

Effectiveness of Interventions Aimed at Improving Women's Employability and Quality of Work

A Critical Review

Petra E. Todd

The World Bank
Poverty Reduction and Economic Management Network
Gender and Development Group
September 2012



Abstract

This paper examines the effectiveness of a variety of policy interventions that have been tried in developing and transition economies with the goal of improving women's employability and quality of work. The programs include active labor market programs, education and training programs, programs that facilitate work (such as childcare subsidies, parental leave programs and land titling programs), microfinance programs, entrepreneurship and leadership programs, and conditional cash transfer programs.

Some of these policy interventions were undertaken to increase employment, some to increase female

employment, and some for other reasons. All of these programs have been subjected to impact evaluations of different kinds and some also to rigorous cost-benefit analyses. Many were found to be effective in increasing women's quantity of work as measured by increased rates of labor market participation and number of hours worked. In some cases, the programs also increased women's quality of work, for example, by increasing the capacity for women to work in the formal rather than the informal sector where wages are higher and where women are more likely to have access to health, retirement, and other benefits.

This paper is a product of the Gender and Development Group, Poverty Reduction and Economic Management Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at petra@athena.sas.upenn.edu or ptodd@ssc.upenn.edu.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Effectiveness of Interventions Aimed at Improving Women's Employability and Quality of Work: A Critical Review

Petra E. Todd*

JEL: J21, J08, D04, J16

Keywords: labor supply, impact evaluation, gender, active labor market policies, training programs, family policies, microfinance, land titling programs, entrepreneurship programs, conditional cash transfers.

* Todd is Professor of Economics at the University of Pennsylvania, an Associate of the Population Studies Center at the University of Pennsylvania, NBER, and IZA. She wrote this paper as a consultant to the World Bank. Elena Bardasi, Jere Behrman, Emmanuel Skoufias and two anonymous reviewers provided very helpful comments. Marco Cosconati provided excellent research assistance.

Contents

1	Introduction	1
2	Alternative Evaluation Approaches	4
2.1	The Evaluation Problem and Key Parameters of Interest	4
2.2	Solutions to the Evaluation Problem	6
2.2.1	Randomization	6
2.2.2	Nonexperimental Estimators	9
3	Employment Creation and Job Training Strategies	11
3.1	Active labor market (ALMP) policies	11
3.1.1	ALMP programs in Latin America	12
3.1.2	ALMP programs in Transition Economies	19
3.1.3	ALMP programs in China	23
3.2	Sectoral policies: Commodity commercialization in Nepal	24
3.3	Summary	25
4	Programs to increase entrepreneurship, access to credit	27
4.1	Microcredit/grant programs in Asia	28
4.1.1	Group-based microcredit programs in Bangladesh	28
4.1.2	Gender-targeted grants in Sri Lanka	31
4.1.3	Microcredit programs in Pakistan	32
4.1.4	Microcredit in India	33
4.1.5	Expanding credit in Thailand: the Thai Million Baht Village Fund Program	34
4.2	Relaxing consumer credit in South Africa	35
4.3	Opening savings accounts for the poor in rural Kenya	36
4.4	Entrepreneurship, decision-making, leadership training in Peru	37
4.5	Summary	37
5	Programs that facilitate work, through lowering costs of work or improving working conditions	39
5.1	Childcare programs	40
5.1.1	Programs to increase availability of childcare in Latin America and Africa	40
5.1.2	Child-care subsidies in transition economies	43
5.2	Summary	46
6	Conditional Cash Transfer (CCT) programs	47
6.1	The Mexican <i>Oportunidades</i> program: Impacts on mothers' time use	47
6.1.1	Impact of PROGRESA on time use	48
6.1.2	Effects of <i>Oportunidades</i> in rural areas on alcohol abuse and domestic violence	50
6.1.3	Effects on marriage and fertility-related behaviors of adolescents	50
6.2	Nicaragua's Red de Proteccion Social program	51
6.3	Summary	52
7	Other types of programs	53
7.1	Family friendly policies in Western Europe	53
7.2	Community association programs in Kenya	54

7.3	Land titling programs in Peru and Argentina	55
8	Synthesis and Directions for Future Work	57
9	Appendix A: Nonexperimental Evaluation Estimators	71
9.0.1	Sources of bias in estimating $E(\Delta X, D = 1)$ and $E(\Delta X)$	72
9.1	Matching Methods	75

1 Introduction

Policy interventions in developing countries often aim to improve the employment prospects of low-income individuals and their families, particularly those facing severe labor market shocks or credit constraints. Some interventions specifically target women, under the view that women have an inequitable share of power in household decision-making, that women are more likely to face barriers to labor market entry, and/or that women spend resources more effectively than men, for example, on investment in children. Other policies, such as active labor market programs, do not explicitly target women but may affect men and women in different ways.

This paper examines the effectiveness of a variety of policy interventions that have been tried in developing and transition economies with the goal of improving women's employability and quality of work. The programs include active labor market programs, education and training programs, programs that facilitate work (such as childcare subsidies, parental leave programs and land titling programs), microfinance programs, entrepreneurship and leadership programs, and conditional cash transfer programs. The programs studied are heterogeneous, differing in scope, content, targeting and eligibility criteria, as well as in objectives. Common objectives are to reduce unemployment, increase wages, provide social protection, and/or increase women's empowerment in the household. The populations served by the programs are also heterogeneous in terms of demographics and labor markets. They reside in rural and urban settings in Latin America, Africa, Europe and Asia. Because the coverage of this survey is broad, we consider only a subset of relatively recent programs that have rigorous impact evaluations. Also, we focus on programs that aim to directly affect employment and earnings outcomes and do not consider programs oriented toward education or health, which may also influence labor market prospects. In reviewing the literature, we discuss the merits of alternative evaluation studies and attempt to synthesize results across multiple studies.

There are many questions that are potentially of interest in evaluating the effects of a social policy intervention. Foremost is the question of whether people affected by the policy or program intervention benefit from it, at least on average. Most of the evaluation literature, including most of the studies examined in this paper, focus on estimating the so-called average effect of treatment on the treated, which is the average program impact for people who were exposed to some treatment. The treatment may represent active participation in some program, such as a job training program, or passive exposure to some policy, such as being eligible for a

subsidy. Another question of interest is how program benefits are distributed across people. For example: What fraction of people experience a positive benefit from the intervention and how do the benefits vary according to the demographic characteristics of the participant? A third question that usually merits consideration is whether the program benefits outweigh the costs. A program that generates benefits that are less than its costs might be deemed unsuccessful, unless there are other compensating factors.¹ A fourth question is how program impacts and costs would differ if some features of the program were changed. For example, if a policy intervention provides childcare subsidies, we may want to know how an increase in the level of the subsidy would affect mothers' labor supply. Answering these types of questions requires assessing the effects of programs that have never been tried, by extrapolating from previous experience with an existing program. Relatedly, it might be of interest to explore how program benefits would vary if the program were extended to new segments of the population. For example, the program eligibility criteria might be relaxed to be more inclusive or the program may be introduced to new regions. Answering questions about the impacts of hypothetical treatments, such as changes in a subsidy level, usually requires a more fully specified behavioral model.

Most program evaluation studies compare the performance of a group that participates in a program to that of a group that does not participate. It is common, however, to target social programs selectively at families or individuals that are deemed most in need of them or that are likely to get the largest benefit from participating. For example, a family planning program might be targeted at high-fertility regions or an unemployment program at areas with high unemployment rates. In addition, individuals cannot usually be coerced to participate in programs and typically self-select into them, raising the possibility that only more motivated individuals or those that expect to benefit most from the programs participate. Selective targeting and self-selection into programs can promote efficient use of program resources; but these mechanisms also pose challenges for evaluating the impact of the program, because they generate differences between the groups that participate and do not participate in the program. Such differences need to be taken into account in evaluating the program's impact through a comparison of program participants and nonparticipants.

¹ For example, some training programs in the U.S. that are targeted at older displaced workers generate income benefits that are less than the cost of the program, but it may still be desirable to provide training and employment services for such workers if a high value is placed on their employment per se.

There are two main approaches for evaluating social programs in a way that addresses the problem of noncomparability between participants and nonparticipants. One approach, considered by many to be the gold standard, is to use a randomized experimental design.²

Under a randomized design, some fraction of individuals that satisfy program eligibility criteria are randomly excluded from the program and serve as the control group. Randomization ensures that the group that is offered the program is comparable to the group that is not offered the program along both observable and unobservable dimensions, which is the major virtue of experiments. There are, however, some potential limitations to randomized designs that will be discussed in Section 2.

An alternative evaluation approach is the nonexperimental approach, which uses nonexperimental data (sometimes called observational data) on program participants and nonparticipants and employs statistical methods to adjust for noncomparability between the groups. There is an extensive literature that develops nonexperimental methods for evaluating the impact of social interventions and assesses their efficacy. Section 2 and Appendix A describe some commonly used methods. The statistical estimator used and the quality of the data available for modeling program participation decisions are important factors affecting the reliability of inferences from nonexperimental evaluation studies.

This paper is organized as follows. Section 2 provides an overview of experimental and nonexperimental evaluation methods, focusing on the methods most commonly used in the evaluation studies reviewed in this paper. Section 3 summarizes the results of several evaluations of so-called Active Labor Market Policy (ALMP) programs that have been tried in Latin America, Eastern Europe and China. The aim of ALMP programs is usually to mitigate the effects of severe macroeconomic shocks. Most programs are not explicitly targeted at women, but women participate in them. The types of programs considered in this paper include wage subsidy programs, public works programs, occupational retraining programs, and internship programs. Participation in ALMP programs is sometimes mandatory to receive unemployment benefits.

Section 4 reviews evidence on the effectiveness of microcredit, group lending and entrepreneurship programs, which are sometimes targeted at women. Microcredit programs

² For a discussion of the use of randomized social experiments in developing country contexts, see Duflo and Kremer (2004). For a more critical discussion of the use of randomization in economic development, see Deaton (2009)

provide loans to small businesses or to groups of borrowers under either an individual or a joint liability arrangement as well as other training or banking services. Such programs have been strongly advocated on the basis of high demand and a high rate of loan repayment, but relatively few of them have been subjected to rigorous impact evaluations of the rates of return. The papers we review here provide mixed evidence on the effectiveness of microcredit programs and suggest trade-offs between program rates of return and more selective targeting at poor or female borrowers.

Section 5 considers the effectiveness of programs that are designed to facilitate women's work by increasing the availability and affordability of childcare. Section 6 reviews the evidence on how conditional cash transfer (CCT) programs influence the lives of women based on evidence drawn from evaluations of the well-known Mexican *Oportunidades* program and the Nicaraguan RED program. The cash transfers are given to women in the household. Lastly, section 7 discusses the effectiveness of other kinds of programs, including a community association program in Africa, land titling programs, which the latter facilitate work by reducing the need to stay at home to protect one's property, and family friendly leave policies. Section 8 concludes.

2 Alternative Evaluation Approaches

2.1 The Evaluation Problem and Key Parameters of Interest

We begin by defining some notation for describing the evaluation problem and common parameters of interest. For simplicity, suppose there are two states of the world, corresponding to the state of being with and without some treatment intervention. For example, the outcome of interest could be an indicator for whether employed or unemployed and the treatment could be participating in a job training program.

Let $D = 1$ for persons who receive the intervention and $D = 0$ for persons who do not receive it. Associated with each state is a potential outcome, which may or may not be realized. Y_0 denotes the potential outcome in the untreated state and Y_1 the potential outcome in the treated state. Each person is associated with a (Y_0, Y_1) pair that represents the outcomes that would be realized in the two states of the world. Because a person can only be in one state at a time, at most one of the two potential outcomes is observed at any point in time. The observed outcome can be written as

$$Y = DY_1 + (1 - D)Y_0.$$

The gain from moving an individual from the state “without treatment” to the state “with treatment” is the treatment effect for that individual:

$$\Delta = Y_1 - Y_0.$$

Because only one of the states is observed, the gain from treatment is not directly observed for anyone. Inferring gains from treatment therefore requires solving a missing data problem. The evaluation literature has developed a variety of different approaches to solve this problem. The program evaluation literature has focused mainly on estimating direct effects of the program on program participants under the assumption that the indirect effects of the program on nonparticipants are negligible. This allows nonparticipants to be used as a source of comparison group data and to represent the “no treatment” state.

Solving the evaluation problem requires solving a difficult missing data problem. Because treatment impacts are not directly observed for individuals, researchers usually aim instead to uncover some features of the treatment impact distribution, such as the mean or median program impact. Much of the evaluation literature focuses on methods for estimating two key parameters of interest:³

(i) The average gain from the program for persons with characteristics X , commonly referred to as the *average impact of treatment (ATE)*

$$E(Y_1 - Y_0|X) = E(\Delta|X).$$

(ii) The average gain from the program for program participants with characteristics X , known as the *average impact of treatment on the treated (TT)*:

$$E(Y_1 - Y_0|D = 1, X) = E(\Delta|D = 1, X).$$

The ATE parameter is the gain from the program that would be experienced on average if a randomly chosen person with characteristics X were assigned to participate in the program. The TT parameter is the average gain for individuals who actually participated in the program (for

³ See, e.g., Rosenbaum and Rubin (1985), Heckman and Robb (1985), or Heckman, Lalonde and Smith (1999) for discussions of different parameters of interest.

whom $D = 1$). If individuals who participate in the program tend to be ones that receive the greatest benefit from it, then we would expect $TT(X) > ATE(X)$.

A comparison of the average gain accruing to participants, expressed in monetary terms, and the average costs of a program is informative on whether the program covers its costs. In determining the average gain to participants, any opportunity costs of participating in a program need to be taken into account. For example, while a worker is participating in a 3-month job training program, she may not be able to work. The gain from the program might be calculated as the sum of earnings in the 18 months following program participation, inclusive of the zero earnings during the 3-month training period, minus the predicted sum of 18 months of earnings that the individual would have experienced in the absence of the program. This net gain can then be compared with the cost of the program to come up with a benefit-cost ratio that is informative on whether the program at least covers its costs.

2.2 Solutions to the Evaluation Problem

2.2.1 Randomization

Randomized social experiments are considered by some to be the ideal design for evaluating the effects of a treatment. Under a randomized experimental design, a group of individuals is randomly selected to receive a treatment and another group is randomly denied the treatment and serves as the control group. The main advantage of random assignment is that it ensures that program participants and nonparticipants are comparable both in terms of observables and unobservables. Also, random assignment usually takes place conditional on being eligible for a program, which ensures that both the treatment and control groups satisfy eligibility criteria.

In terms of the previously described parameters of interest, randomization provides a way of estimating the average effect of treatment on the treated (TT). To see why, let $D = 1$ denote having applied and been deemed eligible for a program; otherwise $D = 0$. Also, let $R = 1$ if randomly assigned to the treatment group and $R = 0$ if randomly assigned to the control group. From the treatment group, we obtain

$$E(Y_1|R = 1, D = 1, X)$$

and from the control group

$$E(Y_0|R = 0, D = 1, X).$$

The difference in means gives

$$\begin{aligned} & E(Y_1|R = 1, D = 1, X) - E(Y_0|R = 0, D = 1, X) \\ &= E(Y_1|D = 1, X) - E(Y_0|D = 1, X) \\ &= TT \end{aligned}$$

where the conditioning on R can be dropped by virtue of random assignment (R is uninformative about Y_1 or Y_0). Thus, a well designed randomized experiment delivers one of the key parameters of interest in evaluations.

So far, we have discussed randomization that is performed after assessing eligibility for a program. That is, people apply and are deemed eligible for the program and are then randomly included. An alternative randomized design randomizes eligibility for the program. For example, some randomly determined fraction of the population is told that they are eligible for the program. Those who are eligible may then choose whether to apply or not. This alternative randomization strategy, which is less commonly used, is discussed in Heckman, Lalonde and Smith (1999).⁴

Although there are many advantages to using randomized experimental designs to evaluate effects of program interventions in terms of assuring comparability between the treatment and control groups, there are also some potential drawbacks to randomized experiments. The following types of problems may arise:

(i) *Randomization bias* or so-called *Hawthorne effects*: This problem occurs when introducing randomization changes the way the program operates. For example, individuals might choose not to apply to a program if they know they will be subject to randomization, which could change the mix of individuals receiving treatment and therefore change the observed treatment effects. Individuals may also behave differently when they know they are being observed as part of an experiment.⁵ In these cases, the outcomes that are observed do not necessarily represent the outcomes that would be observed in the absence of the treatment intervention, calling into question the external validity of the experiment.

⁴ For example, it may be difficult to deny pregnant women who express interest in a smoking cessation program access to it. On the other hand, it may be easy to randomly inform women about their eligibility to apply for the program.

⁵ Landsberger (1968).

(ii) *Contamination or cross-over effects*: The problem of contamination occurs if some of the controls that were randomly excluded from treatment are nonetheless able to receive the treatment and/or some members of the treatment group do not receive treatment.

(iii) *Dropout*: Some of the treatment group may drop out before completing the program, sometimes at a very early stage. In that case, the offer of treatment was randomized but not whether individuals completed the treatment.

(iv) *Sample attrition*: Random assignment ensures that the treatment and control groups are comparable at the start of the experiment. However, people cannot usually be compelled to participate in the program over a longer term or to respond to surveys. For this reason, the treatment and control groups may become less comparable over time due to nonrandom attrition. Oftentimes, people receiving the treatment have higher response rates on surveys than people who were excluded from the treatment, because they are happy to have been included in the program. When there is nonrandom program attrition that differentially affects the treatment and control groups, a nonexperimental evaluation method usually needs to be used to address bias concerns.

Another important issue concerning experimental evaluations is the issue of internal versus external validity. If the experimental protocol was followed and the potential problems described by (i)-(iv) are not that significant, then the experiment could be considered to be internally valid. However, extrapolating the results of the experiment to a larger population of interest requires external validity. That is, the sample participating in the experiment should be representative of the population of interest, especially if it is expected that people will respond to treatment in different ways. If the sample participating in the experiment is not similar, for example, if the sample in the experiment is younger, poorer or more likely to be female, then statistical adjustment can sometimes be used to extrapolate from the experimental results to the larger population of interest. A difficulty arises, though, if the sample in the experiment is dissimilar in unobservable ways, for example, if the sample that signed up to participate in the subsidy experiment is more motivated.

Lastly, most field experiments in economics are run for fairly short periods of time (at most 2-3 years). They usually do not permit an evaluation of programs for longer periods, and may, in addition, be affected by “pioneer” effects, stemming from the program not having been in operation for some time. (On this point, see Behrman and King, 2008, and King and Behrman,

2009). For a recent critical view on the value of randomized control trials in economic development, see Deaton (2009). In this paper, we present evidence from both experimental and nonexperimental studies, in recognition of the fact that both approaches have relative strengths and limitations.

2.2.2 Nonexperimental Estimators

In the absence of a randomized experiment, evaluations must be based on nonexperimental (or observational) data. As discussed above, if the randomized experimental design was compromised in some way, nonexperimental methods can also be used to increase comparability between the treatment and control groups. Nonexperimental estimators of program impacts typically use two types of data to impute the missing counterfactual (Y_0) outcomes for program participants: data on participants at a point in time prior to entering the program and data on nonparticipants. The following types of methods are commonly used in evaluation work:

(i) *Cross-section or difference-in-difference regression estimators*: These estimators evaluate the effects of the program by comparing the outcomes for a treated group to those of a nonexperimental comparison group, using regression adjustment to control for preexisting differences in observed characteristics that are thought to be determinants of the outcomes. The cross-section estimator compares participant and nonparticipant outcomes at some point in time after the program start date, whereas the difference-in-difference estimator compares the difference in outcomes between a post-program and pre-program time period (for example, the change in earnings).

(ii) *Cross-section or difference-in-difference matching estimators*: These estimators evaluate the impacts of the program by matching individuals in the treatment group to observably similar individuals in the control group on the basis of a set of observable characteristics.

The overall effect of the program is obtained by averaging over the differences in participant and matched nonparticipant outcomes. A commonly used metric for matching the individuals is the propensity score, which is the predicted probability of participating in the program conditional on a set of observed characteristics.

(iii) *Control function estimators*: These estimators evaluate the impacts of the program by a comparison of treated individuals and comparison group individuals, using statistical adjustment to control for both observed and potentially unobserved differences between the groups. These estimators usually require some assumptions on the distribution of unobservables and on their

relationship to observables. For example, it might be assumed that the unobservables affecting the program participation decisions and outcomes are jointly normally distributed and statistically independent of the observables.

(iv) *Instrumental Variables* or *LATE estimators*: These estimators require that there be some factor that influences the program participation decision but not the outcome directly, for example, an administrative rule that affects whether individuals are admitted in the program but that is not correlated with individual outcomes. They provide an estimate of the average impact of the program for the subgroup whose participation status is affected by the factor.

(v) *Behavioral modeling*: In some contexts, particularly in evaluating ex ante the effects of hypothetical treatments that have never been tried, researchers proceed by fully specifying a behavioral model that can then be used to extrapolate from historical observations on behavior to a new environment. For example, the model representing women's choices about labor supply given current availability and pricing of daycare could be estimated and then used to analyze the effect of changing the availability of pricing.

Appendix A describes these methods and the assumptions needed to justify them in greater detail. The performance of nonexperimental methods depends on whether the assumptions necessary to apply them are justified as well as on other factors, such as the quality of the data used in implementing them. For example, matching-on-observables estimators usually only perform well in situations where the observables in the data are rich enough to capture the key determinants of the program participation process. The models that need to be developed usually need to be tailored to the application.

A commonly used evaluation method in the studies reviewed in this paper is the method of matching. Heckman, Ichimura and Todd (1997) and Heckman, Lalonde and Smith (1999) study how various aspects of data quality relate to the performance of matching estimators and establish some guidelines for best practice in nonexperimental evaluations of job training programs:

(i) Program participants and nonparticipants should be situated in the same local labor markets (to control for unobserved labor market attributes affecting their employment prospects).

(ii) The questionnaires used to gather data on participants and nonparticipants should be comparable.

(iii) The dataset should include information on labor force and earnings histories as well as demographics, because earnings and employment dynamics are key predictors of decisions to enter training programs.

Different studies face different kinds of data limitations and the choice of estimator should take into account the particular limitations that need to be overcome. For example, if participants and nonparticipants are drawn from different local labor markets or if the survey questionnaires used to gather the data are different across groups, then difference-in-difference approaches tend to be more reliable than cross-sectional methods, because they allow for fixed unobservable differences between groups. (See Heckman, Ichimura and Todd, 1997, and Smith and Todd, 2005.) If there is good reason to believe that selection in the program is based on unobservable attributes, then control function methods, instrumental variables, or behavioral modeling that explicitly allow for program selection on unobservables may be most appropriate.⁶

3 Employment Creation and Job Training Strategies

3.1 Active Labor Market (ALMP) Policies

In this section, we review the evidence on the effectiveness of a variety of Active Labor Market Policy (ALMP) programs in affecting the employment, wages and poverty status of participants. Two predominant types of ALMP programs are wage subsidy programs that subsidize wages in either public or private sector jobs, and training programs that provide formal training designed to help participants develop occupational skills. Other types of ALMP programs are basic education programs, or short-term interventions that enhance job search skills. ALMP programs may also provide other sorts of benefits, such as work clothing, childcare and transportation expenses.

There are multiple channels through which ALMP programs might be expected to influence workers' employment and earnings outcomes. One is that they may increase the productivity of the worker and therefore her offered wage by augmenting her skill set or by providing opportunities to gain work experience (e.g. through internships). ALMP programs may also affect the process by which workers match with firms, for example, by reducing the costs of searching for a job or by increasing the arrival rate of job offers. Wage subsidy programs could

⁶ See Heckman, Lalonde and Smith (1999) for a discussion of the use of control function methods in program evaluation.

induce some worker-firm matches to take place that might otherwise not take place. For example, in the absence of any subsidy, a worker might only be willing to accept a wage offer above a certain threshold or else keep searching. With the subsidy, the worker might accept wage offers that would otherwise have been deemed too low. When the subsidy is removed, the worker-firm relationship might dissolve, unless the worker has gained enough experience on the job to increase the wage offer and depending on the costs of searching.

In this section, we first consider a number of ALMP programs in Latin America (the majority in Argentina) and then discuss other programs in Eastern European transition economies and in China. Many of these programs do not focus exclusively on women, but women are included among the participants and we highlight how the programs affect women.

3.1.1 ALMP Programs in Latin America

Proempleo Program in Argentina: A number of large-scale ALMP programs have been introduced in Latin America, as a way of alleviating the effects of severe labor market shocks affecting the region. One of these is the *Proempleo* program in Argentina, which is studied by Galasso, Ravallion and Salvia (2001) using a randomized experimental design. The program provided vouchers for workfare participants to give to prospective employers. The voucher entitled employers to a sizable wage subsidy, \$150 per month for workers age 45 and older and \$100 per month for younger workers, which lasted for up to 18 months. The experiment randomly allocated individuals into two treatment groups and one control group. One treatment group received only the voucher program, whereas the other received the vouchers plus had an additional option of skill training. The controls received neither. The program was made available to beneficiaries of temporary employment programs managed by the Ministry of Labor, the main program being *Trabajar*.

Galasso et al. (2001) find that the voucher program reduced the probability of unemployment, despite the fact that few firms actually took up the voucher subsidy. Voucher recipients had a significantly higher probability of employment but had no higher current income. Women and younger workers experienced the largest treatment impacts. Only 30% of those assigned to the voucher plus training treatment arm took advantage of the training option. Impact estimates based on a comparison of the two treatment groups indicate that the additional option to take training had no additional impact.

Galasso et al. (2001) hypothesize that the treatment effect of the vouchers may have been an "empowerment effect" in that workers who received vouchers seem to have been more comfortable in approaching private employers. They might also have been perceived by employers to be different from regular *Trabajar* workers. One possible reason for the low employer take-up rate is that taking up the voucher requires formalizing the employment arrangement, which could imply additional costs for the firm (such as severance payments to fire the worker).

The Galasso et al. (2001) study reports both intent-to-treat estimates and local average treatment effect (LATE) estimates of program impacts. The intent-to-treat estimates give the effect of the program offer to participants (irrespective of whether participants took advantage of the program). The LATE estimates use the randomized group assignment as an instrument for program participation status. The LATE estimate represents the average program effect for the group induced to participate in the voucher program as a result of being assigned to the treatment group (known in the literature as the group of "compliers").⁷ The outcome variables are the changes in earnings and employment between the last follow-up wave (May, 2000) and a baseline survey in December 1998. A limitation of the analysis is that there was some attrition in the experimental samples, with only 77.5% of those interviewed at baseline staying until the fourth round. Galasso et al. (2001) find that private sector employment improved among voucher recipients, with an employment rate of 14% for voucher recipients and 9% for the control group. However, the program had no impact on incomes measured 18 months after the program. Nevertheless, the program is deemed cost-effective, because it yielded employment impacts at very low cost, because of the low take-up of the subsidy by employers.

Trabajar II program in Argentina: In response to the macroeconomic crisis in the mid-1990s, the Government of Argentina introduced in May 1997 the *Trabajar II* program, which provided short-term work opportunities at relatively low wages and targeted unemployed workers from poor families. Under the program, local governmental and nongovernmental organizations submitted proposals for socially useful projects, such as projects to repair local infrastructure. The proposals had to be viable with respect to a set of criteria and were given priority according to how well they targeted poor areas, what benefits they were likely to bring to the community,

⁷ For discussion and development of LATE estimation methods, see Imbens and Angrist (1994) and Heckman and Vytlačil (2005).

and how much the area had already benefitted from the program. To be eligible for program benefits, workers had to be hired on to a successful proposal project and could not be receiving unemployment benefits or participating in another employment or training program. The projects lasted a maximum of six months but a worker could continue in the program if he/she switched to working on a new project. The wage rate was set at a maximum of \$200 per month, which was deemed low enough to ensure good targeting and to help ensure that workers preferred regular work when it became available.

Jalan and Ravallion (2003) analyze the impacts of the *Trabajar II* program on household income using a nearest neighbor propensity score matching methodology. The average gain accruing to program participants was \$103 dollars, about half the average *Trabajar* wage. The gains for female participants were not much different from the gains of male participants, but female participants tended to be from less poor backgrounds. Income gains were greatest for younger people (in the 15-24 age range).

Jalan and Ravallion (2003) do not report a benefit-cost analysis of the program. If the productivity of the workers on the socially useful projects exceeded the wages, then the program could be considered to have provided a benefit that exceeded the cost. However, a clear aim of the program was also redistribution toward the poor. In that case, the program might be deemed a success even if the worker productivity did not exceed the program's expenses, depending on how the government valued the income redistribution achieved through the program.

Jefe program in Argentina: A subsequent study by Galasso and Ravallion (2004) analyzes effects of another more recent Argentinean program, called *Jefes*, that replaced the earlier *Trabajar* program and was designed to provide direct income support for heads of households with dependents who became unemployed as a result of Argentina's economic crisis of 2002. At that time, the poverty rate soared from 37% to 58%. The *Jefes* program had work requirements, instituted to ensure that it reached those in greatest need, and it covered about 2 million households. Program participants were required to do 20 hours per week of community work, training, school attendance or employment in a private company with a wage subsidy. A major concern with regard to program implementation was program leakage, because administrators did not closely monitor whether the people signing up for the program were truly heads of households. It was also difficult to verify unemployment status, because many Argentines work in the undocumented informal sector.

The impact analysis carried out by Galasso and Ravallion (2004) is based on the October 2001 and October 2002 rounds of the *Encuesta Permanente de Hogares* survey, which covers urban areas. With respect to leakage, the study finds that one-third of those receiving the program were ineligible and that 80% of individuals who were eligible did not receive the program. In particular, more than half of the program participants were women who were probably not heads of households. Despite the problems in imposing eligibility criteria, however, the program was fairly well targeted at poor households. About half of program participants came from the poorest fifth and 80% came from the poorest 20% of the population.

Galasso and Ravallion (2004) evaluate program impacts using cross-sectional and difference-in-difference propensity score matching approaches. The treatment group includes those who applied and were admitted into the program, and the comparison group includes persons who applied for the program but had not yet joined. Galasso and Ravallion (2004) find it difficult to predict participation status among program applicants, and the predictive power of the propensity score model is not very high, raising some concerns as to whether the observables included in the propensity score model adequately control for differences between the treatment and comparison groups.

The matching analysis reveals that program participants experienced a smaller drop in real income on average than the comparison group, suggesting net gains on average between half and two-thirds of the gross wage, depending on the estimator used. Galasso and Ravallion (2004) argue that, given the level of income support, the observed income gains should have been in the range (0,150) and that negative estimates or estimates that exceed 150 should therefore be excluded. On these grounds, they prefer the impact estimates derived from the cross-sectional estimator, which indicate that 26% of *Jefes* participants would have been unemployed were it not for the program and 23% would have been inactive (primarily women).⁸ On the whole, Galasso and Ravallion (2004) find that the program reduced Argentina's aggregate unemployment rate by about 2.5% and contributed to social protection during the economic crisis by supplementing the income of poor families.

⁸ Ruling out negative estimates or estimates that exceed 150 is potentially problematic. Income support could have been used for productive purposes, such as a small business, making gains in excess of 150 possible. See, for example, the discussion of the Kaboski and Townsend (2007) study in section 4. Also, negative income gains are also feasible, if, given the transfer, some women might have withdrawn from the labor force.

Assessing the cost effectiveness of the *Jefes* program requires comparing the program costs to the value associated with the income redistribution under the program.

PROBECAT program in Mexico: Revenga, Riboud, Tan (1994) evaluate the effects of short-term vocational training in Mexico provided by the PROBECAT program, which was offered to more than 250,000 unemployed people. Program participants were selected according to an eligibility index that gave weight to factors such as the number of economic dependents, whether the individual attained a basic level of education, whether the individual was unemployed for less than 3 months, and prior work experience. Also, to be eligible, individuals had to be within the age range of 20 and 55 and be registered at the unemployment office. The impact analysis is based on longitudinal data on PROBECAT trainees combined with a separate dataset on a control group of unemployed people who did not join PROBECAT, drawn from national labor force surveys, which survey individuals for 5 quarters. Women were 49% of the trainee group but only 33.8% of the comparison group. The average female trainee was 29 years old and 46% were married. Women were less likely than men to have completed secondary education (grades 10-12).

The key outcomes analyzed in the study are employment, monthly salary, and number of hours worked. Impacts of the program on the length of unemployment spells are assessed using a Cox proportional hazards model, estimated on the treatment group and on a subgroup of controls predicted to have a high probability of participating in the program based on their characteristics. Potential drawbacks of the analysis are that participants and controls were given different surveys, so that measurements across surveys may not be comparable, and the control sample was relatively small. On average, program trainees found jobs more quickly (program participation reduced unemployment spell by 1.9 months for females and 2.5 months for males). Subgroup analyses reveal that the positive program impacts in terms of shortening unemployment spells were concentrated on trainees older than 25 and those with work experience. Female trainees with work experience were more likely to be employed at 3, 6 and 12 months after the training than were similar controls. Training increased the number of hours worked for both women and men, but only increased monthly earnings for men. Revenga, Riboud and Tan (1994) conclude that the program was cost effective, in the sense of covering its costs, for women over age 25 but not for younger women.

Programa Joven in Argentina: Aedo and Nunez (2004) study the effectiveness of another training program introduced in Argentina called *Programa Joven* that was targeted at low-income individuals aged less than 35. The program targeted young people from poor households with low education levels, little or no working experience, who were either unemployed or inactive. It provided an average of 200 hours of training, a monetary subsidy for females with young children, transportation expenses, medical checkups, books, material and work clothing. The duration of training varied from 14-20 weeks and was divided into a technical knowledge phase, in which participants were taught occupational skills, and an internship phase in which participants completed an eight-week internship at a firm. The impact evaluation study is based on two analysis samples: (i) a sample of 139,732 so-called “*Acreditados*” who qualified and registered to take training and who were at different points in their training at the time of answering the survey (some had not started, some were in the technical knowledge phase, some in the internship phase and some had finished or dropped out), and (ii) a subsample of 3,340 program beneficiaries and matched comparisons, drawn from the *Acreditados* sample, who were surveyed at the time of registration and then again one year later.

The evaluation study examines whether the program increased the labor income of trainees and their probability of employment, using a cross-sectional propensity score matching methodology to control for preprogram differences between program participants and nonparticipants.⁹ The propensity score model depends on current labor force status of the individual, a poverty measure, sociodemographics, education, marital status, and geographic region, and is estimating separately for four groups: young males, adult males, young females and adult females. A potential drawback of the propensity score model is that there are no historical data available on earnings or employment history at the time of program registration. Recent labor force history is a good predictor of participation in training programs and, lacking such data, the predictive power of the propensity score models is not high.

The impact estimates show statistically significant effects of the program on earnings of adult women (age 21-35) and young males (age less than 21) but not for young females or adult males. Statistically significant effects on employment are found only on adult women, in the range of 9-12 percentage points. Estimates obtained using alternative propensity score models and nearest

⁹ An individual was defined as a program beneficiary if he/she has completed the technical knowledge phase, else was designated a nonbeneficiary.

matching are fairly robust relative to changes in the variables included in the propensity score model, to differences in the source of data used and to variation in the number of neighbors used.¹⁰ The estimated impact for adult women and young males on earnings is around US \$20-\$25 / month. Rate of return estimates, obtained under alternative assumptions on the discount rate, show that the program would have a positive return only if the benefits are fairly long-lasting (9 years or more).

ProJoven program in Peru: Ñopo, Robles and Saavedra (2007) analyze the impact of a Peruvian youth labor training program, called *ProJoven*, on female and male youth living in urban areas. The program provided classroom training and internships lasting about three months for youths from poor families. Trainees received stipends during their training period, with mothers of young children receiving a double stipend. More than 20,000 youth participated in the program. An interesting feature of the program was that one of its explicit goals was to train female youth for traditionally male occupations so as to reduce gender segregation in the labor force.

Ñopo et al.'s evaluation of the program is based on a sample of beneficiaries and a sample of matched controls selected on the basis of gender, age, geographic proximity, poverty status, income, schooling, number of children and employment status. These individuals were administered a baseline survey and three follow-up surveys at 3, 6 and 18 months. The impact estimates are derived from a somewhat unusual two-stage matching procedure that first selects for each treated individual three matched controls on the basis of similarity in hourly wages. The motivation for this two-stage procedure is to closely align the treatment and comparison groups in terms of pre-program earnings so as to account for the so-called "Ashenfelter Dip" problem, namely, that program participants often exhibit a pre-program dip in their earnings that is not seen in comparison group data. The two-stage matching procedure generates a similar pre-program earnings dip pattern in both the treatment and comparison groups, although aligning the groups in terms of preprogram earnings would not necessarily guarantee that post-program comparisons are valid.

¹⁰ Recent research has shown that bootstrapping does not lead to valid inference about standard errors for nearest neighbor matching estimators, so the standard errors reported in this study would not be valid. However, correcting them using the alternative standard error estimators suggested in Abadie and Imbens (2005) or using kernel smoothed estimators for which bootstrapping is valid probably would make little difference.

The outcome measures of interest in the study are labor supply, hourly earnings, monthly earnings, and occupational segregation. Employment impacts for women are found to have been greater than for men, with women having experienced positive impacts of 6% at 12 months and 15% at 18 months and men having experienced negative employment impacts. The impacts on hours worked, hourly earnings and monthly earnings were positive for both women and men. The program had especially pronounced effects on monthly income from the main job. After 18 months, beneficiary females generated 92.9 percent more labor income than their control counterparts, in comparison with an increase of 10.9 percent for males. As a result of participation in *ProJoven*, the levels of occupational segregation, measured by the Duncan Index, were noticeably lower among program beneficiaries.

3.1.2 ALMP Programs in Transition Economies

Transition economies typically undergo large shifts in the demand for different kinds of labor as they move from a centralized to a more market-based economy. Because of labor market frictions and because it takes time to acquire new skills, transition economies typically undergo a period of high unemployment rates and large stocks of long-term unemployed persons. There have been a number of large-scale ALMP programs implemented in transition economies aimed at equipping workers with skills that are in greater demand in the market economy and at facilitating their job search process. Here we review the results of programs implemented in Russia, Romania, Slovakia, and Poland. None of these programs were specifically targeted at women, although women in each case made up a substantial fraction of the program participants.

ALMP programs in Russia and Romania: Benus, Brinza, Cuica, Denisova, and Kartseva (2005) analyze the effects of ALMP programs in Russia and Romania. The program eligibility criteria and the populations served differed somewhat by country. In Russia, training services were available only to individuals who were officially registered as unemployed with an employment center and receipt of unemployment benefits was conditional on making efforts to gain employment and on being available for work and taking suitable jobs as they became available. The unemployment benefit for persons who worked at least 26 weeks over last year was equal to 75% of the former wage at first and declined to 45% or the minimum wage (whichever was greater) over time. The benefit for other categories of workers was the minimum wage. A person who had been unemployed for a year and whose family income did not exceed two minimum wages also qualified for social assistance benefits. The analysis sample used in the

evaluation study consisted of a group of program participants and a control group that were selected on the basis of 2002 administrative data on training program clients.

In Romania, eligibility for ALMP programs depended on being registered unemployed, having income less than 50% of the minimum wage, being unemployed due to layoffs, having been employed at least 6 months of the last 12 months or being a recent graduate from school or university. There were various types of training available to program participants, including a public service component, whereby local government and other eligible organizations could propose public projects with a maximum cost of up to \$50,000 and hire ALMP participants to work on them. The analysis sample was comprised of individuals who entered the register not earlier than January 2001 and got off no later than December 2002.

Benus et al. (2005) evaluate the impacts of both the Russian and Romanian ALMP programs overall and separately by gender, using a propensity score methodology. The propensity score model is based on a fairly limited set of predictors of program participation, that include gender, age and education. Nonparticipants are people who applied for training but were not selected for it, so they would be expected to differ in some respects from participants. The outcomes of interest in the evaluation are the likelihood of being employed at the time of the follow-up survey, the likelihood of being employed at least once after the program, the likelihood of having a high salary, and the length of the current unemployment spell. The impact evaluation finds no significant effects of the ALMP programs in Russia, on the whole. In Romania, however, the program is found to have had a statistically significant impact for three of the four outcomes (the likelihood of employment, the likelihood of being employed at least once, and the level of wages were all higher among participants in Romania). Subgroup analyses reveal some gender, age, and education heterogeneity in the impacts for Romania. Re-training increased the probability of employment and decreased the wage for females. That is, re-training appears to have helped find employment but at a lower wage than the individuals would have found on their own. Middle age and lower education level individuals experienced the biggest program impacts. For men, the re-training was found to have had no effect. The authors hypothesize that the difference in impact findings between Russia and Romania may be attributed to differences in the characteristics of program participants affecting their labor market prospects; the Russian sample was better educated (45% had a university degree), older and had more labor force experience than the Romanian sample. In both countries, the program was not found to have been beneficial

for highly educated workers. Also, program participants may have been more negatively selected in Russia, because not everyone unemployed was registered there.

ALMP programs in Slovakia: Lubyova and Ours (1999) use administrative data from 20 Slovak districts to analyze whether it was beneficial for unemployed workers who wanted a regular job to accept a temporary ALMP job or enter a retraining program. Specifically, they study the effects of two ALMP programs in Slovakia targeted at registered unemployed workers on their exit rate from unemployment. The programs were targeted especially toward older workers, disabled workers and the long-term unemployed. The program provided retraining and counseling services as well as wage subsidies in two types of jobs: socially purposeful jobs (SPJ) and publicly useful jobs (PUJ). The subsidy in SPJ, which could be at private sector firms, had a minimum duration of 2 years and the subsidy at PUJ, which were typically public works jobs, had a maximum duration of 6 months.

Lubyova and Ours (1999) base their analysis on the administrative records of 100,000 individuals who entered unemployment in 1993. The records allow construction of detailed labor market histories. Using multivariate duration analysis, the authors jointly model the duration of unemployment and the duration of staying in an ALMP program, controlling for observable and unobservable heterogeneity among people. The focus of the study is on whether participation in ALMP programs affected the exit rate from unemployment to regular jobs. For women, 40% exited unemployment by finding jobs, 9% exited by entering ALMP and 51% had right-censored spells. For males, 47% exited unemployment by finding a job, 8% exited it by entering an ALMP program and 45% had a right-censored spell. On average, workers that entered the ALMP programs are found to have had a 150% increase in the exit rate into a regular job, with similar estimated program impacts for men and women. From additional analyses allowing the ALMP program effect to depend on the type of program, the authors conclude that there were positive benefits of retraining and publicly useful jobs on exiting unemployment into a job. For socially purposeful jobs, however, they find a negative effect.

ALMP programs in Poland: Kluve, Lehmann and Schmidt (1999) study the effectiveness of ALMP programs in Poland. The program took three forms: publicly financed training and retraining, intervention works (wage subsidies for workers in private or public firms), and public works. The aim of training and retraining was to increase the skill set of individuals in demanded fields such as data processing, accounting, secretarial work and welding, through courses lasting

on average 2-3 months. Individuals received unemployment benefits during the course of their studies. Workers in the training component tended to have higher education levels and to be female. The wage subsidy component of the program was structured so that the subsidy was increasing in the time the worker stayed with the firm. The public works component was targeted at the longer-term unemployed and many of the jobs available were low-skill jobs working on infrastructure improvements. Workers in either the wage subsidy program or the public workers program had an incentive to participate in the wage subsidy program at least 6 months to qualify for another 12 months of unemployment benefits.

The ALMP programs are evaluated using a difference-in-difference matching approach, where the main outcome of interest is labor force status. The samples are drawn from the 18th wave of the Polish Labor Force Survey that included a supplement with four years of historical information on individual labor market histories (monthly from 1992 to 1996). The treated group were individuals who were offered participation in the programs by their local labor office and who accepted the offer. Sample sizes in the three types of programs (training or retraining, intervention works or public works) were 241, 532 and 93. The control group consisted of 7,784 individuals who had been registered at least once as unemployed since January 1992. The matching procedure pairs treated individuals with control individuals who have the exact same labor force history and are matched on certain demographic characteristics (gender, marital status, education, region and age). The matching impact estimates turn out to be sensitive to which variables are used in the matching analysis, for example, to whether the selection of matches also takes into account local labor market conditions.

The impact estimates indicate that the training/retraining program increased the average employment probability for both men and women. Participation in the non-training ALMP programs did not affect women's employment probabilities but had a negative effect on men's employment probabilities, which the authors attribute to benefit churning rather than stigmatization of intervention and public works participants. That is, males appear to have taken intervention works and public works jobs between two spells of unemployment benefit receipt. Overall, the study concludes that ALMP training/retraining programs in Poland raised women's employment rates over the short and medium term.

3.1.3 ALMP Programs in China

Bidani, Goh and O’Leary (2002) analyze the effects of a retraining program in China called the “Reemployment Project,” which was designed to promote labor market entry of so-called *xiagang*, people who were laid off from state-owned enterprises but remained attached to their former employer for unemployment stipends, health insurance, pensions and sometimes also housing. The Reemployment project was administered by labor bureaus in local areas and included a range of active labor market policies such as job search assistance, counseling, training, wage subsidies, tax incentives for firms and assistance for self-employment. Individuals were allowed to be registered with the Reemployment service for up to three years.

The impact evaluation study was carried out in a city with very high unemployment, Shenyang, in northeastern China, and in another city with moderate unemployment, Wuhan, in central China. The laid-off workers were 47% female, tended to be less educated and were concentrated in the <35, and 35-46 year-old range. The training intervention was relatively short-term (one month, 132 hours of classroom training) and included courses in computer training, beauty and massage, hair cutting, sewing, toy making, cooking, repair training and driver education. Class sizes during the training sessions were often large with 200 to 300 workers in a small classroom.

Analysis samples were drawn from a census which required each state-owned enterprise to provide a list of workers laid off at different times. The sample of trainees was selected from the training registers from the training institutes (in the case of Shenyang) and from a master list supplied by the Wuhan Labor Bureau.¹¹ Three different treatment/comparison group samples are analyzed using multiple methodologies that include prop score matching, matching on odds ratios, and OLS. The outcomes of interest are employment and earnings, and, with a few exceptions, most of the impact estimates are robust to the use of different methods. Training is found to have had a negative impact in Shenyang on employment probability and no effect on earnings. In Wuhan, however, training is found to have had a positive impact on employment probability. The estimated impact on earnings is difficult to ascertain in the case of Wuhan, because the estimates are not robust to changes in the sample used for the analysis and/or estimator used.

¹¹ No-shows were excluded from the treatment group and in some cases included in the control group and cross-overs were excluded from the control group and in some cases added to the treatment group.

Subgroup analyses show, somewhat surprisingly, that training impacts did not differ much by age, marital status, gender, educational attainment or home ownership. In Shenyang, training appears to have had a stronger negative impact on men and those with lower education. In both Shenyang and Wuhan, individuals who contributed personally to the cost of training had higher reemployment rates. It is expected that individuals who expect to benefit the most from an intervention would be willing to contribute more to its cost.

To understand the reason for the observed discrepancy in impact estimates across the two program sites, it is useful to consider the details of the program's operation. It appears that the Reemployment program may not have been properly implemented in Shenyang, as many workers seemed not to have received the layoff stipends that they were supposed to have received under the program. In Wuhan, the program appears to have been better implemented. Quality of training remains an issue in both sites, though, because it may have been difficult for workers to learn new skills in large, overcrowded classrooms.

3.2 Sectoral Policies: Commodity Commercialization in Nepal

Paolisso, Hallman, Haddad and Regmi (2001) study the impact of a training program in Nepal that was designed to commercialize fruits and vegetables, improve farm output diversification and improve agricultural productivity. The program, called the Vegetable and Fruit Cash Crop Program (VFC), provided training on how to grow and process vegetables and fruits (e.g. into jams). Paolisso et al. (2001) base their analysis on data collected in 1991-1993 fieldwork on 264 households. The surveys collected detailed information on activity patterns in the household and their study examines how participation in the program affected time use patterns of both male and female program participants.¹² The VFC program was successful in the sense that household participation in the VFC program resulted in increased male and female time spent growing vegetables and fruits. The estimated average increase for women ranged from 3 to 55 minutes and the estimated increase for men ranged from 24 to 64 minutes. Although the income from sales of VFC products was relatively small, in many cases it represented the first opportunities for these women to earn income without leaving the community.

The authors also examine how participation in the program altered time spent on other activities. Although female labor is useful in agricultural production, it is also important to

¹² Some households have multiple wives and only head wife is used in analysis

household production, in tasks such as food preparation and caring for children. This raises a potential concern that programs that reallocate female labor toward agricultural production could have deleterious side-effects on time spent caring for children. Paolisso et al. find that for households with only one preschooler, VFC participation resulted in more time devoted to cash crops for both men and women and less time devoted to caring for preschoolers. The time trade-off was not as apparent, though, in households with more than one preschooler. The study does not evaluate the cost-effectiveness of the program.

3.3 Summary

The studies described above generally find that ALMP programs tried in the context of Argentina, Mexico and Peru have been effective in increasing employment rates and that women were often major beneficiaries of such programs. Some of the programs have been observed to also increase wages and income, but the observed effects on employment tended to be more robust than those on wages. Wage subsidy programs such as the Argentinean *Proemplo* voucher program increased the employment rate of participants on average but did not affect their income levels.¹³ Public works programs, of the kind made available by *Trabajar II*, led to job creation for workers and to increased income, with similar estimated gains for men and women. The *Jefes* income support program also increased employment and income and was highly demanded by women. It provided support to poor families during a particularly difficult economic crisis. The Mexican PROBECAT program was successful in augmenting employment of women over age 25 with previous labor market experience, but not in increasing their earnings.

We reviewed impact findings for two programs that were targeted at youth or young adults. The *Programa Joven* program, which is both a job training and an internship program, led to statistically significant impacts on earnings for adult woman and younger males but not for adult men or younger females. The impacts were relatively small in magnitude, though. The Peruvian *ProJoven* program had fairly large impacts on employment and hours worked of women as well as modest positive impacts on earnings. It also affected women's occupational choices and

¹³ A standard theoretical job search model (such as Burdett and Mortensen (1998)) would predict that firms would be more willing to hire workers with a voucher subsidy but would not necessarily pay a higher wage, which appears to be borne out in the data. Some worker-firm matches would take place with the voucher that would not otherwise be profitable. For the matches that would take place regardless, whether the worker's wage increases would depend on the bargaining between the firm and the worker.

decreased gender segregation, through the focus on training women to work in traditionally male dominated occupations.

On the whole, the empirical evidence suggests that many of the ALMP programs were effective in increasing employment rates. The evidence on whether the programs also increased wages is more mixed. The pattern of higher employment without higher wages might be expected for two reasons. First, it is difficult to bring about large changes in an individual's earnings capacity with any short-term program intervention. Rather, it seems that many ALMP programs operate by facilitating the worker-firm matching process, for example, by introducing workers to firms through internships. The worker-firm matches sometimes result in lower wages than the worker might have obtained independently from a longer job search. Where jobs are very scarce, for example, during a particularly strong downturn in the economy as was the case at the time of the introduction of *Jefes* in Argentina, ALMP programs do appear to have increased wages and incomes of program beneficiaries or to have alleviated poverty.

Only a few of the evaluation studies we reviewed here carried out a rigorous cost effectiveness analysis. One study is Aedo and Nunez (2004), which finds that program impacts have to be sustained over 9 years or more for the program to be cost-effective. Evaluation studies do not typically follow individuals for such long periods of time, so whether benefits can be sustained over such a long time horizon is unclear.

The evaluation studies of ALMP programs in transition economies tend to find positive program benefits for women, although the evidence differs somewhat depending on the country context. The main way that ALMP programs alter women's employment outcomes is by increasing their probability of employment and their exit rate from unemployment into jobs. Again, there is less support for an effect of these programs on the level of wages received by the employed. The few studies that have examined longer-term effects, such as the Kluve, Lehmann and Schmidt (1999) study, find that the positive program effects for women tend to be sustained over an 18-month time frame. More evidence is needed on the costs of the various programs to allow a study of their cost effectiveness.

Some of the ALMP programs we have reviewed operate on a very large scale, and there remains the question of the extent to which people who do not participate in these programs suffer adverse consequences, such as job displacement or lower wages. It is also possible that such large governmental training programs crowd out training that firms or individuals might

carry out privately. These questions warrant further examination before any full accounting of the effectiveness of these programs can be made.

Betcherman, Olivas and Dar (2004) provide an overview of the recent international experience with active labor market programs (ALMPs), focusing on the impacts of ALMPs on the employment and earnings of participants and considering the impacts of ALMPs in developed, developing country and transition economy settings. The evidence is reviewed for seven different ALMP categories: employment services, training for the unemployed, training for workers in mass layoffs, training for youth, wage and employment subsidies, public works, and micro-enterprise development/self-employment assistance.

Betcherman et al. (2004) review five studies of youth training programs in developing countries (Jovenes programs in Argentina, Chile, Peru, and Uruguay), which include the studies described above. All of the evaluations find positive employment impacts and two of the three that compute earnings effects also find positive impacts. The positive impacts found for developing countries contrast with the mainly negative estimated impacts for youth-oriented ALMP programs in developed and transition economies. Betcherman et al. (2004) argue that for youth oriented programs to be effective, they need to offer a comprehensive set of services that include basic education, employment services and social services. In evaluating the effectiveness of any youth-oriented training program, an important question is whether the program funds might be better spent on alternative programs that keep youth from dropping out of school. Formal schooling, especially secondary school, is often found to have a relatively high wage return of 10% or more in developing countries.

Betcherman et al. (2004) also review the evidence on the impact of training and job placement programs for the adult unemployed population and conclude that the programs tend to increase employment but not wages. They also emphasize the importance of incorporating the substitution effects toward beneficiary workers and away from other workers into any analysis of program effects, which is not commonly done.

4 Programs to Increase Entrepreneurship and Access to Credit

In this section, we review the results of some evaluations of microcredit/grant programs in Bangladesh, Sri Lanka, Pakistan, Thailand, Peru, South Africa and Kenya. These programs usually provide liquidity for small businesses through individual or group loans, but sometimes

also for consumers. Micro-credit programs have spread rapidly as a policy instrument for helping poor populations since the 1970s when they were first introduced. Initially, the loans were granted to both men and women, but in the 1980s there has been increased targeting of loans toward women, under the view that women are more likely to be credit constrained than men, that women have an inequitable share of power in household decision-making and that women may face discrimination in the regular labor market. Among the programs described below, some are gender targeted and some are not but are inclusive of women. In many cases, the programs also use poverty-based targeting mechanisms.

The Grameen Bank in Bangladesh is probably the best known of the microcredit programs. It provides loans for nonagricultural self-employment activities for poor borrowers, usually under a group lending arrangement, and has served as a prototype for similar lending programs around the world. In the case of the Grameen Bank, borrowers form small groups that are held jointly liable for repayment of loans, that is, all members of the group are ineligible for further lending if any one member defaults. It is thought that group members can better monitor each other's activities than can bank lenders, so that these programs may be able to overcome screening, monitoring and enforcement problems that can lead to market failures in the lending market. Microcredit programs often also provide other noncredit services, such as savings programs, training for skill development, literacy, investment strategies, health training, and programs aimed at changing women's attitudes. There are also relatively simple microfinance programs that increase access to credit for borrowers, without any group liability arrangement or training components.

The Grameen Bank and similar programs are often considered successful on the basis of high loan recovery rates, typically in excess of 90%, and high observed demand for loans, although these two indicators do not necessarily reflect productive use of funds. There are relatively few evaluation studies that assess directly the rates of return for microfinance programs, some of which are described below.

4.1 Microcredit/Grant Programs in Asia

4.1.1 Group-based Microcredit Programs in Bangladesh

There have been a large number of papers evaluating the effects of micro-lending programs in Bangladesh and the literature has reported some contradictory findings, even for studies based on

the same dataset. Estimated program impacts appear to be sensitive to different econometric approaches used. Also, different studies tend to focus on different treatment effect parameters, with some studies estimating average effects of treatment on the treated and other studies focusing on local treatment effects for specific subpopulations. Whether and to what extent microcredit programs help poor individuals is still a matter of intense debate, but we will summarize here some of the accumulated evidence from randomized field studies and nonexperimental studies.

Pitt and Khandker (1998) is an influential and often cited nonexperimental study that estimates the impact of participation in one of three Bangladesh group-based lending programs (the Grameen Bank, Bangladesh Rural Advancement Committee (BRAC), and Bangladesh Rural Development Board (BRDB)) on women's and men's labor supply, on girls' and boys' schooling, on expenditures and on assets. It explores how the program impacts vary by gender of the borrower. Very few women work in the regular labor market in Bangladesh, because it is a Islamic society that excludes women from many activities. Self-employment is a viable alternative for these women, because the productive activities can be carried out in their home and while taking care of children.

Pitt and Khandker (1998) estimate the effect of the lending program using survey data collected in 87 rural Bangladeshi villages in 1991-92. In recognition of the fact that the program was nonrandomly placed, their analysis does not simply compare individuals that received and did not receive the program. Rather, they use a feature of the program eligibility rule as a way of generating a comparable comparison group. Only landless households, defined as households with less than one-half acre of land, were eligible to participate in the program; households with more land were ineligible. Their identification strategy estimates conditional household demand equations for households that were and were not eligible to participate in the program by virtue of the land ownership restriction but that lived within the same village (the effects of regional location are controlled using village-level fixed effects). An assumption required to justify this approach is that land ownership is uncorrelated with the unobserved household type. The estimation uses data on villages where the microcredit programs were available as well as villages where the programs were not available. Indicators for program availability within the village essentially serve as instruments for program receipt.

The Pitt and Khandker (PK, 1998) study provides estimates of the impact of credit provided to men and women on their labor supply, their children's schooling, household expenditure and nonland assets held by women. They find that it mattered whether credit was provided to the man or woman within the household. Program credit provided to poor households had a larger effect when women were the participants, for example in increasing household expenditure, increasing consumption, and increasing women's labor supply in cash income earning activities.

Their findings concerning the substantial benefits of micro-credit programs have not been without controversy. Roodman and Morduch (2009) attempt to replicate the PK study, but find negative estimated impacts on credit given to females relative to males on the same outcomes. They also perform overidentifying tests for the validity of some of the instruments used in PK's study and reject their validity. Lastly, they apply an alternative regression-discontinuity estimator to the same data, using the discontinuity in program eligibility that occurs when individuals have just under and just over one-half acre of land. The RD results point to little effect of program participation on school enrollment of girls or boys. The RD estimates indicate a negative effect of male borrowing on female labor supply with no association with male labor supply. The results based on the RD analysis are quite different from those based on the more structural analysis presented in PK. They are not necessarily inconsistent, though, as the RD analysis recovers the treatment effects locally, near the point of discontinuity and PK's analysis aims to uncover an average of treatment effects for all individuals participating in the program, including those not at the margin of eligibility by virtue of the amount of land they own. Nevertheless, Roodman and Morduch conclude from their analysis, "We come away from the PK study with doubts about the magnitude, sign and direction of the reported effects of microcredit."¹⁴

Khandker, Samad and Khan (1998) analyze the village level impacts of three micro-credit programs in Bangladesh (the Grameen Bank, Bangladesh Rural Advancement Committee (BRAC), and Bangladesh Rural Development Board's (BRDB) RD-12 project), allowing program effects to differ for the three programs. All three are group-based lending programs, with somewhat different features. For example, the BRAC program also provided literacy and training services in addition to credit. The Grameen Bank covered 40% of Bangladeshi villages, while BRAC and RD-12 covered roughly 20%. According to Khandker et al. (1998), 60% of rural households met the eligibility criteria for the programs but only about 45% participated,

¹⁴ For an analysis on how microcredit programs affect consumption variability, see Morduch (1994, 1995).

which appears to have been due to low demand for the program.¹⁵ Khandker et al. (1998) assess the village-level impacts of the three programs using a regression model applied to village level data, with program placement indicators for whether a particular village has a particular kind of program. The data used came from a household survey administered in 1991-1992. A limitation of the analysis is that it does not allow program placement to be based on unobservable village attributes, which could lead to bias in the estimated program impacts. Also, the study does not distinguish between impacts on male and female borrowers within the village.

Khandker et al. (1998) find that all of the programs have positive impacts on income, production and employment, especially in the non-farm sector. For example, average household income increased by about 20-30% in villages with the programs. However, only the Grameen Bank, which on average provided the largest loans, increased household labor supply (by 7%); the other programs appear to have reduced labor supply (by 11-12%). Also, only the Grameen Bank had an effect on the average village level wage, which the authors hypothesize was due to a general equilibrium effect stemming from the decrease in the supply of wage workers and an increase in self-employment. For this reason, the Grameen Bank appears to have had important positive spillovers on wage workers not directly participating in it.

4.1.2 Gender-targeted Grants in Sri Lanka

A recent study examining the differential impacts of providing grants to women and men is de Mel, McKenzie, and Woodruff (2007), which estimates the returns to capital for Sri Lankan microenterprises owned by women and men using a randomized social experiment. The experiment randomly provided cash or equipment grants in the amount of 10,000 or 20,000 rupees. 10,000 rupees was equivalent to about three months of median profits. The study finds that about 75% of the grants were invested in businesses, both for men and for women.

De Mel. et al. (2008) estimate the average return to capital to have been in the range of 4.6%-5.3% per month, or about 60% per year. But they also document substantial heterogeneity in return across different types of enterprises and find the return to capital for female owned enterprises to have been substantially lower than that for male owned enterprises. In fact, they estimate a return of zero for female owned enterprises and a return in excess of 9 % per month for male owned enterprises. After exploring various possible explanations for the gender

¹⁵ Roodman and Morduch (2009) note that the credit-consumption relationship is "U"-shaped and that the poorest are often excluded (or self-excluded) from micro-credit programs.

disparity, the study concludes that differences in industry accounted for some of the disparity, with female-dominated industries exhibiting negative returns to capital and male-dominated industries having had high returns.

The de Mel et al. (2008) study also challenges the notion that female-owned businesses were poorer and more credit-constrained and shows instead, at least in the context of Sri Lanka, that female entrepreneurs were on average more likely to come from dual earner families and to have had more assets. A somewhat puzzling finding is that the females had on average lower ability, as measured by a digit span recall test that was administered to respondents, which is surprising given that they had on average more years of education. When returns to capital are allowed to vary by ability level of the entrepreneur, it is observed that those with higher measured ability experienced higher returns.

In contrast to PK's reported findings, the de Mel et al. (2008) study finds no effect of the randomized grants on school-going for children, nor on health or education expenditures. In general, their findings call into question the efficiency of targeting credit programs to women rather than to entrepreneurs operating the most productive businesses. They suggest that the benefits of targeting credit toward women need to be weighed against the possible costs of not targeting to the highest return activities or to the poorest households.

4.1.3 Microcredit Programs in Pakistan

Hussein and Hussein (2003) survey a large number of microcredit programs that have been implemented in Pakistan and provide an overview of the results of evaluations of these programs along various dimensions. There is mixed evidence on whether microcredit programs have reached the poorest families and whether the programs have substantially alleviated poverty. Some programs were explicitly targeted at the poor and did have a large proportion of poor borrowers, but most programs did not explicitly target the poor and oftentimes did not keep detailed records on the poverty status of borrowers. Programs in Pakistan tended to target households rather than explicitly targeting women, although a few gave loans only to women. Even for programs with explicit targeting toward women, the household was usually seen as the borrowing unit because it was not uncommon for women to pass the loan on to male relatives.

In an assessment of the gender impact of microcredit programs, a study called the *Rural Financial Markets Study* finds that of borrowing individuals, 78 percent were male. Also, males tended to borrow larger loan amounts, with the average loan size received by women being less

than half the average loan size received by men. In rural areas, loans have been mainly used for agriculture production (including livestock), while in urban areas they have been mainly used for enterprise development. In both rural and urban areas, some loans have been used for consumption purposes, particularly for households with low income potential. One study that gathered detailed consumption expenditure data finds that borrowers reported improvements in their diet.

There are very few studies of the rates of return to investments, but the ones available tend to find returns to enterprise development to be between 8 and 30 percent. One difficulty in accurately assessing rates of return is that families often have multiple sources of income, with family members working elsewhere providing substantial support, some of which is funneled into family businesses. Also, only a few studies examine the impact of microfinance programs on employment. A study of the Kashf microcredit program interviewed 129 female respondents and finds that the program led to more family members participating in family enterprise activities. Also, women borrowers reported higher levels of confidence, self-esteem, and measures of empowerment and demonstrated increased awareness of social issues.

In summary, microfinance programs in Pakistan mainly have had male borrowers unless they have been explicitly targeted at women, and they tend to have benefited the near-poor and middle-income groups rather than the poorest, unless they explicitly targeted the poorest. The rate of return studies suggest that there were relatively high rates of returns to the loans, although the returns are somewhat difficult to measure given multiple household income sources. There is suggestive evidence that the programs have had some nonmonetary benefits, such as an increase in women's confidence and empowerment and improvements in the household's diet.

4.1.4 Microcredit in India

Banerjee, Duflo, Glennerster and Kinnan (2009) evaluate the effects of placing a microfinance institution in a new market using a large-scale randomized experimental design. In 2005, 52 of 104 neighborhoods of Hyderabad (a large city in India) were randomly selected for opening a microfinance lending branch, called Spandana, with the remaining serving as controls. The loans were group liability loans and were targeted at groups of women (women formed their own groups). The eligibility criteria favored women who owned their homes and who had longer residence in the community, and the program did not serve the "poorest of the poor." There was no training component.

The researchers administered a baseline household survey in 2005 and another survey 15 to 18 months after the branch was introduced to approximately 65 households in each neighborhood (6,850 households total) to measure the program's impact on outcomes related to consumption, new business creation, business income, education, health and women's empowerment. It is worth noting that other microfinance institutions operated in the area and some of them opened up in both treatment and control areas. The evaluation strategy is an intent-to-treat analysis that compares averages in treatment and comparison areas, averaging over both borrowers and non-borrowers. The study finds no effect of the program on average monthly expenditure per capita but that expenditure on durable goods increased in treatment areas and the number of new businesses increased by one-third in treatment areas. They find that treatment effects were heterogeneous, with households with a high propensity for starting a business cutting back on non-durable consumption in an apparent effort to save for the fixed costs of starting a business. The study finds that there were no discernible effects on education, health or women's decision-making.

4.1.5 Expanding Credit in Thailand: The Thai Million Baht Village Fund Program

Kaboski and Townsend (2007) study the effectiveness of microcredit programs in the context of rural Thai villages. In particular, they develop and structurally estimate a dynamic model of credit-constrained households deciding on consumption, indivisible investment, and savings and study how these households were affected by the introduction of a large-scale government microfinance program, the Thai Million Baht Village Fund Program. This program, begun in 2001, transferred one million baht (about \$25,000) to each of almost 80,000 villages in Thailand to start village banks that would lend to local households. The total amount of funding to each village was the same regardless of the size of the village, so village size provides a plausibly exogenous source of variation in the amount of credit increase per household. The funds were not directly targeted at women, but women were among the program beneficiaries. Kaboski and Townsend's (2007) analysis samples came from the Townsend Thai project, which gathered panel data on rural and semi-urban households and businesses from 64 villages in four Thai provinces from 1997 to 2007.

The behavioral model is based on the standard buffer stock model of savings behavior under income uncertainty (e.g., Aiyagiri (1994) and Deaton (1991)) with the additional feature that households have an investment option in each period. The household's problem is to maximize

the expected discounted value of utility over an infinite horizon by making decisions on consumption, savings, and on whether to take advantage of the investment opportunity, subject to borrowing constraints. Kaboski and Townsend (2007) derive a steady state solution to the model.

The behavioral model is estimated using five years of "pre-program" data, in particular, information from households and local businesses on their consumption, income, investment, credit, liquid assets, and interest income from these assets in addition to village population data. The validity of the estimated model is assessed by comparing the model's predictions of the effects of the Thai Million Baht program on consumption, investment and the probability of investing to the actual effects observed under the social experiment.¹⁶ Impact estimates obtained using the simulated model are close and not statistically different from impact estimates obtained from reduced form regressions based on post-experiment data, lending support to the model.

One of the notable predictions of the model that is borne out in the data is that the impact of the program on consumption exceeded one million baht and that consumption increased more than credit did because the credit was used toward productive activities. Kaboski and Townsend (2007) use the estimated model to compare the costs of the microfinance program to the costs of a direct cash transfer program that would have provided the same utility benefit. They find that the cost of the microfinance program was 33 percent less, due to the fact that the microfinance program relaxed borrowing constraints that the transfer program did not.¹⁷ In summary, Kaboski and Townsend (2007) demonstrate both theoretically and empirical that microfinance programs are an effective means of increasing the liquidity of credit constrained households and that they positively impact both investment and consumption.

4.2 Relaxing Consumer Credit in South Africa

Many microcredit programs focus on providing credit for entrepreneurs rather than consumers. Providing credit to consumers is a more controversial policy initiative, because of concerns that behavioral biases may lead consumers to take on too much debt. Karlan and Zinman (2007)

¹⁶ The program is introduced into the model as a reduction in borrowing constraints in an amount that would increase the amount of total expected credit (as calculated from the model) in the village by one million baht. This implies a larger reduction in credit constraints in smaller villages (because all size villages received the same amount).

¹⁷ Interestingly, even households that do not use credit can be affected by the relaxation in borrowing constraints, as it lowers their need for a buffer stock of liquidity and allows them to increase consumption.

estimate the impact of expanding the supply of consumer credit in South Africa on consumers' well-being, on their employment outcomes and on lenders' profitability. They implemented a field experiment that randomly encouraged loan officers to give loans to marginally rejected loan applicants. The loans were made to relatively high-risk borrowers, and had average interest rates of around 200% and default rates of around 20%. A survey was administered 6-12 months after the loan to measure borrowing activity, loan uses and household well-being. Typical reported uses of the loans were to pay off other debt or for transportation expenses. Key impact findings are that individuals in the treatment group were significantly more likely to retain their job and had higher incomes. It seems that the loans helped individuals maintain employment by avoiding shocks that might have affected their ability to travel to work. Individuals in the treatment group also reported more positive attitudes toward their future prospects but also had higher rates of depression and stress. Additionally, individuals in the treatment group were less likely to report experiencing hunger in the last 30 days (14% of the sample reported experiencing hunger). The study did not focus on differences between female and male borrowers but found that along most dimensions there were no significant differences by gender. Lastly, the study found that the loans to the marginally rejected applicants were profitable for the lenders, suggesting that both borrowers and lenders might benefit from an expansion of consumer credit.

4.3 Opening Savings Accounts for the Poor in Rural Kenya

Dupas and Robinson (2009) report results from a field experiment in rural Kenya designed to test whether savings constraints, in terms of a lack of opportunities for formal saving, inhibit self-employed individuals from expanding their businesses. The treatment intervention opened interest-free savings accounts for a random subsample of poor, daily income earners. The study finds that the take-up of these savings accounts was high among women but low among men, despite the fact that the accounts were interest-free and had relatively high penalties for early withdrawal. It finds that savings accounts were associated with positive benefits for women, in terms of positive impacts on productive investment levels and expenditures, but that they had no effect on men. The authors hypothesize that investment and expenditures increased with the interest-free savings because the women faced negative private returns to savings when they tried to save informally at home.

4.4 Entrepreneurship, Decision-making, and Leadership Training in Peru

Using a randomized control trial, Karlan and Valdivia (2006) analyze the effect of providing business training as an add-on to a Peruvian group lending program targeted at female microentrepreneurs. The study was conducted with the Foundation for International Community Assistance in Peru (FINCA), a microfinance institution that supports poor, female entrepreneurs in Lima and Ayacucho through village banks. Both the treatment and control groups received loans under the program. In addition to the loans, the treatment groups received 30 to 60 minutes of training during their normal weekly or monthly meetings with the bank over a period of one to two years.¹⁸ The program provided general business skills and strategy training but not client-specific advice. For example, women learned how to target customers, identify competitors, and position their product as well as promotional and planning strategies. The study finds that the business skills training program had significant benefits for both the clients and for the microfinance institution. The clients improved their knowledge of business processes and increased sales, revenues and profits. The largest positive impacts were observed among the subgroup that expressed the least interest in receiving the business training. The study does not find effects on measures of female empowerment, possibly because the women entrepreneurs participating in the study were relatively empowered already. Also, the study finds that children from mothers in the treatment group spent significantly more time on studying and less time on leisure-related activities. The microfinance institution benefitted from increased client retention and from better repayment of loans. The study suggests that poor entrepreneurs may be helped both by alleviating credit constraints as well as by improving their business-related skills.

4.5 Summary

Evaluation studies of microlending programs in different country contexts yield mixed evidence on the effectiveness of the programs as well as on the benefits of gender-specific targeting.

Pitt and Khandker's (1998) study finds substantial benefits of targeting microcredit to women relative to men in the context of the Bangladeshi group-lending program. Program impacts have been larger when targeted to women, as measured by effects on consumption, labor supply and children's school-going. Unfortunately, the recent replication analysis of Roodman and Morduch

¹⁸ The average loan is \$203 and the average level of savings of the individuals in the program is \$233. The recovery rate is 99 percent. For further discussion of the program, see Karlan and Valdivia (2006).

(2009) casts doubt on the robustness of these findings with respect to the functional form specifications adopted and the validity of the instrumental variables. An RD analysis of the same data finds less support for positive program impacts and provides new evidence that providing loans to men has reduced the labor supply of women. In another study based on a fairly simple comparison of villages that do or do not have microlending programs, Khandker, Samad, and Khan (1998) find that these programs have had positive impacts at the village level on income, production and employment. They also find that the Grameen Bank has had positive spillover effects on people not participating in the program.

On the other hand, de Mel, Duresh, McKenzie and Woodruff (2008) question the efficacy of targeting programs toward women. Their experimental evaluation of a program that randomized grants in Sri Lanka finds much lower rates of return on capital for female-owned enterprises in comparison to male-owned enterprises, which they attribute in part to differences in the returns for female and male dominated industries. They also show that the female entrepreneurs tended to come from less poor dual-earner families and that loans targeted at women may not have targeted the poorest families. de Mel et al. (2008) do not find any benefits of the program on children's school-going, but do not examine the full set of outcomes examined in the Pitt and Khandker study. Nevertheless, their evidence calls into question the justification for targeting credit programs at women. Whether micro-lending programs yield higher rates of return when targeted at women is likely to depend on the country context and the types of enterprises in which women are engaged.

Hussein and Hussein's (2003) survey of Pakistani microcredit programs similarly shows that microcredit programs do not necessarily benefit women or the poorest families, unless there is an explicit targeting mechanism in place to reach these groups. Even when the loans are targeted at women, it is not uncommon for women to relinquish control of the loans over to male household members. Hussein and Hussein (2003) provide some evidence, however, that microcredit programs provide some nonpecuniary benefits for women, such as increasing reported levels of confidence and self-esteem.

Kaboski and Townsend's (2007) study of a Thai microcredit program that was not gender targeted and did not impose joint liability finds relatively large impacts of the program on consumption. Their estimates indicate that consumption increased by substantially more than the initial amount of credit due to the productive use of funds. Interestingly, using the dynamic

savings and investment model that they develop, they show that even those who did not borrow through the program would be expected to have benefited from it; because the program reduced the need for buffer stocks to guard against future low consumption.

In the Peruvian context, Karlan and Valdivia's (2006) randomized trial indicates potential benefits of supplementing microcredit programs targeted at female entrepreneurs with additional, general business training services, both in terms of improving the profitability of their firms and improving loan repayment rates.

In summary, these evaluation studies provide valuable evidence that microfinance programs have yielded high rates of returns, but that the returns have been typically lower if the programs have been too narrowly targeted. In designing a microcredit program, an important consideration is whether the cost of more specific targeting warrants the benefits in terms of redistribution toward women or toward the poor.

5 Programs That Facilitate Work through Lowering Costs of Work or Improving Working Conditions

We next review the results of some evaluation studies of programs designed to facilitate women's work, either by providing better access to affordable and reliable childcare, by providing childcare subsidies or through more generous parental leave benefits. A program that subsidizes childcare has both an income and a substitution effect on women's labor supply. The income effect would be expected to reduce the number of hours she works, assuming that leisure is a normal good. The substitution effect would increase consumption of the subsidized good, i.e., increase the use of childcare and increase hours worked.

In this section, we also review evidence on the effects of Peruvian and an Argentinean land titling programs, which provide land titles to household squatting on public or private land. By providing legal protection for property, land titling programs significantly affect the wealth of these households and reduce the incentives for women and children to stay home. For this reason, land titling programs have been found to have substantial impacts on women working.

5.1 Childcare Programs

5.1.1 Programs to Increase Availability of Childcare in Latin America and Africa

The accessibility of affordable and reliable daycare is an important determinant of women's labor force participation decisions. Formal sector jobs, such as factory work, often require long hours of work and do not easily accommodate the presence of children. Informal sector jobs, such as making things at home, typically pay less and are less likely to include health care benefits, but may be more flexible in terms of hours and in terms of allowing mothers to supervise children while working. The availability of childcare not only affects the decision about whether to work but also whether the mother engages in formal market work.

In recognition of the importance of daycare to a mother's working decision, many Latin American countries have introduced programs aimed at increasing the supply and lowering the costs of daycare. One class of programs, called community daycare programs, has been implemented in Peru, Colombia, Bolivia, and Venezuela as well as in most Central American countries. These programs aim to (i) foster human capital accumulation of children through better nutrition, better hygiene and through activities that promote child development and socialization, and (ii) facilitate mothers' working outside the home and at higher wage jobs. The largest program, *Hogares Comunitarios* in Colombia, has nearly one million beneficiaries.

Community daycare programs in Guatemala: Ruel and Quisumbing (2006) present the results of an impact evaluation of Guatemala's Community Day Care Program. The evaluation focuses on Guatemala City, although the program is available throughout the country in both rural and urban areas. Under the program, a group of parents selects a women from the community to serve as an in-home daycare provider for up to 10 children (under the age of seven), Monday through Friday from 6am to 6pm. As part of the program, children are involved in developmental activities and receive food and snacks. The caretaker typically receives furniture, educational materials and money for compensation and for the children's food. The program is designed to provide 80% of the children's nutrition and 40% of the program cost goes to food. In addition to the compensation the caretaker receives from the program providers, families are expected to make some supplemental contributions. In Guatemala, only 3% of eligible families participated in the program, in part because spaces were limited.

The impact evaluation found that the program improved children's diets and increased mothers' incomes. The impact evaluation was carried out using the method of matching,

comparing a cross-section of beneficiaries to a matched set of control households with children in the age 2-5 range. The controls were selected by matching beneficiary children to other children from the same neighborhood, of a similar age and gender and whose mothers were working. Impact estimates were obtained by propensity score matching.

Because all mothers had to be working to be a part of the evaluation, it is impossible to estimate the effect of the program on mothers' rate of working. However, to get an idea of the potential effect, Ruel and Quisumbing (2006) compare the labor force participation rate for a random sample of mothers who have children aged newborn to 6 years with that of women who do not. The comparison suggests that the program could potentially increase labor force participation by 25 percent.

The program's impact on mothers' earnings and job choice is assessed by comparing beneficiary mothers to matched control mothers. This comparison indicates that beneficiary mothers had 30% higher earnings than mothers using alternative childcare arrangements and were more likely to be employed in the formal sector. The program had the largest benefits for younger and older women with low levels of education. A limitation of the analysis is that all mothers in the control group were restricted to be working. It is likely that some of the mothers in the treated group would not have been working had the program not been available to them. For this reason, the evaluation probably understates the effect of the program on mothers' incomes. The results from interviews with focus groups show that the program was very well received and much appreciated by the beneficiaries.

A comparison of Guatemala and Ghana: Quisumbing, Hallman and Ruel (2003) analyze the determinants of mothers' joint labor supply and daycare utilization choices in Guatemala City and in Accra, Ghana. The analysis for Guatemala is based on a random sample of mothers with preschool age children from one zone of Guatemala City. The analysis for Accra is based on a household survey.

The urban settings of Guatemala and Ghana are quite different with regard to the type of work that women typically do. In Guatemala City, the landscape is becoming increasingly urbanized and women often engage in formal work, such as factory work. About one-fifth of households are headed by single women, about half of which are poor or indigent. In Accra,

71.9% of female employment is in the informal sector. For cultural reasons, there is a high percentage of female headed households (35.1%).¹⁹

Quisumbing et al. (2003) estimate a model of mothers' working and childcare decisions for these two different samples of women. From their analysis, they conclude that the supply of daycare is an important influence on mothers' working decisions only in Guatemala City, where women are more likely to work in the formal sector. In Accra, the supply measures, such as the distance to the nearest formal daycare provider, have no effect on labor supply choices. The findings suggest that enhancing formal daycare options for women is likely to be most effective in increasing female labor supply where a large percentage of women work in formal sector jobs.

Hogares Comunitarios in Columbia: Attanasio and Vera-Hernandez (2004) analyze the impacts of a large community daycare program in rural Colombia on children's nutrition, female labor supply and school achievement. The operational aspects of the program were similar to the Guatemalan program described above, except that caretaker mothers could have up to 15 children and food was delivered weekly to their house. The Hogares Comunitarios (HC) program is the largest welfare program in Colombia. It is targeted at poor households, as measured by an eligibility index. As noted above, the program has extensive coverage, but there are still many children not participating that could serve as a comparison group for the purpose of evaluating the impacts of the program.

In evaluating the impact of the HC program both on mothers' labor supply and on child outcomes, Attanasio and Vera-Hernandez (2004) compare beneficiary families with nonbeneficiary families. They argue that the use of cross-sectional matching on observables methods would be inappropriate, because they believe the participation decision to be based in part on unobserved attributes. Indeed, when they apply propensity score matching, they get negative estimated impacts of the program. To allow program selectivity to be based in part on unobservables, they implement an instrumental variables estimator. To do so, they require a variable that influences the program participation decision but has no direct influence on the outcomes. They maintain that distance of the household to the nearest HC and distance averaged at the community level can serve as instruments. This requires an assumption that households are not choosing their location with regard to the location of the HC, but Attanasio and Vera-

¹⁹ Men and women from the indigenous Ga population traditionally live in separate houses after marriage.

Hernandez (2004) present evidence that the location of the nearest HC frequently changes, making this assumption more likely to be satisfied.

The IV estimates indicate that the program has had extremely large, positive impacts on female employment and hours worked as well as beneficial impacts on child outcomes. The probability of female employment increased from 0.12 to 0.37 and the number of hours worked increased by 75 hours per month. The study also finds statistically significant effects of the program on children's height and also, over the longer-term, on school going and grade achievement.

Preschool building program in Argentina: Berlinski and Galiani (2007) analyze the impact of a large pre-primary school building program in Argentina on pre-primary school attendance and maternal labor supply. The program was mainly targeted at middle-income households living in urban areas. It attempted to compensate for geographic differences in the availability of preschool facilities by rolling out the program first in areas with the lowest level of facilities. Between 1994 and 2000, the program created about 175,000 places, which expanded the number of places available at baseline by 18%, with different regions and different cohorts differentially exposed to the program. Berlinsky and Galiani (2007) use a difference-in-difference approach to evaluate the effects of the program on maternal employment and weekly hours worked, exploiting differences across regions in the numbers of facilities built and in exposure of cohorts due to program timing. The study finds evidence of full take-up of new facilities and also that the likelihood of maternal employment increased between 7 and 14 percentage points. The effect of the program on weekly hours is not precisely estimated.

5.1.2 Childcare Subsidies in Transition Economies

In Communist economies, such as the former economies of Russia and Romania, almost all women participated in the labor force. The high rates of participation were feasible in part because of the wide availability of government sponsored childcare centers (nurseries, preschools, kindergartens and after school programs). The transition to a market economy has diminished the availability and increased the cost of childcare in many of these countries. Fong and Lokshin (2000) analyze how mothers' demand for paid childcare, mothers' labor force participation and working hours have responded to changes in the cost of care and to changes in wage offers in Romania. Between 1989 and 1995, Romania saw a sharp decline in public

funding for childcare services. Over the same time period, legislation was passed that provided mothers with 65% of their previous salary if they cared for their own child during the first year.

The approach taken by Fong and Lockshin (2000) to understand the effects of these and other policy interventions is to jointly model households' decisions about childcare and mothers' labor supply. To this end they develop an economic model of household decision making about consumption of childcare quality, of market goods and of leisure. Child care arrangements are classified into six categories defined by combinations of the mother's employment status, mode of care (formal and informal), and employment status of other household members. The model assumes that households pay a flat fee for childcare services, where the fees charged by kindergartens are a function of the quality of care provided and the total level of childcare prices within the locality.²⁰

The empirical model consists of a discrete choice equation for the childcare mode and mother's labor supply, an equation for the mother's hours at work, and an equation for children's hours in paid care. The effects of unobserved variables is incorporated by imposing a factor structure on the error components of the model. The estimation is based on household survey data from the Romanian Child Care and Employment Survey (RCCES) linked with data on childcare providers from the Romania Child Care Facilities Survey (RCCFS). Both datasets were collected by the World Bank in the same geographical areas during the same time period. The surveys were designed to allow matching data on childcare fees and child care quality with the households surveyed in the communities. After the model parameters are estimated, the model is used to simulate responses to changes in the policy environment, for example, to changes in the price of childcare and the level of the mother's wage. The model is used to extrapolate to policy variation that is outside the range of that in the data.

Model simulations indicate that a 10% increase in a mother's wage offers increases the rate of the mother's labor force participation by 10.9% and increases her use of formal care arrangements by 4.3%. Changes in the price of childcare have a smaller effect on the level of maternal employment and on the use of formal care; an increase in the price of care by 10 percent leads to a 1.2 percent decline in the number of working mothers and a 2 percent decrease in the number of households that use formal care. A policy that fully subsidizes formally

²⁰ In actuality, households pay fees that depend on their total household income and the number of children enrolled. However, to avoid complications in estimation, Fong and Lokshin (2000) treat childcare fees as exogenous to the households, using regional variation in fees as the main source of identifying information.

provided childcare, however, would increase the rate of women's labor force participation by as much as 12-15 percent. The elasticity of mothers' labor supply with respect to childcare cost are found to be -0.17, which is in line with estimated elasticities reported in the related literature based on data from the U.S. and Canada.²¹ Fong and Lokshin (2000) conclude that government subsidies for childcare are an effective means of increasing the number of mothers who work, increasing the incomes of poor households and lifting some families out of poverty, but that the effects of such policies are less significant for the poorest households.

A similar study by Lockshin (1999) studies mothers' participation in the labor force, working hours and demand for childcare in Russia. In the 1980s, most women in Russia worked and the government heavily subsidized childcare programs that were widely available. A decline in GDP in the 1990s led to a sharp decrease in the availability of state-run child care facilities and an increase in the cost of sending children to these facilities. As described in Lockshin (1999), Russia moved from a country in which childcare was provided by the government and almost all households with children had access to affordable or free childcare to one in which few households have access and the cost of day care significantly affects labor force participation decisions. Lockshin (1999) builds a static utility maximizing model of households' decisions about labor force participation, working hours, and choice of childcare mode to motivate an econometric model that he uses to assess the effects of three different kinds of policy interventions: family allowances, childcare cost subsidies and wage subsidies. The model is estimated using panel data from the Russian Longitudinal Monitoring Survey (RLMS).

The simulations show that childcare subsidies increase maternal employment by almost twice as much as comparable wage subsidies. Also, childcare subsidies are more effective than wage subsidies or family allowance transfers (transfers to families with children) in increasing family income. Childcare subsidies increase the amount of time working women spend at work and increase the proportion of mothers who choose to work.

A limitation of the analysis is that it does not examine changes in the utility of households under the alternative types of policies. Increasing a mother's work would likely reduce her time spent in leisure, which leads to a loss in utility that offsets the benefit of higher income. It would be useful to compare the alternative policies on a utility basis. Also, in both the Lockshin (1999) and Fong and Lockshin (2000) studies, it is not clear how the quality of formal childcare

²¹ Cleveland, Gunderson, and Hyatt (1996) for Canada, Connelly (1992) and Blau and Robins (1988) for the U.S.

provided compares to the quality of home or informal care, leaving open the question of how children's development is affected by policies that encourage the use of more formal care.

5.2 Summary

The studies of Latin American community day care programs described above generally find the programs to have substantial positive impacts on women's propensity to work, on their number of hours worked, and on the health, nutrition, and educational outcomes of children. Assessing the benefit-cost ratio of these programs requires an assessment of the monetary value of the child outcomes. One assessment for a community daycare program in Bolivia that is reported in Behrman, Cheng and Todd (2004) finds benefit-cost ratios of roughly 2:1 (for a discount rate of 5%), only taking into account the benefits accruing to the child participants and not taking into account any benefits for the mothers. Some additional support for the effectiveness of these programs comes from the evidence that the take-up for them is high.

Berlinski and Galiani's (2007) study of a preschool building program in Argentina also finds full take-up of the new places for children created by the program. Simply expanding the availability of preschool led to an increase in women's labor force participation.

Formerly communist countries have undergone large changes in the pricing and availability of childcare. The modeling frameworks of Lokshin (1999) and Fong and Lokshin (2000) permit comparisons of the relative effectiveness of wage subsidy programs, childcare subsidy programs and income transfer programs in affecting female labor supply. That analysis finds that childcare subsidies are more effective than comparable wage subsidies or family allowance transfers in increasing the proportion of mothers working and the amount of time they spend at work. Unfortunately, not that much is known about whether and to what extent children in the context of these transition economies (Russia, Romania) benefit from being in formal daycare rather than other forms of care (e.g. home care or relative provided care). More evidence needs to be accumulated to better understand how these subsidy programs affect both mothers and their children.

6 Conditional Cash Transfer (CCT) Programs

In recent years, governments in many countries have adopted conditional cash transfer (CCT) programs as a strategy for alleviating poverty and stimulating investment in human capital. These programs typically provide cash grants to poor families if they send their age-eligible children to school as well as subsidies for regularly visiting health clinics for check-ups. Mexico and Brazil adopted the earliest CCT programs in the late 1990s. Subsequently, CCT programs were adopted in many Latin American countries, including Argentina, Chile, Colombia, Costa Rica, El Salvador, Ecuador, Honduras, Nicaragua, Panama, Peru and Uruguay, and similar programs are now available in countries in Asia, Africa and even in the U.S. Fiszbein, Schady, Ferreira, Grosh, and Kelleher (2009) provide a thorough overview of CCT programs around the world and an overview of the measured impacts of these kinds of programs.

Many of the CCT programs give the cash transfers to women in the household, in an attempt to improve the bargaining position of women and because it is thought that targeting transfers toward women shifts expenditures toward investments in children. Below, we summarize impact findings from three evaluation studies on how CCT programs have been found to affect women and adolescent girls, two studies of the Mexican *Oportunidades* program and one of the Nicaraguan *Red de Proteccion Social* (RPS) program.

6.1 The Mexican *Oportunidades* Program: Impacts on Mothers' Time Use

The Mexican *Oportunidades* program (formerly called PROGRESA) is one of the best known of the CCT programs. The program has been rigorously evaluated using both experimental and nonexperimental evaluation designs. An experiment carried out in the first two years of its implementation (1998-1999) in rural areas demonstrated statistically significant impacts of the program on reducing child labor, improving health and nutrition and increasing school enrollment and attainment.²² The experiment was a place-based randomization that randomized 506 villages in or out of the program. Within the treatment villages, about half the families were deemed eligible to participate in the program. Families were visited door to door and told about their eligibility, and most families who were eligible agreed to participate. A subsequent nonexperimental evaluation of the program in urban areas also found statistically significant

²² See, e.g., Schultz (2000,2004), Gertler (2004), Behrman, Sengupta and Todd (2005), Parker and Skoufias (2000), Buddelmeyer and Skoufias (2003), Skoufias and McClafferty (2001), and Todd and Wolpin (2006).

impacts similar in magnitude to those found in rural areas. Today, the Mexican program provides payments to about one-quarter of all families in Mexico that constitute on average 10-20% of those families' household income.

6.1.1 Impact of PROGRESA on Time Use

Parker and Skoufias (2000) analyze the impacts of the PROGRESA program on children and women's time use using the experimental data that was gathered in rural areas of Mexico. We focus here on the results they report pertaining to adults. The time use questions asked about time use in the previous day.

Theoretically, the expected impact of the program on time use is ambiguous. The additional income that families receive under the program would be expected to reduce the labor supply of all family members, assuming that leisure is a normal good. On the other hand, most of the program benefits are tied to children attending school and the highest level of benefits are given to the older children. If children are no longer able to work as much outside school, either at home or at work, then the program could lead to a substitution of mothers' time. The CCT program also required that children attend health clinics and that mothers attend group meetings, which directly impacted mothers' time allocation.

Parker and Skoufias (2000) find that the labor force participation of beneficiary mothers in these rural villages was low, about 18% prior to the program. Working mothers typically work as unpaid workers in family businesses, in self-employment or in non-agricultural work. In comparison, men's labor force participation rate is about 90%, with men often participating in day laborer types of activities and being heavily involved in agricultural work.

Parker and Skoufias (2000) evaluate the impact of the program using double difference regression models that compare the change in the labor force participation rate before and after the program of treatments and controls.²³ They do not find any evidence that the program significantly affected the labor force participation of either women or men. One plausible reason for this finding is that the program eligibility criteria did not create disincentives for working. The criteria did not depend on whether the household members were working, but rather on other factors designed to capture the poverty of the household (such as whether the house had a dirt

²³ Because the data are generated from a random experiment, it is not really necessary to adjust for preprogram differences, but Parker and Skoufias's (2000) approach differences out any preprogram differences in the labor force participation of treatments and controls that may have randomly arisen.

floor). Also, once a family was deemed eligible for the program, it was included in the program for three years before eligibility was reassessed.

Parker and Skoufias (2000) also analyze women's time use related to time spent satisfying program requirements. They find that women spent a significant amount of time engaging in activities such as taking children to school or to health clinics, in order to satisfy program requirements. There was also some evidence that women spent less time engaged in domestic work.

A recent paper by Dubois and Rubio-Codina (2009) further analyzes the same *Oportunidades* data and examines in detail the impact of the program on the time mothers and older sisters devoted to taking care of children under the age of 3, where the analysis is restricted to 4,036 mothers age 18 or over with older daughters (age 12-17) still living in the household, with children younger than three, and without elderly or sick relatives that might require care. The study finds support for the existence of substitution effects, namely that mothers in the treated households were more likely to substitute for their older daughters' time devoted to childcare.²⁴ Dubois and Rubio-Codina (2009) find that daughters devoted more time to schooling and less time to caring for children and that the total amount of time devoted to childcare increased. They maintain that the program augmented the human capital of children both by keeping teenage girls in school and through more and arguably better "mother-provided" childcare. Corroborating evidence that *Oportunidades* increased women's time burden comes from focus group interviews where mothers often said they were doing work that was previously done by children who were now attending school (Adato et al., 2000). Some limitations of the Dubois and Rubio-Codina (2009) study are that it was not established that mothers' care was better in promoting child development. If mothers were more apt to combine their care activities with other household activities (such as cooking), the care might not necessarily have been better. Also, it was unclear whether the increase in time devoted to childcare was accompanied by a decrease in some other productive activity.

Although the Mexican CCT program did not appear to significantly affect mothers' labor supply behavior, Parker and Skoufias (2000) find that it substantially reduced the labor supply of children and adolescent girls and boys in both salaried and nonsalaried work. Behrman, Parker

²⁴ When all mothers are pooled together, there is no significant effect of treatment on maternal participation in childcare; but when treatment is interacted with whether the mother has offspring age 12-17, there is a positive and statistically significant increase in mother's participation in childcare.

and Todd (2011) report similar impacts on children's working behavior in their nonexperimental evaluation of the *Oportunidades* program in urban areas.²⁵

6.1.2 Effects of *Oportunidades* in Rural Areas on Alcohol Abuse and Domestic Violence

Angelucci (2008) evaluates the impact of *Oportunidades* on alcohol abuse and on domestic violence in rural areas. If income shortage is a major cause of domestic arguments, then the provision of extra income through CCT programs may reduce domestic violence. On the other hand, if program funds are used to purchase more alcohol and increased alcohol consumption is positively associated with domestic violence, then the program could increase violence against women. The shift in bargaining power resulting from the income transfers being given to women could also lead to increased violence or to marital instability.

Angelucci (2008) finds that the program on average led to a 13-fold increase in wife income, because wives hardly had any income before the program, given their low rates of formal labor market participation. The program led to a decrease in the share of spousal income earned by the husband from 97 to 62%. For households receiving relatively small amounts of transfers, Angelucci (2008) finds that the program led to a 15% decrease in alcohol abuse and a 37% decrease in drunken violence. However, in households in which the wife was entitled to large transfers (usually due to larger families with age-eligible children) and where husbands had low education levels (about 17% of the sample), she finds evidence of increased violence and aggressive behavior, as reported by the wife. The increase in violence appears to have been larger the older the husband was relative to the wife.

Thus, her findings suggest that CCT programs targeted at women are beneficial for the majority of them, but that there is a subset of families with low-education male heads who tend to hold more traditional values for whom violence increases.

6.1.3 Effects on Marriage and Fertility-related Behaviors of Adolescents

Gulametova-Swan (2009) studies the effect of the Mexican *Oportunidades* program on adolescent girls' decisions about marriage and fertility. Adolescent youth enrolled in the program were required to attend health talks, where topics such as prevention of pregnancy were discussed. Gulametova-Swan's (2009) analysis is based on data from the 2002–2004 urban *Oportunidades* evaluation data samples, which were nonexperimental. Program participation was

²⁵ The impacts of *Oportunidades* in urban areas were assessed using propensity score matching estimators.

not randomly assigned in urban areas, where families self-selected into the program by visiting program sign-up modules. To minimize the potential problem of selection bias, Gulametova-Swann defines treatment as the offer of the program and estimates intent-to-treat parameters.²⁶ The treatment group consisted of individuals who were eligible to receive the program by virtue of living in an area where the program was available and meeting the eligibility criteria. The control group consisted of individuals who met the eligibility criteria but lived in areas where the program had not yet been introduced.

To capture the dynamics of how the program affected decision-making, the program effects are estimated within a multi-state hazard model. The model allows study of how the program affected the timing of first sexual experience, first marriage, and first and second births. The timing of events is allowed to depend on the timing of previous events and unobservables are incorporated into the analysis in the form of a permanent unobserved component that enters into the hazard functions equations with different factor loadings. The key findings are that *Oportunidades* significantly delayed the onset of premarital sex, delayed marriage and delayed the timing of first births. The estimated magnitude of the estimated program effects are somewhat sensitive, though, to the functional form specification of the hazard model.

6.2 Nicaragua's *Red de Proteccion Social* Program

In 2002, world coffee prices were at their lowest point in 50 years, which had a great impact on the economies of coffee producing countries in Latin America. Maluccio (2005) examines the effects of this price decline on poor, rural populations in Nicaragua and the role of a conditional cash transfer program, the *Red de Proteccion Social* (RPS) in mitigating adverse effects of these aggregate price shocks. The RPS program was modeled after the *Progreso/Oportunidades* program in Mexico and provides cash "food security" and education transfers that are conditional on attendance at health clinics and health education workshops and on sending children age 7-13 who have not yet completed fourth grade to school.

The evaluation of the RPS program was based on a randomized-community-based design with measurements taken before and after the program intervention. One-half of 42 *comarcas* were randomly selected into the program and the others served as a control. The data used for the

²⁶ The offer of the program is likely to be more random than the decision to take up the program. Her analysis conditions on observables to take into account potential nonrandomness in the program offer. It is known that the program was first introduced in poorer areas, which would tend to bias the analysis against finding program effects.

evaluation came from a baseline survey administered in 2000 and from follow-up surveys in 2001 and 2002. The sample was representative of a poorer part of the Central Region in Nicaragua. Within the evaluation sample, 21 of the 42 *comarcas* were identified as being areas where coffee was cultivated, 10 in the treatment group and 11 in the control group. The methodology used for the evaluation was double difference estimators, applied to the experimental data samples. The impact analysis finds that household and food expenditures were significantly higher, by around 20%, in treatment localities than in control localities. During the economic downturn, control households increased their number of hours worked as a way of coping with the downturn, whereas treatment households worked about the same numbers of hours. School enrollment rates also rose significantly more in treatment areas than control areas, with a simultaneous decline in child labor. Maluccio (2005) concludes that the RPS program played an important role in helping poor, rural Nicaraguans cope with the economic crisis created by the falling coffee prices.

6.3 Summary

Affecting adult female employment outcomes has not been a primary goal of most CCT programs, which instead have been oriented toward providing incentives for increasing school attendance of children, improving the health of both children and adults, and transferring income to poor families. Nevertheless, the incentives created by these programs have indirectly influenced adult women's time use. The rural evaluation of the Mexican *Oportunidades* program was based on a rigorous experimental design and collected extensive data. The studies we reviewed indicate that the program did not significantly affect mothers' labor force participation status but did affect mothers' time use in several ways. First, mothers spent substantial amounts of time meeting program requirements, such as bringing children to school and to health clinic appointments and attending the required health talks. Second, in families with girls age 12-17, there is evidence that mothers' time engaged in childcare activities increased and that the older girls' time in childcare activities decreased, suggesting a substitution of mothers' time. Focus group interviews also supported the notion that mothers increased time devoted to childcare while their older girls attended school.

7 Other Types of Programs

This section describes other types of interventions that have been found to influence women's labor market outcomes.

7.1 Family Friendly Policies in Western Europe

Most countries in Western Europe have relatively generous childcare and family leave policies in comparison with the rest of the world. Del Boca, Pasqua, Pronzato and Wetzels (2007) study the effects of family friendly social policies in European countries on the labor force participation, hourly wages and fertility of women aged 21-45. Their analysis is based on a random sample of households from 15 European Union member countries, drawn from the European Community Panel Survey. Their goal is to assess the impact of social policies on employment and fertility using a reduced form methodology that relates labor force participation and fertility decisions to a set of determinants of these decisions that include measures of countries' relative generosity in providing childcare services (for children age 0-3) and of countries' parental leave policies. They focus on the provision of childcare for young children, because there is a lot of variation across countries for the 0-3 age group. Del Boca et al. (2007) create indicators of childcare generosity by ranking countries according to their available programs and then grouping them into four generosity categories. A similar grouping is performed for parental leave. Denmark is the country with the most generous childcare programs for children age 0-3 and Italy is the country with the most generous parental leave policies. Spain is the least generous both in terms of childcare and parental leave.

Del Boca et al. (2007) then estimate a statistical model of the joint decisions to work and to have a child. The model is a bivariate probit model, with the outcomes modeled as a function of individual characteristics, such as education and age, household characteristics that include the husband's income, and the social policy generosity indicator variables. The study finds that the generosity indicators are highly predictive of fertility and employment outcomes and in some cases more predictive than are individual or household characteristics. Women who reside in countries in the two most generous childcare categories have a higher probability of working and of having a child. Similarly, the probability of working and having children is higher in countries with more generous parental leave policies. Thus, the authors conclude that family friendly social policies strongly influence women's decision-making with regard to fertility and labor

supply and are a major determinant of the relatively high rates of female labor force participation observed in Scandinavian countries.

7.2 Community Association Programs in Kenya

The poor have lower rates of participation in community associations and some social programs aim to stimulate participation in such organizations. Increasingly, the World Bank has included a community based component in many of its programs. Gugerty and Kremer (2008) examine the effects of a social program, sponsored by a Dutch nongovernmental organization, that supported 80 Kenyan women's community associations through outside funding, training, and in-kind agricultural transfers (seeds, fertilizer, pesticides). In Kenya, rural women's groups engage in many activities, such as group farming and rotating savings activities, in which the members of the group contribute regularly to a pot that is given to each member in turn. They also engage in civic action promoting certain causes, such as prevention of violence against women, or holding fund-raising activities to raise money for local public goods. The social program gave about \$674 to each group, half of which was in-kind agricultural transfers. The program's goals were to strengthen women's community organizations and improve agricultural output and practices. The program implementation was gradual and random assignment was used to decide which groups received the program first.²⁷

Gugerty and Kremer (2008) analyze the impacts of the program using information gathered on one baseline survey and two follow-up surveys. They find that the program induced enrollment effects, with the number of applicants to groups being 40% higher in treatment than in comparison groups. However, the program had only small effects on levels of output from the group-farmed plots; the in-kind farming subsidies appear to have been distributed to individual group members rather than used to farm the group plots. Within groups, the younger, more educated members seem to have been the ones that best availed themselves of the training opportunities. The study also finds little evidence that the treatment groups did more to provide local public goods, though they did receive more visits from government officials. Treatment groups were more likely to experience a change in leadership, with men or better educated women taking on leadership roles. In summary, the program intervention did induce statistically significant differences between the treatment and comparison groups in enrollment rates and in

²⁷ To be precise, the groups were ordered alphabetically within geographic region and every other group was assigned to receive the program first. The comparison group received the program two years later.

education levels of group leaders. The program did not lead to substantial public goods investment, but rather to distribution of some of the program transfers among group members and to their private investment. However, the precise benefits of the induced changes and their rates of return are as yet unclear and deserve further consideration.

7.3 Land Titling Programs in Peru and Argentina

Land titling programs typically transfer property rights to poor households occupying the land. These types of programs do not aim to directly influence female employment. However, in transferring wealth to the household and in securing property rights, these programs often do have a significant influence on women's working decisions.

Field (2003a, 2003b) studies the effect of an urban land titling program in Peru that was targeted at urban squatter households. The program greatly decreased the administrative burden of obtaining a land title, which had required application at a large number of offices, and distributed over 1.6 million property titles over a five-year period. One of the explicit aims of the program was to improve gender inequality of property ownership and to this end the program rules stipulated that, among common law and legally married households, both spouses' names had to appear on government issued property documents. The allocation of the program was not random, although the neighborhoods that received the program early were highly similar in terms of observable characteristics to those that received the program later.²⁸ The evaluation approach taken in the study is a difference-in-difference comparison of eligible women in treated neighborhoods to eligible women in as yet untreated neighborhoods, with an adjustment for differences between noneligible households living in the same neighborhoods. Noneligible households were those that possessed a title prior to the program and therefore had nothing to gain from participating in it.

Field (2003a) finds that program beneficiary women were significantly more likely to appear on property documents and were more likely to report participating in household decisionmaking. Field also examines whether the program affected their fertility behavior. An effect of land titling on birth rates may come through multiple channels and the direction of the effect is theoretically ambiguous. On the one hand, land titling represents a wealth effect that

²⁸ The characteristics include rates of malnutrition, illiteracy, fraction of school-aged children not in school, residential crowding, proportion of the population without access to water, sewer or electricity services.

may increase fertility, assuming children are a normal good. On the other hand, including women's names on property documents can change the balance of power in a relationship and, if women have preferences for lower fertility and fertility decisions are made through family bargaining, can decrease fertility. Land titling can also influence the value of children, for example, by affecting parents' options about where to live in old age. Lastly, when people have a legal right to their property and no longer have to squat to retain control of it, their time can be used for other purposes. Changes in the value of mothers' time may also affect fertility. Field (2003a) finds a strong effect of the land titling program on fertility behavior, with eligible households who had been exposed to the program exhibiting roughly a 20% reduction in annual birth rates in the few years after the program.

In a separate study of the same program, Field (2003b) analyzes the effects of the program on hours of work, location of entrepreneurial activity and child labor force participation, using a similar evaluation strategy. The impact estimates indicate that land titling increased labor hours, shifted labor supply away from work at home and toward work in the outside market and led to substitution of adult for child labor. On average, labor hours increased by 17%, the probability of working inside the home decreased by 47% and the probability of child labor declined by 27%.

Another study of land titling is that of Galiani and Schargrodsky (2009), which examines the effects of land titling in an urban area of Buenos Aires. When the squatters had originally settled on the land, they thought it was public land, but it was actually private land. In 1984, a law was passed expropriating the former owners' land (with compensation) and entitling current occupants. Some original owners accepted the governmental decision and compensation, while others challenged the decision in drawn-out lawsuits (some of which are still pending). Galiani and Schargrodsky exploit the variation in the owners' decision to accept or challenge the law to identify the effects of land titling on occupants' behavior. Although it is conceivable that owners with more favorable land quality would have been more likely to contest the appropriation, Galiani and Schargrodsky show that this was not the case and that the parcels of land for which owners contested or did not contest the appropriation were actually highly comparable and basically next to each other. For this reason, their analysis considers the land titling "treatment" exogenous from the point of view of the squatters.

Galiani and Schargrodsky find substantial effects of the land titling program on household behaviors using data from two surveys, performed in 2003 and 2007, and focusing on 245

families that were identified as having arrived on the land prior to the intervention. Their main findings are that entitled families increased their housing investment, reduced household size, and increased children's education relative to the control group. The entitled households also had fewer children per household head.

Both the Field (2003a,b) and the Galiani and Schargrodsky (2009) evaluations of land titling programs find substantial benefits of the programs that include increases in mothers' working, decreases in fertility, increases in children's education, decreasing child labor, and increases in housing investment. Although not that much evidence has been accumulated yet on the effects of land titling programs, they appear to be a promising way of significantly increasing a poor household's wealth over the short term and inducing an array of changes in behavior. Requiring that the spouse's name appears on the title, as in the Peruvian program, significantly increased the probability that a woman's name was included on the title.

8 Synthesis and Directions for Future Work

This paper has studied the effectiveness of a variety of policy interventions and social programs at improving the quantity and quality of women's work. Some of these policy interventions were undertaken to increase employment, some to increase female employment, and some for other reasons. All of these programs have been subjected to impact evaluations of different kinds and some also to rigorous cost-benefit analyses. Many were found to be effective in increasing women's quantity of work as measured by increased rates of labor market participation and number of hours worked. In some cases, the programs also increased women's quality of work, for example, by increasing the capacity for women to work in the formal rather than the informal sector where wages are higher and where women are more likely to have access to health, retirement, and other benefits.

Active Labor Market Programs are often adopted by countries as a way of ameliorating the effects of economic shocks. These programs are not usually targeted at women, although women often seek out the employment and training services offered by these programs. The Latin American programs surveyed in this paper were implemented in Argentina, Mexico, and Peru. The majority were found to increase women's employment rates and to increase their exit rate out of unemployment. Oftentimes, the program impacts for women exceeded those for men. Most of the programs did not, however, lead to substantial wage increases, with the exception of

the ProJoven program in Peru, which was targeted at younger individuals and was found to increase wages and income.

On the whole, the evidence suggests that many of the ALMP programs in Latin America have been effective in reducing labor market frictions and facilitating firm-worker matches but have not been very effective in increasing worker productivity. It is perhaps not surprising that productivity cannot be greatly enhanced with relatively short-term types of interventions. Only a few ALMP programs have been subjected to rigorous evaluation of rates of return, making it difficult to assess which programs have generated benefits that exceeded program costs. Aedo and Nunez (2004) find that program impacts have to be sustained over the longer term (9 years or more) for the program to be cost-effective.

We also surveyed the results of evaluations of ALMP programs in the transition economy settings of Russia, Romania, Slovakia, and Poland. As with the Latin American programs, the evaluations typically find positive effects of program participation for women on employment outcomes, again with less support for the impact of the program on wage levels. There is some evidence suggesting that ALMP programs have helped workers find jobs more quickly but at the expense of lower wages than they would otherwise be able to obtain searching on their own. It would be useful to more systematically analyze the effects of these programs within a job search model to gain a better understanding of the mechanisms through which the programs operate, that is, how they affect the costs of searching for a job, the arrival rate of offers and the distribution of wage offers.²⁹

The evidence from Russia and Romania suggests that the effects of the program on workers were heterogeneous and that highly educated workers on the whole did not benefit from ALMP programs. The types of skills and job training services offered by these programs may not have been well suited to highly educated and more specialized workers. More evidence is needed on how these programs can be tailored to the needs of the workers participating in them. The few studies that examine longer term program effects, such as that of Kluve, Lehmann and Schmidt (1999), find that the positive program effects for women were sustained over a longer 18-month timeframe.

²⁹ For example, Lise, Seitz and Smith (2003) make some progress towards understanding the effects of a Canadian social program using an equilibrium job search model.

In many countries, ALMP programs operate on a very large scale. An issue that was not addressed in the evaluation studies is to what extent large government-sponsored training programs crowd out private training. Additionally, it is possible that the benefits observed for program beneficiaries are at the expense of nonbeneficiaries, because of displacement, substitution effects or lower wages. For example, a wage subsidy might encourage a firm to hire a particular worker rather than some other worker. These questions warrant more investigation to more fully assess of the effectiveness of these programs on the general population.

Our review of microlending programs found mixed evidence on their effectiveness and on the benefits of gender-targeted lending. For example, Pitt and Khandker's (1998) evaluation of a program in Bangladesh finds substantial benefits of targeting microcredit to women relative to men, as measured by effects on consumption, labor supply and children's school-going. However, Roodman and Morduch's (2009) replication and extension of their analysis indicates that the results are highly sensitive to changes in the specifications used in estimation, questions the validity of some of the instruments and finds estimates to suggest impact estimates in the opposite direction.

The de Mel, Duresh, McKenzie and Woodruff (2007) study also raises questions about the rationale for gender targeting, because the female-owned enterprises exhibited much lower rates of return on capital (given through grants rather than credit) than male-owned enterprises in their study of Sri Lankan microenterprises. Female borrowers also tended not to come from the poorest families, so that gender targeting may have been at odds with poverty targeting. The evidence for Pakistan (Hussein and Hussein (2003)), however, shows that when loans were not targeted at women, they tended to go to men and even when they were targeted at women, male relatives tended to appropriate them. Even so, targeting them toward women appears to have affected their bargaining power within the household. The degree to which gender targeting makes a difference to how funds are used and to female empowerment is likely to depend on the country context.

One of the studies reviewed by Kaboski and Townsend's (2007) uses a theoretical model to study the relative effectiveness of microcredit (non-gender targeted) versus cash transfers as a way of increasing the consumption and utility of the poor. Interestingly, they show that even people who did not avail themselves of the credit benefited from it, in the sense of having been able to better smooth their consumption, knowing that credit was available. They also conclude

that a microcredit program was more effective than a conditional cash transfer program in increasing consumption, because the credit was used for productive purposes. More evidence needs to be accumulated on the relative merits of different kinds of programs, both from a theoretical and empirical perspective.

In summary, many of the evaluation studies of microcredit programs document high rates of return to such programs, but the returns can be compromised by targeting the loans to narrowly targeted specific groups. In implementation, the relative benefits of targeting loans at women or at poor borrowers, in terms of redistribution, need to be weighed against the benefits of allocating funds to the highest return activities.

We also reviewed two evaluation studies of land titling programs that were implemented in Peru and Argentina. Such programs are an innovative approach to increasing the wealth of poor households and are observed to change an array of household behaviors, including the labor supply of household members. The land titling programs we reviewed increased female labor supply, increased the tendency for women to work outside the home, and decreased fertility.

In recent years, conditional cash transfer (CCT) programs have spread throughout the developing world as a new strategy for reducing poverty and encouraging investment in children's schooling. Many of the CCT programs currently in operation give the transfers to mothers, under the assumption that mothers are more apt to spend the transfers on children, are less likely to spend them on alcohol and that providing transfers to women improves their relative bargaining position within the household. The Mexican *Oportunidades* program led to major changes in the share of household income coming from mothers. The program was found to substantially increase school-going and decrease work of children and adolescents. The program does not appear to have had substantial effects on mothers' labor force participation, at least in rural areas where impacts on mothers' work have been examined. An added bonus is that many beneficiary women have reported other kinds of benefits, such as higher rates of participation in household decision-making. For a small fraction of households, though, wives have reported a higher rate of aggressive behavior, suggesting that care needs to be taken in implementing CCT programs to take into account their potential for increasing domestic violence.

Most studies of CCT programs focus on their effects on schooling and health. Thus far, the effects of CCT programs on women's labor force participation have not yet been systematically

studied, particularly in the urban context. In rural areas, the evaluation results of the Mexican program indicate that the program did not substantially impact mothers' labor force participation status. However, the program led to increases in mothers' time spent caring for children, which substituted for less time devoted to childcare by older girls in the household who were now in school. In urban areas, there was some evidence that some families did not take up the program in part because working mothers were unable to meet the time requirements of participating in the CCT program (e.g. attending meetings). This suggests that mothers' work competed with meeting the requirements to participate in the program. Maluccio's (2005) study of the RED program in Nicaragua demonstrates that it provided a valuable source of social protection for vulnerable populations during economic crises and mitigated the effects of economic shocks on household labor supply.

Another class of policies we reviewed that are designed to influence women's working behavior are policies that affect the availability and pricing of childcare. Our review of childcare programs in Guatemala, Colombia and Argentina found strong effects of childcare availability on mothers' rates of working and on the number of hours worked. The community daycare programs also had substantial positive benefits on the nutrition and development of the young children participating in them, that imply high benefit-cost ratios. In the context of Accra, Ghana, though, the local supply of childcare was found not to be a significant determinant of mothers' working decisions, because mothers mainly worked in the informal sector where it was easier to combine work with taking care of children. Thus, the effectiveness of childcare programs in increasing mothers' labor supply and hours worked is likely to depend on the country context.

Some strong evidence for a substantial effect of childcare costs on working behavior comes from formerly Communist countries like Romania and Russia that have undergone very large changes in the costs of childcare. Some countries went from childcare being free and widely available to all women to a situation where childcare was expensive and a major determinant of whether women worked. The two studies we reviewed by Lockshin (1999) and Fong and Lockshin (2000) find women's labor supply to be fairly elastic with respect to the price of childcare. An important finding is that childcare subsidies were more effective than wage subsidies or family income subsidies in increasing family income levels, in part because of better targeting at women whose labor force behavior was potentially affected by the policy. There

needs to be further examination, however, of how these alternative types of programs compare on a utility basis and further study of how children are affected by being placed in formal care.

References

- Abadie, Alberto and Guido W. Imbens (2005): “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, Vol. 74, No. 1, 235-267.
- Adato, Michelle, Bénédicte de la Briere, Dubravka Mindek and Agnes Quisumbing (2000), “The Impact of ProgresA on Women’s Status and Intrahousehold Relations,” IFPRI Report, Washington, DC.
- Aedo, Cristian and Sergio Nunez (2004): “The Impact of Training Policies in Latin America and the Caribbean: The Case of Programa Joven,” IDB working paper #R-483.
- Angelucci, Manuela (2008): “Love on the Rocks: Aggressive Behavior and Alcohol Abuse in Rural Mexico,” Working Paper, University of Arizona.
- Aiyagari, S. Rao, “Uninsured Idiosyncratic Risk and Aggregate Saving,” *Quarterly Journal of Economics*, Vol. 109, No. 3, 659-684.
- Attanasio, Orazio and Marcos Vera-Hernandez (2004): “Medium and Long Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Programme in Rural Colombia,” Institute for Fiscal Studies, Working Paper 04/06, University College, London.
- Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan (2009): “The mirable of microfinance? Evidence from a randomized evaluation,” Working Paper, MIT.
- Behrman, Jere R., Yingmei Cheng and Petra E. Todd (2004): “Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach,” *Review of Economics and Statistics*, Vol. 86, No. 1, 108-132.
- Behrman, Jere R., Piyali Sengupta and Petra E. Todd (2005): “Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment,” *Economic Development and Cultural Change*, Vol. 54, No. 1, 237-275.

Behrman, Jere R., Susan W. Parker and Petra E. Todd (2011): “Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Follow-Up of PROGRESA/*Oportunidades*,” *Journal of Human Resources*, Vol. 46, No. 1, 93-122.

Behrman, Jere R. and Elizabeth M. King (2008): “Program Impact and Variation in Duration of Exposure,” in Samia Amin, Jishnu Das and Markus Goldstein, eds. *Are You Being Served: New Tools for Measuring Service Delivery*, Washington, DC: World Bank, 147-172.

Benus, Jacob, Raluca Catrinel Brinza, Casilica Cuica, Irina Denisova, and Marina Kartseva (2005): “Retraining Programs in Russia and Romania: Impact Evaluation Study,” CEFIR research papers, <http://www.cefir.ru/index.php?l=eng&id=34&yf=2004>.

Berlinski, Samuel and Sebastian Galiani (2007): “The Effect of a Large Expansion of Preprimary School Facilities on Preschool Attendance and Maternal Employment,” *Labour Economics*, Vol. 14, 665-680.

Betcherman, Gordon, Karina Olivas and Amit Dar (2004): “Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries,” Social Protection Discussion Paper Series No. 0402, The World Bank.

Bidani, Benu, Chorching Goh and Christopher J. O’Learly (2002): “Has Training Helped Employ Xiagang in China? A Tale from Two Cities,” World Bank manuscript.

Blau, David M. and Philip K. Robins (1988): “Child-Care Costs and Family Labor Supply,” *Review of Economics and Statistics*, Vol. 70, No. 3, 374-381.

Buddelmeyer, Hilke, and Emmanuel Skoufias (2003): “An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA,” IZA Discussion Paper No. 827 (July), Institute for the Study of Labor (IZA), Bonn, Germany.

Burdett, Kenneth and Dale Mortensen (1998): “Wage Differentials, Employer Size, and Unemployment,” *International Economic Review*, Vol. 39, No. 2, 257-273.

Cleveland, Gordon, Morley Gunderson, and Douglas Hyatt, (1996) “Child Care Costs and the Employment Decision of Women: Canadian Evidence” *Canadian Journal of Economics*, Vol. 29, No. 1, 132-51.

Connelly, Rachel, (1992) “The Effect of Child Care Costs on Married Women’s Labor Force Participation,” *Review of Economics and Statistics*, Vol. 74, No. 1, 83-90.

Deaton, Angus (1991): “Saving and Liquidity Constraints,” *Econometrica*, Vol. 59, No. 5, 1221-1248.

Deaton, Angus (2009): “Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development,” NBER working paper No. W14690.

Del Boca, Daneila, Silvia Pasqua, Chiara Pronzato and Cecile Wetzels (2007): “An Empirical Analysis of the Effects of Social Policies on Fertility, Labour Market Participation, and Hourly Wages of European Women,” in Daniela Del Boca and Cecile Wetzels, eds., *Social Policies, Labour Markets and Motherhood*, Cambridge University Press, 269-292.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2008): “Returns to Capital in Microenterprises: Evidence from a Field Experiment,” in *Quarterly Journal of Economics*, Vol. 123, No. 4, 1329-1372.

Dulfo, Esther and Michael Kremer (2004): “Use of Randomization in the Evaluation of Development Effectiveness” in *Evaluating Development Effectiveness*, ed. George Pitman, Osvaldo Feinstein, Gregroy Ingram, World Bank.

Dupas, Pascaline and Jonathan Robinson (2009): “Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya,” NBER Working Paper No. 14693.

Field, Erica (2003a): “Fertility Responses to Urban Land Titling Programs: The Roles of Ownership Security and the Distribution of Household Assets,” Working Paper, Harvard University.

Field, Erica (2003b): "Entitled to Work: Urban Property Rights and Labor Supply in Peru," manuscript, Harvard University.

Fiszbein, Ariel, Norbert Schady, Francisco H. G. Ferreira, Margaret Grosh, Nial Kelleher (2009): "Conditional Cash Transfers: Reducing Present and Future Poverty," World Bank Publications.

Fong, Monica and Michael Lokshin (2000): "Child Care and Women's Labor Force Participation in Romania," Policy Research Working Paper #2400, the World Bank.

Galasso, Emanuela and Martin Ravallion (2004): "Social Protection in a Crisis: Argentina's Plan Jefes y Jefas," *The World Bank Economic Review*, Vol. 18, No. 3, 367-99.

Galasso, Emanuela, Martin Ravallion and Agustin Salvia (2001): "Assisting the Transition from Workfare to Work: A Randomized Experiment in Argentina," World Bank manuscript.

Galiani, Sebastian and Ernesto Schargrodsky (2009): "Property Rights for the Poor: Effects of Land Titling," Ronald Coase Institute Working Paper Series, No. 7.

Gertler, Paul (2004): "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment," *The American Economic Review*, Vol. 94, No. 2, 336-41.

Gugerty, Mary Kay and Michael Kremer (2008): "Outside Funding and the Dynamics of Participation in Community Organizations," *American Journal of Political Science*, Vol. 52, No. 3, 585-602.

Gulametova-Swann, Michaela (2009): "Evaluating the Impact of Conditional Cash Transfer Programs on Adolescent Decisions about Marriage and Fertility: the Case of *Oportunidades*," Ph.D. Dissertation, University of Pennsylvania.

Heckman, James and Richard Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions," in James Heckman and Burton Singer, eds., *Longitudinal Analysis of Labor Market Data*, Cambridge, England: Cambridge University, 156-246.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd (1998): “Characterizing Selection Bias Using Experimental Data,” *Econometrica* , Vol. 66, No. 5, 1017-1098.

Heckman, James, Hidehiko Ichimura and Petra Todd (1997): “Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program,” *Review of Economic Studies*, Vol. 64, No. 4, 605-654.

Heckman, James, Hidehiko Ichimura and Petra Todd (1998), “Matching As An Econometric Evaluation Estimator,” *Review of Economic Studies*, Vol. 65, No. 2, 261-294.

Heckman, James, Robert Lalonde and Jeffrey Smith (1999): “The Economics and Econometrics of Active Labor Market Programs” in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics* Volume 3A, Amsterdam: North-Holland, 1865- 2097.

Heckman, James and Edward Vytlacil (2005): “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, Vol. 73, No. 3, 669-738.

Hussein, Maliha and Shazreh Hussein (2003): “The Impact of Micro Finance on Poverty and Gender Equity: Approaches and Evidence from Pakistan,” Pakistan Micro Finance Network Working Paper.

Imbens, Guido W., and Joshua D. Angrist (1994): “Identification of Local Average Treatment Effects,” *Econometrica* Vol. 62, No. 2, 467-475.

Jalan, Joytsna and Martin Ravallion (2003): “Estimating the Benefit Incidence of an Anti-Poverty Program using Propensity Score Matching,” *Journal of Business and Economic Statistics*, Vol. 21, No. 1, 19-30.

Kaboski, Joseph P. and Robert M. Townsend (2007): “Testing a Structural Model of Credit Constraints Using a Large-Scale Quasi-Experimental Microfinance Initiative,” manuscript, MIT.

Karlan, Dean and Martin Valdivia (2006): “Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions,” manuscript, Yale University.

Karlan, Dean and Jonathan Zinman (2007): “Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts,” Economic Growth Center Discussion Paper No. 956, Yale University.

Khandker, Shahidur R., Hussain A, Samad, and Zahed H. Khan (1998): “Income and Employment Effects of Micro-credit Programmes: Village-level Evidence from Bangladesh,” *Journal of Development Studies*, Vol. 35, No. 2, 96-124.

King, Elizabeth M. and Jere R. Behrman (2009): “Timing and Duration of Exposure in Evaluations of Social Programs,” *World Bank Research Observer*, Vol. 24, No. 1, 55-82.

Kluve, Jochen, Lehmann, Hartmut and Christoph Schmidt (1999): “Active Labor Market Policies in Poland: Human Capital Enhancement, Stigmatization or Benefit Churning,” *Journal of Comparative Economics*, Vol. 27, No. 1, 61-89.

Landsberger, Henry (1948): *Hawthorne Revisited*, Cornell University Press.

Lise, Jeremy, Seitz, Shannon and Jeffrey Smith (2003): “Equilibrium Policy Experiments and the Evaluation of Social Programs,” IZA discussion paper No. 758.

Lockshin, Michael M. (1999): “Household Child Care Choices and Women’s Work Behavior in Russia,” Policy Research Working Paper #2206, The World Bank.

Lubyova, Martina and Jan C. van Ours (1999): “Effects of Active Labor Market Programs on the Transition Rate from Unemployment into Regular Jobs in the Slovak Republic,” *Journal of Comparative Economics*, Vol. 27, No. 1, 90-112.

Maluccio, John A. (2005): “Coping with the ‘Coffee Crisis’ in Central America: The Role of the Nicaraguan Red de Proteccion Social (RPS),” IFPRI Discussion Paper BRIEFS No. 188.

Morduch, Jonathan (1994): “Poverty and Vulnerability,” *American Economic Review*, Vol. 84, No. 2, 221-25.

Morduch, Jonathan (1995): "Income Smoothing and Consumption Smoothing," *Journal of Economic Perspectives*, Vol. 9, No. 3, 103-14.

Ñopo, Hugo, Miguel Robles and Jaime Saavedra (2007): "Occupational Training to Reduce Gender Segregation: The Impacts of ProJoven," Inter-American Development Bank Working Paper No. 623.

Paolisso, Michael J., Kelly Hallman, Lawrence Haddad and Shibesh Regmi (2001): "Does Cash Crop Adoption Detract from Childcare Provision? Evidence from Rural Nepal," FCND Discussion Paper No. 109, IFPRI, Washington, DC.

Parker, Susan and Emmanuel Skoufias (2000): "The Impact of PROGRESA on Work, Leisure and Time Allocation," International Food Policy Research Institute, Final Report.

Pitt, Mark and Shahidur R. Khandker (1998): "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy*, Vol. 106, No. 5, 958-996.

Quisumbing, Agnes R., Kelly Hallman and Marie T. Tuel (2003): "Maquiladoras and Market Mamas: Women's Work and Childcare in Guatemala City and Accra," IFPRI Discussion Paper No. 153, Food Consumption and Nutrition Division.

Revenge, Ana, Michelle Riboud, and Hong Tan (1994): "The Impact of Mexico's Retraining Program on Employment and Wages," *World Bank Economic Review*, Vol. 8, No. 2, 247-277.

Roodman, David and Jonathan Morduch (2009): "The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence," Working paper, Center for Global Development and New York University.

Rosenbaum, Paul and Donald Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70, No. 1, 41-55.

Rosenbaum, Paul and Donald Rubin (1985): “Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score,” *American Statistician*, Vol. 39, No. 1, 33-38.

Ruel, Marie T. and Agnes R. Quisumbing with Kelly Hallman (2006): “The Guatemala Community Day Care Program,” IFPRI Research Report No. 144, Washington, DC.

Schultz, T. Paul (2000): “Impact of PROGRESA on school attendance rates in the sampled population,” February. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C.

Schultz, T. Paul (2004): “School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program,” *Journal of Development Economics*, Vol. 74, No. 2, 199-250.

Skoufias, Emmanuel and Bonnie McClafferty (2001): “Is PROGRESA Working? Summary of the Results of an Evaluation by IFPRI.” Report submitted to PROGRESA. Washington, D.C.: International Food Policy Research Institute, <http://www.ifpri.org/themes/progres.htm>.

Smith, Jeffrey and Petra E. Todd (2005): “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?” *Journal of Econometrics*, Vol. 125, 305-353.

Todd, Petra E. and Kenneth Wolpin (2006): “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility,” *American Economic Review*, Vol. 96, No. 5, 1384-1417.

Woolcock, Michael J. V. (1999): “Learning from Failures in Micro-Finance: What Unsuccessful Cases Tell Us about How Group-Based Programs Work,” *The American Journal of Economics and Sociology*, Vol. 58, No. 1, 17-42.

Appendix A: Nonexperimental Evaluation Estimators

This appendix describes in greater detail some of the evaluation estimators that are widely used in evaluating effects of programs. Different types of estimators are valid under different assumptions on the processes generating outcomes and participation decisions and on the observability of key data elements.

Consider the following statistical model for outcomes. Suppose the outcomes in the treated and untreated states can be written as an additively separable function of some observables (X) and unobservables (U_0 and U_1):

$$Y_1 = \varphi_1(X) + U_1$$

$$Y_0 = \varphi_0(X) + U_0$$

For example, earnings might depend on some observables such as education level and also on unobservables (such as personality traits).

In this notation, the observed outcome $Y = DY_1 + (1 - D)Y_0$ can be written as:

$$Y = \varphi_0(X) + D(\varphi_1(X) - \varphi_0(X)) + \{U_0 + D(U_1 - U_0)\}.$$

Assume that the unobservables are mean independent of the observables, $E(U_0|X) = E(U_1|X) = 0$. The gain to an individual from participating in the program is $\Delta = D(\varphi_1(X) - \varphi_0(X)) + D(U_1 - U_0)$. Individuals may or may not know their values of U_1 and U_0 at the time of deciding whether to participate in a program. If people self-select into the program based on their anticipated gains from the program, then we would expect that $E(U_0|X, D) \neq 0$ and $E(U_1|X, D) \neq 0$. In other words, if people know their values of U_1 and U_0 , or can forecast the values, then we would expect people to use of any information they have on U_1 and U_0 when they decide whether to participate in the program.

In terms of the statistical model, the *average impact of treatment (ATE)* for a person with characteristics X is

$$E(\Delta|X) = \varphi_1(X) - \varphi_0(X) + E(U_1|X) - E(U_0|X) = \varphi_1(X) - \varphi_0(X).$$

The *average impact of treatment on the treated (TT)* is:

$$E(\Delta|X) = \varphi_1(X) - \varphi_0(X) + E(U_1 - U_0|X, D = 1).$$

For completeness, we can also define the *average impact of treatment on the untreated (UT)* as

$$E(\Delta|X) = \varphi_1(X) - \varphi_0(X) + E(U_1 - U_0|X, D = 0).$$

This parameter gives the impact of a program or intervention on the group that currently does not participate in it. It is usually of more limited interest, but might be informative if there are plans to expand the scope of the program to include nonparticipants. The relationship between TT , ATE and UT is:

$$ATE = \Pr(D = 1|X) TT + \Pr(D = 0|X) UT.$$

If $U_1 = U_0$, then the TT , ATE and UT parameters are all the same. Another special case where the parameters may be the same even if $U_1 \neq U_0$ arises when $E(U_1 - U_0|X, D) = 0$.

Under this case, the participation decision D is uninformative on $U_1 - U_0$, but it need not be the case that $U_1 = U_0$. This assumption might be reasonable if agents making the participation decisions (e.g. individuals, program administrators or others) do not act on $U_1 - U_0$, perhaps because they cannot forecast their own idiosyncratic gain from participating in the program at the time of making participation decisions. In this special case, there is said to be *ex-post* heterogeneity in how people respond to treatment, which is not acted upon *ex-ante*.

As discussed in Heckman, Lalonde and Smith (1999), there are three different types of assumptions that can be made in the evaluation model. In order of increasing generality, the assumptions are:

(A.1) conditional on X , the program effect is the same for everyone ($U_1 = U_0$)

(A.2) the program effect (given X) varies across individuals but $U_1 - U_0$ does not help predict program participation

(A.3) the program effect (given X) varies across individuals and $U_1 = U_0$ predicts who participates in the program.

Ideally, it would be desirable to use an estimation method that allows for the highest level of generality. There are estimators that allow for (A.3), although they usually require making distributional assumptions on the unobservables or assuming the existence of instrumental variables that affect the decision to participate in the program but not the outcome equation.

Most of the evaluation studies reviewed in this paper use matching estimators that are operational primarily under assumption (A.2). However, Panel data versions of matching estimators can to a limited extent accommodate (A.3).

Sources of bias in estimating $E(\Delta|X, D = 1)$ and $E(\Delta|X)$

Consider the previously described model of outcomes:

$$Y = \varphi_0(X) + D(\varphi_1(X) - \varphi_0(X)) + \{U_0 + D(U_1 - U_0)\}.$$

In terms of the TT parameter, $TT = E(\Delta|X, D = 1)$, the model can be written as:

$$Y = \varphi_0(X) + DE(\Delta|X, D = 1) + \{U_0 + D[U_1 - U_0 - E(U_1 - U_0|X, D = 1)]\}.$$

For simplicity, suppose the X variables are discrete and that we estimate the effects of the intervention (D) by the coefficients \hat{b}_x from an ordinary least squares regression:³⁰

$$y = aX + b_xXD + v$$

This model, which is popular in applied work, is known as the *common effect model*. If the goal is to estimate TT , then, under assumptions A.1-A.3, bias arises if $E(U_0|X, D) \neq 0$.

Two widely used simple regression methods for estimating the (TT) parameter, $E(\Delta|X, D = 1)$, using nonexperimental data are the *cross-section* estimator, and the *difference-in-difference* estimator. Extensions to estimating the ATE parameter are straightforward and are discussed in Heckman, Lalonde and Smith (1999).

Consider a panel data regression framework where it is assumed that there is access to data on outcomes and conditioning variables for multiple time periods. Denote the outcome measures by Y_{1it} and Y_{0it} , where i denotes the individual and t the time period of observation,

$$Y_{1it} = \varphi_1(X_{it}) + U_{1it}$$

$$Y_{0it} = \varphi_0(X_{it}) + U_{0it}$$

U_{1it} and U_{0it} are assumed to be distributed independently across persons and to satisfy $E(U_{1it}|X_{it}) = 0$ and $E(U_{0it}|X_{it}) = 0$. X_{it} represents conditioning variables that may either be fixed or time-varying (such as gender or age), but whose distributions are assumed to be unaffected by whether an individual participates in the program.³¹ The observed outcome at time t can be written as

$$Y_{it} = \varphi_0(X_{it}) + D_{it}\alpha^*(X_{it}) + U_{0it},$$

where D_{it} denotes being a program participant and $\alpha^*(X_{it}) = \varphi_1(X_{it}) - \varphi_0(X_{it}) + U_{1it} - U_{0it}$ is the program impact.³² Prior to the program intervention, we observe $Y_{0it} = \varphi_0(X_{it}) + U_{0it}$ for everyone. After the intervention we observe $Y_{1it} = \varphi_1(X_{it}) + U_{1it}$ for those who received the

³⁰ Here, we allow the effects of treatment to differ by the observed X , as reflected in the X subscript on b .

³¹ For example, if the set of conditioning variables X_{it} includes marital status and the program intervention is a job training program, we need to assume that the job training program does not affect marital status.

³² This model is a random coefficient model, because the impact of treatment can vary across persons even after conditioning on X_{it} . Assuming that $U_{0it} = U_{1it} = U_{it}$, so that the unobservable is the same in both the treated and untreated states, yields the fixed coefficient or *common effect* version of the model.

intervention (for whom $D_{it} = 1$, for $t > t_0$, the time of the intervention) and $Y_{oit} = \varphi_0(X_{it}) + U_{oit}$ for those who did not receive it (for whom $D_{it} = 0$ in all time periods).

Cross-section Estimator. A *cross-section* estimator uses data on a comparison group of nonparticipants to impute counterfactual outcomes for program participants. The data requirements of this estimator are minimal, only one post-program cross section of data on $D_{it} = 1$ and $D_{it} = 0$ persons. Define $\hat{\alpha}_{CS}$ as the ordinary least squares solution to α^* in

$$Y_{it} = \varphi_0(X_{it}) + D_{it}\alpha^* + \varepsilon_{it},$$

where the regression is estimated using data measured at time t on persons for which $D_{it} = 1$ and $D_{it} = 0$. Consistency requires that $E(\varepsilon_{it}|D_{it}, X_{it}) = 0$. In a general model, where $U_{oit} \neq U_{1it}$, this restriction rules out the possibility that people select into the program based on expectations about their idiosyncratic gain from the program, and therefore cannot accommodate assumption (A.3).

Difference-in-Differences Estimator. The *difference-in-differences* (*DID*) estimator is probably the most commonly used evaluation estimator. It measures the impact of the program intervention by the difference in the change in outcomes between participants and nonparticipants before and after the program, which requires access to both pre- and post-program data (t and t' data) on program participants and nonparticipants. Define an indicator I_i^D that equals 1 for participants (for whom $D_{it'} = 0$ and $D_{it} = 1$), and 0 otherwise.

The difference-in-differences estimator is the least squares solution for α^* in

$$Y_{it} - Y_{it'} = \varphi_0(X_{it}) - \varphi_0(X_{it'}) + D_{it}\alpha^* + \{U_{it} - U_{it'}\}$$

which allows for individual fixed effects that are differenced-out. Alternatively, the *DID* estimator is often implemented using a regression

$$Y_{it} = \varphi_0(X_{it}) + I_i^D + D_{it}\alpha^* + U_{it} \text{ for } t = t', \dots, t$$

where I_i^D is an intercept that denotes whether a member of the treatment group.³³ This regression is estimated using participant ($D_{it'} = 0, D_{it} = 1$) and nonparticipant ($D_{it'} = 0, D_{it} = 0$) observations. The *DID* estimator allows for time-specific intercepts that are common across

³³ The specification could include individual-specific fixed effects, but estimating them consistently would require an assumption that the panel length T go to infinity.

groups (they can be included in $\varphi_0(X_{it})$). The estimator is consistent if $E(U_{it} - U_{it'} | D_{it} - D_{it'}, X_{it}) = 0$.³⁴

9.1 Matching Methods

Matching is a widely-used method of evaluation that compares the outcomes of program participants with the outcomes of similar, matched nonparticipants. Some of the earliest applications of matching to evaluate economic development programs were carried out to evaluate World Bank programs. One of the main advantages of matching estimators over other kinds of evaluation estimators is that they do not require specifying the functional form of the outcome equation and are therefore not susceptible to misspecification error along that dimension. For example, they do not require specifying that outcomes are linear in observables.

Traditional matching estimators pair each program participant with an observably similar nonparticipant and interpret the difference in their outcomes as the effect of the program intervention (see, e.g., Rosenbaum and Rubin, 1983). More modern methods pair program participants with more than one nonparticipant observation, using statistical methods to estimate the matched outcome. Here, we focus on a class of matching estimators called *propensity score matching estimators*, which are the most commonly used.

Matching estimators typically assume that there exist a set of observed characteristics Z such that outcomes are independent of program participation conditional on Z . That is, it is assumed that the outcomes (Y_0, Y_1) are independent of participation status D conditional on Z ,³⁵

$$(Y_0, Y_1) \perp D | Z. \quad (1)$$

It is also assumed that for all Z there is a positive probability of either participating ($D = 1$) or not participating ($D = 0$) in the program, i.e.,

$$0 < Pr(D = 1 | Z) < 1. \quad (2)$$

This second assumption is required so that a matches for $D = 0$ and $D = 1$ observations can be found. If assumptions (1) and (2) are satisfied, then the problem of determining mean program impacts can be solved simply by substituting the Y_0 distribution observed for the matched non-participant group for the missing Y_0 distribution for program participants.

³⁴ If the error terms follow a fixed effect error structure, then program participation can depend on unobservable, fixed attributes of persons.

³⁵ In the terminology of Rosenbaum and Rubin (1983) treatment assignment is “strictly ignorable” given Z .

Heckman, Ichimura and Todd (1998) show that the above assumptions are overly strong if the parameter of interest is the mean impact of treatment on the treated (TT), in which case a weaker conditional mean independence assumption suffices:

$$E(Y_0|Z, D = 1) = E(Y_0|Z, D = 0) = E(Y_0|Z). \quad (3)$$

Furthermore, identifying the TT parameter requires only

$$Pr(D = 1|Z) < 1. \quad (4)$$

Under these assumptions, the mean impact of the program on program participants can be written as

$$\begin{aligned} \Delta &= E(Y_1 - Y_0|D = 1) \\ &= E(Y_1|D = 1) - E_{Z|D=1}\{E_Y(Y|D = 1, Z)\} \\ &= E(Y_1|D = 1) - E_{Z|D=1}\{E_Y(Y|D = 0, Z)\}, \end{aligned}$$

where the second term can be estimated from the mean outcomes of the matched (on Z) comparison group.³⁶ Assumption (3) implies that D does not help predict values of Y_0 conditional on Z . Thus, selection into the program cannot be based directly on values of Y_0 . However, no restriction is imposed on Y_1 , so the method does allow individuals to be electing into the program on the basis of Y_1 . Thus, it permits to a limited extent assumption (A-3).

With nonexperimental data, there may or may not exist a set of observed conditioning variables for which (1) and (2) hold. A finding of Heckman, Ichimura and Todd (1997) in their application of matching methods in the context of evaluating a job training program is that (2) was not satisfied, meaning that for a fraction of program participants no match could be found. If there are regions where the support of Z does not overlap for the $D = 1$ and $D = 0$ groups, then matching is only justified when performed over the *region of common support*.³⁷ The estimated treatment effect must then be defined conditionally on the region of overlap.

Matching can be difficult to implement when the set of conditioning variables Z is large.³⁸ Rosenbaum and Rubin (1983) provide a theorem that is useful in reducing the dimension of the conditioning problem in implementing the matching method. They show that when Y_0 outcomes are independent of program participation conditional on Z , they are also independent of

³⁶ The notation $E_{Z|D=1}$ denotes that the expectation is taken with respect to the $f(Z|D = 1)$ density.

³⁷ An advantage of randomized experiments noted by Heckman (1997), Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1998), is that they guarantee that the supports are equal across treatments and controls, so that the mean impact of the program can always be estimated over the entire support.

³⁸ If Z is discrete, small cell problems may arise. If Z is continuous and the conditional mean $E(Y_0|D = 0, Z)$ is estimated nonparametrically, then convergence rates will be slow due to the “curse of dimensionality” problem.

participation conditional on the probability of participation, $P(Z) = \Pr(D = 1|Z)$. Thus, when matching on Z is valid, matching on the summary statistic $\Pr(D = 1|Z)$ (the *propensity score*) is also valid. For this reason, much of the matching literature focuses on so-called propensity score matching.³⁹

Using the Rosenbaum and Rubin (1983) theorem, the matching procedure can be broken down into two stages. In the first stage, the propensity score $\Pr(D = 1|Z)$ is estimated, using a binary discrete choice model such as a logit or probit. In the second stage, individuals are matched on the basis of their predicted probabilities of participation, obtained from the first stage.

The literature has developed a variety of matching estimators. The most common matching estimator is the *cross-sectional matching estimator*. For notational simplicity, let $P = P(Z)$. A typical cross-section matching estimator takes the form

$$\hat{\alpha}_M = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} [Y_{1i} - \hat{E}(Y_{0i}|D = 1, P_i)]$$

$$\hat{E}(Y_{0i}|D = 1, P_i) = \sum_{j \in I_0} W(i, j) Y_{0j},$$

where I_1 denotes the set of program participants, I_0 the set of non-participants, S_P the region of common support (see below for ways of constructing this set). n_1 denotes the number of persons in the set $I_1 \cap S_P$. The match for each participant $i \in I_1 \cap S_P$ is a weighted average over the outcomes of non-participants, where the weights $W(i, j)$ depend on the distance between P_i and P_j . Alternative matching estimators (discussed below) differ in how the neighborhood is defined and in how the weights $W(i, j)$ are constructed. See Todd (2008) for a discussion of various approaches.

The cross-sectional matching estimator assumes that after conditioning on a set of observable characteristics, mean outcomes are conditionally mean independent of program participation. For a variety of reasons, though, there may be systematic differences between participant and nonparticipant outcomes, even after conditioning on observables. Such differences may arise, for example, because of program selectivity on unmeasured characteristics, or because of levels

³⁹ The reduction in dimensionality only occurs if $P(Z)$ can be estimated parametrically or semiparametrically at a rate faster than the nonparametric rate, otherwise high dimensionality would be of concern in estimating the propensity score.

differences in outcomes across different labor markets in which the participants and nonparticipants reside.

A difference-in-differences (*DID*) matching strategy, as defined in Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1998), allows for temporally invariant differences in outcomes between participants and nonparticipants. This type of estimator is analogous to the standard differences-in-differences regression estimator defined, but it reweights the observations according to the weighting functions used by the propensity score matching estimators defined above. The *DID* matching estimator requires that

$$E(Y_{0t} - Y_{0t'} | P, D = 1) = E(Y_{0t} - Y_{0t'} | P, D = 0),$$

where t and t' are time periods after and before the program enrollment date. This estimator also requires the support condition given in (7), which must now hold in both periods t and t' . The estimator is given by

$$\hat{\alpha}_{KDM} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} \left\{ (Y_{1i} - Y_{0t'}) - \sum_{i \in I_0 \cap S_P} W(i, j) (Y_{0tj} - Y_{0t'j}) \right\},$$

If repeated cross-section data are available, instead of longitudinal data, the estimator can be implemented as

$$\hat{\alpha}_{KDM} = \frac{1}{n_1} \sum_{i \in I_{1t} \cap S_P} \left\{ (Y_{1ti} - \sum_{i \in I_{0t} \cap S_P} W(i, j) Y_{0tj}) \right\} - \frac{1}{n_{1t'}} \sum_{i \in I_{1t'} \cap S_P} \left\{ (Y_{1t'i} - \sum_{i \in I_{0t'} \cap S_P} W(i, j) Y_{0t'j}) \right\},$$

where I_{1t} , $I_{1t'}$, I_{0t} , and $I_{0t'}$ denote the treatment and comparison group datasets in each time period.

An advantage of difference-in-difference matching over cross-sectional matching is that it to some extent allows selection into the program to be based on anticipated gains from the program, in the sense of assumption (A-3) described earlier. That is, D can help predict the value of Y_1 given P . However, the method assumes that D does not help predict changes in the value of $Y_0(Y_{0t} - Y_{0t'})$ conditional on P . Thus, individuals who participate in the program may be the ones who expect the highest values of Y_1 but they may not be systematically different from nonparticipants in terms of their changes in Y_0 .