can be a means of showing that a program can reduce the cost of credit. It is, thus, reasonable to believe that the firms that received the flyers are more likely to participate in the program compared with those that were not included in the distribution. Given that the incentive was randomly distributed, there is no reason to suppose that the promotion mechanism is correlated with the outcomes variable, which thereby makes it a reasonable instrument.[19]

Given the difficulties to identify effective instruments, most literature adopting this particular method has concentrated on doing so through random stimulus. One of the best known examples of this approach is presented in Karlan and Zinman (2008). These authors test the hypotheses of inelastic demand for microcredits using data from a randomized field experiment carried out in South Africa. The data include information about previous borrowers from an important for-profit institution that provided

[18] For more details about the characteristics of the IV method and its limitations, see Angrist and Pischke (2009).

[19] Another limitation of the IV approach is that it can only estimate the Local Average Treatment Effect (LATE), which means that its results are relevant only for those enterprises whose behavior is affected by the instrument (Imbens and Angrist, 1994). For example, in the previous case, the results are valid only for those enterprises that participate in the program because of the reduced costs and that, if there were no discount, would not participate. However, the results are not valid for the enterprises that do exploit the discount, but would participate even if there were no discount. Furthermore, it is important to consider the problem of instrument weakness (e.g., Bound, Jaeger, and Baker, 1995): when an instrument is weak, it can generate biases and increase the standard errors of the IV estimation.

micro-consumer loans to poor workers. Karlan and Zinman first calculate the price elasticity of demand for consumer loans by offering, through mailing, a random interest rate to each one of the more than 50,000 previous customers. Subsequently, they calculate the time period elasticity by, again, mailing with randomly assigned suggestions to draw a selection of certain time periods.

Although this type of design has not been fully implemented in the study of PDBs, it can be easily adapted. For example, the evaluation plan for a loan that the Inter-American Development Bank (IDB) recently provided to a PDB in Mexico adopts a random stimulus design that includes a random assignment of publicity campaigns concerning new financial products in a given region. Similarly, a project in Ecuador, supported by the IDB, intends to randomize information about the availability of a line of credit to the passive clients of microcredit institutions, the latter relating to the national Tier 2 microcredit fund. In both cases, if the publicity campaigns prove sufficiently effective to influence the acceptance of lines of credit, they could be used as a powerful IV approach for the evaluations of these lines of credit.

Regression discontinuity

Regression discontinuity (RD) is another powerful approach for identifying the impact of a PDB program on firm performance. It is based on the idea that, in a world highly governed by regulation, some of these regulations are arbitrary and, thus, provide natural experiments. In this framework, the approach measures the average effect of a treatment on the discontinuity that determines which enterprises are assigned to the treatment (receive the program) and which ones are assigned to the control group (do not receive the program). The perception behind this approach is that the treated units just above the cut-off point are very similar to the control units just below it, which enable the results to be compared without incurring any bias. Regression discontinuity designs are presented in two forms: sharp and fuzzy. The former are based on a selection of observables, whereas the latter suggest the use of instrumental variables (Angrist and Pischke, 2009).

A good example of a sharp regression discontinuity is a PDB program that provides lines of credit for firms, according to their specific credit history: those that are found above the threshold can benefit from the program and those located below form part of the control group. This scheme has the advantage that the credit rating can be determined outside of the financial institution providing the loan, by a central authority or other entity, thereby enhancing the transparency of the selection process.

A fuzzy regression discontinuity differs from a sharp one in that there is no single value that perfectly determines the treatment and control groups. Rather, there is a variable that influences the probability of treatment. In this case, the variable that influences program participation can be used as an instrumental variable to predict the treatment. Since this type of regression discontinuity can be seen as a special element within the IV model, its advantages and limitations are the same as the latter.

For example, Bubb and Kaufman (2009) argue that investment banks in the United States adopted issuer selection rules (with cutoff points), based on credit ratings, in response to the Fannie Mae and Freddie Mac subscription directives. The authors offer a simple model that rationalizes this general rule of origin, and suggest that the increase in defaulted loans is not sufficient proof that securitization has led to lax screening. They analyze the data relating to the loans in detail and, based on a regression discontinuity design, they discover that the evidence is, on the one hand, inconsistent with the theory of the automatic securitization rule and, on the other hand, consistent with the theory of the automatic rule of origin. They also document an increase in the number of loans and in the rate of defaults at the credit rating cutoff point, while there is no corresponding increase in the securitization rate. Finally, they conclude that the cutoff point rules, based on credit ratings, provide evidence that the major securitizers are, to a certain point, capable of modifying the behavior of the investment banks.

Furthermore, on the basis of a regression discontinuity framework, Skiba and Tobacman (2007) benefit from a credit-rating process, used to approve or deny payday loan applications, in order to study the causal impact of access to these loans on payday loan uptake, bankruptcies, and misdemeanors. They present evidence that those employees who were approved for payday loans requested on average 8.8 more payday loans, until their debt reached US$2,400 (with an additional US$350 in financing charges). Based on this evidence, it is unlikely that the behavior associated with payday loans is determined by temporary shocks to consumer needs. Approval of these payday loans reduces the incidence of short-term collateral loans, but this reduction dissipates after a few weeks.

Structural models

When selecting the best empirical approach for analyzing economic data, it is key for an analyst to establish which questions need answers. Explicit economic models facilitate the formulation of economic questions. Defenders of nontheoretical approaches to analyzing economic data suggest randomization as a model, and invoke the IV, PSM, or regression discontinuity methods as substitutes for randomization. However, even perfectly executed randomizations fail to respond to all economic questions. There are clear examples that show that structural models generate more information on preferences than experiments do.

Structural models seek to utilize data to define the parameters of an underlying economic model, based on individual choice models, or on the aggregate relationships deriving from them. Structural calculus enjoys a long tradition in economics, but it is only recently that better and wider databases have become available, in parallel to more powerful computers, perfected modeling methods, faster computing techniques, and new econometric models (e.g., those mentioned above), which have enabled significant progress. Based on a group of assumptions, these kinds of models permit the calculation of the contribution of a given policy change to the economy. The works of Todd and Wolpin (2006), Keane and Wolpin (1997), and Attanasio, Meghir, and Santiago (2010) present examples of this methodology, although not necessarily applied to PDBs.

RESOURCES

To be comprehensive, an evaluation plan must clearly identify the resources needed for its execution, which include: (i) choosing the evaluation team and defining the respective responsibilities and tasks, (ii) setting the budget and the work plan, and (iii) identifying the source of financing.

The Evaluation Team

Ideally, a combination of external evaluators and expert managers should make up the evaluation team (in other words, professionals who are involved in implementing the program). The external evaluators guarantee both greater independence, because they are much more involved in the success of the evaluation than in the success of the program (this way, a high degree of objectivity and credibility can be obtained), and higher concentration, because they are exclusively dedicated to the evaluation, rather than to the implementation of the project.

The professionals involved in implementation are crucial to ensuring that (i) the program's objectives and its execution mechanisms are clearly understood; (ii) there is easy and timely access to data and information about the project; and (iii) there is a fluid dialogue with the authorities and greater recognition for the results of the evaluation.

The evaluation plan must specify the capacities and technical knowledge required for a successful evaluation. Although it is difficult to generalize the exact composition of an ideal team (which depends on the program and the available resources), the team should, at least, be able to collectively offer knowledge of (and experience in) the following:

1. Design of evaluations, including the evaluation method and interpretation of the statistical power.
2. Negotiation of the evaluation design with the main stakeholders.
3. Design and administration of data gathering, which ranges from designing the questionnaire, developing sample plans, and collecting information in contexts that are relevant to the project to be evaluated.[20]

[20] If the project team decides to contract external individuals or firms to contribute to managing the impact evaluation, it would be useful to consider the evaluation and data gathering as two separate elements that can be executed, theoretically, by two separate individuals or agencies. For this model to succeed, the two individuals or agencies have to work cohesively, and it is recommended that the entity that solicits the evaluation play a role in its coordination.

4. Design of systems to protect the integrity of the evaluation.
5. Review of the statistical analysis for estimating the impacts.
6. Presentation of the conclusions to a wide spectrum of audiences, including academics and policy-makers.

Irrespective of the exact composition of the team, the management of the program and implementation of the evaluation are interrelated, and should not function as independent and separate activities.

Financial Resources

The evaluation plan should include a detailed calculation of the resources necessary to finance the evaluation. Therefore, the design of the plan should include a work plan that describes who will carry out what activity, and when. It is recommended to allocate a budget to each activity, in order to accurately define the financing needs, mobilize resources, and ensure that available funding levels are adequate (Gorgens and Kusek, 2009). It is also important to distinguish the cost between monitoring and evaluation activities. Table 3.1 presents an example of a budget for an impact evaluation work plan.

A significant component related to the cost of any evaluation is the combination of resources needed for data gathering. A recent study regarding World Bank impact evaluations concludes that more than half of the resources earmarked for an evaluation go toward data collection (see Gertler et al., 2011). The cost of compiling information depends on various factors. However, the two key factors are sample size and the number of data-gathering rounds. It is, therefore, essential to carefully consider these two factors during the early phases of evaluation design.

TABLE 3.1: IMPACT EVALUATION AND WORK PLAN BUDGET

DESCRIPTION OF TASKS	YEAR 1				YEAR 2				YEAR 3				PERSON RESPONSABLE	COST (US$)	SOURCE OF THE RESOURCES
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4			
I. Impact Evaluation															
Staff															
Program evaluation personnel (evaluation manager, etc)															
International and local consultants (chief researcher)															
Research assistant															
Statistics expert															
Field study coordinator															
Trips															
International and domestic flights															
Local land transport															
Expenses (hotels and sundries)															
Data Gathering[a]															
Instrument Design															
Pilot test															
Training															
Travel and expenses															
Survey material and equipment															
Printed questionnaires															
Fieldwork staff															
Survey staff															

continued →

TABLE 3.1: IMPACT EVALUATION AND WORK PLAN BUDGET *(continued)*

DESCRIPTION OF TASKS	YEAR 1				YEAR 2				YEAR 3				PERSON RESPONSABLE	COST (US$)	SOURCE OF THE RESOURCES
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4			
Supervisors															
Transport (vehicles and fuel)															
Drivers															
Data Digitalization															
Data cleaning and digitalization															
Data Analysis and Dissemination															
Workshops															
Documents, reports															
Other															
Office space															
Communications															
Software															
Impact Evaluation Report															
Final report															

[a] Calculation of the cost of data-gathering should reflect assumptions such as the number of rounds necessary, how long data-gathering will last, number of communities in the sample, number of households per community, questionnaire length, duration of field trips, etc.

CONCLUSIONS

To be successful, the design of an impact evaluation of a PDB program must incorporate the following key aspects. First, it must account for the externalities of the beneficiary firms, given that economies of scale can arise. Moreover, impacts of PDB programs can take some time before they become apparent. Therefore, for any impact evaluation, it is fundamental to establish the distribution of a program's effects over time. Furthermore, it could be the case that firms take differing amounts of credit from the program, or that they participate by taking out loans at different times. It is, thus, vital to consider treatment intensity and dosage effects. Finally, two additional elements should be considered for evaluating a PDB program: (i) the potential multiple treatments that arise, whenever a beneficiary firm accepts additional credit from other institutions in the market; and (ii) the heterogeneous nature of the impact, when there are varying effects for different beneficiary groups.

Second, in an analysis of the effectiveness of a PDB program, the use of quality data can make all the difference in the evaluation outcome. The data used should be available, accurate, and reliable. In this sense, the quality of the data, whether primary or secondary, is also an indispensable element for a successful evaluation.

Finally, it is possible to apply different methodologies—both experimental and quasi-experimental—to the evaluation of PDB programs. As a general rule, an experimental methodology guarantees the quality of both the counterfactual and the outcomes. However, the general challenge is to select the methodology that best suits the particular circumstances of each program.

In Latin America and the Caribbean, despite the fact that the empirical evidence is still scarce, researchers have begun to document the effectiveness of PDB-related programs. Those impact evaluations are based on rigorous methodologies and reliable data and, in general, seek to control for several of the previously mentioned relevant factors. However, a clear—but also stimulating—challenge remains in the future, given the wide variety of PDB programs and the methodologies currently available.

BIBLIOGRAPHICAL REFERENCES

Aivazian, V., D. Mazumdar, and E. Santor. 2003. "Financial Constraints and Investment: Assessing the Impact of a World Bank Loan Program on Small and Medium-Sized Enterprises in Sri Lanka." Working Paper No. 2003–37. Ottawa, Canada: Bank of Canada.

Angrist, J. D. and J-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Arráiz, I., F. Henríquez, and R. Stucchi. 2011. "Impact of the Chilean Supplier Development Program on the Performance of SME and their Large Firm Customers." Working Paper No. OVE/WP-04/11. Washington, DC: IDB.
Available at: http://idbdocs.iadb.org/WSDocs/getdocument.aspx?docnum=36182301.

Attanasio, O., C. Meghir, and A. Santiago. 2010. "Education Choices in Mexico: Using a Structural Model and Randomized Experiment to Evaluate Progress." London: Institute of Fiscal Studies. Unpublished.

Bach, L. 2011. "Are Small Businesses Worthy of Financial Aid? Evidence from a French Targeted Credit Program." *Review of Finance* (2013) doi: 10.1093/rof/rft022.

Banerjee, A. V. and E. Duflo. 2004. "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program." CEPR Discussion Paper No. 4681. London: CEPR.
Available at: http://ideas.repec.org/p/cpr/ceprdp/4681.html.

———. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics, Annual Reviews*, 1(1): 151–78.

Bartelsman, E. J. and M. Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature* 38(3): 569–94.

Benavente, J. M., G. Crespi, and A. Maffioli. 2007. "Public Support to Firm-Level Innovation: An Evaluation of the FONTEC Program." OVE Working Paper No. OVE/WP-05/07. Washington, DC: IDB.
Available at: http://idbdocs.iadb.org/WSDocs/getdocument.aspx?docnum=1323102.

Bound, J., D. A. Jaeger, and R. Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variables Is Weak." *Journal of the American Statistical Association* 90: 443–50.

Bubb, R. and A. Kaufman. 2009. "Securitization and Moral Hazard: Evidence from a Lender Cutoff Rule." Public Policy Discussion Document No. 09–5. Boston, MA: Federal Reserve Bank of Boston.

Caliendo, M. and S. Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Wiley Blackwell* 22(1): 31–72.

Card, D. and A. B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84(4): 772–93.

Castillo, V., A. Maffioli, S. Rojo, and R. Stucchi. 2011. "Innovation Policy and Employment: Evidence from an Impact Evaluation in Argentina." IDB Technical Notes No. IDB-TN-341. Washington, DC: IDB. Available at: http://www.iadb.org/en/publications/publication-detail,7101.html?id=29206.

Chudnovsky D., A. López, M. Rossi, and D. Ubfal. 2010. "Evaluating a Program of Public Funding of Private Innovation Activities: An Econometric Study of FONTAR in Argentina." OVE Working Paper No. OVE/WP-03/10. Washington, DC: IDB. Available at http://www.iadb.org/en/publications/publication-detail,7101.html?id=19292.

Cotler, P. and C. Woodruff. 2008. "The Impact of Short-Term Credit on Microenterprises: Evidence from the Fincomun-Bimbo Program in Mexico." *Economic Development and Cultural Change* 56: 829–49.

De Negri, J. A., A. Maffioli, C. M. Rodríguez, and G. Vázquez. 2011. "The Impact of Public Credit on Brazilian Firms." IDB Working Paper No. IDB-WP-293. Washington, DC: IDB. Available at: http://www.iadb.org/en/publications/publication-detail,7101.html?id=31526.

Eslava, M., A. Maffioli, and M. Meléndez Arjona. 2012a. "Second-Tier Government Banks and Firm Performance: Micro-Evidence from Colombia." IDB Working Paper Series No. IDB-WP-294. Washington, DC: IDB. Available at: http://www.iadb.org/en/publications/publication-detail,7101.html?id=31446.

———. 2012b. "Second-Tier Government Banks and Access to Credit: Micro-Evidence from Colombia." IDB Working Paper Series No. IDB-WP-308. Washington, D.C.: IDB. Available at: http://www.iadb.org/en/publications/publication-detail,7101.html?id=36046.

Frölich, M. and B. Melly. 2009. "Unconditional Quantile Treatment Effects under Endogeneity." IZA Discussion Paper No. 3288. Bonn, Germany: IZA.

Gertler, P. J., S. Martínez, P. Premand, L. B. Rawlings, and C. M. J. Vermeersch. 2011. *La evaluación de impacto en la práctica*. Washington, DC: World Bank.

Gorgens, M. and J. Z. Kusek. 2009. *Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit*. Washington, DC: World Bank.

Hall, B. and A. Maffioli. 2008. "Evaluating the Impact of Technology Development Funds in Emerging Economies: Evidence from Latin America." NBER Working Paper No.13835. Cambridge, MA: NBER.

Heckman, J. J., J. Smith, and N. Clements. 1997. "Making the Most Out of Program Evaluations and Social Experiments: Accounting for Heterogeneity in Program Impacts." *The Review of Economic Studies* 64(4): 487–535 (Special Issue: Evaluation of Training and Other Social Programs).

Heinrich, C., A. Maffioli, and G. Vázquez. 2010. "A Primer for Applying Propensity-Score Matching." IDB Technical Note No IDB-TN-161. Washington, DC: IDB. Available at: http://www.iadb.org/en/publications/publication-detail,7101.html?id=8240.

Henderson, V., A. Kuncoro, and M. Turner. 1995. "Industrial Development in Cities." *Journal of Political Economy* 103(5): 1067–90 (October).

Holland, P. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81(396): 945–60.

Hulten, Ch. R. 2001. “Total Factor Productivity: A Short Biography.” In: Ch. Hulten, E. R. Dean and M. Harper (eds.), *New Developments in Productivity Analysis, Studies in Income and Wealth*. Chicago, Illinois: University of Chicago Press.

Imbens, G. W. and J. D. Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62(2): 467–75.

Imbens, G. W. and J. M. Wooldridge. 2009. “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature* 47(1): 5–86.

Karlan, D. S. and J. Zinman. 2008. “Credit Elasticities in Less-Developed Economies: Implications for Microfinance.” *American Economic Review* 98(3): 1040–68..

Keane, M. P. and K. I. Wolpin. 1997. “The Career Decisions of Young Men.” *The Journal of Political Economy* 105(3): 473–522.

López Acevedo, G. and H. Tan. 2010. *Impact Evaluation of SME Programs in Latin America and the Caribbean*. Washington DC: World Bank.

Marshall, A. 1920. *Principles of Economics* (revised edition). London: Macmillan (reissued by Prometheus Books).

Paravisini, D. 2008. “Local Bank Financial Constraints and Firm Access to External Finance.” *Journal of Finance* 63(5): 2161–93.

Ribeiro, E. P. and J. A. De Negri. 2009. “Public Credit Use and Manufacturing Productivity in Brazil.” Unpublished.

Rosenbaum, P. and D. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70: 41–55.

Skiba, P. M. and J. Tobacman. 2007. “Measuring the Individual-Level Effects of Access to Credit: Evidence from Payday Loans.” Job Market Paper. Berkeley, CA: University of Berkeley.
Available at: http://www.clevelandfed.org/research/conferences/2007/october2/skibajmpaper.pdf.

Todd, P. and K. Wolpin. 2006. “Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico.” *American Economic Review* 96(5): 1384–1417.

Van Biesebroeck, J. 2009. “Disaggregate Productivity Comparisons: Sectorial Convergence in OECD Countries.” *Journal of Productivity Analysis* 32(2): 63–79.

Volpe Martincus, C. and J. Carballo. 2008. “Is Export Promotion Effective in Developing Countries? Firm-Level Evidence on the Intensive and the Extensive Margins of Exports.” *Journal of International Economics* 76(1): 89–106.

Zecchini, S. and M. Ventura. 2006. “Public Credit Guarantees and SME Finance.” Institute for Studies and Economic Analyses (ISAE) Working Paper No. 73. Rome, Italy: ISAE.
Available at: http://lipari.istat.it/digibib/Working_Papers/WP_73_2006_Ventura.pdf.

APPENDIX 3.1[21]

The idea of a counterfactual can be formalized using the Rubin Causal Model (RCM) (Holland, 1986), as follows: Y1 and Y0 denote the potential outcomes for an individual with and without treatment, respectively. The result Y observed for an individual is Y1 if the individual is treated and Y0 if not. The binary variable T shows the status of the treatment of the individuals, with T=1 for those that participate and T=0 for those that do not participate. The result can therefore be expressed as:

$$Y = Y_0.(1-T) + Y_1.T$$

In this context, Y0 is the counterfactual outcome for the units treated and Y1 is the result for the untreated ones. The impact of the program for the individual I, which cannot be observed, is defined as the difference between the two potential outcomes:

$$\delta_i = Y_{1i} - Y_{0i}$$

In general, impact evaluations focus on calculating the average effect of the treatment, rather than the individual effect. In practice, various "average effects" can be calculated.

First, the average treatment effect (ATE), which is the average impact of the treatment on the population as a whole:

$$ATE = E(\delta) = E(Y_1 - Y_0)$$

Second, the ATT is the average impact of the treatment on the treated population:

$$ATT = E(\delta | T = 1) = E(Y_1 - Y_0 | T = 1)$$

Third, the average effect on the untreated (ATU) is the impact that the program would have had on the population that did not participate in the program:

$$ATU = E(\delta | T = 0) = E(Y_1 - Y_0 | T = 0)$$

However, none of these parameters can be observed. For example, the ATT can be rewritten as:

[21] This Appendix is based on Heinrich, Maffioli and Vázquez (2010).

$$ATT = E(Y_1 \mid T=1) - E(Y_0 \mid T=1)$$

where the second term is not observable, given that it measures the average result that the treated population would have obtained without treatment. One possibility is to exchange the second term for $E(Y_0 \mid T=0)$, which is the average observed result for the untreated population. Therefore:

$$\Delta = E(Y_1 \mid T=1) - E(Y_0 \mid T=0)$$

$$\Delta = E(Y_1 \mid T=1) - E(Y_0 \mid T=1) + E(Y_0 \mid T=1) - E(Y_0 \mid T=0)$$

$$\Delta = ATT + SB$$

Where the final term is usually called the selection bias (SB). This term reflects the difference in the counterfactual between the individuals treated and the results observed in the untreated individuals. Unless the bias is zero (which is very unlikely in practice), econometric techniques will have to be used to correctly calculate the average impact of the program.