Comparative Advantage, Segmentation And Informal Earnings: A Marginal Treatment Effects Approach*

Omar Arias, World Bank†

Melanie Khamis, London School of Economics‡

Preliminary Draft: May 18, 2007, Comments Welcome

Abstract

This paper uses recently developed econometric models of essential heterogeneity (Heckman and Vytlacil 2001, 2005; Heckman, Urzua and Vytlacil 2006) to analyze the relevance of labor market comparative advantage and segmentation in the participation and earnings performance of workers in formal and informal jobs in urban Argentina. Our results offer evidence for both labor market comparative advantage and segmentation. We find no significant differences between the earnings of formal salaried workers and the self-employed once we account for positive selection bias into formal salaried work based on tastes. This is consistent with compensating differentials and comparative advantage based on tastes as the main driver of choice between salaried work and self-employment. On the contrary, informal salaried employment carries significant earnings penalties. There is a considerable negative selection bias into formal relative to informal salaried

*The authors are grateful to Sergio Urzua for invaluable help with the implementation of the marginal treatment effects code (available at http://jenni.uchicago.edu/underiv/) and to Pedro Carneiro for early discussions. We would also like to thank seminar participants at the World Bank and the LSE for helpful comments. Melanie Khamis would like to thank the LSE for financial support. The opinions expressed in this paper are our own and should not be attributed to the World Bank, its Executive Directors or the countries they represent. All errors are our own.

†Senior Economist at the World Bank. Email: oarias@worldbank.org

‡PhD candidate at the London School of Economics. Email: m.khamis@lse.ac.uk

work and only modest positive sorting based on expected earnings gains. These results are more consistent with labor market segmentation. The results are robust to different empirical specifications and are consistent with individuals' reported reasons for being formal and informal salaried or self-employed.

1 Introduction

A key question in the labor markets literature of developing countries is the extent to which informal employment results from segmentation or reflects voluntary choice. Following on the Harris and Todaro (1970) tradition, the conventional 'exclusion' view sees informal workers, either self-employed or informal employees (salaried workers lacking mandated labor benefits), as the disadvantaged class of a segmented labor market (Piore, 1979). Workers would prefer the higher wages and benefits of the formal sector but many are rationed out due to labor market rigidities (unions, minimum and efficiency wages), overly generous labor benefits (pension, health protection), unequal market power arising from lax state enforcement of regulations.

The competitive markets or 'voluntary' view of informality sees it as resulting from workers and firms weighing the private costs and benefits of operating informally (Maloney, 2004). Many informal salaried and self-employed workers, for instance youth, married women and the unskilled, may voluntarily choose these occupations as a labor market entry point and to enjoy non-pecuniary benefits such as more flexible hours, exploit entrepreneurial abilities to improve mobility, and escaping taxing regulations and/or inadequate formal social protection systems.

The labor literature on compensating differentials and occupational choice based on workers' comparative advantage provides a framework that encompasses these two views. The basic idea, first advanced by Adam Smith (1776), is that wages paid to various types of labor must, in general, equalize total *advantages and disadvantages*, pecuniary and non-pecuniary, and that workers select occupations that yield the highest net advantage for their tastes and skills. Comparative advantage can arise because individuals gain by choosing the jobs that better fit their range of talents including cognitive, social, and mechanical skills (Lucas, 1978; Rosen, 1978; Willis and Rosen, 1979; Heckman and Sedlacek, 1985). These elements are central to the decision of becoming an entrepreneur (Lucas, 1978; Lazear, 2005; Blanchflower and Oswald, 1998;

Ñopo and Valenzuela, 2007). Recent studies indeed show that comparative advantage in the labor market is a central determinant of occupational choice, human capital investments and earnings performance (e.g., Carneiro, Heckman and Vytlacil, 2005; Heckman and Li, 2003; Carneiro and Heckman, 2002;). Jobs that are more desirable (due to amenities such as fringe benefits, stability, safety, autonomy and flexibility) or that require relatively abundant skills should have *lower-than-average* wages while jobs that are less desirable or demand scarce skills should pay *higher* wages. A competitive labor market determines an implicit (hedonic) wage for each type of labor and in equilibrium labor mobility leads to a set of relative wages that makes workers indifferent between the various types of jobs. The differences in these implicit wages are called *compensating differentials*. Given the heterogeneity in workers' preferences and skills, both supply and demand for particular jobs determine the size of the compensating differential between jobs with different working conditions.

This paper uses recently developed econometric methodologies by Heckman and Vytlacil (2001, 2005, 2007) and Heckman Urzua and Vytlacil (2006) to analyze the relevance of labor market comparative advantage in the participation and earnings performance of workers in the self-employed and formal and informal salaried sectors in Argentina. These methods allow the investigation of the links between heterogeneous ability, earnings, and occupational choice connecting the treatment effects literature with conventional Mincer earnings analysis and the generalized Roy model (1951) of selectivity in occupational choice.

The approach addresses two key implications of the theory in the estimation of informal-formal earnings gaps and their policy interpretation. First, the 'treatment' impact, becoming formal, might be heterogeneous so workers could benefit differently depending on both their observed and unobserved characteristics. Second, estimation should address two types of selectivity bias, selection bias and sorting on the gain, generated by the correlation between unobserved characteristics that affect both earnings and job choice and the fact that the latter depends on the expected return to the observed and unobserved characteristics of the individual. In this case, conventional methods, OLS nor IV, do not provide consistent estimates of the earnings premium of formality for a randomly selected worker. Moreover, there is no single representative impact of formality on wages, that is, conventional mean regression estimates do not provide a full description of the presumed earnings gains that any given worker would derive from

getting a formal job.

The paper uses local instrumental variables, semi-parametric and polynomial methods to estimate a distinct set of treatment parameters, derived from the *marginal treatment effect*, to address different policy questions (Heckman and Vytlacil 2001, 2005; Heckman, Urzua and Vytlacil, 2006). In particular, we estimate the average treatment effect, the treatment on the treated, and the treatment on the untreated, for comparing earnings between formal salaried, informal salaried and independent workers.

Argentina presents a very suitable context to study these questions. Despite a half century of relative stagnation and macroeconomic volatility, it remains among the richest countries in the Latin American and Caribbean region, has one of the highest levels of human capital and a strong productive base. The country experienced the largest dramatic upward trend in informal salaried employment rates over the 1980s and 1990s, not limited to small firms only, while the share of self-employment remained relatively constant. This occurs in the context of volatile, modest economic growth, a sharp surge in unemployment, the erosion of the manufacturing base and union power, and arguably among the most rigid labor regulations in the region. Earnings analysis reveals that the self-employed and informal salaried seem to face an earnings disadvantage with respect to formal salaried workers suggesting the existence of segmentation. However, sociological survey work and related economic research has identified a significant importance of entrepreneurship and non-pecuniary motives for self-employment (World Bank, 2007).

The plan of the paper is as follows. First, we summarize the relevant empirical literature. Then, we outline a simple model of occupational choice, which highlights the case for the empirical strategy of marginal treatment effects estimation following Heckman, Urzua and Vytlacil (2006). Next, the econometric methodology, data and estimation specifications are discussed. This is followed by the discussion of the empirical results and their implications for the underlying questions that motivate the paper. The paper concludes with a summary of the findings and related policy implications.

2 Empirical tests of the 'exclusion' and 'competitive' labor market views

An extensive literature has examined empirically the two views of informal employment, the traditional 'exclusion' and the 'competitive' views. Here we only summarize some illustrative studies. Dickens and Lang (1985) used a switching earnings model with unknown regimes to test empirically the presence of dual labor markets in the USA. Their analysis suggests the presence of labor market segmentation and dual labor markets. In the Latin America context, Heckman and Hotz (1986) present evidence of selection-corrected earnings regressions that are consistent with labor market segmentation among males in Panama. Gindling (1991) also argues for labor market segmentation using selection corrected wage equations in Costa Rica. A study by Basch and Paredes-Molina (1996) employed a switching regression model with three equations with unknown sample separation to test the hypothesis of segmented labor markets for Chile, and finds support for the segmentation hypothesis. Fields (1990) argued that informal employment largely reveals the presence of segmentation in developing countries although he posits that a minority upper tier may conform to voluntary motives.

On the contrary, Magnac (1991) analyses segmented and competitive labor market in Colombia with an extended four-sector model. He concludes that comparative advantage in this case seems to be a more prevalent feature and finds support for sector choice being determined by tastes and not ability. More importantly, he argues that simply assessing the differences in earnings between formal and informal jobs cannot be used to test segmentation in the labor market. In a similar spirit, Pisani and Pagan (2004) test the notion of 'negative selection' and 'positive selection' in informal and formal sector participation in Nicaragua. For instance, workers with low skill levels participate in the informal sector while workers with high skill levels choose formal work. Using a three-equation switching model, they find positive selection for the formal sector and also for the informal sector, which suggests an element of individual choice. Pianto, Tannuri-Pianto and Arias (2004) propose quantile earnings regressions with selectivity bias corrections based on multinomial choice models of the choice between formal, informal salaried and self-employed in Bolivia. Their findings suggest segmentation at the lower quantiles of the earnings distribution (which they ascribe to workers with lower unobserved productivity) and little difference in earnings between formal and informal

jobs at higher quantiles of the earnings distribution, which they interpret as consistent with voluntary choice by higher productivity workers. Guenther and Launov (2006) test the proposition of segmented and competitive informal labor markets with an econometric model that accounts for unobservable sector affiliation and selection bias, and also found evidence of a two-tier structure in informal employment in Côte d'Ivoire. Yamada (1996), Maloney (1999), and Saavedra and Chong (1999), have also argued with evidence from Peru, Mexico, and Brazil, that informal self-employed workers are largely voluntary.

In the case of Argentina, two recent studies have tested the hypothesis of segmentation between informal and formal labor as the defining feature of the labor market. Pratap and Quintin (2006) use labor force survey data for 1993-1995 to test whether workers with similar observable characteristics earn more in the formal sector than in the informal sector. After controlling for selection on observables with propensity score matching and accounting for unobservables through a difference-in-difference matching estimator they find no significant difference between formal and informal earnings, evidence against the segmentation hypothesis. On the contrary, Alzua (2006) applies an endogenous switching regression model without ex-ante definition of sector and finds evidence in favor of segmentation of the labor market during 1970-1990 and 1991-2000.

As emphasized by Heckman, Urzua and Vytlacil (2006), the considerable lack of consensus in much of the empirical literature on labor market performance reflects the fact that the causal parameters being estimated are ill-defined. When earnings performance is heterogeneous and workers sort into different jobs on the basis of expected gains, conventional OLS, matching and IV estimation does not estimate a well-defined causal parameter that allow to extrapolate the impact of changes in an individual employment status on his earnings. Not only observable characteristics, but unobservable heterogeneity determine the returns and people sort according to their perceived individual returns in each sector, that is, their comparative advantage.

As noted by Magnac (1991) and stressed recently by Maloney (2004), informal-formal earnings gaps cannot offer unambiguous tests of segmentation. In a market with no rigidities, informal earnings should be higher to compensate workers for the lost value of benefits and whatever risk they may be facing. On the other hand, they may be lower to compensate for taxes evaded, greater independence and flexibility, or, perhaps for young workers, on-the job training. Even in the absence of compensating differentials, Galiani

and Weinschelbaum (2006) recently show that the efficient allocation of more productive labor and entrepreneurship can lead to a natural matching of lower productivity workers and informal small firms. Thus, selection biases and sorting based on gains and tastes are likely to be very relevant empirical drivers of formal and informal sector participation.

When choosing between informal and formal employment, workers weigh the advantages and disadvantages of each potential job, subject to the availability of jobs with their desired attributes. Informal and formal jobs differ by more than labor protections, and formal benefits are just one ingredient in workers' calculations. Workers equilibrate utilities–not just earnings–in choosing between jobs in the two sectors. Comparative advantage could make the informal sector a better match for many labor market participants. Lucas (1978) argued that individuals choose between salaried work and self-employment, depending on whether they are relatively more talented as an entrepreneur or as a salaried employee. Some workers might find that their observed and unobserved skills are better rewarded in occupations, which have a higher propensity to be informal (such as those in construction). Informal jobs may offer an entry point to the labor market for youth and unskilled middle-age workers that partially remedies deficient or obsolete skills through on-the-job training unavailable to them in formal salaried jobs. Women, particularly of young age with children, might be willing to forgo some of the benefits of formality in exchange for the flexibility of informal employment.

A novelty of our paper is the application of the recently developed marginal treatment methods for models of essential heterogeneity developed by Heckman, Urzua and Vytlacil (2006) to examine the links between earnings performance and the choice of formal and informal salaried work and self-employment. This method allows to account for observable and unobservable characteristics of the individuals that affect their decision to participate in different occupations. This is done through the explicit estimation of the marginal return of an individual indifferent between a formal and informal job or between self-employment and dependent worker status. From this one can derive the standard treatment parameters, average treatment effect, treatment on the treated and treatment on the untreated. This is a unique feature of this paper compared to the previous literature, which does not properly estimate these treatment parameters for the different sectors in the labor market. From these it is possible to draw conclusion whether an individual at the margin of indifference between different job types would gain or loose in terms of wages given his observed and unobserved characteristics. Depending on the

margin of comparison, this in turn would give an indication whether the segmentation or integration, or equivalently whether 'exclusion' or 'competitive' forces, are the important defining features of the labor market.

3 A model of occupational choice

To formally spell out the issues outlined earlier, consider a simple parametric formulation of selectivity in occupational choice, based on the Roy model (1951), that connects the comparative advantage framework and the treatment effects literature.

Suppose there are two types of occupations indexed by two labor market sectors s: 1 for dependent salaried work and 2 for self-employment.[1] Workers choose their occupation by comparing the utility W_s they derive from each occupation, which is given by the sum of the income Y_s and non-pecuniary benefits in the sector ε_s net of costs c_s (pecuniary and non-pecuniary) of sector participation. Adopting a latent index formulation we have:

$$W_{si}^* = Y_{si} + \varepsilon_{si} - c_{si} = Z_i'\gamma_s + \eta_{si} \tag{1}$$

where W_{si}^* depends linearly on the vectors of observed Z (e.g. human capital, demographics) and unobserved characteristics η (tastes for work, intrinsic abilities) of the worker i. A worker chooses a formal occupation when the net benefits of being formal, in welfare terms, are positive:

$$W_1 \geq W_2 \Longleftrightarrow (Y_1 - Y_2) + (\varepsilon_1 - c_1) - (\varepsilon_2 - c_2) \geq 0 \quad \text{iff} \quad Z'(\gamma_1 - \gamma_2) \geq \eta_1 - \eta_2 \tag{2}$$

Since we only observe participation choices we shall consider the probability of sector participation conditional on $Z = z$, or in the language of impact evaluation the probability of receiving treatment or the propensity score, in this case $s = 1$ or being formal, given by $P(z)$:

$$P(s = 1|Z = z) = P(\Delta W \geq 0) \Longleftrightarrow P(Z'(\gamma_1 - \gamma_2) \geq \eta_1 - \eta_2) \tag{3}$$

[1] Other margins of choice such as formal salaried versus informal salaried worker are also represented in this model.

where the $\eta_s$have a common distribution F_{η_s}.

We only observe earnings after participation choices are made, so we should consider two potential outcomes for any given worker. For a given choice of hours of work, the potential earnings of any given individual in each sector can be written as:

$$\ln y_1 = \alpha_1 + X'\beta_1 + \mu_1 \text{ and } \ln y_2 = \alpha_2 + X'\beta_2 + \mu_2 \tag{4}$$

where X is a subset of Z and (μ_1, μ_2) are freely correlated and independent of some components of Z, the 'instruments'. The μ_s can depend on η_s in a general way.

In this context the *average treatment effect* (ATE) or mean earnings differential between dependent salaried and self-employed work conditional on $X = x$ is given by:

$$E(\ln y_1 - \ln y_2 | X = x) \equiv \bar{\beta}(X = x) = (\alpha_1 - \alpha_2) + x'(\beta_1 - \beta_2) \tag{5}$$

This yields a random coefficient earnings model with self-selection. There are two key implications of the theory for the estimation of earnings gaps in this model with attending policy implications (Heckman and Vytlacil, 2001, 2005):

(i) There is no single 'representative' impact of dependent salaried work on wages, i.e. estimates of the ATE do not provide a full description of the earnings gains that any given worker would derive from getting a salaried job. The 'treatment' impact, becoming a salaried worker, is heterogeneous, so workers would benefit differently depending on their observed and unobserved characteristics.

(ii) The estimation should address selection bias and sorting on the gain generated by the fact that the decision to participate in the salaried worker sector depends on the expected earnings return for the individual. Conventional methods such as OLS and IV do not provide an unbiased consistent estimate of the ATE for a randomly selected worker in the presence of heterogeneity and selection (Heckman and Li, 2003).

In the context of this paper, the formal-informal earnings gaps can be affected by the spurious correlation induced by unobserved worker characteristics that affect earnings and cause selection (either by choice or rationing) into formal, informal salaried or independent work. The most talented individuals may be more likely to obtain formal salaried employment because of better prospects for mobility in a career as wage earner. Individuals with more entrepreneurial ability are more likely to succeed as independent workers. On the other hand, those with low work attachment and little adherence to

authority or rigid work schedules may be excluded from formal salaried employment or voluntarily seek the flexibility of self-employment even at the cost of lower earnings. In general, the occupational structure in part reflect differences in individual tastes for work (e.g., industriousness, preference for flexible work schedules and/or being one's own boss), the value attached to social protection (quality of health, unemployment, old age benefits), as well as constraints to being in either sector (lack of capital, connections) and the costs of non-compliance with state regulations (e.g., penalties, social stigma).

In this context it is possible to estimate a wide ranging set of parameters that may answer different policy questions (Heckman and Li, 2003; Heckman, Urzua and Vytlacil, 2006; Heckman and Vytlacil, 2007). To investigate the role of comparative advantage in occupational choice as opposed to the segmentation hypothesis the following treatment parameters are of particular interest, with implicit conditioning on $X = x$:

The *treatment on the treated* (TT), the mean wage gain from dependent salaried work for those who are currently in salaried employee jobs,

$$E(\ln y_1 - \ln y_2 | s = 1) \equiv \bar{\beta}(s = 1) = (\alpha_1 - \alpha_2) + x'(\beta_1 - \beta_2) + E(\mu_1 - \mu_2 | s = 1) \quad (6)$$

The *treatment on the untreated* (TUT), the mean wage gain for those in self-employment were they to switch to salaried jobs,

$$E(\ln y_1 - \ln y_2 | s = 0) \equiv \bar{\beta}(s = 0) = (\alpha_1 - \alpha_2) + x'(\beta_1 - \beta_2) + E(\mu_1 - \mu_2 | s = 0) \quad (7)$$

These treatment parameters and other can be derived as weighted averages from an estimate of the *marginal treatment effect* (MTE),

$$E(\ln y_1 - \ln y_2 | \Delta W = 0) \equiv \bar{\beta}(X = x, \eta = \ddot{\eta}) =$$

$$(\alpha_1 - \alpha_2) + x'(\beta_1 - \beta_2) + E(\mu_1 - \mu_2 | X = x, \eta = \ddot{\eta}) \quad (8)$$

This is the mean wage gain from having a dependent salaried occupation for those workers who are indifferent between salaried and self-employed job conditional on $X = x$ and at the level of unobservable characteristics $\eta = \ddot{\eta}$. As noted by Heckman and Vytlacil (2001, 2005) equivalently this can be derived from conditioning on the propensity scores

given the monotonicity of the latent variable model. The MTE can be also interpreted as a 'willingness to pay' measure (Heckman, 2001). For instance, in the case of formal salaried and self-employment it gives a measure of the earnings a self-employed worker is willing to forgo in exchange for non-pecuniary benefits such as more flexibility in the job or being independent.

From these parameters we can derived measures of two types of biases: selection bias and the bias that arises from the sorting of workers based on expected gains (Heckman and Li, 2003). The selection bias is determined by the difference of the OLS estimate and TT . Meanwhile the differences TT-ATE and TUT-ATE yield the sorting gains, say, how salaried and self-employed-like workers gain from participating in the salaried and self-employed sectors, respectively, compared to randomly sampled workers. Presence of large, positive sorting gain indicate that comparative advantage considerations of workers are a feature of the labor market (Heckman and Li, 2003). In this paper we take the following as evidence of *comparative advantage* in the labor market: There are differences in returns to unobserved characteristics η across the labor market sectors and people self-select into different occupations or job types based on these returns or tastes.

4 Estimation and data

This section outlines the empirical method for the estimation of the marginal treatment effect and related parameters following Heckman, Urzua and Vytlacil (2006). Thereafter the specific data collected for this study, the estimation specifications, and in particular the 'instruments', are presented.

4.1 Empirical methodology

The MTE outlined in equation (8) can be estimated with parametric, polynomial and semiparametric techniques (Heckman, Urzua and Vytlacil, 2006).[2] The key term for the estimation is

$$E(\mu_1 - \mu_2 | X = x, \eta = \ddot{\eta}) = K'(z) \tag{9}$$

[2]This paper employs the recently developed MTE software by Heckman, Urzua and Vytlacil (2006). We are very grateful to Sergio Urzua for invaluable help with the implementation of the routine

where $K'(z) = \frac{\partial K(z)}{\partial z}\Big|_{z=\eta}$ is the function of unobservables given the particular propensity score z and treatment decision. In the standard Heckit method this term would be equivalent to the inverse Mills ratio. The MTE to be estimated is the following

$$MTE = (\alpha_1 - \alpha_2) + x'(\beta_1 - \beta_2) + K'(z) \tag{10}$$

The parametric estimator estimated the MTE with the standard normal distribution for the error terms/unobservables. This implies that it is possible to estimate the term $K'(z)$ as a function of the standard normal random variable. This results in a flat MTE across unobservables (Heckman, 2001).

Heckman, Urzua and Vytlacil (2006) show that the MTE method in the semiparametric case relaxes the assumption of homogeneity of the MTE and assumes essential heterogeneity. Here, wage outcomes of the occupational choice are heterogeneous and individuals participate with partial knowledge of their individual gain or loss from the labor market status, which differs among individuals (Heckman, Urzua and Vytlacil, 2006).

Heckman and Vytalcil (2001ab, 2005) show that the local instrumental variable (LIV) estimator yields a semiparametric MTE. Following Heckman, Urzua and Vytlacil (2006) $(\beta_1 - \beta_2)$ and $K'(z)$ need to be estimated. Values for $(\beta_1 - \beta_2)$ are obtained through a semiparametric double residual regression procedure (Robinson, 1998; Heckman, Ichimura, Smith and Todd, 1998). Local linear regressions of regressors x on $P(z)$ and of outcomes y on $P(z)$ provide the residuals, from which $(\beta_1 - \beta_2)$ is obtained through double residual regression. Then the term $K'(z)$ is estimated with standard nonparametric techniques. So, contrary to the parametric case, which exploits a known functional form for the estimation of $K'(z)$, here a more general form in the semiparametric case is estimated through nonparametric technique. From the results of $(\beta_1 - \beta_2)$ and $K'(z)$ the semiparametric MTE is computed over the common support of the propensity scores z (Heckman, Urzua and Vytlacil, 2006). Contrary to the parametric MTE the estimates of the semiparametric MTE, using the local instrumental variables, does not result in a flat MTE across all unobservables. The treatment effect at the margin is not homogeneous, but heterogeneous across different levels of unobservables, which determine participation in the occupation.

4.2 Data and empirical specification[3]

The paper exploits unique labor force survey data together with a supplementary informality survey and administrative data on enforcement of labor laws. We use the Argentine national household survey, the Continuous Permanent Household survey (EPH-C), for the second semester and fourth trimester 2005. This household survey covers about 31 urban areas in the country and thereby about 60 percent of the Argentine population. The survey collects data on demographics, education, income, employment, benefits and social security contribution of individuals.

In addition to the standard questionnaires of the EPH-C, the Argentine national statistical office (INDEC), with support from the World Bank, implemented a one-time informality module for the Greater Buenos Aires area, which was attached to the regular EPH-C in the fourth trimester 2005. This survey collects new, unique data on the intrinsic preferences of workers for salaried work or self-employment, the multiple motivations for formal and informal salaried work and for self-employment, participation in the social security system, individual occupational histories, degree of informality of firms and private arrangements to insure against old-age risks.

Moreover, we collect data from the Argentine Ministry of Labor on the number of workers inspected for violation of labor laws (including social security contributions) per province for the year 2005. In the presence of large informality, especially after the Argentine crisis in 2001/02, the Argentine government stepped up the enforcement of labor legislation, through the "Plan Nacional de Regularizacion del Trabajo" (PNRT) in September 2003 (Ministerio de Trabajo, 2004ab). Under this plan labor inspections examined the level of compliance with labor laws, including social security registration of workers by firms. At the time of the inspection visit, inspectors would cross-check the databases of the tax agency with whether the employees are registered or not. Fines for non-registration are imposed. A main goal of the PNRT is the registration of workers to the social security system (Ministerio de Trabajo, 2004ab). The allocation of the number of labor inspectors, hence also the number of inspected workers and firms, under the PNRT varies between provinces and largely depends on the population size of the province and the levels of informality measured. In order to account for these factors in the allocation of workers, the analysis also controls for population and GDP per capita

[3]Descriptive summary statistics and variable descriptions can be found in the appendix 2.

per province from the 2001 national census and the Province of Buenos Aires Ministry of Economy.

Three different groups of labor market participants are employed in the estimations and they provide the basis for the different occupational choice margins. These are: *Formal salaried workers* are workers working in a dependent employee relationship with social security contribution through automatic pay reduction or voluntarily; *Informal salaried workers* are workers working in a dependent employee relationship without social security contribution; and *Self-employed or independent workers* constitute the group of independent workers with no employees and microentrepreneurs of small firms with 1 to 5 employees. The margins of choice and earnings comparisons are the following: dependent salaried work (formal or informal) versus self-employment (margin 1 and margin 2 respectively) and formal versus informal salaried work (margin 3).

The dependent variable in the probit model of participation is coded 1 if the individual works in the treated status and 0 if the individual works in the comparison work status. The treated and comparison work status depends on the margin of comparison estimated: For margin 1 and 2 the dependent worker status (for margin 1 formal workers and for margin 2 informal workers) is the treatment group and the self-employed are the non-treated. For margin 3 formal salaried workers form the treatment group while informal salaried workers are the comparison group. The dependent variable in the outcome equation is the natural logarithm of labor income per hour in the main occupation. The earnings model follows a standard Mincer equation with additional controls (Mincer, 1974).[4] The margin 1 model is estimated only for Greater Buenos Aires given the availability of variables that could serve as instruments (see below).

Initial tests of the data show that the marginal treatment effect estimation under essential heterogeneity proposed by Heckman, Urzua and Vytlacil (2006) is applicable to the margins of choice between self-employment, formal and informal salaried workers. Essential heterogeneity implies that outcomes of choices, here the wages for the different sectors, are heterogeneous in a general way while the choices itself are not heterogeneous in a general way (Heckman, Urzua and Vytlacil, 2006). Individuals make their choices with partial knowledge of the outcomes. In our initial tests of the data, using quantile wage regressions with selectivity correction terms estimated with a multinomial choice model (as in Tannuri, Pianto and Arias, 2004), we found that this was reflected in the

[4]For the variable descriptions, including the base category for the dummy variables, see appendix 2.

differential magnitudes and significance of the selection-correction terms.[5]

In the estimations the participation/choice model for the different margins of comparisons includes the observable characteristics that are also included in the outcome/wage model and most crucially the 'instruments' that are not included per se in the wage model and only enter through the estimation of the propensity score. The actual instruments, which entered in the estimation for the propensity of participation equation differed for the specifications of the different margins of occupational choice. In order to get consistent estimates of the MTE and related parameters, we need correct specification of the instruments in the propensity scores and outcome equations (Heckman, Urzua and Vytlacil 2006). We find strong suggestive evidence that these conditions are satisfied since the instruments enter significantly in the choice model but not in the Mincer equations.

For the dependent worker (formal or informal)-self-employed margins the propensity scores were estimated using as instruments the workers' reported intrinsic *preference for being self-employed or a salaried worker*, from responses to the question "if you were able to choose, would you rather be a salaried worker or an independent worker?" in the supplementary informality survey in Greater Buenos Aires. This was found to be a significant determinant of occupational choice as can be seen by the significance in the choice model, and other results show that it does not enter significantly in the earnings Mincer model. This is in line with other research on self-employment and motivations for self-employment which point at this being driven by largely idiosyncratic motives (Oswald, Blanchflower and Stutzer, 2001; Cunningham and Maloney, 2001). Similar results hold for variables constructed from the workers' reported motivations to be self-employed (i.e., flexibility, desire of independence, or inability to find salaried employment). Other individual-level instruments included having the spouse of other relatives employed in the formal sector, which as suggested by Pratap and Quintin (2006) affects sector participation and is uncorrelated with wage outcomes.

For the formal-informal salaried margin the main instruments included to estimate the propensity score were the number of inspected workers at the province of residence as a proxy for the *cost of informality*, (De Soto, 1989). Workers living in provinces

[5]In our initial tests of the data we employed a three-way choice model (formal salaried, informal salaried and self-employed) for the quantile selection-correction. However, this is not possible as of yet with current estimation routines in the MTE framework which only allow estimation of the marginal treatment effects in a two-way choice model.

with a higher number of inspected workers have a higher propensity to be employed as formal salaried. We also included the indicators for having the spouse of other relatives employed in the formal salaried sector. These also entered significantly in the propensity scores regressions. This follows Heckman and Li (2003), who also include both regional and individual-level instruments, such as the provincial unemployment rate, parental education and income, as the determinants of the probability of going to college.

5 Results and implications[6]

The results are presented in Figures 1 to 9 and Tables 1 to 16. The tables, in particular, present a distinct set of summary parameters to answer different policy questions: (i) The *average treatment effect* (ATE), i.e., the mean earnings gain from formality for a randomly selected worker; (ii) The *treatment on the treated* (TT), i.e., the mean earnings gain from formality derived by those workers selecting into formal jobs; (iii) The *treatment on the untreated* (TUT), i.e., the mean earnings gain (or loss) for those in informal (salaried on independent) jobs were they to switch to formal salaried jobs. As shown by Heckman and Vytlacil (2001, 2005) these parameters can be derived from an estimate of the *marginal treatment effect* (MTE) using local instrumental variables (LIVs). The tables show the estimates obtained with parametric, semi-parametric and polynomial estimators (see Heckman, Urzua and Vytlacil, 2006).[7] These are alternative measures of the mean earnings gain from having a formal occupation for workers with the same set of observed and unobserved characteristics, who are indifferent between a formal and an informal job and are found participating in different sectors. The figures present the full MTE estimates from which these are derived.

The results corroborate the mixed view of the Argentine labor market and support the importance of both comparative advantage and segmentation in workers selection into formal and informal salaried work and self-employment. On the one hand, the results reveal little difference in the earnings of formal salaried and independent workers once one fully account for the sorting of workers based on preferences and the returns

[6] In this paper the results for the parametric and semiparametric LIV estimation are emphasized. Results for the polynomial and an alternative semiparametric estimator are in the appendix. Detailed results for these estimations are available upon request.

[7] The results shown here are robust to different empirical specifications and alternative IVs.

to their observed and unobserved skills. All three treatment parameters are statistically insignificant. When compared with informal salaried workers, the self-employed are in a clear advantage. All treatment parameters are negative and of very similar magnitude in the semi-parametric estimations, while the polynomial results suggest that TT>ATE> TUT. That is, workers with independent-like characteristics (observed and unobserved) would receive much lower earnings were they to move to informal salaried jobs.

On the contrary, for informal salaried workers all treatment parameter estimates are positive and large, and TT>ATE>TUT with only slight differences. That is, although there is evidence of some heterogeneity in the earnings gains that informal salaried workers would derive from formal employment, the differences are not big. Informal salaried work carry very large earnings penalties compared to formal salaried work regardless of the propensity to select into formal salaried employment. That is, workers with informal-like characteristics (observed and unobserved) would experience roughly similar earnings gains were they to move to formal salaried jobs.

The results indicate that selection and sorting biases are important features of these data. Table 16 present the estimated selection and sorting biases derived from the estimated parameters as in Heckman and Li (2003) for each estimation approach. There is positive selection bias into formal salaried work compared to self-employment, but little evidence of sorting based on gains. Those entering self-employment in Argentina appear to be driven by differences in tastes for type of work and not so much for differences in the returns to their observed or unobserved skills in the two sectors. This again underscores the importance of considering differences in the non-pecuniary qualities of independent work. On the other hand, there is a considerable negative selection bias into formal relative to informal salaried work and modest positive sorting based on expected earnings gains—resulting in an overall large downward biased in conventional OLS formal-informal earnings gaps. That is, formal salaried workers would lose out considerably were they to become informal salaried. Unobserved salary work attributes are rewarded modestly more in formal jobs.

To the extent that these are derived from comparing identical workers at the margin of indifference between the two sectors, they provide measures of differences in earnings arising from non-pecuniary characteristics of jobs that affect sector choice or from labor market disequilibria or segmentation. In particular, the MTE has the interpretation of a willingness-to-pay measure, for instance, the earnings that a self-employed worker at the

margin of indifference would be willing to forego in exchange for the labor benefits of a formal salaried job. The absence of compensating differentials between formal salaried work and independent work suggests that the perceived amenities (i.e., flexibility) and disamenities (e.g., risk) of self-employment tend to cancel out as predicted by the generalized Roy (1951) model. This and other evidence points to compensating welfare differentials as the main driver of the choice between salaried work and self-employment in Argentina.

In the case of the formal-informal salaried margin, however, the magnitude of earnings gaps seems very large to arise from compensating earnings differentials and suggest the presence of segmentation between informal and formal salaried employment. As argued by Magnac (1991), the test of the competitive model of comparative advantage with micro-data is not capable of properly accounting for this type of disequilibria in the labor market. Overall, our results are less consistent with informal salaried work resulting from choice driven by compensating welfare differentials and seem more consistent with labor market segmentation.

These results are entirely consistent with workers' reported motivations to be independent and in informal salaried jobs in Argentina. In responses to the special informal employment survey, most of the self-employed state primarily voluntary motivations to be independent: 70 percent of independent workers prefer to be independent than to work as salaried workers, citing reasons like flexibility, better mobility opportunities and being their own bosses as the main reasons for that preference. On the contrary, the vast majority of informal salaried workers are so involuntarily: more than 90 percent report that the main reason for being informal is that their employer would not hire them with regulated benefits rather than reflecting a consensual agreement for them to obtain higher earnings, and a majority say that their current job is the only employment they could get.

6 Conclusions

This paper uses recently developed econometric models of essential heterogeneity (e.g., Heckman and Vytlacil, 2001, 2005; Heckman, Urzua and Vytlacil, 2006) to analyze the relevance of labor market comparative advantage and segmentation in the participation and earnings performance of workers in formal and informal jobs in urban Argentina.

The paper estimates the *marginal treatment effect* (the mean earnings gain from having a formal job for workers at the margin of indifference between the sectors), the *average treatment effect* (the mean earnings gain for a randomly selected worker), the *treatment on the treated* (the mean earnings gain from formality for those who select into formal jobs), and the *treatment on the untreated* (the mean earnings gain for those selecting into informal jobs).

The results support the importance of both comparative advantage and segmentation in Argentina's informal-formal employment composition. On the one hand, there are not significant differences between the earnings of formal salaried workers and the self-employed regardless of the propensity to select into each sector (all treatment parameters are 0), but there is positive selection bias into formal salaried work based on tastes. This and other evidence points to compensating welfare differentials as the main driver of the choice between salaried work and self-employment in Argentina. Workers sort into formal salaried and self-employment occupations according to labor market comparative advantage. That is, some workers find advantageous niches for their observed and unobserved skills in sectors or occupations where jobs have a different propensity to be exercised as formal salaried or independent.

On the other hand, for the formal-informal salaried margin all treatment parameters are positive and large, and TT>ATE>TUT with only slight differences. That is, informal salaried employment carries significant earnings penalties regardless of the propensity to select into formal salaried employment. There is a considerable *negative selection bias* into formal relative to informal salaried work and modest positive sorting based on expected earnings gains—resulting in an overall large *downward bias* in conventional OLS formal-informal earnings gaps. That is, formal salaried workers would lose out considerably were they to become informal salaried. Overall, these results are less consistent with choice driven by compensating welfare differentials and seem more consistent with segmenting forces. The results are robust to different empirical specifications and are consistent with individuals' reported reasons for being formal and informal salaried or self-employed.

Thus, the paper lends credence to both the 'exclusion' and 'voluntary' nature of informal employment. Independent workers are largely voluntary and implicitly attach significant value to the non-pecuniary benefits of autonomous work. Meanwhile, informal salaried workers tend to be excluded from more desirable jobs either formal salaried or

self-employment.

The existence of a sizeable earnings differential between informal and formal salaried workers, unrelated to compensating differentials, has implications for the functioning of labor markets in developing countries like Argentina. This can reflect 'queues' for formal salaried sector jobs given that they are comparatively better-paid across the spectrum of low and high paid jobs in the labor market and have social benefits. This is a product of the labor market not being flexible and competitive enough to equalize earnings through arbitrage. This may reflect numerous sources of labor segmentation, including evasion of general (income, VAT), labor market frictions, which must be addressed with tighter enforcement of improved labor and tax laws and improved collective bargaining.

The results suggest that independent workers reveal no willingness to pay for the social protection benefits (social security, health) that formal wage earners enjoy. This highlights the issue of how to engineer incentives for voluntary participation in the social security system of workers with different preferences regarding job flexibility, with different concerns with respect to their future, with different intertemporal discount rates and who may derive different levels of welfare from a particular benefit package. Workers may have a different willingness to pay or accept lower take-home earnings in exchange for such benefits depending on their preferences, the cost and quality of the services (real and perceived) provided by the public and private sectors and the characteristics of alternative sources of services and benefits not related to the labor contract (e.g. informal insurance, social networks, etc.). Analyses like those provided in this paper for other developing country contexts may serve to inform this important policy question.

References

[1] Arias, Omar, Kevin F. Hallock and Walter Sosa-Escudero. 2001. "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data." Empirical Economics, 26, pp. 7-40.

[2] Alzua, Maria Laura. 2006. "Are informal workers secondary workers?: Evidence from Argentina." Working Paper. Boston University.

[3] Basch, Michael and Ricardo D. Paredes-Molina. 1996. "Are there dual labor markets in Chile?: empirical evidence." Journal of Development Economics, 50, pp.297-312.

[4] Blanchflower, David G. and Andrew J. Oswald. 1998. "What makes an entrepreneur?" Journal of Labor Economics, 16, pp.26-60.

[5] Blanchflower, David G., Andrew J. Oswald and Alois Stutzer. 2001."Latent entrepreneurship across nations." European Economic Review, 45, pp. 680-691.

[6] Carneiro, Pedro and James J. Heckman. 2002. "The Evidence on Credit Constraints in Post-secondary schooling." Economic Journal, 112, pp. 705-734.

[7] Carneiro, Pedro, James J. Heckman and Edward Vytlacil. 2005. "Estimating the Return to Education When It Varies Among Individuals.", Working Paper. Unpublished. University of Chicago.

[8] Cunningham, Wendy V. and William F. Maloney. 2001. "Heterogeneity among Mexico's Microenterprises: An application of factor and cluster analysis." Economic Development and Cultural Change, 50, pp.131-156.

[9] De Soto, Hernando. 1989. The Other Path. Harper and Row.

[10] Dickens, William T. and Kevin Lang. 1985. "A test of dual labor market theory." American Economic Review, 75, pp.792-805.

[11] Fields, Gary S..1990."Labour market modelling and the urban informal sector: theory and evidence." in The Informal Sector Revisited. David Thurnham, Bernard Salomé and Antoine Schwarz ed: OECD. Paris.

[12] Galiani, Sebastian and Federico Weinschelbaum. 2006. "Modeling Informality Formally: Households and Firms." Mimeo.

[13] Gindling, T.H.. 1991. "Labor market segmentation and the determination of wages in the public, private-formal, and informal sectors in San José, Costa Rica." Economic Development and Cultural Change, 39, pp.585-605.

[14] Guenther, Isabel and Andrey Launov. 2006. "Competitive and Segmented Informal Labor Markets." Ibero-America Institute for Economic Research Discussion Paper Nr. 153. Georg-August-Universitaet Goettingen.

[15] Harris, J.R. and M.P. Todaro. 1970. "Migration, Unemployment and Development: A two sector analysis." American Economic Review, 60, pp.126-142.

[16] Heckman, James J.. 2001. "Microdata, heterogeneity and the evaluation of public policy."Journal of Political Economy, 109, pp.673-748.

[17] Heckman, James J. and Joseph Hotz. 1986. "An investigation of the labor market earnings of panamanian males: evaluating the sources of inequality." Journal of Human Resources, 21, pp.507-542.

[18] Heckman, James J., H. Ichimura, J. Smith and P.E. Todd. 1998. "Characterizing selection bias using experimental data." Econometrica, 66, pp. 1017-1098.

[19] Heckman, James J. and Xuesong Li. 2003. "Selection bias, comparative advantage and heterogeneous returns to education: evidence from China in 2000." NBER Working Paper Number 9877.

[20] Heckman, James J. and Guilherme Sedlacek. 1985. "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market." Journal of Political Economy, 93, pp.1077-1125.

[21] Heckman, James J., Sergio Urzua and Edward J. Vytlacil. 2006. "Understanding instrumental variables in models with essential heterogeneity." Review of Economics and Statistics, 88, pp.389-4320.

[22] Heckman, James J. and Edward J. Vytlacil. 2001a. "Local Instrumental Variables." in Nonlinear Statistical Modeling: Proceedings of the Thirteenth International

Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya. C.Hsiao, K. Morimune, and J.L. Powell ed.:Cambridge University Press. New York.

[23] Heckman, James J. and Edward J. Vytlacil. 2001b. "Policy-relevant treatment effects." American Economic Review, 91, pp.107-111.

[24] Heckman, James J. and Edward J. Vytlacil. 2005. "Structural equations, treatment effects and econometric policy evaluation." Econometrica, 73, pp. 669-738.

[25] Heckman, James J. and Edward J. Vytlacil. 2007a. "Econometric Evaluation of Social Programs, Part I: Causal models, structural models and econometric policy evaluation." in Handbook of Econometrics, 6. J.J. Heckman and E.Leamer ed.: Elsevier. Amsterdam

[26] Heckman, James J. and Edward J. Vytlacil. 2007b. "Econometric Evaluation of Social Programs, Part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments." in Handbook of Econometrics, 6. J.J. Heckman and E.Leamer ed.: Elsevier. Amsterdam.

[27] Lazear, Edward P.. 2005. "Entrepreneurship." Journal of Labor Economics, 23, pp.649-680.

[28] Lucas, Robert E. Jr.. 1978. "On the size distribution of business firms." The Bell Journal of Economics, 9, pp.508-523.

[29] Maloney, William F..1999."Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico." World Bank Economic Review, 13, pp.275-302.

[30] Maloney, William F.. 2004. "Informality Revisited." World Development, 32, pp.1159-1178.

[31] Magnac, Th.. 1991. "Segmented or Competitive Labor Markets." Econometrica, 59, pp.165-187.

[32] Mincer, Jacob. 1974. Schooling, Experience and Earnings. Columbia University Press. New York.

[33] Ministerio de Trabajo (MTSS). 2004a. "Plan Nacional de Regularización del Trabajo, 2 de enero de 2004, Informe." Ministry of Labor, Argentina.

[34] Ministerio de Trabajo (MTSS). 2004b. "Plan Nacional de Regularización del Trabajo, Período: 1° septiembre de 2003 al 23 marzo de 2004, Informe." Ministry of Labor, Argentina.

[35] Ñopo, Hugo and Patricio Valenzuela. 2007."Becoming an entrepreneur." IZA Discussion Paper Series No.2716.

[36] Pianto, Donald M., Maria Tannuri-Pianto and Omar Arias. 2004. "Informal employment in Bolivia: a lost proposition?" Background Paper for the 2005 World Bank Bolivia Poverty Assessment.

[37] Piore, Michael J.. 1979. Unemployment and inflation.M.E. Sharpe. New York.

[38] Pisani, Michael J. and José A. Pagán. 2004. "Sectoral Selection and Informality: a Nicaraguan Case Study." Review of Development Economics, 8, pp.541-556.

[39] Pratap, Sangeeta and Erwan Quintin. 2006. "Are labor markets segmented in developing countries? A semiparametric approach." European Economic Review, 50, pp.1817-1841.

[40] Robinson, P.M.. 1988. "Root-N-Consistent semiparametric regression." Econometrica, 56, pp.931-954.

[41] Rosen, Sherwin. 1978. "Substitution and Division of Labour." Economica, 45, pp.235-250.

[42] Roy, Andrew D. "Some thoughts on the distribution of earnings." Oxford Economic Papers, 3, pp. 135-146.

[43] Saavedra, Jaime and Alberto Chong. 1999. "Structural Reform, Institutions and Earnings: Evidence from the Formal and Informal Sectors in Urban Peru." Journal of Development Studies, 35, pp.95-116.

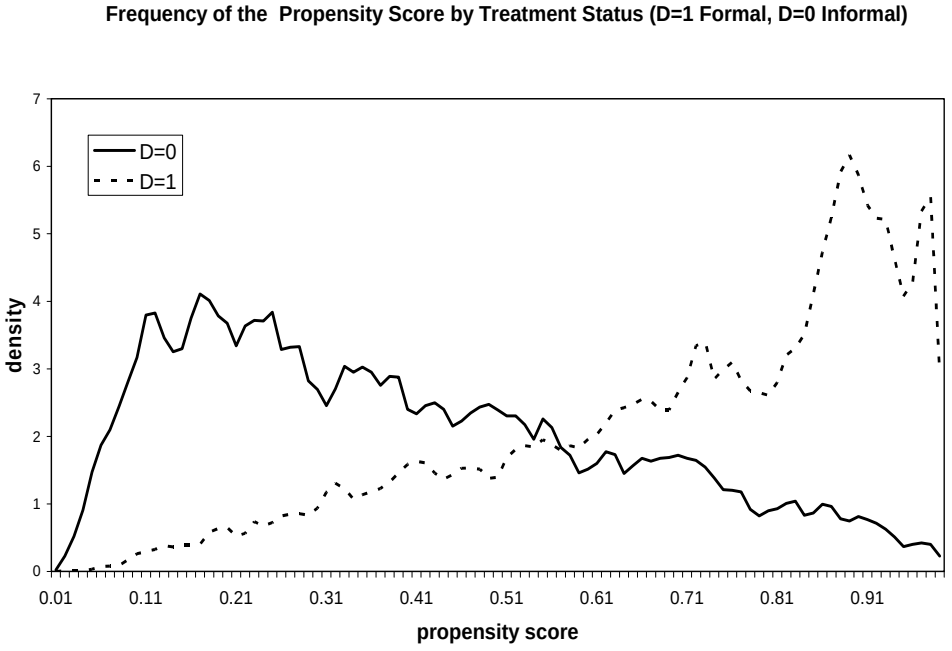
[44] Smith, Adam. 1776. The Wealth of Nations. Reprinted, New York: Modern Library, 1937.

[45] Willis, Robert J. and Sherwin Rosen. 1979. "Education and Self-selection." Journal of Political Economy, 87, pp.7-36.

[46] Yamada, Gustavo. 1996. "Urban Informal Employment and Self-Employment in Developing Countries: Theory and Evidence." Economic Development and Cultural Change, 44, pp. 289-314.

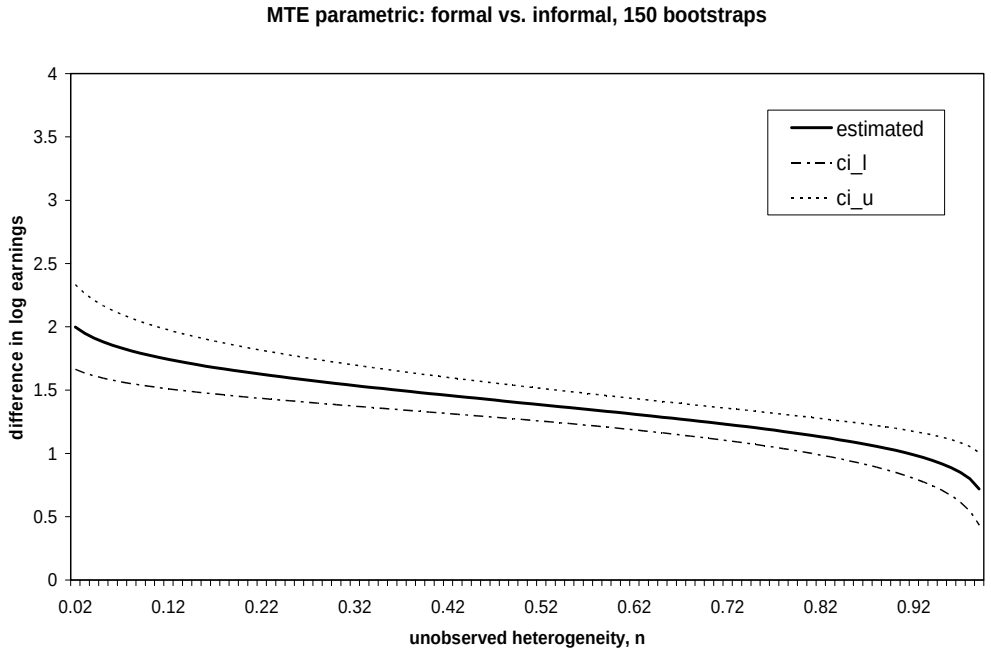
Appendix 1: Figures and Tables

Figure 1: Formal and Informal salaried workers: Common Support



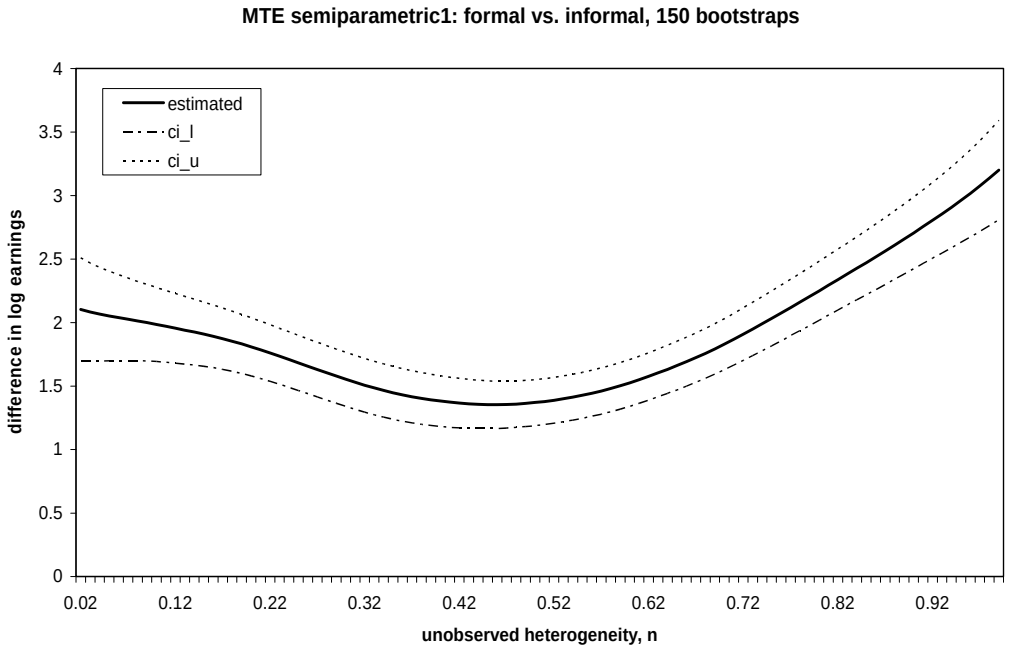
Source: Author's estimations based on EPH-C, INDEC.

Figure 2: Formal and Informal salaried workers: MTE - parametric



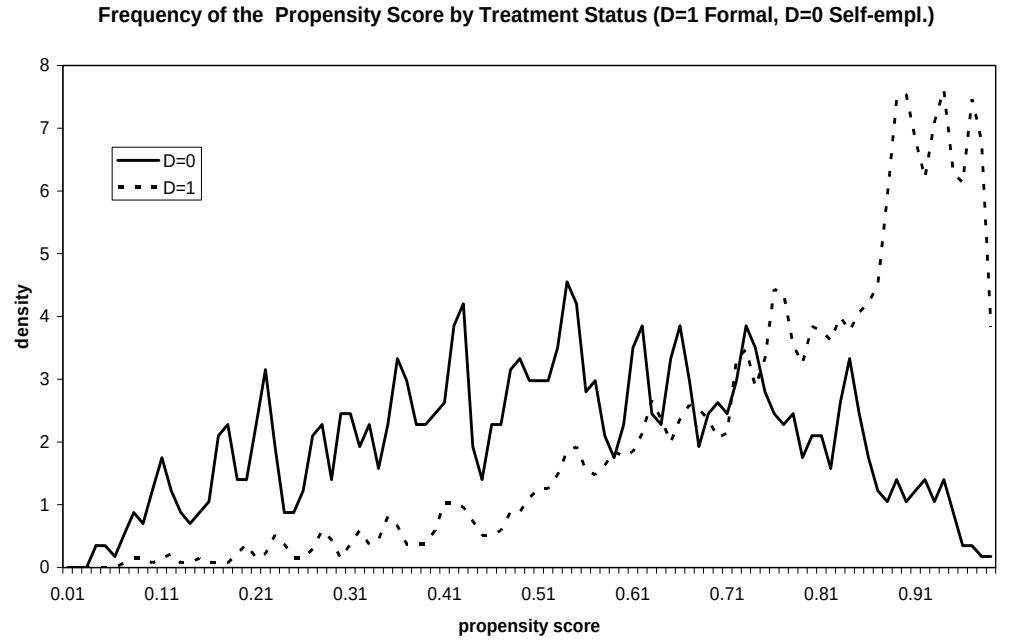
Note: ci_l, ci_u: lower and upper confidence interval. Source: Author's estimations based on EPH-C, INDEC.

Figure 3: Formal and Informal salaried workers: MTE – semiparametric1



Note: ci_l, ci_u: lower and upper confidence interval. Source: Author's estimations based on EPH-C, INDEC.

Figure 4: Formal salaried workers and Self-employed: Common Support

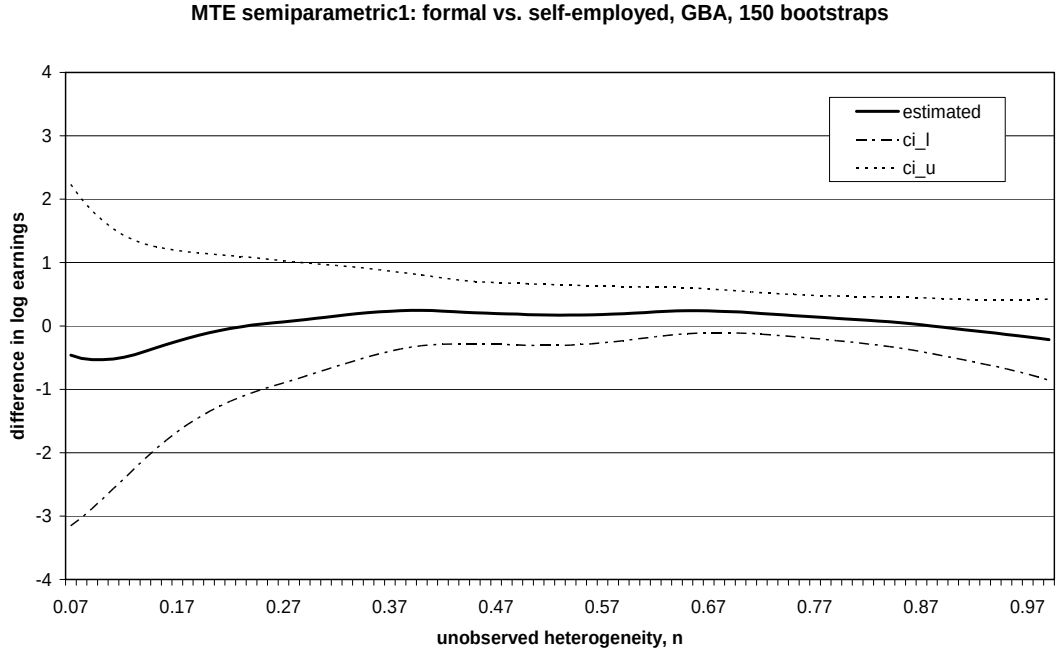


Source: Author's estimations based on EPH-C, INDEC.

Figure 5: Formal salaried workers and Self-employed: MTE - parametric

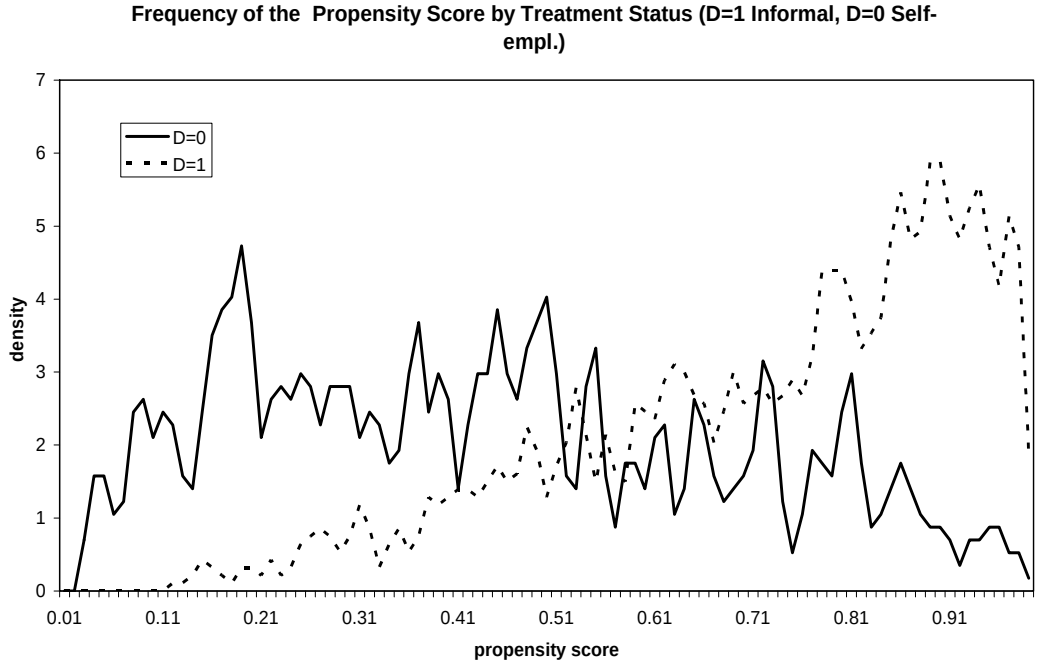
Note: ci_l, ci_u: lower and upper confidence interval. Source: Author's estimations based on EPH-C, INDEC.

Figure 6: Formal salaried workers and Self-employed: MTE – semiparametric1



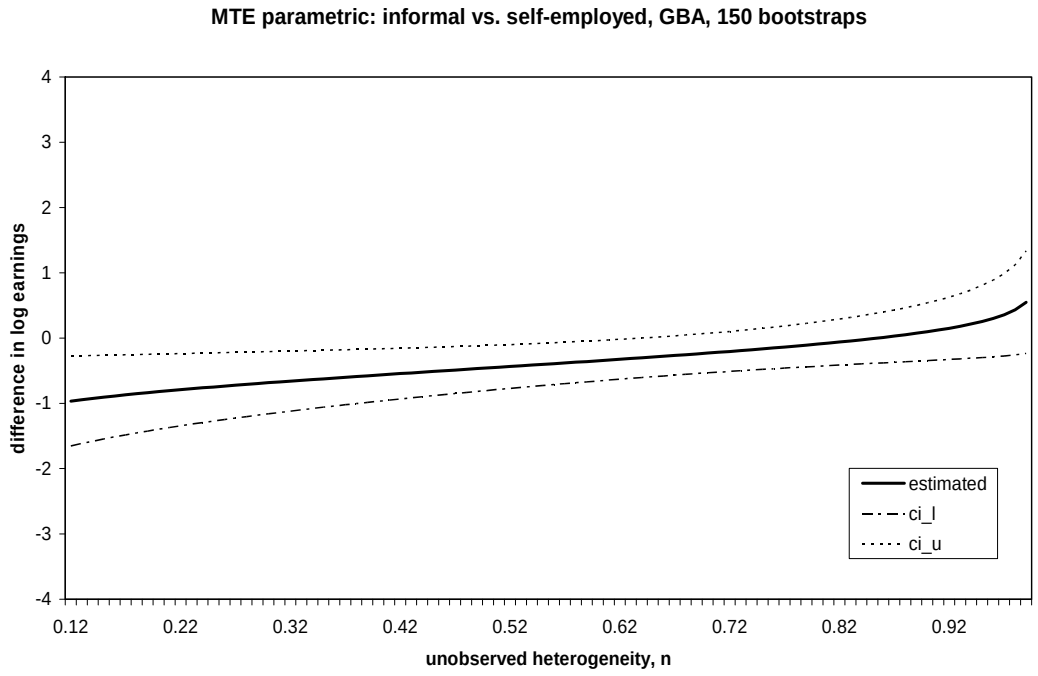
Note: ci_l, ci_u: lower and upper confidence interval. Source: Author's estimations based on EPH-C, INDEC.

Figure 7: Informal salaried workers and Self-employed: Common Support



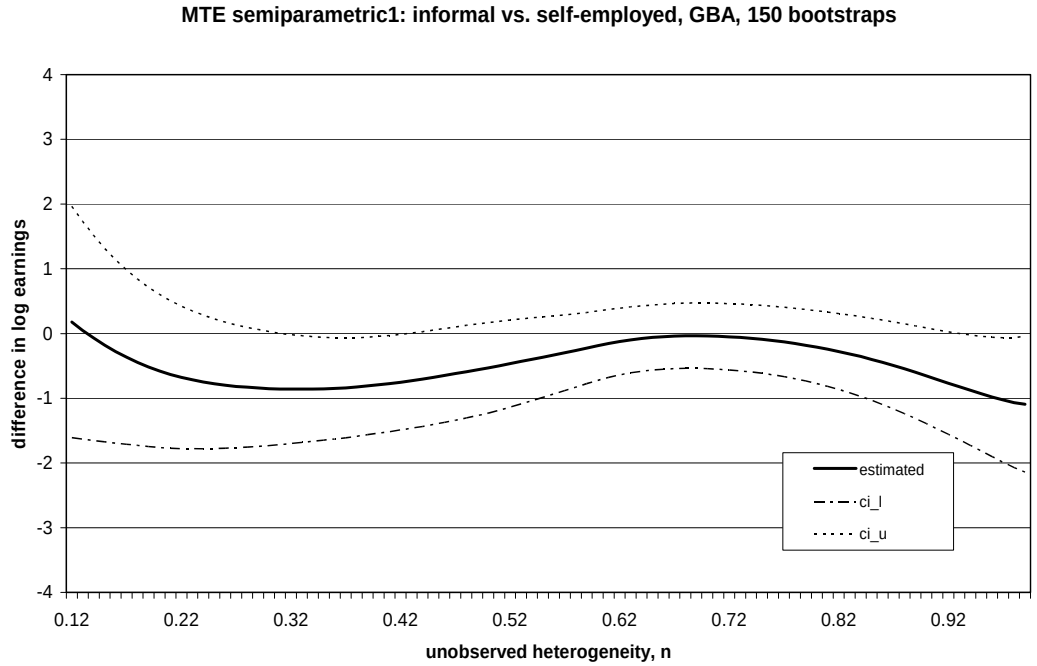
Source: Author's estimations based on EPH-C, INDEC.

Figure 8: Informal salaried workers and Self-employed: MTE – parametric



Note: ci_l, ci_u: lower and upper confidence interval. Source: Author's estimations based on EPH-C, INDEC.

Figure 9: Informal salaried workers and Self-employed: MTE – semiparametric1



Note: ci_l, ci_u: lower and upper confidence interval. Source: Author's estimations based on EPH-C, INDEC.

Table 1: Formal and Informal salaried workers: Choice Model

Choice model - Probit

	Probit	Probit, 150 replics	Marginal effects
secondary education	0.471***	0.471***	0.178***
	[17.76]	[18.68]	[17.76]
tertiary education	0.976***	0.976***	0.342***
	[29.24]	[30.90]	[29.24]
experience	0.028***	0.028***	0.011***
	[9.75]	[8.90]	[9.75]
experience^2	-0.001***	-0.001***	-0.000***
	[10.94]	[10.08]	[10.94]
female	-0.456***	-0.456***	-0.175***
	[12.86]	[13.44]	[12.86]
primary sector	-0.041	-0.041	-0.016
	[0.53]	[0.55]	[0.53]
construction/trade/utility/transport sector	-0.366***	-0.366***	-0.142***
	[11.25]	[11.05]	[11.25]
finance sector	-0.100**	-0.100*	-0.039**
	[2.17]	[1.93]	[2.17]
public and social services sector	-0.134***	-0.134***	-0.052***
	[3.94]	[4.08]	[3.94]
Pampeana	-0.056*	-0.056*	-0.022*
	[1.91]	[1.78]	[1.91]
Cuyo	-0.070*	-0.070*	-0.027*
	[1.87]	[1.77]	[1.87]
NOA	-0.185***	-0.185***	-0.072***
	[4.90]	[4.82]	[4.90]
Patagonia	0.599***	0.599***	0.207***
	[9.58]	[8.81]	[9.58]
NEA	-0.156***	-0.156***	-0.061***
	[3.63]	[3.64]	[3.63]
tenure: less than 1 year	-1.375***	-1.375***	-0.508***
	[49.77]	[48.87]	[49.77]
tenure: 1-5 years	-0.834***	-0.834***	-0.321***
	[33.96]	[33.48]	[33.96]
single	-0.273***	-0.273***	-0.106***
	[6.88]	[7.26]	[6.88]
single_female	0.305***	0.305***	0.113***
	[6.60]	[6.64]	[6.60]
children<=6	-0.014	-0.014	-0.005
	[0.74]	[0.81]	[0.74]
children<=6_female	0.006	0.006	0.002
	[0.23]	[0.24]	[0.23]
hhs.size	-0.034***	-0.034***	-0.013***
	[5.50]	[5.75]	[5.50]
pension_hh	0.279***	0.279***	0.105***
	[7.66]	[7.68]	[7.66]
hhs.head	0.144***	0.144***	0.055***
	[4.96]	[5.32]	[4.96]
pension_head	-0.053	-0.053	-0.021
	[1.16]	[1.20]	[1.16]
single parent	-0.050*	-0.050*	-0.019*
	[1.77]	[1.79]	[1.77]
hhs.human capital	0.016***	0.016***	0.006***
	[5.72]	[5.69]	[5.72]
gdp	-0.007	-0.007	-0.003
	[1.62]	[1.56]	[1.62]
check 2005	0.008**	0.008*	0.003**
	[1.98]	[1.79]	[1.98]
Constant	0.506***	0.506***	
	[6.29]	[6.08]	
Observations	21865	21865	21865
Pseudo R-squared	0.2581	0.2581	0.2581

Absolute value of z statistics in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Author's estimations based on the EPH-C.

Table 2: Formal and Informal salaried workers: Outcome Equation - parametric

Coefficients in the outcome equation - parametric				
		coefficients	stdv.	sig.
	D=1			
$\alpha_1+\varphi$	intercept	1.394	0.040	***
β_{11}	secondary education	0.152	0.018	***
β_{21}	tertiary education	0.482	0.029	***
β_{31}	experience	0.018	0.002	***
β_{41}	experience^2	0.000	0.000	***
β_{51}	female	0.017	0.013	**
β_{61}	Pampeana	-0.072	0.013	***
β_{71}	Cuyo	-0.149	0.015	***
β_{81}	NOA	-0.197	0.015	***
β_{91}	Patagonia	0.194	0.020	***
β_{101}	NEA	-0.264	0.018	***
β_{111}	tenure less than 1 year	0.078	0.035	**
β_{121}	tenure 1-5 years	0.007	0.020	
β_{131}	primary	0.262	0.047	***
β_{141}	construction/trade/utility/transport	-0.015	0.015	
β_{151}	finance	0.020	0.021	
β_{161}	public and social services	0.090	0.014	***
	σ_1	0.353	0.045	***
	D=0			
α_0	intercept	0.117	0.071	*
β_{10}	secondary education	-0.069	0.023	***
β_{20}	tertiary education	0.137	0.046	***
β_{30}	experience	0.008	0.002	***
β_{40}	experience^2	0.000	0.000	*
β_{50}	female	-0.003	0.022	
β_{60}	Pampeana	-0.091	0.022	***
β_{70}	Cuyo	-0.286	0.028	***
β_{80}	NOA	-0.479	0.025	***
β_{90}	Patagonia	-0.129	0.039	***
β_{100}	NEA	-0.484	0.029	***
β_{110}	tenure less than 1 year	0.448	0.051	***
β_{120}	tenure 1-5 years	0.312	0.037	***
β_{130}	primary	-0.004	0.054	
β_{140}	construction/trade/utility/transport	0.096	0.024	***
β_{150}	finance	0.261	0.045	***
β_{160}	public and social services	0.258	0.026	***
	σ_0	0.645	0.055	***

Note: sig.(significance): * significant at 10%; ** significant at 5%; *** significant at 1%

stdv.: standard deviation

Source: Author's estimations based on the EPH-C.

Table 3: Formal and Informal salaried workers: Outcome Equation – semiparametric1

Coefficients in the outcome equation - semiparametric1				
		coefficients	stdv.	sig.
β_{10}	secondary education	-0.131	0.027	***
β_{20}	tertiary education	0.053	0.053	
β_{30}	experience	0.004	0.003	*
β_{40}	experience^2	0.000	0.000	
β_{50}	female	-0.037	0.027	*
β_{60}	Pampeana	0.020	0.031	
β_{70}	Cuyo	-0.170	0.042	***
β_{80}	NOA	-0.407	0.034	***
β_{90}	Patagonia	0.027	0.060	
β_{100}	NEA	-0.380	0.043	***
β_{110}	tenure less than 1 year	0.250	0.063	***
β_{120}	tenure 1-5 years	0.062	0.059	
β_{130}	primary	-0.292	0.084	***
β_{140}	construction/trade/utility/transport	0.114	0.033	***
β_{150}	finance	0.171	0.059	***
β_{160}	public and social services	0.183	0.040	***
difference between betas (treatment betas-non-treatment betas)				
	secondary education	0.213	0.040	***
	tertiary education	0.211	0.058	***
	experience	0.008	0.004	**
	experience^2	0.000	0.000	
	female	0.215	0.036	***
	Pampeana	-0.161	0.044	***
	Cuyo	-0.042	0.056	
	NOA	0.216	0.050	***
	Patagonia	-0.066	0.075	
	NEA	0.068	0.063	
	tenure less than 1 year	0.197	0.086	**
	tenure 1-5 years	0.265	0.070	***
	primary	0.757	0.110	***
	construction/trade/utility/transport	-0.089	0.048	**
	finance	-0.087	0.076	
	public and social services	-0.020	0.057	

Note: sig.(significance): * significant at 10%; ** significant at 5%; *** significant at 1%

stdv.: standard deviation

Source: Author's estimations based on the EPH-C.

Table 4: Formal salaried workers and Self-employed: Choice Model

Choice model - Probit

	Probit	Probit, 150 replics	Marginal effects
secondary education	0.158	0.158*	0.049
	[1.64]	[1.68]	[1.64]
tertiary education	0.068	0.068	0.021
	[0.62]	[0.56]	[0.62]
experience	-0.026***	-0.026***	-0.008***
	[2.64]	[2.78]	[2.64]
experience^2	0.000	0.000	0.000
	[0.05]	[0.05]	[0.05]
tenure: less than 1 year	-0.527***	-0.527***	-0.185***
	[4.94]	[4.54]	[4.94]
tenure: 1-5 years	-0.311***	-0.311***	-0.102***
	[3.85]	[3.80]	[3.85]
female	-0.107	-0.107	-0.034
	[0.90]	[0.92]	[0.90]
single	0.321**	0.321*	0.096**
	[1.98]	[1.90]	[1.98]
single_female	0.146	0.146	0.044
	[0.76]	[0.75]	[0.76]
children<=6	-0.029	-0.029	-0.009
	[0.42]	[0.38]	[0.42]
children<=6_female	0.026	0.026	0.008
	[0.24]	[0.23]	[0.24]
hhs.size	-0.051**	-0.051**	-0.016**
	[2.14]	[2.07]	[2.14]
pension_hh	0.209	0.209	0.064
	[1.29]	[1.24]	[1.29]
hhs.head	0.057	0.057	0.018
	[0.53]	[0.56]	[0.53]
pension_head	-0.021	-0.021	-0.007
	[0.12]	[0.11]	[0.12]
single parent	-0.041	-0.041	-0.013
	[0.35]	[0.35]	[0.35]
hhs.human capital	-0.007	-0.007	-0.002
	[0.70]	[0.70]	[0.70]
taste	1.294***	1.294***	0.411***
	[8.23]	[7.67]	[8.23]
prefer	1.041***	1.041***	0.326***
	[14.70]	[15.73]	[14.70]
Constant	0.587**	0.587**	
	[2.22]	[2.38]	
Observations	1924	1924	1924
Pseudo R-squared	0.2126	0.2126	0.2126

Absolute value of z statistics in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Author's estimations based on the EPH-C.

Table 5: Formal salaried workers and Self-employed: Outcome equation – parametric

Coefficients in the outcome equation - parametric				
		coefficients	stdv.	sig.
	D=1			
$\alpha_1+\varphi$	intercept	1.072	0.065	***
β_{11}	secondary education	0.341	0.041	***
β_{21}	tertiary education	0.907	0.044	***
β_{31}	experience	0.027	0.004	***
β_{41}	experience^2	0.000	0.000	***
β_{51}	tenure less than 1 year	-0.247	0.050	***
β_{61}	tenure 1-5 years	-0.163	0.035	***
β_{71}	female	-0.124	0.027	***
	σ_1	-0.035	0.059	
	D=0			
α_0	intercept	1.260	0.291	***
β_{10}	secondary education	0.239	0.101	***
β_{20}	tertiary education	1.024	0.098	***
β_{30}	experience	0.026	0.010	***
β_{40}	experience^2	0.000	0.000	***
β_{50}	tenure less than 1 year	-0.355	0.118	***
β_{60}	tenure 1-5 years	-0.347	0.102	***
β_{70}	female	-0.360	0.078	***
	σ_0	-0.245	0.128	**

Note: sig.(significance): * significant at 10%; ** significant at 5%; *** significant at 1%

stdv.: standard deviation

Source: Author's estimations based on the EPH-C.

Table 6: Formal salaried workers and Self-employed: Outcome equation – semiparametric1

Coefficients in the outcome equation - semiparametric1				
		coefficients	stdv.	sig.
β_{10}	secondary education	0.297	0.163	**
β_{20}	tertiary education	1.268	0.175	***
β_{30}	experience	0.043	0.014	***
β_{40}	experience^2	-0.001	0.000	***
β_{50}	tenure less than 1 year	-0.244	0.161	*
β_{60}	tenure 1-5 years	-0.387	0.145	***
β_{70}	female	-0.563	0.119	***
difference between betas (treatment betas-non-treatment betas)				
	secondary education	0.006	0.216	
	tertiary education	-0.465	0.230	**
	experience	-0.023	0.018	
	experience^2	0.000	0.000	
	tenure less than 1 year	-0.081	0.219	
	tenure 1-5 years	0.228	0.183	
	female	0.522	0.148	***

Note: sig.(significance): * significant at 10%; ** significant at 5%; *** significant at 1%

stdv.: standard deviation

Source: Author's estimations based on the EPH-C.

Table 7: Informal salaried workers and Self-employed: Choice Model

Choice model - Probit

	Probit	Probit, 150 replics	Marginal effects
secondary education	-0.236**	-0.236**	-0.088**
	[2.46]	[2.42]	[2.46]
tertiary education	-0.713***	-0.713***	-0.274***
	[5.74]	[5.70]	[5.74]
experience	-0.021**	-0.021*	-0.008**
	[2.06]	[1.93]	[2.06]
experience^2	0.000	0.000	0.000
	[0.01]	[0.01]	[0.01]
tenure: less than 1 year	0.733***	0.733***	0.253***
	[7.32]	[6.91]	[7.32]
tenure: 1-5 years	0.404***	0.404***	0.145***
	[4.48]	[4.59]	[4.48]
female	0.327***	0.327***	0.120***
	[2.71]	[2.84]	[2.71]
single	0.409**	0.409**	0.144**
	[2.44]	[2.42]	[2.44]
single_female	-0.3	-0.3	-0.115
	[1.49]	[1.41]	[1.49]
children<=6	0.029	0.029	0.011
	[0.39]	[0.36]	[0.39]
children<=6_female	-0.065	-0.065	-0.024
	[0.59]	[0.60]	[0.59]
hhs.size	-0.033	-0.033	-0.012
	[1.35]	[1.22]	[1.35]
pension_hh	-0.204	-0.204	-0.077
	[1.16]	[1.09]	[1.16]
hhs.head	-0.313***	-0.313***	-0.115***
	[2.84]	[2.75]	[2.84]
pension_head	0.27	0.27	0.096
	[1.31]	[1.26]	[1.31]
single parent	0.036	0.036	0.013
	[0.30]	[0.29]	[0.30]
hhs.human capital	-0.01	-0.01	-0.004
	[0.89]	[0.85]	[0.89]
taste	-0.769***	-0.769***	-0.284***
	[4.33]	[4.59]	[4.33]
prefer	0.615***	0.615***	0.223***
	[7.99]	[8.56]	[7.99]
Constant	0.784***	0.784**	
	[2.85]	[2.49]	
Observations	1505	1505	1505
Pseudo R-squared	0.2324	0.2324	0.2324

Absolute value of z statistics in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Author's estimations based on the EPH-C.

Table 8: Informal salaried workers and Self-employed: Outcome equation - parametric

Coefficients in the outcome equation - parametric				
		coefficients	stdv.	sig.
	D=1			
$\alpha_1+\varphi$	intercept	0.764	0.108	***
β_{11}	secondary education	0.108	0.051	**
β_{21}	tertiary education	0.731	0.078	***
β_{31}	experience	0.027	0.006	***
β_{41}	experience^2	0.000	0.000	***
β_{51}	tenure less than 1 year	-0.213	0.071	***
β_{61}	tenure 1-5 years	-0.015	0.067	
β_{71}	female	-0.014	0.053	
	σ_1	-0.030	0.113	
	D=0			
α_0	intercept	1.413	0.265	***
β_{10}	secondary education	0.128	0.108	
β_{20}	tertiary education	0.777	0.130	***
β_{30}	experience	0.024	0.010	***
β_{40}	experience^2	0.000	0.000	***
β_{50}	tenure less than 1 year	-0.040	0.145	
β_{60}	tenure 1-5 years	-0.185	0.102	**
β_{70}	female	-0.235	0.086	***
	σ_0	-0.463	0.157	***

Note: sig.(significance): * significant at 10%; ** significant at 5%; *** significant at 1%

stdv.: standard deviation

Source: Author's estimations based on the EPH-C.

Table 9: Informal salaried workers and Self-employed: Outcome equation – semiparametric1

Coefficients in the outcome equation - semiparametric1				
		coefficients	stdv.	sig.
β_{10}	secondary education	0.456	0.136	***
β_{20}	tertiary education	1.066	0.158	***
β_{30}	experience	0.023	0.014	**
β_{40}	experience^2	0.000	0.000	*
β_{50}	tenure less than 1 year	-0.139	0.220	
β_{60}	tenure 1-5 years	-0.124	0.149	
β_{70}	female	-0.139	0.157	
difference between betas (treatment betas-non-treatment betas)				
	secondary education	-0.536	0.191	***
	tertiary education	-0.569	0.233	***
	experience	-0.001	0.017	
	experience^2	0.000	0.000	
	tenure less than 1 year	-0.042	0.295	
	tenure 1-5 years	0.066	0.242	
	female	0.105	0.206	

Note: sig.(significance): * significant at 10%; ** significant at 5%; *** significant at 1%

stdv.: standard deviation

Source: Author's estimations based on the EPH-C.

Table 10: Treatment Parameters: Parametric

Treatment Parameters: Parametric

	F vs I	F vs SE	I vs SE
Treatment on the Treated	1.624***	-0.030	-0.581***
	[0.093]	[0.177]	[0.209]
Treatment on the Untreated	1.079***	0.231***	-0.040
	[0.073]	[0.098]	[0.187]
Average Treatment Effect	1.392***	0.049	-0.369**
	[0.066]	[0.135]	[0.160]

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

standard deviations in brackets

F: Formal salaried, I: Informal salaried, SE: self-employed

Source: Author's estimations based on the EPH-C.

Table 11: Treatment Parameters: Semiparametric1

Treatment Parameters: Semiparametric1

	F vs I	F vs SE	I vs SE
Treatment on the Treated	1.724***	0.033	-0.496*
	[0.096]	[0.303]	[0.309]
Treatment on the Untreated	2.122***	0.034	-0.522**
	[0.118]	[0.150]	[0.249]
Average Treatment Effect	1.893***	0.044	-0.486**
	[0.089]	[0.215]	[0.211]

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

standard deviations in brackets

F: Formal salaried, I: Informal salaried, SE: self-employed

Source: Author's estimations based on the EPH-C.

Table 12: Treatment Parameters: Polynomial

Treatment Parameters: Polynomial

	F vs I	F vs SE	I vs SE
Treatment on the Treated	2.088***	0.187	-0.449
	[0.187]	[0.443]	[0.426]
Treatment on the Untreated	1.892***	-0.122	-0.989**
	[0.204]	[0.245]	[0.510]
Average Treatment Effect	2.002***	0.105	-0.600***
	[0.105]	[0.291]	[0.244]

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

standard deviations in brackets

F: Formal salaried, I: Informal salaried, SE: self-employed

Source: Author's estimations based on the EPH-C.

Table 13: Treatment Parameters: Semiparametric2

Treatment Parameters: Semiparametric2

	F vs I	F vs SE	I vs SE
Treatment on the Treated	1.972***	0.069	-0.468*
	[0.161]	[0.354]	[0.354]
Treatment on the Untreated	1.788***	0.014	-0.599**
	[0.168]	[0.170]	[0.319]
Average Treatment Effect	1.892***	0.063	-0.496**
	[0.098]	[0.242]	[0.220]

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

standard deviations in brackets

F: Formal salaried, I: Informal salaried, SE: self-employed

Source: Author's estimations based on the EPH-C.

Table 14: OLS regressions

OLS regressions			
dependent variable: log hourly wage	F vs I	F vs SE	I vs SE
	(1)	(2)	(3)
choice 1/	**0.507*** **	**0.284*** **	**0.010**
	[57.33]	**[8.53]**	**[0.23]**
secondary education	0.184***	0.309***	0.153***
	[18.22]	[7.33]	[3.16]
tertiary education	0.619***	0.942***	0.872***
	[52.42]	[21.06]	[15.15]
experience	0.023***	0.028***	0.031***
	[23.76]	[7.66]	[7.42]
experience^2	-0.000***	-0.000***	-0.000***
	[17.51]	[6.09]	[6.04]
tenure less than 1 year	-0.152***	-0.268***	-0.297***
	[13.55]	[5.78]	[5.53]
tenure 1-5 years	-0.105***	-0.206***	-0.166***
	[10.92]	[5.88]	[3.37]
female	-0.085***	-0.200***	-0.142***
	[10.21]	[6.49]	[3.60]
primary	0.143***	…	…
	[5.02]	…	…
construction/trade/utility/transport	-0.060***	…	…
	[4.78]	…	…
finance	0.085***	…	…
	[4.80]	…	…
public and social services	0.127***	…	…
	[9.83]	…	…
Pampeana	-0.093***	…	…
	[8.85]	…	…
Cuyo	-0.223***	…	…
	[16.18]	…	…
NOA	-0.379***	…	…
	[32.08]	…	…
Patagonia	0.211***	…	…
	[13.97]	…	…
NEA	-0.404***	…	…
	[27.98]	…	…
Constant	0.744***	0.834***	0.821***
	[35.36]	[10.91]	[9.14]
Observations	21865	1924	1505
R-squared	0.468	0.306	0.217

Note: Absolute value of t statistics in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

F: Formal, SE: self-employed, I: informal

1/ Choice dummy: estimates the average treatment effect

column 1: Choice: formal=1, informal=0

column 2: Choice: formal=1, self-employed=0

column 3: Choice: informal=1, self-employed=0

Source: Author's estimations based on the EPH-C.

Table 15: Comparison of treatment parameters

Comparison of Different Parameters
Treatment on the Treated (TT)
Treatment on the Untreated (TUT)
Average Treatment Effect (ATE)
Selection Bias: OLS-TT
Sorting Gain: TT-ATE
Bias: OLS -ATE or Selection Bias + Sorting Gain

Source: Heckman and Li (2003)

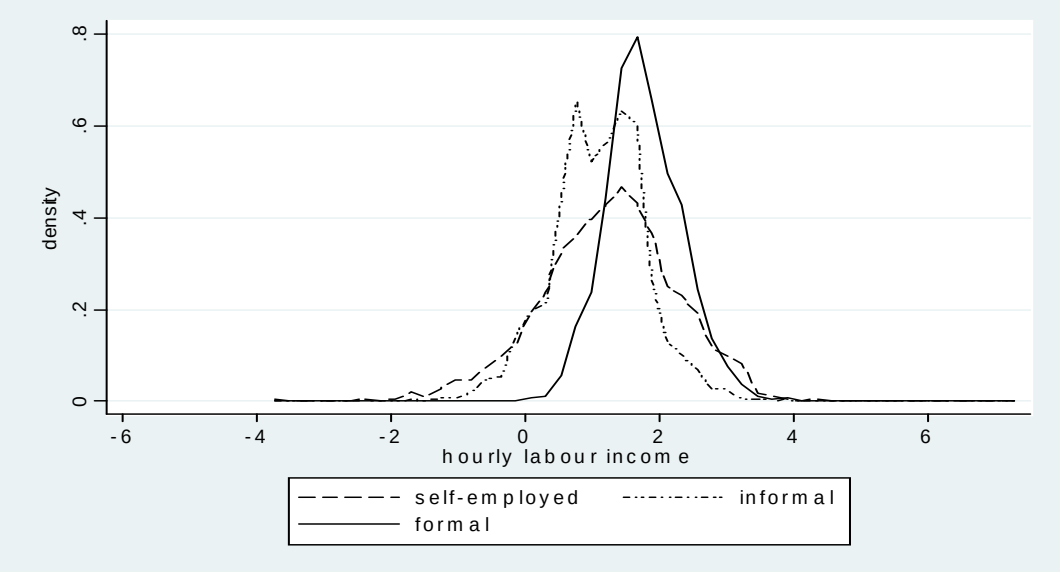
Table 16: Comparison of Bias and Gains

Treatment Parameters: Bias and Gains				
	Parametric	**Semiparametric1**	**Polynomial**	**Semiparametric2**
Formal vs. Informal				
OLS	0.507	0.507	0.507	0.507
Selection Bias	-1.117	-1.217	-1.587	-1.465
Sorting Gain	0.232	-0.169	0.086	0.080
Bias	-0.885	-1.386	-1.495	-1.385
Formal vs. Self-employed				
OLS	0.284	0.284	0.284	0.284
Selection Bias	0.314	0.251	0.097	0.215
Sorting Gain	-0.079	-0.011	0.082	0.006
Bias	0.235	0.240	0.179	0.221
Informal vs. Self-employed				
OLS	0.010	0.010	0.010	0.010
Selection Bias	0.591	0.506	0.459	0.478
Sorting Gain	-0.212	-0.010	0.151	-0.964
Bias	0.379	0.496	0.610	-0.486

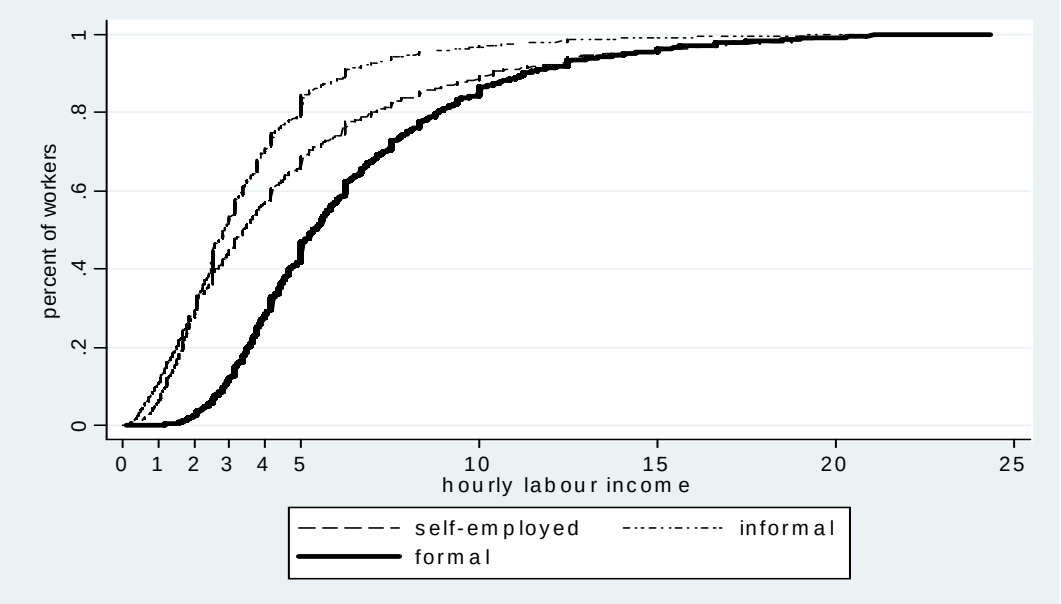
Note: Based on Treatment Parameter tables from Author's estimations of the EPH-C.
OLS compared with treatment parameters from MTE estimations.

Appendix 2: Descriptive statistics and variable description

all urban areas (2nd semester 2005)

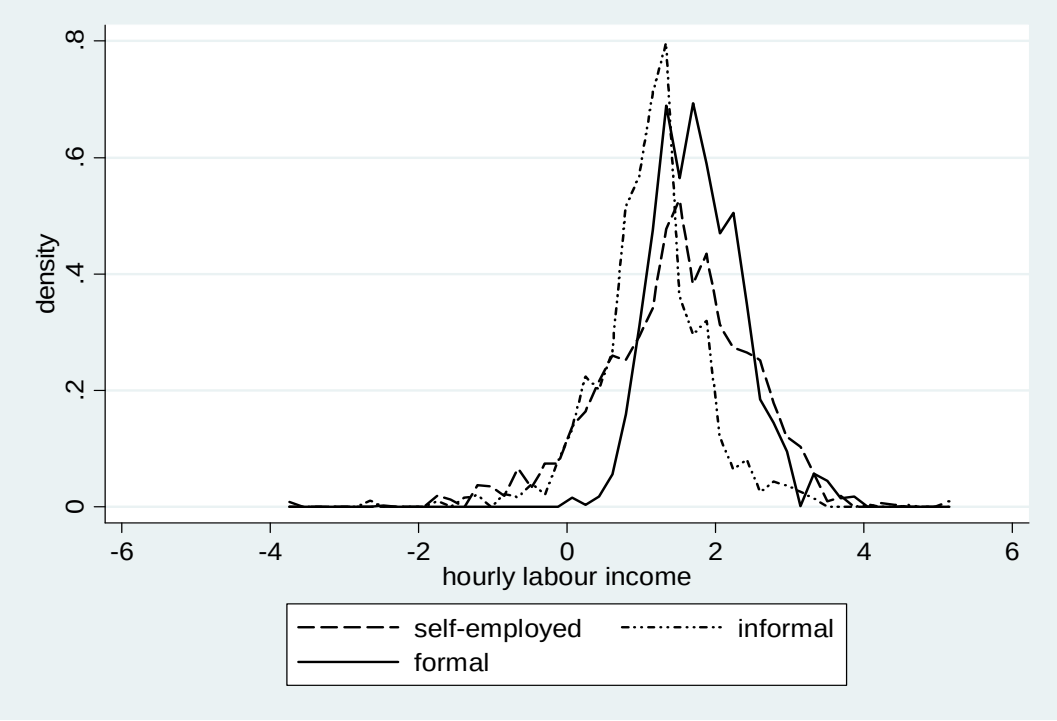


Source: Author's estimations based on EPH-C, INDEC.

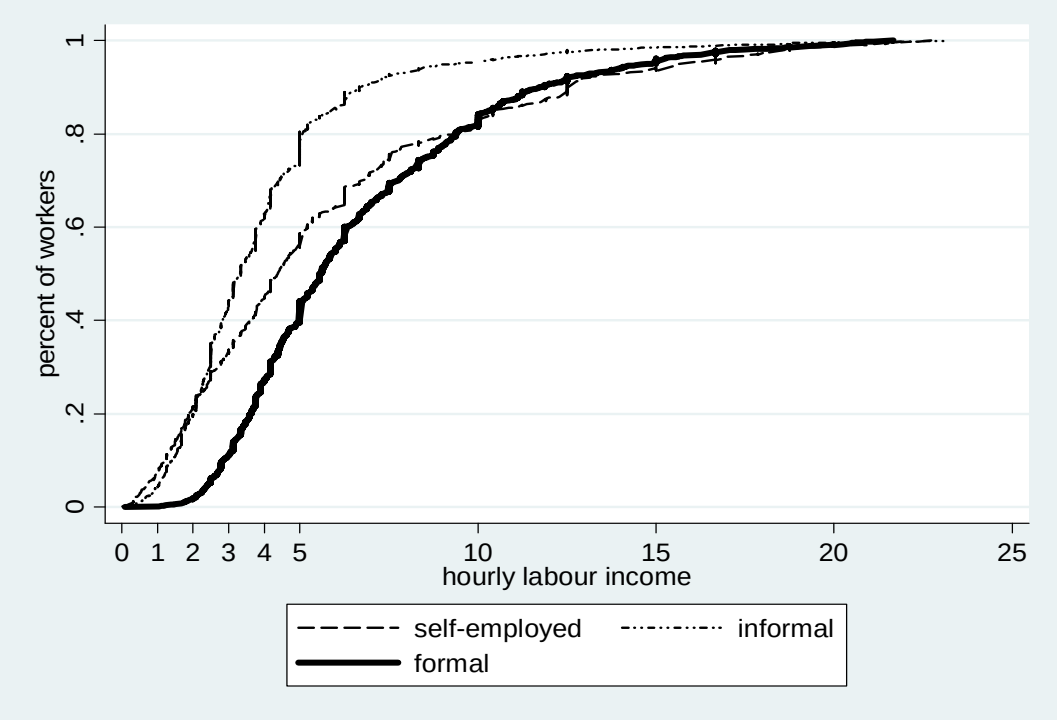


Source: Author's estimations based on EPH-C, INDEC.

GBA (4th trimester 2005)



Source: Author's estimations based on EPH-C, INDEC.



Source: Author's estimations based on EPH-C, INDEC.

Variable description

Variable Description	
Variables	Explanations
indicator	value 1 if not missing data in sample, intercept
choice	choice/participation variable (1= , 0=)
lnwage	log of wage/hourly labour income
primary	primary education (complete/incomplete)
secondary	secondary education (complete/incomplete), base primary
tertiary	tertiary education (complete/incomplete), base primary
exp	experience=age - years of education - six
exp2	experience squared
female	gender variable (1=female, 0=male)
pampa	Pampeana, base GBA
cuyo	Cuyo, base GBA
noa	Noroeste, base GBA
pata	Patagonia, base GBA
nea	Nordeste, base GBA
gba	Gran Buenos Aires
te1	less than 1 year' tenure, base 'more than 5 years' tenure
te2	1 year to 5 years' tenure, base 'more than 5 years' tenure
te3	more than 5 years' tenure
sea1	primary sector, base manufacturing
sea2	manufacturing
sea3	construction/trade/utility/transport, base manufacturing
sea4	finance, base manufacturing
sea5	public and social services, base manufacturing
single	marital status (1=single, 0=married/separated/widow)
single_female	single*female interaction term
children <=6	children under or equal 6 in household
children <=6_female	children under or equal 6 in household*female interaction term
hhs. size	household size
pension_hh	hhs.head/spouse with pension
hhs. head	houshold head (1=if household head, 0=otherwise)
pension_head	hhs.head/spouse with pension* hhs.head/spouse interaction
single parent	lives in household with only household head and no spouse
hhs.human capital	maximum education level in the household
gdp	provincial GDP per capita
check05	number of inspected workers per 1000 people, 2005
taste	preference for occupation (1=choice/opportunity reasons, 0=involuntary/income reasons) 1/
prefer	preference for working dependent (1=prefers dependence, 0=prefers independence)

Note:1/ See main text for explanation.

Descriptive statistics

Summary statistics, weighted averages, urban Argentina, 2nd semester 2005 -Part I of II

Variable	all	formal	informal	self-employed
log of wages	1.386	1.733	1.041	1.205
	[0.801]	[0.585]	[0.708]	[1.005]
primary education	0.305	0.198	0.396	0.383
	[0.460]	[0.398]	[0.489]	[0.486]
secondary education	0.382	0.376	0.411	0.348
	[0.486]	[0.484]	[0.492]	[0.476]
tertiary education	0.314	0.427	0.193	0.268
	[0.464]	[0.495]	[0.395]	[0.443]
experience	22.507	20.923	20.377	29.076
	[14.777]	[13.265]	[15.244]	[15.125]
experience^2	724.924	613.735	647.582	1074.176
	[815.653]	[675.396]	[809.454]	[977.537]
Regions				
Pampeana	0.225	0.224	0.220	0.234
	[0.417]	[0.417]	[0.414]	[0.423]
Cuyo	0.070	0.068	0.071	0.074
	[0.256]	[0.251]	[0.257]	[0.262]
NOA	0.092	0.079	0.101	0.105
	[0.289]	[0.270]	[0.301]	[0.307]
Patagonia	0.026	0.036	0.018	0.015
	[0.158]	[0.186]	[0.134]	[0.123]
NEA	0.043	0.039	0.046	0.047
	[0.203]	[0.194]	[0.209]	[0.211]
GBA	0.544	0.554	0.544	0.525
	[0.498]	[0.497]	[0.498]	[0.499]
Tenure				
less than 1 year	0.250	0.137	0.450	0.171
	[0.433]	[0.343]	[0.498]	[0.377]
1-5 years	0.332	0.317	0.369	0.303
	[0.471]	[0.465]	[0.483]	[0.460]
more than 5 years	0.419	0.546	0.181	0.525
	[0.493]	[0.498]	[0.385]	[0.499]
Economic Sectors				
primary	0.011	0.011	0.013	0.009
	[0.106]	[0.104]	[0.114]	[0.095]
manufacturing	0.145	0.163	0.132	0.126
	[0.352]	[0.370]	[0.339]	[0.332]
construction/trade/utility/transport	0.395	0.275	0.409	0.622
	[0.489]	[0.446]	[0.492]	[0.485]
finance	0.091	0.106	0.063	0.103
	[0.287]	[0.308]	[0.243]	[0.303]
public and social services	0.358	0.445	0.383	0.140
	[0.479]	[0.497]	[0.486]	[0.347]
Sample Size	27947	12616	9249	6082
Population	6947446	3104906	2336598	1505942

Note: Standard deviation in brackets.

Source: Author's estimations based on the EPH-C.

Summary statistics, weighted averages, urban Argentina, 2nd semester 2005 -Part II of II

Variable	all	formal	informal	self-employed
Household and individual characteristics				
female	0.413	0.402	0.478	0.337
	[0.492]	[0.490]	[0.500]	[0.473]
single	0.274	0.264	0.376	0.135
	[0.446]	[0.441]	[0.484]	[0.342]
single_female	0.123	0.127	0.166	0.049
	[0.328]	[0.333]	[0.372]	[0.216]
children <=6	0.469	0.435	0.529	0.446
	[0.771]	[0.722]	[0.830]	[0.768]
children <=6_female	0.161	0.140	0.211	0.129
	[0.486]	[0.443]	[0.560]	[0.438]
hhs. size	4.018	3.791	4.343	3.979
	[2.052]	[1.831]	[2.275]	[2.055]
pension_hh	0.233	0.294	0.194	0.170
	[0.423]	[0.455]	[0.395]	[0.376]
hhs. head	0.534	0.556	0.416	0.669
	[0.499]	[0.497]	[0.493]	[0.470]
pension_head	0.157	0.210	0.100	0.136
	[0.363]	[0.407]	[0.299]	[0.343]
single parent	0.282	0.258	0.341	0.239
	[0.450]	[0.438]	[0.474]	[0.427]
hhs.human capital	8.510	9.367	7.535	8.254
	[5.341]	[5.422]	[4.945]	[5.478]
Provincial characteristics				
gdp	9.115	9.823	8.512	8.592
	[6.167]	[6.693]	[5.600]	[5.708]
check05	9.632	10.271	9.078	9.175
	[6.151]	[6.651]	[5.605]	[5.753]
Sample Size	27947	12616	9249	6082
Population	6947446	3104906	2336598	1505942

Note: Standard deviation in brackets.

Source: Author's estimations based on the EPH-C.

Summary statistics, weighted averages, GBA, 4th trimester 2005

Variable	all	formal	informal	self-employed
lnwage	1.515	1.788	1.175	1.453
	[0.783]	[0.615]	[0.699]	[0.996]
primary education	0.319	0.213	0.417	0.400
	[0.466]	[0.409]	[0.493]	[0.490]
secondary education	0.370	0.374	0.410	0.294
	[0.483]	[0.484]	[0.492]	[0.456]
tertiary education	0.311	0.414	0.173	0.305
	[0.463]	[0.493]	[0.378]	[0.461]
experience	22.921	20.954	21.588	29.541
	[14.961]	[13.894]	[15.008]	[15.373]
experience^2	749.120	631.958	691.042	1108.569
	[837.771]	[720.908]	[803.967]	[1020.164]
Tenure				
less than 1 year	0.250	0.143	0.446	0.171
	[0.433]	[0.350]	[0.497]	[0.377]
1-5 years	0.339	0.330	0.354	0.337
	[0.474]	[0.470]	[0.478]	[0.473]
more than 5 years	0.411	0.527	0.200	0.492
	[0.492]	[0.499]	[0.400]	[0.500]
Household and individual characteristics				
female	0.416	0.395	0.494	0.335
	[0.493]	[0.489]	[0.500]	[0.472]
single	0.263	0.289	0.314	0.121
	[0.441]	[0.454]	[0.464]	[0.326]
single_female	0.120	0.136	0.137	0.055
	[0.325]	[0.343]	[0.344]	[0.227]
children <=6	0.429	0.398	0.499	0.387
	[0.725]	[0.678]	[0.791]	[0.708]
children <=6_female	0.144	0.124	0.199	0.101
	[0.438]	[0.410]	[0.498]	[0.382]
hhs. size	3.877	3.655	4.197	3.858
	[1.896]	[1.679]	[2.103]	[1.928]
pension_hh	0.230	0.278	0.195	0.179
	[0.421]	[0.448]	[0.397]	[0.384]
hhs. head	0.535	0.549	0.425	0.682
	[0.499]	[0.498]	[0.495]	[0.466]
pension_head	0.154	0.186	0.117	0.141
	[0.361]	[0.390]	[0.322]	[0.349]
single parent	0.271	0.267	0.311	0.216
	[0.445]	[0.443]	[0.463]	[0.412]
hhs.human capital	8.823	9.543	7.784	8.898
	[5.445]	[5.558]	[4.908]	[5.753]
Preference for occupation				
taste	0.271	0.347	0.170	0.263
	[0.230]	[0.217]	[0.215]	[0.218]
prefer	0.539	0.619	0.594	0.269
	[0.499]	[0.486]	[0.491]	[0.444]
Sample Size	2858	1353	934	571
Population	3767646	1738791	1258720	770135

Note: Standard deviation in brackets.

Source: Author's estimations based on the EPH-C and EPH-C Informality module.

Formal vs informal salaried workers (all urban)

Descriptive Statistics

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
indicator	21865	5441504	1	0	1	1
choice	21865	5441504	0.570597	0.495002	0	1
lnwage	21865	5441504	1.436128	0.726339	-2.639057	6.620073
secondary	21865	5441504	0.390903	0.487964	0	1
tertiary	21865	5441504	0.326328	0.46888	0	1
exp	21865	5441504	20.68899	14.15074	0	77
exp2	21865	5441504	628.2686	736.1313	0	5929
female	21865	5441504	0.434615	0.495718	0	1
pampa	21865	5441504	0.222147	0.415699	0	1
cuyo	21865	5441504	0.069182	0.25377	0	1
noa	21865	5441504	0.088635	0.284223	0	1
pata	21865	5441504	0.028418	0.166167	0	1
nea	21865	5441504	0.041934	0.200444	0	1
te1	21865	5441504	0.271251	0.444616	0	1
te2	21865	5441504	0.339597	0.473583	0	1
sea1	21865	5441504	0.011922	0.108535	0	1
sea3	21865	5441504	0.332145	0.470993	0	1
sea4	21865	5441504	0.087519	0.2826	0	1
sea5	21865	5441504	0.418431	0.493313	0	1
single	21865	5441504	0.312438	0.463498	0	1
single_female	21865	5441504	0.143394	0.350482	0	1
children<=6	21865	5441504	0.474934	0.771284	0	8
children<=6_female	21865	5441504	0.170221	0.497661	0	6
hhs.size	21865	5441504	4.028145	2.051635	1	21
pension_hh	21865	5441504	0.250692	0.433421	0	1
hhs.head	21865	5441504	0.495971	0.499995	0	1
pension_head	21865	5441504	0.162398	0.368824	0	1
single parent	21865	5441504	0.293943	0.455577	0	1
hhs.human capital	21865	5441504	8.580382	5.300461	0	20
gdp	21865	5441504	9.260121	6.280539	3.077	28.201
check05	21865	5441504	9.758572	6.25128	4.038	24.323

Source: Author's estimations based on the EPH-C.

Formal salaried vs. Self-employed (GBA)

Descriptive Statistics

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
indicator	1924	2508926	1	0	1	1
choice	1924	2508926	0.693042	0.461352	0	1
lnwage	1924	2508926	1.68548	0.767982	-3.713572	5.115996
secondary	1924	2508926	0.349286	0.476868	0	1
tertiary	1924	2508926	0.380286	0.485583	0	1
exp	1924	2508926	23.58982	14.89607	0	75
exp2	1924	2508926	778.2573	852.9468	0	5625
te1	1924	2508926	0.151362	0.358495	0	1
te2	1924	2508926	0.332155	0.471109	0	1
female	1924	2508926	0.376618	0.484664	0	1
single	1924	2508926	0.23772	0.425797	0	1
single_female	1924	2508926	0.110959	0.314164	0	1
children<=6	1924	2508926	0.394805	0.687252	0	4
children<=6_female	1924	2508926	0.116797	0.401884	0	3
hhs.size	1924	2508926	3.716904	1.761451	1	15
pension_hh	1924	2508926	0.247663	0.431767	0	1
hhs.head	1924	2508926	0.590181	0.491928	0	1
pension_head	1924	2508926	0.172547	0.377953	0	1
single parent	1924	2508926	0.251495	0.433985	0	1
hhs.human capital	1924	2508926	9.344717	5.624728	0	17
taste	1924	2508926	0.320939	0.220641	0	1
prefer	1924	2508926	0.511676	0.499994	0	1

Source: Author's estimations based on the EPH-C.

Informal vs. Self-employed (GBA)

Descriptive Statistics

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
indicator	1505	2028855	1	0	1	1
choice	1505	2028855	0.620409	0.485446	0	1
lnwage	1505	2028855	1.280614	0.834954	-3.713572	5.115996
secondary	1505	2028855	0.366352	0.481967	0	1
tertiary	1505	2028855	0.223016	0.416407	0	1
exp	1505	2028855	24.60685	15.62678	0	75
exp2	1505	2028855	849.531	914.6319	0	5625
te1	1505	2028855	0.341741	0.474451	0	1
te2	1505	2028855	0.347437	0.476314	0	1
female	1505	2028855	0.433723	0.495753	0	1
single	1505	2028855	0.240934	0.427792	0	1
single_female	1505	2028855	0.105796	0.307679	0	1
children<=6	1505	2028855	0.456234	0.762161	0	4
children<=6_female	1505	2028855	0.161848	0.460041	0	4
hhs.size	1505	2028855	4.068319	2.04445	1	15
pension_hh	1505	2028855	0.189193	0.391792	0	1
hhs.head	1505	2028855	0.52243	0.499663	0	1
pension_head	1505	2028855	0.126249	0.33224	0	1
single parent	1505	2028855	0.274908	0.446616	0	1
hhs.human capital	1505	2028855	8.206481	5.270804	0	17
taste	1505	2028855	0.205383	0.220982	0	1
prefer	1505	2028855	0.470936	0.499321	0	1

Source: Author's estimations based on the EPH-C.