INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

CENTRO DE INVESTIGACIÓN ECONÓMICA

Discussion Paper Series

Are Labor Markets Segmented in Argentina? A Semiparametric Approach

Sangeeta Pratap
Instituto Tecnológico Autónomo de México

and

Erwan Quintin
Federal Reserve Bank of Dallas

February 2002
Discussion Paper 02-02

Av. Camino a Santa Teresa # 930
Col. Héroes de Padierna
México, D.F. 10700
M E X I C O

Are Labor Markets Segmented in Argentina? A Semiparametric Approach

Sangeeta Pratap Erwan Quintin*

February 13, 2002

Abstract

We use data from Argentina's household survey to evaluate the hypothesis that informal workers would expect higher wages in the formal sector. Using various definitions of informal employment we find that, on average, formal wages are higher than informal wages. Parametric tests suggest that a formal premium remains after controlling for individual and establishment characteristics. However, this approach suffers from several econometric problems, which we address with semiparametric methods. The resulting formal premium estimates prove either small and insignificant, or negative. In other words, we find no evidence that Argentina's labor markets are segmented along formal/informal lines.

Keywords: Segmented Labor Markets; Wages; Semiparametric Methods.

JEL classification: C14; J42

*Instituto Tecnológico Autónomo de México and Federal Reserve Bank of Dallas, respectively. Email: pratap@itam.mx and erwan.quintin@dal.frb.org.
We wish to thank Steve Bronars and all the participants of the ITAM-UT workshop, as well as Daniel Hammermesh, Hugo Hopenhayn, David Kaplan and Torsten Persson for valuable comments. We are also grateful to Fernanda Fenton and Eric Millis for valuable research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of Dallas or the Federal Reserve System.
Corresponding author: Erwan Quintin, Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, TX 75201.

1 Introduction

Dualistic models of labor markets have pervaded the economic development literature since the seminal work of Lewis (1954). According to this view, some workers are unable to find jobs in the formal, regulated sector and must work in firms where wages and benefits are inferior to what they could expect in the formal sector given their personal characteristics (see, for instance, Mazumdar, 1975). In this paper, we evaluate the premise that informal workers would expect higher wages in the formal sector with data from Argentina's permanent household survey for the 1993-1995 time period.

We follow Castells and Portes (1989) and define informal activities as unregulated activities in a context where similar activities are regulated. As a practical matter, we consider various definitions of informal employment based on benefits mandated by Argentina's labor laws. For all our benefits-based definitions, average informal wages are significantly lower than their formal counterparts. The question we ask is whether a formal sector premium remains after controlling for observable differences between workers and jobs. In particular, formal employees tend to be more educated and experienced than informal employees. Furthermore, the proportion of women is higher in the informal sector. Finally, informal employees are more likely to work in small establishments than formal employees.

Regression analysis continues to suggest a formal premium for many subgroups, even after controlling for size and industry effects. Nonetheless, ordinary least square estimates are biased and inconsistent in this context for several reasons, as discussed by Heckman and Hotz (1986). First, individuals may self-select into a given sector based on observed and unobserved characteristics that also affect earnings. Moreover, those estimates are conditional on a given specification of earning functions.

We proceed to use semiparametric estimators to control for the potential misspecification of earning functions and the endogeneity of wage and sectoral employment outcomes. We begin by matching formal workers with informal workers with similar observed characteristics to obtain an average formality premium. The resulting estimate of the formal sector premium does not significantly differ from zero in any of the three years we consider. We then use the panel structure of the data to compute a difference-in-difference estimate of the formal sector premium that partially control for selection effects due to unobserved characteristics. Again, we find no compelling evidence of a positive formal sector premium.

We also produce estimates of the formal sector premium for various subgroups, including women, young workers, and uneducated workers. Formal earnings are not significantly higher than informal earnings for any of those subgroups. Finally, we discuss the potential impact of controlling for the pecuniary value of formal benefits, notably social security. Simple tests, we argue, provide some indication that our findings are robust to accounting for this potential source of bias.

Our analysis confirms the findings of Magnac (1991) for Columbia, Tannen (1991) for Brazil, and Maloney (1999) for Mexico. Those papers argue that the assumption that labor markets are competitive cannot be rejected. But most studies of labor markets in developing nations find that the relationship between earnings and worker characteristics differs across sectors (see, for instance, Mazumdar, 1981, Heckman and Hotz, 1986, Roberts, 1989, Pradhan and van Soest, 1995, Tansel, 1999, and Gong and van Soest, 2001.) Even in the United States, Dickens and Lang (1985, 1988) find "strong" evidence that there are two distinct labor markets with different earning functions.

All these papers rely exclusively on parametric techniques and, therefore, the interpretation of these results is limited by the potential misspecification of earnings functions. Our semiparametric approach partially circumvents those limitations.

In addition, we point out that existing estimates of the formal sector premium are biased upward because they do not take account of establishment and firm size effects. All else equal, larger establishments or firms pay higher wages in Argentina as in most economies, including economies where the informal sector, by all accounts, is small (see Brown and Medoff, 1989, and Oi and Idson, 1999.) Since large establishments tend to emphasize formal employment, the formal sector premium previous studies report could be no more than a size-wage premium. This bias is likely to be particularly strong in studies based on scale-based definitions of informal employment.

The paper also provides a list of facts with which a satisfactory theory of informal economic activities in Latin America should be consistent. Most existing models of the informal sector predict some wage dualism, or rely on the hypothesis that labor markets are segmented. In a direct extension of a model of Harris and Todaro (1970), Fields (1975) assumes agents can either work in the informal sector or devote their time to searching a higher paying formal job. Rauch (1991) describes a general equilibrium model where firms can choose to violate a minimum wage requirement provided they operate a scale smaller than a given detection threshold. Some workers find jobs in large formal firms while a fraction of the labor force must accept lower-paying informal jobs. Fortin et al. (1997) extend Rauch's framework in several directions and evaluate numerically the quantitative impact of various public policies on the size and characteristics of the informal sector. Models of informal activities that, in contrast, do not assume any segmentation between sectors include Loayza (1996) and Sarte (1999).

Developing nations resort to a vast array of public policies to try and reduce tax evasion and improve compliance with labor laws. A better understanding of the causes and consequences of informal economic activities is necessary to measure the impact of those policies. Our results suggest that modeling the informal sector as the disadvantaged end of dualistic labor markets is likely to yield misleading conclusions.

2 Our approach

It is useful to begin by formalizing the wage segmentation hypothesis. To do this, consider an economy populated by heterogenous agents. Those agents differ in terms of a finite list X of personal characteristics. They are employed either in the formal (F) sector or the informal (I) sector. Each sector offers a menu of jobs that differ along several characteristics described by vector Y. In our empirical analysis, those characteristics include industry and establishment size.

Let $w^F(X, Y, \epsilon)$ and $w^I(X, Y, \epsilon)$ denote integrable random variables that give the agent's log earnings in, respectively, the formal and the informal sector, as a function of their personal and job characteristics and exogenous sources of uncertainty denoted by ϵ.[1] The wage segmentation hypothesis can now be stated as:

$$S: \quad E(w^F(X,Y,\epsilon) - w^I(X,Y,\epsilon)|X,Y \in A) > 0$$

for a non-negligible subset A of characteristics.

We test this hypothesis in various ways. We begin by reporting the results of standard wage regression analysis. As mentioned in the introduction, this approach suffers from several econometric problems. To address those problems, we rely on semiparametric methods. The intuition behind this approach can be described quite simply. Given a subsample of interest, each formal worker in the subsample is matched with an informal worker with similar personal and occupational characteristics. We then compute log wage differences between formal workers and their informal counterparts to obtain an average formality premium. Note that one need no longer specify earning functions. Furthermore, the validity of the resulting estimates is not affected by sample selection on the basis of observed characteristics since wage comparisons are only made for workers with similar characteristics.

However, this method requires that the distributions of characteristics across sectors have similar supports. We argue below that this necessary condition is met in the context of this paper. In particular, although informal workers are more likely to work in small establishments, informal employment is present in all establishment size categories. Another practical concern is the difficulty of matching people directly along several characteristics in a finite sample. Instead, we match workers according to their probability of participation in the formal sector. We check directly that this probability of participation is an effective proxy for earning relevant characteristics in our sample.

We first apply those semiparametric methods to evaluate hypothesis S for our entire

[1] Our formal premium estimates are based on log wages so that they can be interpreted in relative terms.

sample. In other words, we test whether:

$$\int (w^F(X, Y, \epsilon) - w^I(X, Y, \epsilon)) d\mu^F(X, Y) \times d\epsilon > 0,$$

where μ^F is the joint distribution of personal and job characteristics in the formal sector. Rejecting this hypothesis would mean that the systematic formal sector premium predicted by most dualistic models of labor markets[2] is not borne out by our data. Next we apply our semiparametric approach to several subgroups of workers. In particular, we do so for subgroups of workers and jobs where wage segmentation is often thought more likely to occur, namely women, uneducated workers and younger workers.

3 The data

We rely on Argentina's biannual household survey for the 1993-1995 time period. This survey collects socio-economic information from a rotating panel of urban households, in May and October of each year. In the survey, households refer to a group of persons who live in the same dwelling and contribute to and/or draw from the same budget.

Each household is sampled four consecutive times. The information is collected via individual visits. A first questionnaire is used to record the basic demographic and dwelling characteristics of the household. Individual questionnaires are then used to collect each household member's basic demographic data, employment status, the revenues and benefits they derive from their primary and secondary occupation,[3] as well as the size of the establishment and the industry in which they work. Between 1993 and 1995, the survey covered over 30,000 households in 25 urban centers. We concentrate on the "Gran Buenos Aires" area, i.e., Buenos Aires and its suburbs. City size and location are important determinants of wages that would complicate the interpretation of our results. Each wave, approximately 4,500 households are surveyed in the Buenos Aires area.

[2]See e.g. Fields (1975), Rauch (1991) and Fortin et al. (1997).

[3]The survey generally abides by the recommendations of the International Labor Organization. In particular, hours worked are reported for a recent week, income is reported by source for a recent month.

We use the survey's information on job benefits to classify salaried workers as formally or informally employed. We consider definitions of informal employment based on various combinations of the mandated benefits included in the survey.

All the results we report in the sequel pertain to real wages, using Argentina's consumer price index as a deflator. We only consider earnings in individuals' primary occupation. While the survey includes some information on any secondary occupation an individual may have, we find that in many cases the breakdown of total hours worked or total labor income by occupation is missing. We also discard individuals whose answers to key questions (given our purposes) appear miscoded or inconsistent with other answers. In particular, we use the fact that the survey asks for income information in two different sections to check for the consistency of income related answers. Appendix A explains how two methods can be used to compute real hourly wages when the individual reports that they have exactly one occupation. Finally, we discard wage earners who report that they work more than 80 hours a week. Our final sample consists of 15,693 observations.

We choose to compare before-tax wages across sectors. In reality, most informal workers are able to evade income taxation. Comparing before-tax wages thus strongly favors the segmentation hypothesis. Accounting for income taxation should only strengthen our results.[4] By comparing wages directly, we also implicitly ignore the pecuniary value of those benefits that only formal workers receive.[5] Section 7.3 discusses the sensitivity of our findings to this assumption.

4 Characteristics of formal and informal workers

Most existing empirical studies use firm or establishment size as a basis for their definition of informal employment.[6] A worker is considered informal if they work for a firm or estab-

[4]Doing this may be difficult however because the appropriate tax rate depends on the household's overall income. Although the survey inquires about income from various sources, that information is often missing and is likely to be unreliable when it is available.

[5]We do control for whether the worker receives an "aguinaldo" (year-end bonus). See appendix A.

[6]See, for instance, ILO (1972), Maloney (1999), Gong and van Soest (2001).

lishment whose employment-size is below an arbitrary threshold. As we argue below, the notion that large firms operate formally while small firms operate informally is a questionable approximation for Argentina. We choose instead to use government mandated benefits to classify workers. The activities of employees who fail to receive those benefits violate labor regulations, and are thus informal according to most definitions. In this section, we compare the characteristics of formal employees and formal jobs to their informal counterparts.

We begin by comparing the average real hourly earnings of formal and informal employees. Table 1 in the appendix shows that average hourly earnings are significantly higher in the formal sector than in the informal sector for all possible benefits-based definitions of informal employment. The first four sections of the table correspond to four types of benefits included in the survey. The first row of each section gives the average hourly wage of workers who receive a given benefit, the second gives the same statistic for workers who do not receive the benefit. The last row of each section provides a t-statistic based on the differences in means for the two subgroups. In all cases, mean equality can be rejected at the 1% level. Mean wages for the entire period are significantly higher for those individuals who receive mandated benefits than for individuals who do not receive them. These findings appear broadly consistent with the segmented view. The question we ask is the extent to which differences in individual and establishment characteristics can account for this pattern.

Henceforth, to simplify the exposition, an employee is considered informal if they do not receive pension or unemployment insurance benefits. Along with severance pay, those two benefits generate the most significant mean differences in table 1. In this sense, this definition of informality favors the segmentation hypothesis. In addition, social security tends to be the typical basis for a definition of informal work in studies that do not rely exclusively on firm size. Results were qualitatively similar for all definitions. According to this definition, informal employment accounts for roughly a third of our sample.

Table 2 in the appendix gives the average characteristics of employed individuals in our sample. Several key differences between sectors emerge. Formal employees tend to be more experienced and educated than informal employees. Observe in addition that the proportion of women is higher among informal employees. Table 2 also shows that formal employees tend to work in larger establishments that informal employees. The fact that small establishments emphasize informal employment suggests that part of the differences in means shown in table 1 stems from the size-wage premium one finds in most nations, including nations where the informal sector is small by most accounts, such as the United States (See Oi and Idson, 1999.)

Note that a large fraction of employment in both sectors is found in small establishments. This feature of the organization of production in emerging nations such as Argentina is well documented (see e.g. Tybout, 2000.) The fraction of employment in small establishments is particularly large in the informal sector. Nevertheless, a non-trivial proportion of informal employees work in establishments with 25 or more employees. This indicates that scale-based definitions of informality, such as the definition commonly used by the International Labor Organization (ILO),[7] would be questionable approximations for Argentina.

The panel structure of our data also enables us to compare the characteristics of individuals who change occupations and sectors to those whose employment status remains the same from one sampling period to the next. Table 3 in the appendix shows that, on average, roughly 10% of formal employees transit to informal employment from one sampling period to the next in our sample, while over 25% of informal employees become formally employed. Table 4 compares the characteristics of employees who change sectors to those who remain in the same sector. Employees who switch from the formal to the informal sector are younger and less educated on average than employees who remain in the formal sector. Conversely, employees who remain in the informal sector tend to be younger and less educated than employees who enter the formal sector. In addition, workers who enter the formal sector see the highest rise in their gross wages.

For the purpose of this paper, two aspects of these mobility patterns are important. First, there is a significant degree of mobility between formal and informal employment, in both directions. In section 6.2, we exploit this feature to partially control for unobserved earnings

[7] See e.g. ILO, 1972.

determinants. Second, people who switch sectors differ significantly from people who stay put along earning-relevant characteristics such as age and education. Section 6.2 estimates of the formal sector premium control for those differences.

It is important to note, however, that the mobility patterns shown in tables 3 and 4 cannot be interpreted as direct evidence or counter-evidence of labor market segmentation. The fact that individuals who enter the formal sector tend to be older and more educated than their counterparts who remain in the informal sector could be the result of barriers to entry for certain subgroups, but it could simply reflect the fact that the two sectors emphasize different skills for other reasons. On the other hand, should one find that people with similar earning relevant characteristics are compensated differently, the hypothesis that labor markets are competitive would be clearly rejected. In this sense, the fact that average wages are higher in the formal sector, together with the large average increase in wages for workers who enter the formal sector, bode well for the segmentation hypothesis.

5 Parametric tests of the segmentation hypothesis

In this section, we argue that several standard parametric tests would lead one to conclude that our data provides compelling evidence that labor markets are segmented in Argentina. Table 5 in the appendix shows the outcome of regressing log real hourly wages on year dummies, individual, establishment and industry characteristics, as well as a dummy variable called Sector which takes value 1 if the individual is formally employed, 0 otherwise. Variables are defined in more details in appendix A. The table shows that in a baseline specification that does not include any interaction terms, the impact of the sector variable is positive and significant even after controlling for establishment, industry and educational characteristics. Education, size and industry effects are large and significant.[8] The second specification shown in table 5 includes as regressors individual and establishment variables interacted

[8]In particular, this confirms that the positive relationship between size and wages documented for many countries is also present in Argentina. For instance, in our 1993 sample, the average wage of employees in establishments with more than 500 workers is 1.6 times greater than the average wage of employees with 25 workers or fewer.

with the Sector variable. This amounts to estimating separate earning functions for each sector. While the Sector dummy is no longer significant, some of the interacted terms have a significant impact on wages, notably age and some industry dummies. Simple calculations based on those coefficients continue to show a significantly positive formal premium for many subgroups, and this remains true for all basic variations of the baseline specification shown in table 5.[9] In other words, results shown in table 5 support hypothesis S.

So far the analysis has ignored the endogeneity of the selection decision into the formal or informal sector. To control parametrically for self-selection we implement a test suggested by Heckman and Hotz (1986). We split our sample into two subsamples along formal/informal lines and then estimate wage regressions with a two-step correction for selection separately for each subsample. Under the assumption that labor markets are integrated, estimated coefficients should not differ significantly in the two subsamples.

We assume that the selection decision of individuals depends on age, gender, education and whether or not they have a relative in the formal sector. The last variable does not appear to affect wages but has a significant impact on sector assignments. Results are shown in table 6. Several coefficients in the estimated earning functions turn out to be very different in the two samples. Consider for instance the impact of age, a variable which is highly significant in both regressions. The absolute value of the coefficient of the age squared term is almost double in the informal sector, suggesting that age-earning profiles tend to be much more concave in the informal sector. Once again, simple calculations based on these results show a significant formal sector premium for many subgroups. Thus strong evidence of segmentation remains even after controlling for potential selection bias.

Note, however, that this approach is based on strong parametric assumptions, both about the form of the selection bias and the form of wage functions. It is also unable to account for unobserved heterogeneity that may affect the selection decision and earnings. We now turn to semiparametric methods to address those shortcomings.

[9]This includes specifications where all individual variables are interacted with the Gender variable. Findings for each year taken separately were similar, although specific coefficients can differ markedly from year to year. To be concise, we only report results for the pooled sample. Other results are available from the authors upon request.

6 Semiparametric estimators

Matching estimators are typically used to evaluate the impact of specific programs or treatments.[10] The idea is to compare outcomes of the "treated" with what they would have been had they not undergone the treatment. Since this counterfactual is unobservable, it is estimated by using data on the outcomes of an "untreated" or control group, which is similar to the treatment group in observed characteristics, but has not undergone the treatment.

Formal sector employees constitute our treatment group, while informal sector employees will serve as the control group. As in section 2, let w^F denote the log wage outcomes of formal employees and w^I denote the log wage outcomes of informal employees, and denote the sets of individual and job characteristics by X and Y, respectively. Let s equal 1 if the individual is formally employed, 0 otherwise.[11]

The average formal sector premium (or average effect of the treatment on the treated) can now be defined as

$$\alpha = E\left(w^F|X, Y, s = 1\right) - E\left(w^I|X, Y, s = 1\right), \tag{1}$$

The second term on the right-hand side of equation (1) is unobserved and must be replaced by $E\left(w^I|X, Y, s = 0\right)$. This is a valid substitution only if w^I is independent of s once we condition on X and Y. That is, provided

$$w^F, w^I \perp s|X, Y, \tag{2}$$

we have

$$E\left(w^I|X, Y, s = 1\right) = E\left(w^I|X, Y, s = 0\right),$$

[10]For applications to job training programs, see LaLonde, 1986, Heckman et. al. 1997, or Blundell et. al., 2000. Also see Heckman et. al., 1999, for a detailed list of references on matching techniques.

[11]In the program evaluation literature, variables such as s are referred to as the participation variable. In the context of this paper, this is the Sector variable we introduced in the parametric section.

and, therefore,

$$\alpha = E\left(w^F|X,Y,s=1\right) - E\left(w^I|X,Y,s=0\right).$$

Condition (2) says that wage outcomes in both sectors must be independent of participation once we condition on X and Y. Put another way, if selection occurs, it can only occur on the basis of characteristics spanned by X and Y.

In non-experimental studies such as ours, assignment to treatment is non random. Therefore, the covariates may vary systematically between groups. In such cases, Dehejia and Wahba (1999, forthcoming) suggest that matching estimators based on propensity scores may perform better. The use of such estimators is justified by the Rosenbaum and Rubin (1983, 1984) result that if

$$w^F, w^I \perp s|X,Y \quad and \quad 0 < P(s=1|X,Y) < 1$$

then

$$w^F, w^I \perp s|P(s=1|X,Y) \tag{3}$$

where $P(s=1|X,Y)$ is the probability of selection into the formal sector and is known as the propensity score. In other words, the matching estimator remains valid if we condition on the propensity score rather than on the covariates themselves. Relying on propensity scores also enables one to get around the practical difficulty of matching individuals directly along several characteristics with a finite sample.

The matching estimator of the formal sector premium is then given by:

$$\alpha^M = \sum_{i\in F}\left(w_i^F - \sum_{j\in I}\eta_{ij}w_j^I\right)\eta_i \tag{4}$$

where i indexes formal workers and j indexes informal workers. Each formal observation is weighted by $0 \leq \eta_i \leq 1$. For each formal worker, a control wage is formed as a weighted average of informal wages. We denote the weight assigned to informal worker j in building a comparison wage for formal worker i by η_{ij}. The comparison observations in the informal sector are selected on the basis of the proximity of their propensity score to the corresponding formal observation.

Condition (3) is violated if selection into the formal sector depends on unobserved heterogeneity which affects wages but cannot be included as a conditioning variable in estimating the propensity score. This potential problem can be partially addressed by combining the matching estimator with a difference-in-difference estimator (see e.g Blundell and Costa Dias 2000.) Denote by $I \rightarrow F$ the set of workers who move from the informal sector to the formal sector from one period to the next, and denote by $I \rightarrow I$ the set of workers who remain in the informal sector. The difference-in-difference estimator of the average treatment effect is given by

$$\alpha^{MD} = \sum_{i \in I \rightarrow F} \left(\left(w_i^{F,t+1} - w_i^{I,t} \right) - \sum_{j \in I \rightarrow I} \eta_{ij} \left(w_j^{I,t+1} - w_j^{I,t} \right) \right) \eta_j$$

where t and $t+1$ denote two consecutive periods. Assuming that the effect of unobservables on wages is constant through time, differencing removes the component of wages which is attributable to unobserved heterogeneity. This estimator is based on the assumption that wages in the control group sector evolve in the same way as wages in the treatment would have, had they not been treated. Correspondingly, condition (3) becomes

$$\left(w^{F,t+1} - w^{I,t} \right), \left(w^{I,t+1} - w^{I,t} \right) \perp s^{t+1} | P(s^{t+1} = 1 | X, Y).$$

The changes in wages for both movers and stayers must be independent of whether a change in sector occured, conditioning on the probability of the individual being in the formal sector at time $t+1$. We now present results for both estimators.

6.1 The matching estimator

The validity of the matching estimator we use depends on the ability of propensity scores to account for cross-sector differences. Propensity scores turn out to be an effective proxy for individual and establishment characteristics in our application, as we now argue.

We begin by estimating the propensity score with a probit specification. The dependent variable is 1 if the individual is in the formal sector and 0 otherwise. The independent variables are age, gender, an indicator variable which takes the value 1 if any other family member was employed in the formal sector in that year, and dummies for establishment size and education. Not surprisingly, table 7 shows that the propensity score rises with establishment size, age and education and that men are more likely to be formally employed than women. Table 8 gives the relative frequency of the propensity score for individuals in the formal and in the informal sector for each year. Naturally, the proportion of formal (treated) workers rises with the propensity score. What is important for our estimation technique is that there be enough overlap in all strata, which is the case here.[12]

Table 2 shows the distribution of establishment characteristics for formal and informal sector employees for each year. As we mentioned, it is quite obvious that the two groups are very different. But conditioning on propensity scores significantly reduces those differences. Tables 9 to 13 compares employees in the two sectors for 5 subsamples corresponding to 5 different propensity scores intervals. Coarse as these propensity categories may be, individual and job characteristics become markedly closer than in Table 2. Consider, for instance, the sample of workers whose propensity score falls between 0.20 and 0.40. Almost all employees, formal and informal, work in establishments with less than 6 individuals every year. About 59% of all individuals, whether formal or informal sector employees, have at least one other family member in the formal sector. The average age ranges between 33 and 35 years in both sectors. In contrast, table 13 shows that most individuals whose propensity score falls between 0.8 and 1 tend to work in large establishments, and a large fraction of those individuals have some tertiary education, in both sectors. One characteristic for which large differences remain in those tables is gender, particularly for low propensity scores. Below we present separate estimates for males and females to address this concern.

To compute our first matching estimator, each formal worker is matched with the informal worker who has the closest propensity score. The propensity score, and the matching

[12]The fact that treated observations are over-represented at high propensity scores raises our estimated standard errors. As discussed in footnote 13, in matching with replacement, standard errors increase when certain controls are repeatedly used.

estimator are both estimated separately for each year. The resulting version of expression (4) is:

$$\alpha^M = \frac{1}{N_F}\left(\sum_{i\in F} w_i^F - \sum_{i\in I} n_i w_i^I\right) \tag{5}$$

where n_i is the number of times informal worker i is used as a match for a formal worker and N_F is the total number of formal workers.

Table 14 presents the results. In contrast with our parametric results, this method yields a small estimate for the wage premium in the formal sector which does not significantly differ from zero in any year.[13] Thus no systematic formal sector premium can be found in our sample.

Naturally, these numbers could hide significant variations in wages for specific types of individuals in the sample. Table 15 splits the sample on the basis of different ranges of propensity scores. Interestingly, workers with low propensity scores show a (significantly) negative premium. These subcategories comprise low skill individuals working in poorly paid occupations. This suggests that the formal sector does not offer higher wage expectations to low income workers. As the propensity score rises, the wage premium usually goes up. It only becomes positive and (marginally) significant in the 0.8-1.0 range. As for other specific subgroups, table 16 shows that the formal sector premium is negative for women two out three years, and it is not statistically significant in any year. For males, the premium is small and positive, but never significantly different from zero. There is also no evidence that the returns to age or education are different in the formal and informal sector.

[13] The approximate variance of the estimator can be expressed as

$$\frac{1}{N_F}\left(Var\left(w^F\right) + \frac{\sum_{i\in\{I\}} n_i^2}{N_F}.Var\left(w^I\right)\right).$$

There is therefore a high penalty for using certain controls often. Indeed, $\sum_{i\in\{I\}} n_i^2$ is small when informal workers are all used a comparable number of times, which occurs when the composition of the treated (formal) and the control (informal) group is similar. The standard errors are approximate because they do not reflect the fact that the weights are functions of our estimated propensity scores.

6.2 The difference-in-difference matching estimator

Our analysis so far has not taken into account individual heterogeneity. To control for fixed but unobserved earning determinants, we divide our sample into 5 subperiods and, in each period, compare the change in wages for individuals who moved from the informal to the formal sector with the corresponding change for comparable individuals who have stayed in the informal sector. Workers are matched on the basis of their propensity scores at the end of the period.[14] The details of our sample splits are shown in table 17. The second column shows the number of transitions from the informal to the formal sector in each subperiod. The third column shows the number of individuals who stayed in the informal sector. Note that the number of transitions in each period is comparatively small due to a high level of sample attrition, which reduces the precision of the difference-in-difference estimator.

The results are indicative nevertheless. Table 18 shows that the difference-in-difference estimator, much like our matching estimator, is small or negative in early time periods. However, it becomes marginally significant in the last two periods. Results thus become more mixed, but, overall, there continues to be no compelling evidence of a formal premium. For completeness we also compute this estimator for various sub-groups, even though the small size of the corresponding subsamples results in large standard errors. Results are once again mixed, and in many cases estimates turn out to be negative.

7 Sensitivity analysis

7.1 Specification of propensity scores

Given that the estimated propensity score is central to our analysis, it is important to gauge the extent to which our results depend on the specification of this propensity score.

Table 20 in the appendix presents matching estimators obtained with of a logit specification of the propensity score. These results follow a pattern similar to those based on

[14] Using the beginning of period propensity score would bias our results since individuals who transit to the formal sector tend to move to bigger establishments. The change in wages would include a size premium.

the probit specification. There is no significant formal wage premium for the entire sample for any of the three years we consider. As before, the wage premium is negative for the lower propensity score categories, and rises with the propensity score. However, it does not become significantly positive. In fact, we do not find a significant formal premium for any subgroup.

7.2 Calliper matching

Instead of matching each formal worker with the informal worker with the nearest propensity score, one can form comparison wages as the weighted average of the wages of all informal workers whose propensity score is within $\delta > 0$ of the score of the formal worker under consideration, an approach called Calliper matching.

Table 21 in the appendix shows the Calliper matching estimate one obtains with our data for $\delta = 0.0001$. Our basic claim that no formal premium can be found continues to hold. In fact, estimates tend to become smaller in all categories. The tendency for the premium to rise with the propensity score remains. As before, no split of our sample yields a compelling premium. Results for $\delta = 0.001$ and $\delta = 0.00001$ were similar.

7.3 Pecuniary value of benefits

While we find no significant difference in gross wages across sectors, formal employment may still dominate informal employment when one takes into account the pecuniary value of benefits. Indeed, by definition, informal workers do not receive pension or unemployment insurance benefits.

Since we compare before-tax wages, the value of those benefits would have to offset the fact that informal workers become subject to income taxation when they enter the formal sector. This is unlikely since, as discussed by Pessino (1997), it is a common view that in Argentina “workers regard most [social security] contributions as taxes” given the level of uncertainty in the administration of retirement pensions. Nevertheless, directly testing whether accounting for benefits would alter our findings requires some independent evidence on the perceived value of benefits, which we do not have.

In this section, we suggest two indirect tests of the importance of formal benefits. Take, first, unemployment insurance. In Argentina as in most economies, large establishments fail less frequently than small establishments. All else equal, the likelihood of an involuntary separation is lower for large establishment employees than for individuals who work for small establishments. The value of unemployment insurance is therefore likely to be smaller for large establishment workers. Accounting for the value of unemployment insurance should have less impact on our wage premium estimate for the subsample consisting of workers employed in large establishments. The last rows of tables 16, 20 and 21 show no compelling evidence of a formal premium in the subsample of individuals employed in establishments with 100 or more employees.

As for pension benefits, one can argue that younger people are unlikely to value those benefits much since their remaining active lifetime gives them ample time to qualify for those benefits. Until July of 1994, employees were eligible for ordinary retirement at age 60 for men, 55 for women, provided they had made 20 years of contributions. As of July 1994, workers were given the option to invest their contributions in private capitalization accounts and were eligible for ordinary retirement as of age 65 provided they had made contributions for 30 years.[15] Again, tables 16, 20 and 21 show that no formal sector can be found for individuals under 40 years of age. Results were similar for workers under 35 years of age, and workers under 30 years of age.

Those tests indicate that even in subgroups where employees are likely to assign below average value to formal benefits, no formal sector premium can be found. This makes it unlikely that accounting for formal benefits will alter our findings.

[15]See Barreto de Oliveira (1994) for a detailed description of Argentina's social security system as well as a summary of the 1994 reform. Until July 1994, employers had to contribute 10% of their wages, while employers contributed 16% of wages.

8 Conclusion

We find no evidence of a formal sector wage premium in Buenos Aires and its suburbs with data from the Permanent Household Survey for the 1993-1995 time period. While wages are higher on average in the formal sector, this apparent premium disappears after controlling semiparametrically for individual and establishment characteristics. In fact, we find that groups often thought to be queuing for formal sector jobs such as young and uneducated workers would expect lower wages in the formal sector. These findings are all the more striking that we do not take into account the fact that informal employees usually become subject to income taxation when they enter the formal sector.

The hypothesis of segmentation along formal/informal lines constitutes a natural application of semiparametric methods. Matching estimators do not rely on a specific earnings function and, by comparing wages only for similar workers, they minimize the impact of selection effects. Our paper illustrates how results that rely on a given specification of earning functions can be fragile.

The analysis yields several ancillary results of interest. By using a definition of informal employment based on government-mandated benefits we are able to evaluate the correlation between small establishment employment and informality. We find that the informal sector emphasizes small scale production, but a non-trivial share of informal employment is found large firms, and a large fraction of formal workers work in small firms. A satisfactory theory of informal economic activities should be consistent with these facts. This evidence also suggests that by equating small firm employment and informal employment, much of the existing literature on informal economic activities in Latin America is likely to yield biased earnings inferences.

Our data also confirm that the distribution of age, gender and education characteristics differs markedly across sectors. There remains to explain how these differences can arise in a context where labor markets appear to be competitive. There are many potential explanations. To cite but one, it may be optimal for informal firms to concentrate in sectors where technological opportunities are biased towards unskilled labor. This is plausible because informal firms have limited access to outside financing, which forces them to concentrate on activities where physical and working capital requirements are small. Formalizing and testing this and other conjectures are natural avenues for future work.

A Definition of the variables

Real hourly wages

The household survey inquires about total hours worked in a recent week, and wages earned in each occupation in a recent month. We only consider earnings derived from primary occupations. Hourly wages are calculated by dividing monthly income derived from primary occupations by $\frac{52}{12}$ times weekly hours. Argentina's Consumer Price Index is used to obtain real wages. The wages of individuals who receive an "aguinaldo" are multiplied by $\frac{13}{12}$. The aguinaldo or "Christmas bonus" refers to two payments equivalent to half a month worth of wages that employers are required by law to make to their employees. The survey also asks for total labor income in a recent month in a different section and we use this information to check for possible inconsistencies. When individuals report that they have exactly one occupation, this reported income divided by total hours worked should coincide with hourly earnings as computed above. Individuals for which wages computed by these two methods differ by more than 5% are discarded.

Sectoral assignments

The Sector variable takes value 1 if the individual receives both pension and unemployment insurance benefits, 0 otherwise.

Establishment size

Establishments are classified in one of 8 possible employment-size categories and we created a dummy for each of these categories.

Industry

Establishments are also classified according to the three-digit International Standard Industrial Classification (ISIC). We created a dummy variable for each two-digit category.

Education levels

The survey classifies the highest educational level achieved by individuals in eight mutually exclusive categories. A dummy variable was created for each category.

Household members in the formal sector

The dummy variable Fhousehold takes value 1 if a member of the individual's household (other than the individual him or herself) is formally employed, 0 otherwise.

B Tables

Table 1: Differences in average real wages, Buenos Aires and its suburbs

	1993		1994		1995		All Years	
	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs	Mean
Severance pay	3344	4.2665	3416	4.6221	3340	4.4074	10100	4.4334
No severance pay	1922	3.2501	1845	3.4864	1826	3.1652	5593	3.3003
T-statistic		9.1279		9.8017		10.3865		16.9708
Paid vacations	3732	4.1983	3743	4.5514	3614	4.3385	11089	4.3632
No paid vacations	1534	3.1590	1518	3.4162	1552	3.1063	4604	3.2260
T-statistic		8.8042		9.2931		9.8703		16.1781
Retirement benefits	3528	4.2431	3601	4.5916	3469	4.3688	10598	4.4027
No retirement benefits	1738	3.1899	1660	3.4260	1697	3.1496	5095	3.2534
T-statistic		9.2402		9.7969		10.0080		16.8271
Unemployment insurance	3283	4.2832	3420	4.6076	3364	4.3967	10067	4.4314
No unemployment insurance	1983	3.2536	1841	3.5108	1802	3.1685	5626	3.3105
T-statistic		9.3093		9.4550		10.2356		16.8072
At least one benefit	3784	4.1859	3798	4.5418	3677	4.3265	11259	4.3518
No benefit	1482	3.1543	1463	3.3985	1489	3.08369	4434	3.2111
T-statistic		8.6469		9.2553		9.8354		16.0463

Notes: All wages are deflated by the quarterly CPI and corrected for bonuses (aguinaldo).

Table 2: Individual and job characteristics of formal and informal sector employees

	1993		1994		1995	
	Formal	Informal	Formal	Informal	Formal	Informal
Education						
None	0.004	0.006	0.003	0.009	0.003	0.008
Primary	0.311	0.476	0.307	0.476	0.328	0.465
High-school	0.414	0.377	0.413	0.390	0.374	0.364
Superior	0.069	0.037	0.086	0.026	0.090	0.034
University	0.202	0.104	0.192	0.099	0.204	0.128
Establishment size (employees)						
Less than 6	0.126	0.592	0.145	0.587	0.141	0.623
6 to 25	0.273	0.244	0.275	0.262	0.271	0.246
26 to 50	0.159	0.055	0.148	0.055	0.144	0.036
51 to 100	0.120	0.045	0.133	0.040	0.133	0.030
101 to 500	0.181	0.041	0.168	0.033	0.190	0.045
More than 501	0.142	0.023	0.131	0.024	0.121	0.020
Gender						
Male	0.652	0.544	0.644	0.573	0.627	0.532
Female	0.348	0.456	0.356	0.427	0.373	0.468
Another family member in the formal sector						
Yes	0.471	0.359	0.462	0.378	0.444	0.344
No	0.529	0.641	0.538	0.622	0.556	0.656
Average age	37.427	33.626	37.195	33.432	37.333	33.265
Observations	3261	2005	3406	1855	3344	1822

Notes: Entries give the fraction of employees in each category. Age is measured in years.

Table 3: Transitions among occupations and sectors

From \ To	Out of labor force	Unemployed	Formal employee	Informal employee	Employer	Own-account worker	Unpaid worker
Unemployed	51	208	63	94	5	114	3
	(9.5)	(38.7)	(11.7)	(17.5)	(0.9)	(21.2)	(0.6)
Formal employee	161	58	4876	638	38	156	5
	(2.7)	(1.0)	(82.2)	(10.8)	(0.6)	(2.6)	(0.1)
Informal employee	77	122	737	1469	39	347	26
	(2.7)	(4.3)	(26.2)	(52.1)	(1.4)	(12.3)	(0.9)
Employer	13	9	57	46	402	212	12
	(1.7)	(1.2)	(7.6)	(6.1)	(53.5)	(28.2)	(1.6)
Own-account worker	64	133	153	382	182	1722	23
	(2.4)	(5.0)	(5.7)	(14.4)	(6.8)	(64.8)	(0.9)
Unpaid worker	2	5	16	25	12	43	42
	(1.4)	(3.4)	(11.0)	(17.2)	(8.3)	(29.7)	(29.0)

Notes: Sample consists of the 5 inter-survey periods between 1993 and 1995. The table records the number of transitions to and from each possible employment status between sampling periods. The corresponding percentages are in parenthesis.

Table 4: Characteristics of workers who switch sectors

Initial/Terminal Occupation	Age	Tertiary education	% change in gross wage
Formal employee/Formal employee	38.31	19.76	9.93
	(0.19)	(0.66)	(1.30)
Formal employee/Informal employee	36.04	17.36	10.29
	(0.64)	(1.82)	(3.42)
Informal employee/Formal employee	34.65	15.81	16.70
	(0.54)	(1.59)	(3.33)
Informal employee/Informal employee	32.72	9.07	6.50
	(0.39)	(0.82)	(1.36)

Notes: Sample consists of the 5 inter-survey periods between 1993 and 1995. Standard errors are in parenthesis.

Table 5: OLS regressions

Dependent variable is log real hourly wages

	Baseline specification	Specification 2: all variables interacted with Sector	
Age	0.0464 (10.74)	0.0543 (8.34)	-0.0215 (-2.35)
Age^2	-0.0005 (-9.82)	-0.0006 (-7.91)	0.0003 (2.53)
Gender†	0.0690 (2.66)	0.0670 (1.41)	0.0249 (0.44)
Sector††	0.2530 (9.67)	0.2641 (1.14)	
Hours	-0.0162 (-21.71)	-0.0167 (-15.49)	0.0020 (1.35)
Marital Status *	0.1796 (7.01)	0.2191 (5.04)	-0.0634 (-1.18)
Establishment Size			
2 to 5 emp.	-0.0625 (-1.21)	-0.0593 (-0.86)	0.1867 (1.56)
6 to 15	-.0361 (0.66)	0.0423 (0.55)	0.1454 (1.16)
16 to 25	0.0781 (1.31)	0.1631 (1.73)	0.0351 (0.25)
26 to 50	0.1269 (2.18)	0.0226 (0.21)	0.2705 (1.84)
51 to 100	0.1275 (2.12)	0.1266 (1.09)	0.1434 (0.94)
101 to 500	0.1739 (2.92)	0.1910 (1.70)	0.1220 (0.81)
$\geq$ 501	0.2661 (4.24)	0.3841 (2.89)	0.0135 (0.08)
Education Levels			
Primary	0.1566 (1.91)	0.0819 (0.68)	0.0571 (0.27)
National	0.3117 (3.61)	0.0871 (0.66)	0.3119 (1.42)
Commercial	0.3038 (3.65)	0.2008 (1.62)	0.1066 (0.50)
Normal	0.1583 (1.31)	-0.2775 (-1.09)	0.5856 (1.82)
Technical	0.3522 (4.07)	0.2026 (1.49)	0.1806 (0.81)
Other	0.3352 (2.98)	0.1131 (0.57)	0.3028 (1.08)
Superior	0.4863 (5.31)	0.2868 (1.79)	0.2652 (1.10)
University	0.5705 (6.60)	0.4614 (3.41)	0.1178 (0.53)
Industry			
Mining	0.0863 (2.15)	0.0449 (0.55)	0.0542 (0.58)
Manufacturing	0.1611 (3.45)	0.1951 (2.05)	-0.0507 (-0.46)
Electricity, Gas, Water	0.1079 (1.95)	0.0056 (0.06)	0.2080 (1.87)
Construction	0.0109 (0.29)	-0.0114 (-0.18)	0.0235 (0.30)
Retail	0.1497 (3.56)	0.0182 (0.22)	0.1832 (1.92)
Transport	-0.0058 (-0.13)	0.0049 (0.50)	-0.0700 (-0.63)
Finance	-0.1330 (-2.98)	0.1087 (1.23)	-0.2846 (-2.77)
Services	0.1408 (2.91)	0.1650 (2.19)	-0.0908 (-0.90)
Year 1994 dummy	0.1076 (4.50)	0.1088 (4.54)	
Year 1995 dummy	0.0028 (0.11)	0.0034 (0.14)	
R^2	0.4150	0.4192	

Notes: T-statistics based on heteroskedasticity consistent standard errors are in parenthesis. In the second specification, the right-hand panel shows coefficients and t-statistics for variables interacted with the sector variable. † 1=Male, 0=Female, †† 1=Formal Sector, 0=Informal Sector, * 1=Married, 0=Single. Omitted education dummy is no education, omitted establishment size is less than 2 employees, omitted industry dummy is agriculture.

Table 6: OLS regressions with two-step correction for selection bias

Dependent variable is log real hourly wages

	Formal sector	Informal sector
Age	0.0338 (3.81)	0.0569 (5.42)
Age^2	-0.0003 (-3.13)	-0.0007 (-5.75)
Gender	0.1365 (2.69)	0.0880 (0.87)
Hours	-0.0147 (-8.65)	-0.0168 (-10.91)
Marital Status	0.1913(3.92)	0.2255 (2.81)
Establishment Size		
2 to 5 emp.	0.1363 (0.18)	-0.053 (-0.59)
6 to 15	0.1992 (1.17)	0.0500 (0.47)
16 to 25	0.2073 (1.21)	0.1668 (1.00)
26 to 50	0.3057 (1.80)	0.0191 (0.09)
51 to 100	0.2801 (1.65)	0.1250 (0.52)
101 to 500	0.3225 (1.91)	0.1988 (0.74)
$\geq$ 501	0.4096 (2.42)	0.3870 (0.98)
Education Levels		
Primary	0.0695 (0.27)	0.1069 (0.62)
National	0.3767 (1.58)	0.1283 (0.67)
Commercial	0.3033 (1.32)	0.2507 (1.29)
Normal	0.3448 (1.46)	-0.2174 (-0.57)
Technical	0.3620 (1.52)	0.2431 (1.16)
Other	0.4354 (1.71)	0.1816 (0.49)
Superior	0.6161 (2.86)	0.3692 (1.14)
University	0.6113 (2.80)	0.5313 (2.21)
Industry		
Mining	0.1004 (1.66)	0.0541 (0.35)
Manufacturing	0.1419 (1.75)	0.1943 (0.98)
Electricity, Gas, Water	0.2113 (1.91)	0.0138 (0.11)
Construction	0.0113 (0.17)	-0.0033 (-0.03)
Retail	0.1989 (3.21)	0.0304 (0.22)
Transport	-0.0236 (-0.39)	0.0551 (0.32)
Finance	-0.1744 (-2.74)	0.1157 (0.60)
Services	0.0747 (0.80)	0.1818 (1.61)
Year 1994 dummy	0.1077 (4.50)	0.0621 (0.89)
Year 1995 dummy	0.0030 (0.12)	-0.1043 (-1.49)
ρ	0.2111 (1.98)	-0.0587 (-0.83)

Notes: T-statistics are in parenthesis. The selection equation is: $Prob(Sector = 1) = .0138Age + .3646Gender + .2501Mstatus - .9460Primary - .6264National - .4928Commercial - .1618Normal - .6271Technical - .3219Other + .1228Superior - .1855University + .2981Fhousehold$. All variables are significant at the 1% level, except for Normal whose t-statistic is -1.80. The last row of the table gives the estimated correlation between the error term of the selection equation and the error term of the wage equation. Omitted variables are the same as in table 5.

Table 7: Results of Probit estimation of propensity scores

	1993		1994		1995	
Age	0.0131	(0.0015)	0.0131	(0.0015)	0.0142	(0.0016)
Gender	0.2211	(0.0418)	0.1326	(0.0420)	0.1956	(0.0422)
FHousehold	0.2699	(0.0407)	0.1854	(0.0407)	0.2255	(0.0421)
Establishment Size						
6 to 25	0.9556	(0.0496)	0.8205	(0.0489)	0.9237	(0.0421)
26 to 50	1.4603	(0.0698)	1.3443	(0.0705)	1.6176	(0.0790)
51 to 100	1.3933	(0.0757)	1.4124	(0.0776)	1.6570	(0.0835)
101 to 500	1.6614	(0.0726)	1.6772	(0.0779)	1.6605	(0.0719)
$\geq$ 501	1.8249	(0.0875)	1.6564	(0.0889)	1.8154	(0.0962)
Education						
Primary	-1.4917	(0.0792)	-1.2771	(0.0792)	-1.3298	(0.079)
High-school	-1.2060	(0.0732)	-0.9563	(0.0717)	-1.1409	(0.0740)
Superior	-1.0489	(0.1055)	-0.4027	(0.1097)	-0.7940	(0.1081)
University	-1.0793	(0.0846)	-0.8686	(0.0851)	-1.1594	(0.0845)

Note: The dependent variable is 1 if the individual is in the formal sector. The High-school dummy includes normal, technical and commercial high school education. Asymptotic standard errors are in parentheses.

Table 8: Frequency distribution of propensity scores

	1993		1994		1995	
$P(s=1\|X,Y)$	Formal	Informal	Formal	Informal	Formal	Informal
0.00 to 0.20	0.016	0.166	0.003	0.071	0.007	0.088
0.20 to 0.40	0.097	0.401	0.108	0.437	0.113	0.476
0.40 to 0.60	0.082	0.127	0.094	0.166	0.060	0.127
0.60 to 0.80	0.293	0.192	0.245	0.199	0.234	0.171
0.80 to 1.00	0.512	0.115	0.550	0.127	0.586	0.137

Table 9: Individual and establishment characteristics, $0.0 < P(s = 1|X, Y) \leq 0.2$

	1993		1994		1995	
	Formal	Informal	Formal	Informal	Formal	Informal
Education						
None	0.00	0.00	0.00	0.00	0.00	0.00
Primary	0.80	0.85	1.00	1.00	0.64	0.76
High-school	0.20	0.15	0.00	0.00	0.23	0.11
Superior	0.00	0.00	0.00	0.00	0.01	0.00
University	0.00	0.00	0.00	0.00	0.14	0.13
Establishment size (employees)						
Less than 6	1.00	1.00	1.00	1.00	1.00	1.00
6 to 25	0.00	0.00	0.00	0.00	0.00	0.00
26 to 50	0.00	0.00	0.00	0.00	0.00	0.00
51 to 100	0.00	0.00	0.00	0.00	0.00	0.00
101 to 500	0.00	0.00	0.00	0.00	0.00	0.00
More than 501	0.00	0.00	0.00	0.00	0.00	0.00
Gender						
Male	0.47	0.33	0.67	0.37	0.05	0.29
Female	0.53	0.67	0.33	0.63	0.95	0.71
Another family member in the formal sector						
Yes	0.12	0.12	0.00	0.16	0.00	0.05
No	0.88	0.88	1.00	0.84	1.00	0.95
Average age	27.04	26.37	20.22	21.42	22.86	20.66
Observations	51	332	9	132	22	160

Table 10: Individual and establishment characteristics, $0.2 < P(s=1|X,Y) \leq 0.4$

	1993		1994		1995	
	Formal	Informal	Formal	Informal	Formal	Informal
Education						
None	0.00	0.00	0.00	0.00	0.00	0.01
Primary	0.40	0.41	0.46	0.52	0.45	0.52
High-school	0.45	0.46	0.46	0.42	0.42	0.39
Superior	0.03	0.03	0.00	0.00	0.01	0.02
University	0.12	0.11	0.08	0.06	0.12	0.08
Establishment size (employees)						
Less than 6	1.00	0.99	1.00	1.00	1.00	1.00
6 to 25	0.00	0.01	0.00	0.00	0.00	0.00
26 to 50	0.00	0.00	0.00	0.00	0.00	0.00
51 to 100	0.00	0.00	0.00	0.00	0.00	0.00
101 to 500	0.00	0.00	0.00	0.00	0.00	0.00
More than 501	0.00	0.00	0.00	0.00	0.00	0.00
Gender						
Male	0.68	0.49	0.63	0.48	0.59	0.48
Female	0.32	0.51	0.37	0.52	0.41	0.52
Another family member in the formal sector						
Yes	0.42	0.43	0.40	0.39	0.39	0.37
No	0.58	0.57	0.60	0.61	0.61	0.63
Average age	35.98	35.68	34.35	33.12	35.65	33.21
Observations	315	805	367	811	378	868

Table 11: Individual and establishment characteristics, $0.4 < P(s = 1|X, Y) \leq 0.6$

	1993		1994		1995	
	Formal	Informal	Formal	Informal	Formal	Informal
Education						
None	0.00	0.00	0.00	0.00	0.00	0.00
Primary	0.60	0.63	0.60	0.49	0.46	0.46
High-school	0.28	0.28	0.28	0.36	0.39	0.41
Superior	0.04	0.03	0.05	0.07	0.04	0.05
University	0.09	0.05	0.07	0.08	0.11	0.08
Establishment size (employees)						
Less than 6	0.16	0.18	0.35	0.41	0.34	0.40
6 to 25	0.83	0.82	0.65	0.59	0.66	0.60
26 to 50	0.00	0.00	0.00	0.00	0.00	0.00
51 to 100	0.01	0.00	0.00	0.00	0.00	0.00
101 to 500	0.00	0.00	0.00	0.00	0.00	0.00
More than 501	0.00	0.00	0.00	0.00	0.00	0.00
Gender						
Male	0.61	0.65	0.71	0.60	0.51	0.57
Female	0.39	0.35	0.29	0.40	0.49	0.43
Another family member in the formal sector						
Yes	0.25	0.30	0.33	0.36	0.30	0.29
No	0.75	0.70	0.67	0.64	0.70	0.71
Average age	33.39	31.08	36.53	34.65	35.00	35.88
Observations	269	254	322	308	199	232

Table 12: Individual and establishment characteristics, $0.6 < P(s = 1|X, Y) \leq 0.8$

	1993		1994		1995	
	Formal	Informal	Formal	Informal	Formal	Informal
Education						
None	0.00	0.02	0.00	0.03	0.00	0.04
Primary	0.37	0.35	0.36	0.31	0.39	0.32
High-school	0.41	0.45	0.44	0.49	0.35	0.44
Superior	0.07	0.06	0.04	0.02	0.09	0.04
University	0.15	0.12	0.16	0.15	0.17	0.16
Establishment size (employees)						
Less than 6	0.00	0.02	0.01	0.04	0.00	0.04
6 to 25	0.64	0.69	0.75	0.77	0.92	0.91
26 to 50	0.16	0.12	0.15	0.12	0.03	0.03
51 to 100	0.15	0.12	0.07	0.06	0.02	0.00
101 to 500	0.05	0.05	0.01	0.01	0.03	0.02
More than 501	0.00	0.00	0.01	0.00	0.00	0.00
Gender						
Male	0.61	0.65	0.64	0.73	0.64	0.71
Female	0.39	0.35	0.36	0.27	0.36	0.29
Another family member in the formal sector						
Yes	0.44	0.37	0.43	0.40	0.41	0.36
No	0.56	0.63	0.57	0.60	0.59	0.64
Average age	34.83	34.32	35.36	33.18	36.18	35.24
Observations	957	384	835	369	784	312

Table 13: Individual and establishment characteristics, $0.8 < P(s = 1|X, Y) \leq 1.0$

	1993		1994		1995	
	Formal	Informal	Formal	Informal	Formal	Informal
Education						
None	0.01	0.02	0.01	0.03	0.00	0.01
Primary	0.20	0.20	0.20	0.28	0.26	0.28
High-school	0.44	0.40	0.42	0.38	0.38	0.31
Superior	0.08	0.10	0.13	0.08	0.11	0.10
University	0.27	0.28	0.24	0.23	0.25	0.30
Establishment size (employees)						
Less than 6	0.00	0.01	0.00	0.02	0.00	0.01
6 to 25	0.03	0.05	0.06	0.08	0.03	0.10
26 to 50	0.22	0.27	0.20	0.25	0.23	0.23
51 to 100	0.15	0.20	0.21	0.22	0.22	0.22
101 to 500	0.33	0.28	0.30	0.25	0.31	0.30
More than 501	0.27	0.19	0.23	0.18	0.21	0.14
Gender						
Male	0.69	0.75	0.64	0.71	0.65	0.60
Female	0.31	0.25	0.36	0.29	0.35	0.40
Another family member in the formal sector						
Yes	0.55	0.49	0.52	0.46	0.49	0.46
No	0.45	0.51	0.48	0.54	0.51	0.54
Average age	40.16	38.58	38.77	40.04	38.52	36.64
Observations	1669	230	1873	235	1961	250

Table 14: Matching estimator

Period	α^M	Std. error
1993	0.056	0.100
1994	0.113	0.099
1995	0.067	0.124

Table 15: Matching estimator for propensity scores subcategories

	1993		1994		1995	
$P(s=1\|X,Y)$	α^M	Std. error	α^M	Std. error	α^M	Std. error
0.0 to 0.2	-0.534	0.342	-0.389	0.370	-1.400	0.489
0.2 to 0.4	-0.252	0.145	-0.423	0.131	-0.227	0.128
0.4 to 0.6	-0.313	0.138	-0.183	0.172	0.008	0.227
0.6 to 0.8	-0.052	0.128	0.059	0.114	-0.034	0.130
0.8 to 1.0	0.255	0.161	0.294	0.136	0.184	0.148

Table 16: Matching estimator for various subgroups

	1993		1994		1995	
	α^M	Std. error	α^M	Std. error	α^M	Std. error
Females	-0.087	0.131	0.181	0.123	-0.121	0.116
Males	0.130	0.100	0.068	0.099	0.106	0.124
Age $\leq$ 40	0.100	0.098	0.041	0.089	-0.009	0.097
Low education	-0.053	0.108	0.001	0.095	0.143	0.132
Large establishments	0.097	0.188	0.257	0.199	0.128	0.190

Note: Low education individuals have some primary education or less.

Table 17: Sample transitions

Period	Movers	Stayers
5-1993 to 10-1993	116	205
10-1993 to 5-1994	103	206
5-1994 to 10-1994	104	221
10-1994 to 5-1995	63	170
5-1995 to 10-1995	73	230

Table 18: Difference-in-difference matching estimator

Period	α^{MDD}	Std. error
5-1993 to 10-1993	-0.050	0.381
10-1993 to 5-1994	0.006	0.267
5-1994 to 10-1994	0.150	0.425
10-1994 to 5-1995	0.356	0.271
5-1995 to 10-1995	0.598	0.324

Table 19: Difference-in-difference matching estimator for subgroups

Period	Males	Females	Age ≤ 40	Low education
5-1993 to 10-1993	0.128 (0.403)	-0.037 (0.726)	0.237 (0.407)	-0.725 (1.154)
10-1993 to 5-1994	1.021 (1.203)	-0.656 (0.986)	0.881 (1.108)	0.885 (1.290)
5-1994 to 10-1994	0.140 (0.362)	0.202 (0.663)	-0.004 (0.417)	-0.022 (0.383)
10-1994 to 5-1995	0.528 (0.408)	0.004 (0.333)	0.461 (0.346)	-0.212 (0.204)
5-1995 to 10-1995	0.431 (0.460)	0.921 (0.489)	0.541 (0.430)	0.983 (0.366)

Note: Standard errors in parenthesis. Age ≤ 40 refers to individuals below 40 years of age at the end of the period.

Table 20: Matching estimators using a logit specification of the propensity score

	1993		1994		1995	
Full sample	0.0230	(0.0991)	0.0458	(0.0980)	0.0087	(0.1262)
$0.0 < P(s=1\|X,Y) \leq 0.2$	-0.5118	(0.3230)	-0.8697	(0.2732)	-1.3609	(0.4781)
$0.2 < P(s=1\|X,Y) \leq 0.4$	-0.3605	(0.1428)	-0.3071	(0.1398)	-0.3369	(0.1315)
$0.4 < P(s=1\|X,Y) \leq 0.6$	-0.2874	(0.1313)	-0.0540	(0.1761)	0.1177	(0.2277)
$0.6 < P(s=1\|X,Y) \leq 0.8$	-0.1047	(0.1285)	0.0201	(0.1160)	-0.0823	(0.1335)
$0.8 < P(s=1\|X,Y) \leq 1.0$	0.2284	(0.1539)	0.1904	(0.1359)	0.1097	(0.1434)
Men	0.0487	(0.0991)	-0.0187	(0.3799)	0.0144	(0.1262)
Women	-0.0610	(0.1268)	-0.1536	(0.1164)	-0.1251	(0.1211)
Age $\leq$ 40	0.1156	(0.0953)	0.0186	(0.3233)	-0.0108	(0.0949)
Primary or less education	-0.0384	(0.1088)	-0.0868	(0.0920)	0.0868	(0.1284)
Large establishments	-0.0403	(0.1844)	0.2687	(0.2209)	0.1061	(0.1908)

Note: Standard errors are in parenthesis.

Table 21: Calliper matching estimators

	1993		1994		1995	
Full sample	-0.0611	(0.0973)	-0.2090	(0.0775)	-0.1772	(0.1043)
$0.0 < P(s=1\|X,Y) \leq 0.2$	-0.5231	(0.3452)	-0.3894	(0.3301)	-1.3997	(0.4119)
$0.2 < P(s=1\|X,Y) \leq 0.4$	-0.2910	(0.1492)	-0.4507	(0.1113)	-0.4736	(0.1285)
$0.4 < P(s=1\|X,Y) \leq 0.6$	-0.3378	(0.1493)	-0.2332	(0.1276)	-0.0501	(0.1916)
$0.6 < P(s=1\|X,Y) \leq 0.8$	-0.2329	(0.1374)	-0.0917	(0.1165)	-0.2345	(0.1023)
$0.8 < P(s=1\|X,Y) \leq 1.0$	0.4330	(0.1728)	-0.1339	(0.1258)	0.0261	(0.1508)
Men	-0.0911	(0.0973)	-0.1772	(0.0773)	-0.1277	(0.0863)
Women	-0.0270	(0.1255)	-0.2927	(0.1191)	-0.4851	(0.1111)
Age$\leq$ 40	-0.0585	(0.0935)	-0.1772	(0.0773)	-0.1277	(0.0863)
Primary or less education	-0.2404	(0.1088)	-0.2973	(0.0920)	-0.1304	(0.1284)
Large establishments	-0.4311	(0.2169)	-0.0669	(0.1897)	-0.1258	(0.1700)

Note: The propensity score is estimated using a probit specification. Comparison wages are formed by weighing equally the wages of all informal workers whose propensity score is within 0.0001 of that of the formal employee under consideration. Standard errors are in parenthesis.

References

Barreto de Oliveira, F.E. (ed.), "Social Security Systems in Latin America", (Baltimore: Johns Hopkins University Press, 1994).

Blundell, R. and Costa Dias, M., "Evaluation Methods for Non Experimental Data," *Fiscal Studies* 21 (2000), 427-468.

Blundell, R., Costa Diaz, M., Meghir, C. and Van Reenan, J., "Evaluating the Employment Impact of Mandatory Job-search Assistance: the UK New Deal Gateway," Institute of Fiscal Studies manuscript (2000).

Brown, C, Medoff, J., "The Employer Size Wage Effect," *Journal of Political Economy* 97 (1989), 1027-1059.

Dehejia, R.H, and Wahba, S., "Causal Effects in Non Experimental Studies: Re-evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association* 94 (1999),1053-1062.

Dehejia, R.H, and Wahba, S., "Propensity Score Matching Methods for Non Experimental Causal Studies", *Review of Economics and Statistics*, forthcoming.

Fields, G. S., "Rural-Urban Migration, Urban Unemployment and Under-Development, and Job-Search Security in LDCs," *Journal of Development Economics* 2 (1975), 165-87.

Fortin, B., Marceau, N. and Savard, L., "Taxation, Wage Controls and the Informal Sector," *Journal of Public Economics* 66 (1997), 239-312.

Gong, X. and van Soest, A., "Wage Differentials and Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico", IZA, Bonn discussion paper No. 329 (2001).

Harris, J. R. and Todaro, M. P., "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review* 60 (1970), 126-142.

Heckman, J., Ichimura, H., Smith, J., Todd, P., "Characterizing Selection Bias Using Experimental Data," *Econometrica* 66 (1998), 1017-1098.

Heckman, J.J. Ichimura, H, and Todd, P.E., 1"Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies* 64 (1997), 605-654.

Heckman, J.J., LaLonde, R. and Smith, J., "The Economics and Econometrics of Active Labor Market Programs", in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, v3 (1999).

ILO, "Employment, Incomes and Equality: a Strategy for Increasing Productive Employment in Kenya," (Geneva: ILO, 1972).

LaLonde, R., "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review* 76 (1986), 604-620.

Lewis, W. A., "Economic Development with Unlimited Supplies of Labour," *Manchester School* 22 (1954), 139-191.

Loayza, N.V., "The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America," *Carnegie-Rochester Conference Series on Public Policy* 45 (1996), 129-162.

Maloney, W. F., "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico," *The World Bank Economic Review* 13 (1999), 275-302.

Magnac, Th., "Segmented or Competitive Labor Markets," *Econometrica* 59 (1991), 165-187.

Mazumdar, D., "The Theory of Urban Employment in Less Developed Countries," *World Development* 4 (1975), 655-679.

Mazumdar, D., "The Urban Labor Market Income Distribution: A Study of Malaysia" (Oxford University Press, 1981).

Oi, W. Y., Idson, T. L., "Firm Size and Wages," in Ashenfelter, O. C., Card, D. (eds), *Handbook of Labor Economics*, v3b (1999).

Persson, T., Tabellini, G and Trebbi, F., "Electoral Rules and Corruption," IEES manuscript (2000).

Portes, A., Castells, M., and Benton, L.A., (eds.), "The Informal Economy: Studies in Advanced and Less Developed Countries," (Baltimore: Johns Hopkins University Press, 1989).

Pradhan, M. and van Soest, A., "Formal and Informal Sector Employment in Urban Areas of Bolivia," *Labor Economics* 2 (1995), 275-297.

Rauch, J.E., "Modeling the Informal Sector Formally," *Journal of Development Economics* 35 (1991), 33-48.

Roberts, B.R., "Employment Structure, Life Cycle, and Life Chances: Formal and Informal Sectors in Guadalajara," in Portes, A., Castells, M., and Benton, L.A. (eds.), *The Informal Economy: Studies in Advanced and Less Developed Countries*, (Baltimore: Johns Hopkins University Press, 1989).

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

Rosenbaum, P. and Rubin, D.B., "Reducing Bias in Observational Studies using Sub Classification on the Propensity Score," *Journal of the American Statistical Association* 79 (1984), 516-524.

Sarte, P.G., "Informality and Rent-Seeking Bureaucracies in a Model of Long-Run Growth," *Journal of Monetary Economics* (2000).

Tansel, A., "Formal versus Informal Sector Choice of Wage Earners and Their Wages in Turkey," Economic Research Forum Working Paper No. 9927 (1999).

Tybout, J., "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?," *Journal of Economic Literature* 28 (2000), 11-44.