

IZA DP No. 3916

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

Omar Arias
Melanie Khamis

December 2008

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

Omar Arias

World Bank

Melanie Khamis

IZA and London School of Economics

Discussion Paper No. 3916
December 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Comparative Advantage, Segmentation and Informal Earnings: A Marginal Treatment Effects Approach^{*}

This paper employs recently developed econometric models of marginal treatment effects 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 Argentina. A novel household data set on informality and self-employment and information on labor inspections targeting informal work was collected for this purpose. We account for endogeneity and selectivity issues in our estimations. Our results offer evidence for both comparative advantage and segmentation. No significant differences between the earnings of formal salaried workers and the self-employed are found, once accounted for positive selection bias into formal work. This is consistent with labor market comparative advantage considerations. On the contrary, informal salaried employment carries significant earnings penalties, alongside negative selection bias and modest positive sorting. These results are more consistent with segmentation.

JEL Classification: C31, J24, J31, J42, O17

Keywords: marginal treatment effects, occupational choice, segmentation, earnings, comparative advantage, informality, labor markets

Corresponding author:

Melanie Khamis
IZA
P.O. Box 7240
53072 Bonn
Germany
E-mail: khamis@iza.org

^{*} 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 Randy Akee, Oriana Bandiera, Armando Barrientos, Deborah Cobb-Clark, Maitreesh Ghatak, Dan Hamermesh, Doug Krupka, William Maloney, Arthur van Soest, Diana Weinhold, Zhong Zhao and seminar participants at the World Bank, the LSE, IZA, the IZA/World Bank Employment and Development Conference 2007, LAMES 2007, RES 2008, SOLE 2008, the Nordic Conference in Development Economics 2008 and IPEA for helpful comments. Melanie Khamis would like to thank the LSE for financial support and IZA for its 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.

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 and comparative advantage considerations. Following on the Harris and Todaro (1970) tradition, the conventional ‘exclusion’ view sees informal workers, either self-employed or informal employees who lack 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 possibly due to labor market rigidities such as unions, minimum and efficiency wages, generous labor benefits and unequal market power arising from low 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 the young, married women and the unskilled, may voluntarily choose these occupations for various reasons: as a labor market entry point, to enjoy non-pecuniary benefits such as more flexible hours, to exploit entrepreneurial abilities, to improve mobility, to obtain training opportunities and to escape taxing regulations and 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. 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 (Carneiro, Heckman and Vytlačil, 2005; Heckman and Li, 2003; Carneiro and Heckman, 2002). Jobs that are more desirable in terms of their amenities such as fringe benefits, stability and flexibility or that require relatively abundant skills should have lower-than-average wages. Jobs that are less desirable or demand scarce skills should pay higher wages. A competitive labor market determines an implicit wage for each type of labor. In equilibrium labor mobility leads to a set of relative wages that makes workers indifferent between the various types of jobs. Given the heterogeneity in workers' preferences and skills, both supply and demand for particular jobs determine the size of the compensating differential, the difference in implicit wages, between jobs with different working conditions.

The presence of segmentation or comparative advantage considerations in the labor market also have different policy implications:

- In a segmented labor market entry barriers and labor market rigidities prevent informal worker to participate in the formal sector. Policies to tackle entry barriers and rigidities could encompass labor market policies and also policies addressing tax evasion and the design of the social security system that prevent the shift from informality into formal work.
- In a labor market, where labor market comparative advantage consideration prevail,

informal salaried workers and the self-employed chose and remain voluntary in certain jobs and thereby maximize their utility. Policies for comparative advantage reasons would support informal sector participants' decision to work in the informal sector instead of shifting them to formality or penalizing them for not being in the formal labor market. Support for alternative measures of social protection and social security for informal workers and the self-employed could be a possible policy option.

This paper uses recently developed econometric methodologies by Heckman and Vytlačil (2001, 2005, 2007) and Heckman Urzua and Vytlačil (2006) to analyze the relevance of labor market comparative advantage and segmentation in the participation and earnings performance of workers in the self-employed and formal and informal salaried sectors in Argentina.¹ These methods allow the investigation of the links between heterogeneous ability, earnings, and occupational choice. We link the treatment effects literature with conventional Mincer earnings analysis and the generalized Roy model (1951) of selectivity in occupational choice.

We estimate parametric and semiparametric regressions with local instrumental variables to obtain a distinct set of treatment parameters, derived from the marginal treatment effect (Heckman and Vytlačil, 2001, 2005; Heckman, Urzua and Vytlačil, 2006).² 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

¹Informal salaried workers in our paper are workers without social security affiliation and no pensions. The self-employed do not participate in the main social security system. In the empirical estimation section of this paper the three groups in the data are described in detail.

²We also estimated polynomial regressions and a different set of semiparametric regressions, based on the polynomial. In this paper we only report the treatment parameters and the comparison of bias and gains and not other results of these regressions. The remaining results are available upon request from the authors.

worker, informal salaried workers and the self-employed and to understand the main labor market features present.

This approach addresses key implications of the theory in the estimation of informal-formal earnings gaps and solves the empirical problems relating to selection.

First, the ‘treatment’ impact, becoming formal, might be heterogeneous across individuals. Workers could benefit differently depending on both their observed and unobserved characteristics. There is no single representative impact of formality on wages. 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.

Second, estimation addresses two types of selectivity bias, selection bias and sorting on the gain. The standard selection bias arises when only part of the outcomes can be explained by observables. The unobservables give rise to the problem of selection bias. Outcomes of occupational choice are heterogeneous and individuals chose jobs with partial knowledge of their idiosyncratic gain. This second bias, sorting on the gain is generated by the correlation between unobservables 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, result in biased and non-consistent estimates of the earnings premium of formality for a randomly selected worker. The semiparametric marginal treatment effect estimation with local instrumental variables can estimate treatment parameters when sorting on the gain is present (Heckman, Urzua and Vytlačil, 2006).

Argentina presents a very suitable context to study these questions. 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. Initial tests with unconditional earnings gaps, which do not account for selection on observables and unobservables, reveal that the self-employed and informal salaried workers seem to face an earnings disadvantage with respect to formal salaried workers suggesting the existence of segmentation.³ 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). In this context a novel data collection on motivations and constraints for formality, informality and self-employment was attached to the national household survey, which we are able to use in our estimations. In addition to this, we collect data on the number of labor inspectors, who are employed under a new labor regulation plan to tackle informality in the labor market.

Our results offer evidence for both labour market comparative advantage and segmentation. We find no significant differences between the earnings of formal waged workers and the self-employed, once accounted for selection. This is consistent with comparative advantage considerations. On the contrary, informal waged work carries significant earnings penalties. There is a considerable negative selection bias into formal relative to informal waged work and only modest positive sorting based on expected earnings gains. These results are more consistent with labour market segmentation.

The plan of the paper is as follows. First, we discuss some of the most 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 Vytlačil (2006). Next, the data and estimation specifications are

³At the 20th quantile of the earnings distribution the informal-formal salaried worker earnings gap is -53.33 percent, at the 50th -47.77 percent and at the 80th -45.12 percent. For the self-employed-formal salaried worker earnings gap the numbers are -58.00 percent at the 20th, -37.32 percent at the 50th and -19.20 percent at the 80th quantile. The numbers are based on the authors' estimation of the EPH-C.

discussed. This is followed by the empirical results and their implications. The paper concludes with a summary of the findings and related policy implications.

2 Empirical tests of the ‘exclusion’ and ‘competitive’ labor market views⁴

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, in particular from the developing country context. However, an early study in this area was conducted in the U.S. by Dickens and Lang (1985) who used a switching earnings model with unknown regimes to test empirically the presence of dual labor markets. 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.

In contrast 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

⁴In this paper we use the terms of exclusion and segmentation interchangeably for each other.

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 as an alternative model/hypothesis of compensating differentials is necessary to do so. 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. They find 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 find 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 Vytlačil (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 allows to extrapolate the impact of changes in an individual employment status on his earnings. Not only observable characteristics, but unobservable heterogeneity determine the returns. 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 as one needs a model that allows for compensating differentials/comparative advantage considerations as an alternative hypothesis. 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. 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 sector. 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.

Contrary to the previous literature we apply the recently developed marginal treatment methods for models of essential heterogeneity developed by Heckman, Urzua and Vytlačil (2006) to examine the links between earnings performance and the choice of formal and informal salaried work and self-employment.⁵To our knowledge, this is the first application of the marginal treatment effects approach to the analysis of occupational choice and earnings performance, in particular to characterize informal and formal labor

⁵Sorting on the gain is the feature of models of essential heterogeneity. Outcomes of 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 Vytlačil 2006).

markets. This method allows to account for observable and unobservable characteristics of the individuals that affect their decision to participate in different occupations while employing less restrictive econometric methods than, for instance, Magnac (1991). 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 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 comparative advantage considerations, 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.⁶ Workers choose their occupation by comparing the utility W_s they derive from each occupation, which is

⁶Other margins of choice such as formal salaried versus informal salaried worker can also be represented by this model.

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:

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

where W_{si}^* , an individual's utility of a particular occupational choice, 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:

$$W_{1i} \geq W_{2i} \iff (Y_{1i} - Y_{2i}) + (\varepsilon_{1i} - c_{1i}) - (\varepsilon_{2i} - c_{2i}) \geq 0 \quad \text{iff} \quad Z_i'(\gamma_1 - \gamma_2) \geq \eta_{1i} - \eta_{2i} \quad (2)$$

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 $P(z)$:

$$P(s = 1|Z = z) = P(\Delta W \geq 0) \iff P(Z'(\gamma_1 - \gamma_2) \geq \eta_1 - \eta_2) \quad (3)$$

where the η_s have a common distribution F_{η_s} .

We only observe earnings y 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:

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

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:

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

This yields an earnings model with self-selection. There are two key implications of the theory for the estimation of earnings gaps in this model (Heckman and Vytlačil, 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 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 reflects 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 Vytlačil, 2006; Heckman and Vytlačil, 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 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 derive measures of two types of biases: selection bias and sorting on the gain, which is the bias that arises from the sorting of workers based on expected gains (Heckman and Li, 2003).⁷ The selection bias is determined by the difference of the OLS estimate and TT. Meanwhile the difference of TT-ATE and TUT-ATE yield the sorting on the gain or loss respectively. For instance we can determine the gain of salaried and self-employed-like workers from participating in the salaried and self-employed sectors compared to a randomly sampled worker.

According to Heckman and Li (2003) the presence of large, positive sorting gains indicate that comparative advantage considerations of workers are a feature of the labor

⁷The total bias is defined as OLS-ATE or selection bias + sorting gain.

market. However, 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 different labor market sectors and people self-select into different occupations or job types based on these returns or their tastes.

To understand the prevalence of labor market comparative advantage considerations and segmentation, we estimate the marginal treatment effect and account for the selection bias and sorting on the gain in the estimation. At the margin an individual is indifferent between participation in the two different labor market sectors, conditional on observables and unobservables. A considerable wage gap at the margin and significant treatment parameters would imply that earnings penalty or gain exists between the two sectors, indicating the presence of segmentation if coupled with modest sorting on the gain and selection bias. While no significant earnings gap would give an indication of labor market comparative advantage considerations and a competitive labor market, in particular if sorting on the gain and selection bias are accounted for.

4 Data and Empirical specification

This section outlines the specific data collected for this study and the estimation specifications, in particular the instruments, in the application of the empirical method of marginal treatment effect following Heckman, Urzua and Vytlačil (2006).

4.1 Data⁸

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),

⁸Descriptive summary statistics and variable descriptions can be found in the appendix 3.

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, constraints and preferences, 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 previous levels of informality measured. In order to account for these factors, the analysis also controls for population size and GDP per capita per province from the 2001 national census and the Province of Buenos Aires Ministry of Economy.

4.2 Empirical specification⁹

Three different groups of labor market participants are employed in the participation and wage estimations, which follow a Heckman selection correction setup. These three groups provide the basis for the different occupational choice margins of the marginal treatment effects and other treatment parameters.

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: (1) dependent salaried work (formal) versus self-employment, (2) dependent salaried work (informal) versus self-employment and (3) formal versus informal salaried work.

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 the dependent worker status (for the formal workers' margin and for the informal

⁹Details of the empirical MTE method by Heckman, Urzua and Vytlačil (2006) can be found in the appendix 1.

workers' margin) is the treatment group and the self-employed are the non-treated. The decision of work in self-employment, which are not covered by social security arrangements, or a dependent work relationship, either in the formal or informal relationship, was the motivation for the estimation of these two margins. We also investigate the participation choice of a job in formal and informal dependent salaried work, so within the dependent employee relationship. For this margin, 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).¹⁰ The margins of dependent salaried work (formal and informal) versus self-employment are estimated only for Greater Buenos Aires given the availability of variables that could serve as instruments, which we compiled from our novel data set on informality and self-employment (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 themselves 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

¹⁰For the variable descriptions, including the base category for the dummy variables, see appendix 3.

in the differential magnitudes and significance of the selection-correction terms.¹¹

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 marginal treatment effect and related parameters, we need correct specification of the instruments in the propensity scores and outcome equations (Heckman, Urzua and Vytlačil 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 working in a dependent relationship*, 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 for 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 the variable constructed to capture workers' reported *preference for the occupation*. This dummy was coded 1 if the reasons for the current occupation are

¹¹The results are available upon requests from the authors.

relating to choice and opportunity reasons and 0 if the reasons for the current occupation are involuntary and income reasons. Other individual-level instruments also included having the spouse or 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*, which was highlighted by De Soto (1989). Workers living in provinces 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 (Gasparini and Weinschelbaum, 2006). 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 implications¹²

The results are presented in Figures 1 to 9 and Tables 1 to 15. In this section we discuss the results of the participation and wage models for the three different margins and also the marginal treatment effect (MTE) estimations, which are performed over the common support. Thereafter, a discussion of the treatment parameters for the different margins follows.

¹²In this paper the results for the parametric and semiparametric LIV estimation, which are the most and the least restrictive in terms of assumptions, are illustrated in detail. The results for the polynomial and an alternative semiparametric estimator are discussed, but the remaining results are available upon request from the authors.

For the margin of formal and informal salaried workers, we find that variables for the probit model for the probability of participation in the formal sector are mostly significant and exhibit the a priori expected signs (Table 1). For example, females are less likely to participate in the formal sector. Contrary to this, having higher education levels makes participation more likely. The individual level instruments, such as whether the household head has a pension plan or the household human capital, are mostly significant and have the expected signs. The cost of informality, the number of inspected workers per province, is positive and significant. A higher number of inspections in a province increases the probability of participation in the formal sector. In Table 2 and 3 the wage outcome equations are reported for the parametric and semiparametric version with local instrumental variables. For both formal and informal salaried workers most of the coefficients are significant even at the 1 percent level. The selection correction term σ for the parametric estimation (Table 2) is significant for both formal and informal wage regressions. Overall, the magnitude and signs of the coefficients seem to make sense. For instance, returns to education increased with higher levels of education completed (Table 2). This is consistent with the semiparametric outcome for the wage regression (Table 3). The difference between treatment (formal) and non-treatment (informal) betas is significant. For the returns to education the result remains consistent with the parametric version. Most coefficients of the non-treated, the informal salaried workers, are also significant at the 1 percent level.

In order to estimate the MTE for this margin, the common support condition has to be full-filled. The MTE estimation is only performed over the area of common support as to make the two groups as comparable as possible in terms of their observable characteristics. The density of the propensity score z , the probability of selection into formality is displayed in Figure 1. These results for formal and informal salaried workers

demonstrate a large common support. It is also possible to see that people who participate in the formal sector tend to have a higher propensity to participate in the formal sector.

This explains the higher density of the propensity score for the formal workers at the right-hand side of the graph, where the higher propensity scores are. Contrary to this, the density of the propensity score for the informal sector is higher at lower propensities. Informal workers, who are presently in the informal sector, have a lower propensity to be formal. Still, this cannot be generalized for all informal workers. In the medium-range the concentration of the propensity score is less pronounced for either sample, which implies that according to observed characteristics roughly similar amounts of individuals have the propensity to be in either sector.

Restricting the estimation over the area of common support, in Figure 2 and Figure 3 the marginal treatment effect over the range of unobserved heterogeneity is plotted. The unobservables η are plotted in the reverse to the earlier propensity score z , the observables. Low values of η , the left-hand side of the MTE graph, imply low unobservable heterogeneity and high possibility of participation in the formal sector. High values of η , the right-hand side, imply a high measure of unobservable heterogeneity and low possibility of participation in the formal sector.

A positive MTE signals a positive return in terms of earnings to formality for a person who is indifferent, at the margin, between formality and informality, given observables and unobservables. This indicates that a wage gap exists between formal and informal salaried work. If the graphed MTE line is not flat or completely parallel to the horizontal axis across the entire range of unobserved heterogeneity η it indicates that persons, accounting for their observable characteristics, have different returns to participation across different values of unobservables.

The Figure 2 shows that the MTE from the parametric estimation does not vary very much across unobservables. It is a near flat line, which we would expect from the parametric case. A person at the margin between formality and informality were to gain almost twice as much in terms of wage if they were to become a formal instead of informal waged worker.

In Figure 3 the parametric assumptions are relaxed and the semiparametric estimator with local instrumental variable for the MTE is implemented. Different results to Figure 2 are apparent: The MTE is almost a straight line until higher values of η exhibit an increasing MTE. An increase in the MTE at higher values at the measure of the unobservables implies that a person that is more likely to be informal has a higher marginal return to formality than a person that is more likely to be formal. This could mean that a person at the margin between formality and informality were to gain even more than twice as much in terms of wage if they were to become formal (even higher increase than suggested by the parametric case). In either case though, informal salaried workers across different levels of unobservables would gain in terms of earnings were they to become formal.

For the participation margin for the formal salaried worker and the self-employed the same graphs and tables as previously are displayed and discussed. Contrary to the earlier estimation here the estimation was performed only for Greater Buenos Aires due the informality survey at our disposal. The selection problem relating to the participation in self-employment and formal salaried workers is different on several dimensions and requires other instruments to be included than the formal and informal waged work margin.

In Table 4 the probit model for the propensity of participation as a formal salaried worker displays several significant coefficients at the 1, 5 and 10 percent level. Most

notable the instruments for the preference for working in a dependent relationship and the preference for the occupation are positive and significant at the 1 percent level. For instance, higher preference for dependence increases the probability to be a formal employee instead of self-employed.

In Table 5 and Table 6 the outcome equation results exhibit mostly significant coefficient. Especially, in Table 5 for the formal workers and the self-employed all coefficients are significant, except the selection correction term for the formal workers. All of them display the expected signs and coefficient sizes. For instance, higher levels of education increase wages. More experience and higher tenure in the profession, here the base category is tenure being higher than 5 years, result in higher returns. Being female decreases the wage compared to being male for both formal workers and the self-employed.

The semiparametric results in Table 6 are significant for all betas for the non-treated, the self-employed. The results for the difference in the beta coefficients between the formal workers and the self-employed show that only tertiary education and being female are significant. The self-employed have a higher wage pay-off with higher education. Self-employed women, however, earn significantly less. This could be an indication for the different self-employed segments: for example, the high-end male entrepreneur on the one end and the female self-employed who produces and sells small products for subsistence on the other end.

As previously for the formal-informal salaried worker margin, enough common support of the propensity scores needs to be present to estimate the marginal treatment effect. In Figure 4 nearly the whole range of the propensity score is covered by both subsamples with the exception at the very lower end of the propensity score.

For formal salaried workers the propensity score has a higher density at high values as they have a high probability of selection into formality according to their observable

characteristics (as was seen in Figure 1 as well). For the self-employed though a broad range of the propensity is spread out quite evenly across all ranges of the propensity. Observed characteristics of the self-employed are quite heterogeneous.

The MTE estimation over the range of common support is displayed in Figure 5 and Figure 6. In Figure 5 the parametric estimation of the MTE indicates that across all unobservables the margin of return to formality is close to zero. The self-employed were not to gain, or if at all very little, in terms of earnings if they were to be dependent workers with social security contributions.

Relaxing the parametric assumptions that yield a flat MTE, one can observe that at the lower values and higher values of the unobservables the results differ to the parametric case (Figure 6): The self-employed would loose out in terms of wages if they were to become dependent formal workers. At the lower values of unobservables, the ones with ‘low-taste’ for self-employment or high propensity to be formal dependent workers, have a negative return and at higher values, the ones with ‘high-taste’ for self-employment, would also experience a negative return if they were to switch to the formal worker labor market. The middle-range values of unobservables show a small positive return.

Given the previous results from the estimation of the margin of formal salaried workers and the self-employed, which gave an idea about the behavior of the self-employed, we turn to the informal salaried workers and the self-employed margin to complete the picture on the labor market. In the choice model of participation in the informal sector the education variables, tenure, experience, marital status, household head status and gender are significant and display expected signs (Table 7). Higher education, more experience and being the head of a household decrease the likelihood of participation in the informal sector (or increase the likelihood to be self-employed). Women, individuals with less than 5 years tenure in the occupation, and singles are more likely to work as a

dependent worker in the informal sector.

The instruments constructed from the informality survey are both significant at the 1 percent level and have the expected signs. Preference for dependence increases the chance for participation in the dependent informal sector. Preference for the occupation in terms of opportunities and choices decreases the propensity to participate in the informal dependent sector (Table 7).

The wage regressions in Table 8 have mostly significant coefficients, with the selection correction term being significant for the self-employed. The signs and magnitudes are as expected. Returns to tertiary education are higher than secondary education. More experience and tenure increases the wage pay-off. Being female decreases the return to both informal work and self-employment.

The semiparametric outcome equation results are not as strong in terms significance. Education and experience are significant. For the difference between the beta coefficients for both groups secondary and tertiary education are significant and indicate that the self-employed have a higher return to education (Table 9).

Testing for the common support, the frequency of propensity scores are graphed for both groups (Figure 7). The propensity scores for the informal dependent workers, the treatment group, have a high density at higher propensity values, which indicate more likelihood to participate in the informal sector. The density for the self-employed is more evenly spread across the whole range, but at the higher values it indicates fewer self-employed. Estimating the parametric MTE across, the subsample of the common support a small negative MTE equal across all unobservables is found (Figure 8). With the semiparametric MTE the results change slightly (Figure 9): the MTE is still negative for all individual across the different values of unobservables, but with some variation across the unobservables.

The summary tables 10 to 14 present a distinct set of summary parameters to understand the different labor market features and the accompanying policy questions posed in this paper: (i) the average treatment effect (ATE); (ii) the treatment on the treated (TT) and (iii) the treatment on the untreated (TUT). As shown by Heckman and Vytlačil (2001, 2005) these parameters can be derived from an estimate of the MTE. The tables show the estimates obtained with parametric, semi-parametric (local instrumental variables and another semi-parametric method based on the polynomial) and polynomial estimators (see Heckman, Urzua and Vytlačil, 2006).¹³ 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 discussed earlier 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 accounts for the sorting of workers based on preferences and the returns 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 observed and unobserved characteristics 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

¹³The results shown here are robust to different empirical specifications and alternative IVs.

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 carries 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 observed and unobserved characteristics 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 14 presents 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 bias in conventional OLS (Table 15) 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.

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 average treatment effect, the treatment on the treated, the treatment on the untreated and two biases, the selection

bias and sorting on the gain from participation and wage outcomes. For the estimations and our instruments in the participation regression we employ a novel data set on informality and self-employment and the national household survey. We also collect data on labor inspections, which were specifically put in place to tackle the lack of social security registration and thereby informality in the labor market.

The results support the importance of both comparative advantage and segmentation in Argentina’s informal-formal employment composition. On the one hand, there are no significant differences between the earnings of formal salaried workers and the self-employed regardless of the propensity to select into each sector, but there is positive selection bias into formal salaried work. 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. 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 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.

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-employed.

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 might be the result 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 taxes and labor market frictions. This may be addressed with tighter enforcement of reformed labor and tax laws and improved collective bargaining.

The results suggest that independent workers reveal no willingness to pay for the social protection benefits, for instance social security or health coverage, that formal wage earners enjoy. This highlights the issue of 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 heterogeneous willingness to pay or accept lower take-home earnings in exchange for such benefits depending on their preferences, the cost and quality of the real and perceived services provided by the public and private sectors and the characteristics of alternative sources of services and benefits not related to the

labor contract. Analyses like those provided in this paper for other developing country contexts inform whether labor market segmentation or comparative advantage considerations are features of the labor market and the choice of policy to address informality in the labor market.

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 Vytlačil. 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: Empirical methodology

The MTE outlined in equation (8) from Section 3 can be estimated with parametric, polynomial and semiparametric techniques (Heckman, Urzua and Vytlacil, 2006).¹⁴ The key term for the estimation is

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

where $K'(z) = \left. \frac{\partial K(z)}{\partial z} \right|_{z=\eta}$ 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

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

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).

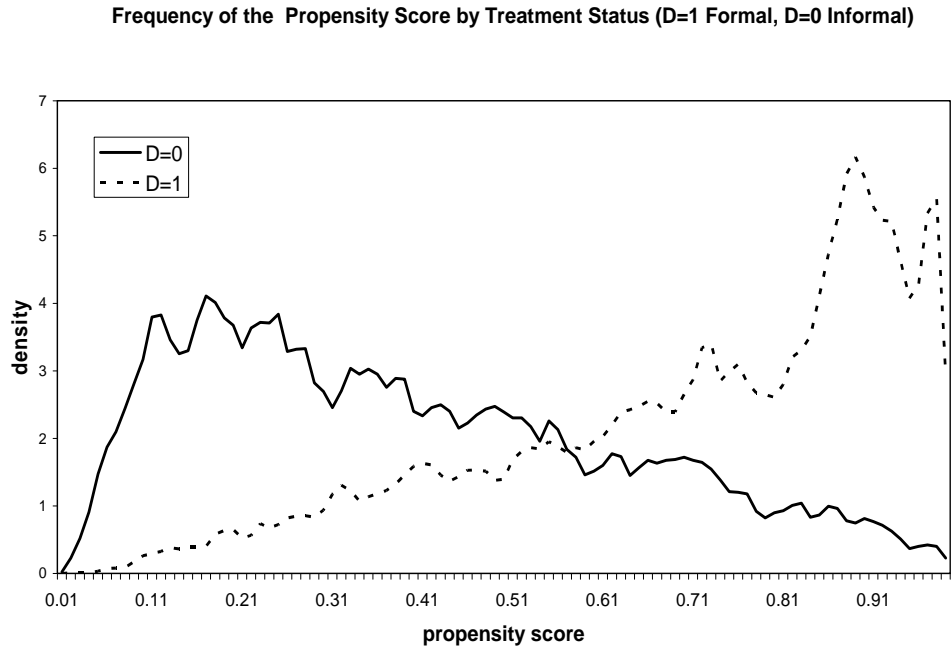
Heckman and Vytlacil (2001ab, 2005) show that the local instrumental variable

¹⁴This paper employs the recently developed MTE software by Heckman, Urzua and Vytlacil (2006). Please refer to their paper and manual for a detailed description of the method. We are very grateful to Sergio Urzua for invaluable help with the implementation of the routine.

(LIV) estimator yields a semiparametric MTE. Following Heckman, Urzua and Vytlačil (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 Vytlačil, 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.

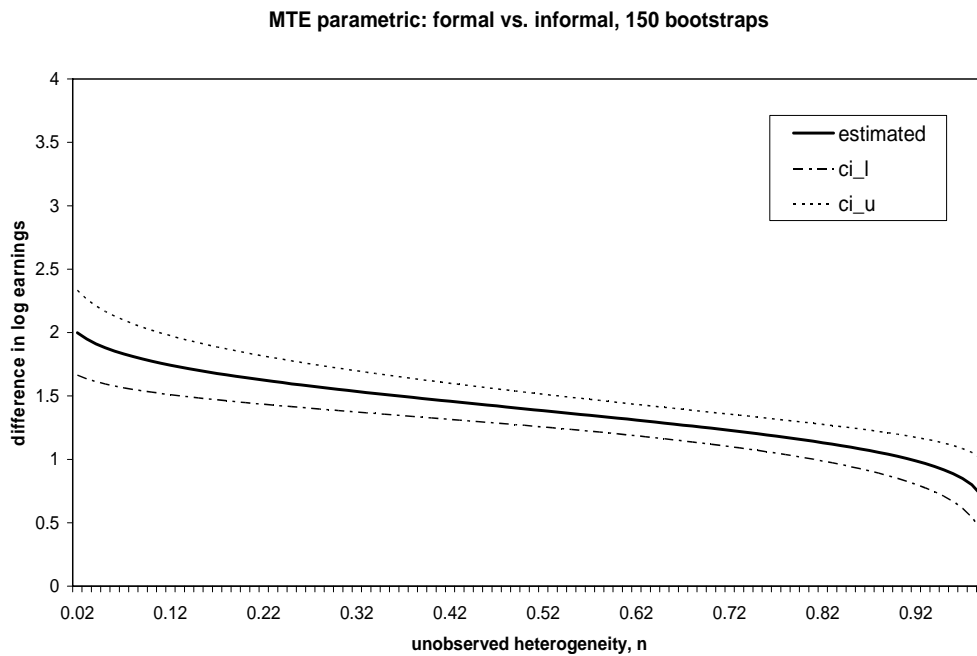
Appendix 2: Figures and Tables

Figure 1: Formal and Informal salaried workers: Common Support



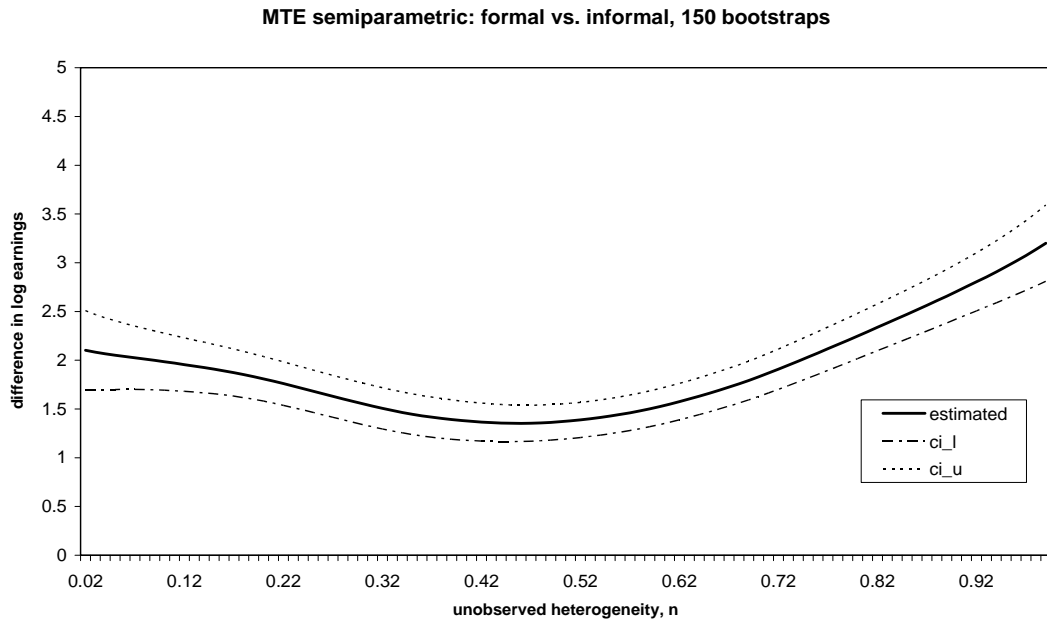
Source: Authors' estimations based on EPH-C, INDEC.

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



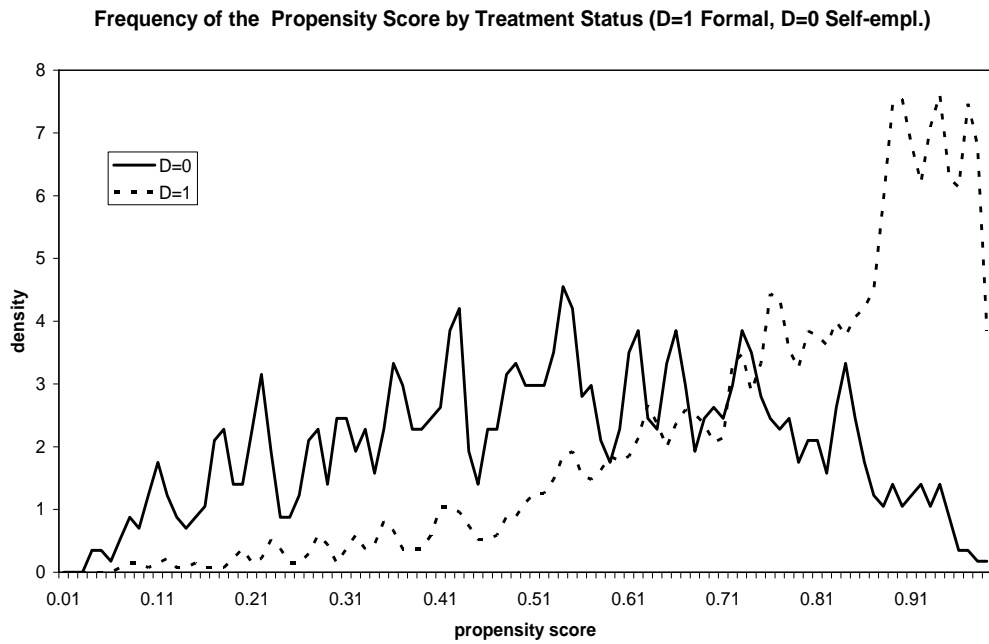
Note: ci_l, ci_u: lower and upper confidence interval. Source: Authors' estimations based on EPH-C, INDEC.

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



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

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



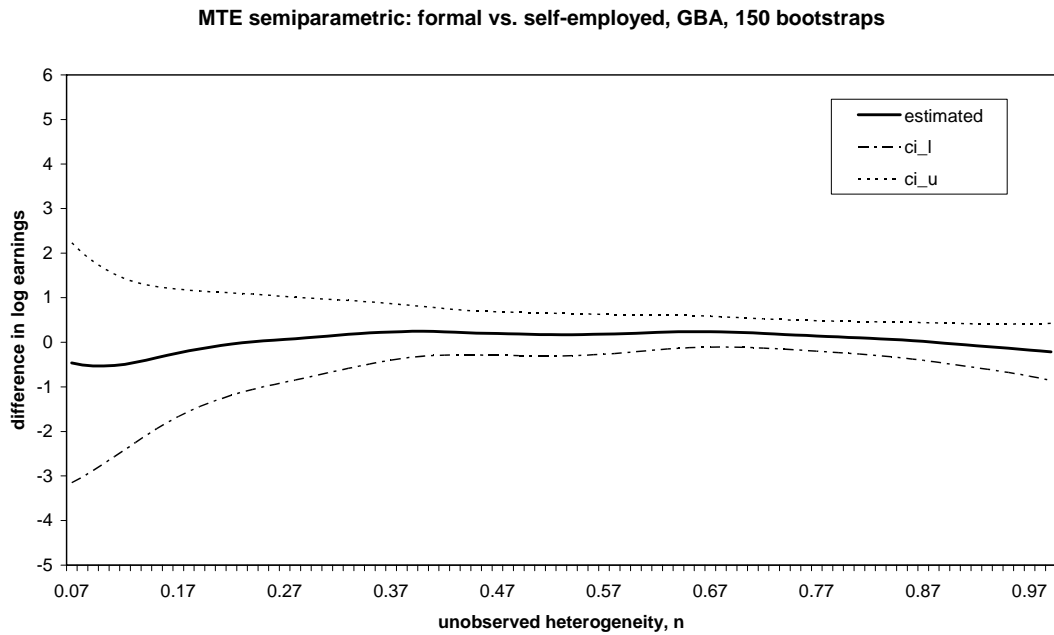
Source: Authors' estimations based on EPH-C, INDEC.

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



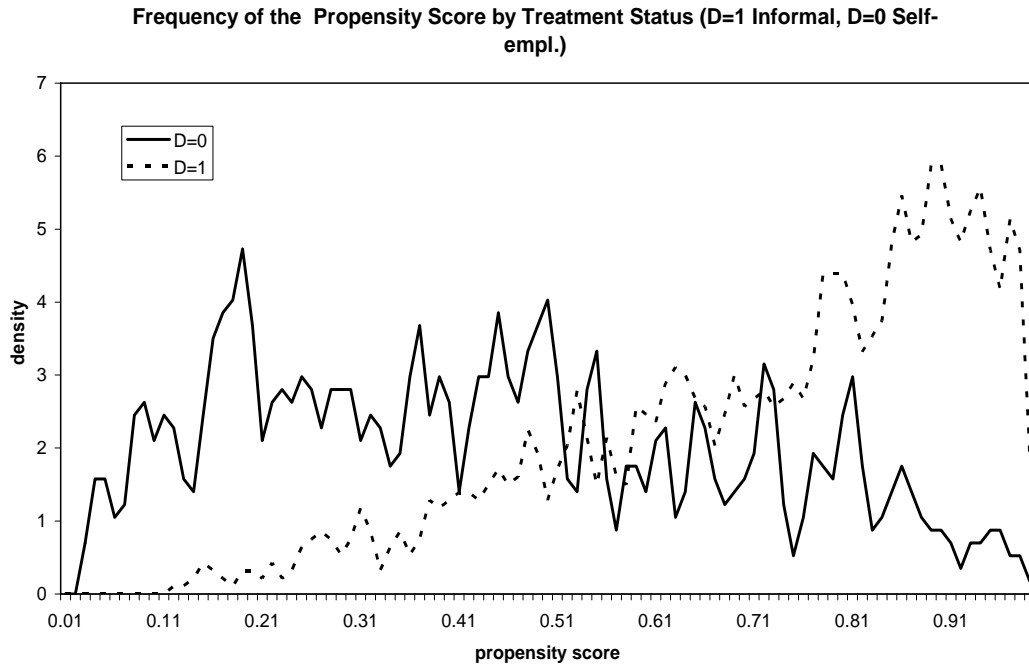
Note: ci_l, ci_u: lower and upper confidence interval. Source: Authors' estimations based on EPH-C, INDEC.

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



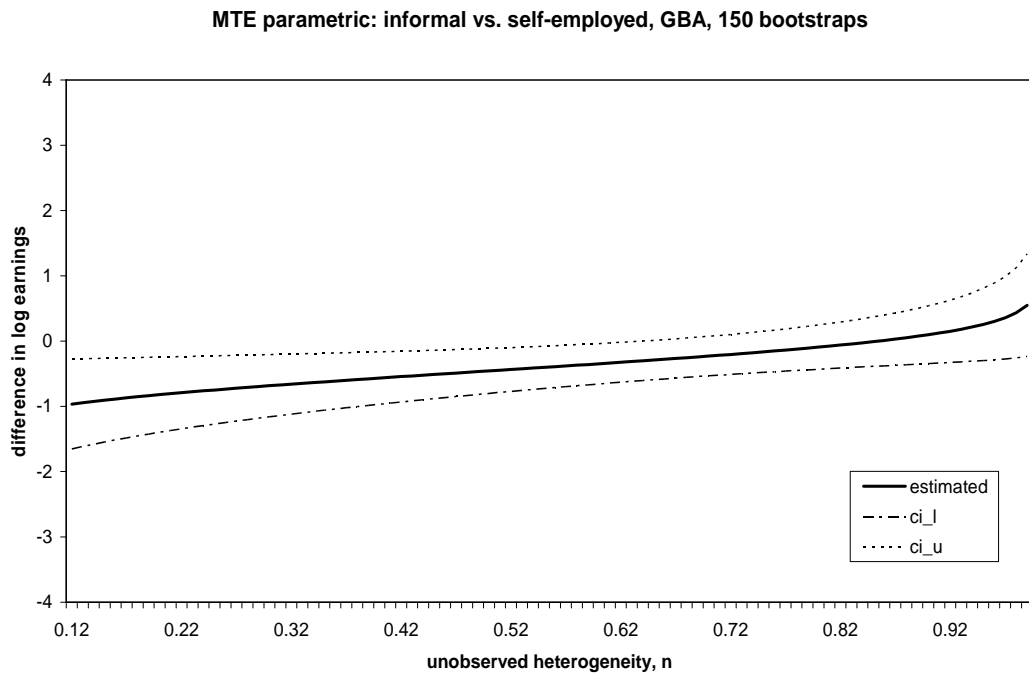
Note: ci_l, ci_u: lower and upper confidence interval. Source: Authors' estimations based on EPH-C, INDEC.

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



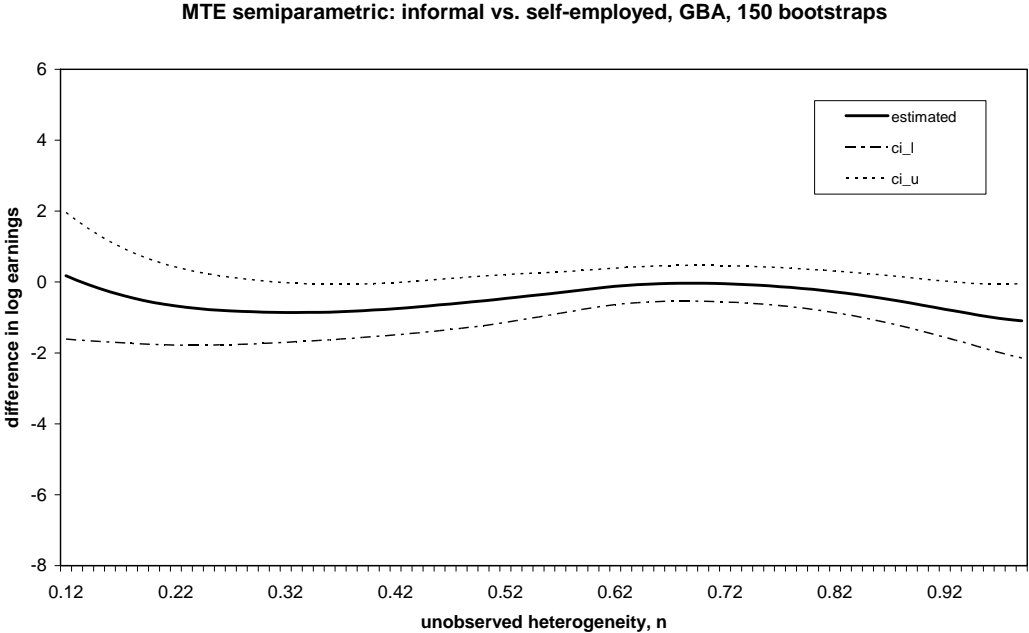
Source: Authors' estimations based on EPH-C, INDEC.

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



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

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



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

Table 1: Formal and Informal salaried workers: Choice Model

Choice model		
	Probit	Marginal effects
secondary education	0.471*** [0.026]	0.178*** [0.010]
tertiary education	0.976*** [0.032]	0.342*** [0.010]
experience	0.028*** [0.003]	0.011*** [0.001]
experience^2	-0.001*** [0.000]	-0.000*** [0.000]
female	-0.456*** [0.038]	-0.175*** [0.013]
primary sector	-0.041 [0.074]	-0.016 [0.030]
construction/trade/utility/transport sector	-0.366*** [0.033]	-0.142*** [0.013]
finance sector	-0.100** [0.046]	-0.039** [0.018]
public and social services sector	-0.134*** [0.034]	-0.052*** [0.013]
Pampeana	-0.056* [0.029]	-0.022* [0.011]
Cuyo	-0.070* [0.039]	-0.027* [0.015]
NOA	-0.185*** [0.042]	-0.072*** [0.015]
Patagonia	0.599*** [0.067]	0.207*** [0.018]
NEA	-0.156*** [0.045]	-0.061*** [0.017]
tenure: less than 1 year	-1.375*** [0.026]	-0.508*** [0.009]
tenure: 1-5 years	-0.834*** [0.024]	-0.321*** [0.009]
single	-0.273*** [0.038]	-0.106*** [0.015]
single_female	0.305*** [0.045]	0.113*** [0.016]
children<=6	-0.014 [0.018]	-0.005 [0.007]
children<=6_female	0.006 [0.026]	0.002 [0.010]
hhs.size	-0.034*** [0.006]	-0.013*** [0.002]
pension_hh	0.279*** [0.037]	0.105*** [0.013]
hhs.head	0.144*** [0.030]	0.055*** [0.011]
pension_head	-0.053 [0.045]	-0.021 [0.018]
single parent	-0.050* [0.025]	-0.019* [0.011]
hhs.human capital	0.016*** [0.003]	0.006*** [0.001]
gdp	-0.007* [0.005]	-0.003 [0.002]
check 2005	0.008** [0.004]	0.003** [0.002]
Constant	0.506*** [0.080]	
Observations	21865	21865
Pseudo R-squared	0.2581	0.2581

Standard errors in brackets. 150 bootstrapping replications for probit regression.

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

Source: Authors' 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: Authors' estimations based on the EPH-C.

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

Coefficients in the outcome equation - semiparametric				
		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: Authors' estimations based on the EPH-C.

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

Choice model		
	Probit	Marginal effects
secondary education	0.158 [0.101]	0.049* [0.029]
tertiary education	0.068 [0.110]	0.021 [0.035]
experience	-0.026*** [0.010]	-0.008*** [0.003]
experience^2	0.000 [0.000]	0.000 [0.000]
tenure: less than 1 year	-0.527*** [0.110]	-0.185*** [0.040]
tenure: 1-5 years	-0.311*** [0.084]	-0.102*** [0.027]
female	-0.107 [0.116]	-0.034 [0.038]
single	0.321* [0.173]	0.096** [0.045]
single_female	0.146 [0.218]	0.044 [0.056]
children<=6	-0.029 [0.071]	-0.009 [0.022]
children<=6_female	0.026 [0.123]	0.008 [0.035]
hhs.size	-0.051** [0.023]	-0.016** [0.008]
pension_hh	0.209 [0.167]	0.064 [0.048]
hhs.head	0.057 [0.099]	0.018 [0.035]
pension_head	-0.021 [0.194]	-0.007 [0.059]
single parent	-0.041 [0.112]	-0.013 [0.037]
hhs.human capital	-0.007 [0.010]	-0.002 [0.003]
taste	1.294*** [0.168]	0.411*** [0.050]
prefer	1.041*** [0.064]	0.326*** [0.021]
Constant	0.587** [0.258]	
Observations	1924	1924
Pseudo R-squared	0.2126	0.2126

Standard errors in brackets. 150 bootstrapping replications for probit regression.

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

Source: Authors' 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: Authors' estimations based on the EPH-C.

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

Coefficients in the outcome equation - semiparametric				
		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: Authors' estimations based on the EPH-C.

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

Choice model		
	Probit	Marginal effects
secondary education	-0.236** [0.098]	-0.088** [0.036]
tertiary education	-0.713*** [0.140]	-0.274*** [0.047]
experience	-0.021** [0.010]	-0.008** [0.004]
experience^2	0.000 [0.000]	0.000 [0.000]
tenure: less than 1 year	0.733*** [0.108]	0.253*** [0.031]
tenure: 1-5 years	0.404*** [0.095]	0.145*** [0.031]
female	0.327*** [0.119]	0.120*** [0.043]
single	0.409** [0.192]	0.144*** [0.055]
single_female	-0.300 [0.236]	-0.115 [0.079]
children<=6	0.029 [0.076]	0.011 [0.028]
children<=6_female	-0.065 [0.115]	-0.024 [0.041]
hhs.size	-0.033 [0.024]	-0.012 [0.009]
pension_hh	-0.204 [0.201]	-0.077 [0.068]
hhs.head	-0.313*** [0.112]	-0.115*** [0.040]
pension_head	0.270 [0.228]	0.096 [0.069]
single parent	0.036 [0.127]	0.013 [0.044]
hhs.human capital	-0.010 [0.011]	-0.004 [0.004]
taste	-0.769*** [0.172]	-0.284*** [0.066]
prefer	0.615*** [0.077]	0.223*** [0.027]
Constant	0.784*** [0.300]	
Observations	1505	1505
Pseudo R-squared	0.2324	0.2324

Standard errors in brackets. 150 bootstrapping replications for probit regression.

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

Source: Authors' 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: Authors' estimations based on the EPH-C.

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

Coefficients in the outcome equation - semiparametric			
		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: Authors' 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*** [0.093]	-0.030 [0.177]	-0.581*** [0.209]
Treatment on the Untreated	1.079*** [0.073]	0.231*** [0.098]	-0.040 [0.187]
Average Treatment Effect	1.392*** [0.066]	0.049 [0.135]	-0.369** [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: Authors' estimations based on the EPH-C.

Table 11: Treatment Parameters: Semiparametric

Treatment Parameters: Semiparametric			
	F vs I	F vs SE	I vs SE
Treatment on the Treated	1.724*** [0.096]	0.033 [0.303]	-0.496* [0.309]
Treatment on the Untreated	2.122*** [0.118]	0.034 [0.150]	-0.522** [0.249]
Average Treatment Effect	1.893*** [0.089]	0.044 [0.215]	-0.486** [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: Authors' 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]	0.187 [0.443]	-0.449 [0.426]
Treatment on the Untreated	1.892*** [0.204]	-0.122 [0.245]	-0.989** [0.510]
Average Treatment Effect	2.002*** [0.105]	0.105 [0.291]	-0.600*** [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: Authors' estimations based on the EPH-C.

Table 13: Treatment Parameters: Semiparametric Version 2

Treatment Parameters: Semiparametric 2			
	F vs I	F vs SE	I vs SE
Treatment on the Treated	1.972*** [0.161]	0.069 [0.354]	-0.468* [0.354]
Treatment on the Untreated	1.788*** [0.168]	0.014 [0.170]	-0.599** [0.319]
Average Treatment Effect	1.892*** [0.098]	0.063 [0.242]	-0.496** [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: Authors' estimations based on the EPH-C.

Table 14: Comparison of Bias and Gains

Treatment Parameters: Selection Bias, Sorting Gain and Total Bias				
	Parametric	Semiparametric	Polynomial	Semiparametric 2
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
Total 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
Total 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
Total Bias	0.379	0.496	0.610	-0.486

Note: Based on Treatment Parameter tables from Authors' estimations of the EPH-C.

OLS compared with treatment parameters from MTE estimations.

Selection Bias: OLS-TT, Sorting Gain: TT-ATE, Total Bias: OLS-ATE or Selection Bias + Sorting Gain

Semiparametric 2: Semiparametric regression based on polynomial

Table 15: 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: Authors' estimations based on the EPH-C.

Appendix 3: Variable description and descriptive statistics

Variable description

Variable Description	
Variables	Explanations
indicator	value 1 if not missing data in sample, intercept
choice	choice/participation variable (0 or 1, depending on margin)
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)
prefer	preference for working dependent (1=prefers dependence, 0=prefers independence)

Descriptive statistics

Summary statistics, 2nd semester 2005 - Part I of II

Variable	Sample 1	Formal	Informal
log of wages	1.436 [0.726]	1.733 [0.585]	1.041 [0.708]
primary education	0.283 [0.450]	0.198 [0.398]	0.396 [0.489]
secondary education	0.391 [0.488]	0.376 [0.484]	0.411 [0.492]
tertiary education	0.326 [0.469]	0.427 [0.495]	0.193 [0.395]
experience	20.689 [14.151]	20.923 [13.265]	20.377 [15.244]
experience ²	628.269 [736.131]	613.735 [675.396]	647.582 [809.454]
Regions			
Pampeana	0.222 [0.416]	0.224 [0.417]	0.220 [0.414]
Cuyo	0.069 [0.254]	0.068 [0.251]	0.071 [0.257]
NOA	0.089 [0.284]	0.079 [0.270]	0.101 [0.301]
Patagonia	0.028 [0.166]	0.036 [0.186]	0.018 [0.134]
NEA	0.042 [0.200]	0.039 [0.194]	0.046 [0.209]
GBA	0.550 [0.498]	0.554 [0.497]	0.544 [0.498]
Tenure			
less than 1 year	0.271 [0.445]	0.137 [0.343]	0.450 [0.498]
1-5 years	0.340 [0.474]	0.317 [0.465]	0.369 [0.483]
more than 5 years	0.389 [0.488]	0.546 [0.498]	0.181 [0.385]
Economic Sectors			
primary	0.012 [0.109]	0.011 [0.104]	0.013 [0.114]
manufacturing	0.150 [0.357]	0.163 [0.370]	0.132 [0.339]
construction/trade/utility/transport	0.332 [0.471]	0.275 [0.446]	0.409 [0.492]
finance	0.088 [0.283]	0.106 [0.308]	0.063 [0.243]
public and social services	0.418 [0.493]	0.445 [0.497]	0.383 [0.486]
Sample Size	21865	12616	9249
Population	5441504	3104906	2336598

Note: Standard deviation in brackets. Weighted Averages. Urban Argentina.

Sample 1: Formal and Informal salaried workers

Source: Authors' estimations based on the EPH-C, Census 2001 and Ministry of Labor data.

Summary statistics, 2nd semester 2005 - Part II of II

Variable	Sample 1	Formal	Informal
Household and individual characteristics			
female	0.435 [0.496]	0.402 [0.490]	0.478 [0.500]
single	0.312 [0.463]	0.264 [0.441]	0.376 [0.484]
single_female	0.143 [0.350]	0.127 [0.333]	0.166 [0.372]
children <=6	0.475 [0.771]	0.435 [0.722]	0.529 [0.830]
children <=6_female	0.170 [0.498]	0.140 [0.443]	0.211 [0.560]
hhs. size	4.028 [2.052]	3.791 [1.831]	4.343 [2.275]
pension_hh	0.251 [0.433]	0.294 [0.455]	0.194 [0.395]
hhs. head	0.496 [0.500]	0.556 [0.497]	0.416 [0.493]
pension_head	0.162 [0.369]	0.210 [0.407]	0.100 [0.299]
single parent	0.294 [0.456]	0.258 [0.438]	0.341 [0.474]
hhs.human capital	8.580 [5.300]	9.367 [5.422]	7.535 [4.945]
Provincial characteristics			
gdp	9.260 [6.281]	9.823 [6.693]	8.512 [5.600]
check05	9.759 [6.251]	10.271 [6.651]	9.078 [5.605]
Sample Size	21865	12616	9249
Population	5441504	3104906	2336598

Note: Standard deviation in brackets. Weighted Averages. Urban Argentina.

Sample 1: Formal and Informal salaried workers.

Source: Authors' estimations based on the EPH-C, Census 2001 and Ministry of Labor data.

Summary statistics, 4th trimester 2005

Variable	Sample 2	Sample 3	Formal	Informal	Self-employed
log of wages	1.685 [0.768]	1.281 [0.835]	1.788 [0.615]	1.175 [0.699]	1.453 [0.996]
primary education	0.270 [0.444]	0.411 [0.492]	0.213 [0.409]	0.417 [0.493]	0.400 [0.490]
secondary education	0.349 [0.477]	0.366 [0.482]	0.374 [0.484]	0.410 [0.492]	0.294 [0.456]
tertiary education	0.380 [0.486]	0.223 [0.416]	0.414 [0.493]	0.173 [0.378]	0.305 [0.461]
experience	23.590 [14.896]	24.607 [15.627]	20.954 [13.894]	21.588 [15.008]	29.541 [15.373]
experience^2	778.257 [852.947]	849.531 [914.632]	631.958 [720.908]	691.042 [803.967]	1108.569 [1020.164]
Tenure					
less than 1 year	0.151 [0.358]	0.342 [0.474]	0.143 [0.350]	0.446 [0.497]	0.171 [0.377]
1-5 years	0.332 [0.471]	0.347 [0.476]	0.330 [0.470]	0.354 [0.478]	0.337 [0.473]
more than 5 years	0.516 [0.500]	0.311 [0.463]	0.527 [0.499]	0.200 [0.400]	0.492 [0.500]
Household and individual characteristics					
female	0.377 [0.485]	0.434 [0.496]	0.395 [0.489]	0.494 [0.500]	0.335 [0.472]
single	0.238 [0.426]	0.241 [0.428]	0.289 [0.454]	0.314 [0.464]	0.121 [0.326]
single_female	0.111 [0.314]	0.106 [0.308]	0.136 [0.343]	0.137 [0.344]	0.055 [0.227]
children <=6	0.395 [0.687]	0.456 [0.762]	0.398 [0.678]	0.499 [0.791]	0.387 [0.708]
children <=6_female	0.117 [0.402]	0.162 [0.460]	0.124 [0.410]	0.199 [0.498]	0.101 [0.382]
hhs. size	3.717 [1.761]	4.068 [2.044]	3.655 [1.679]	4.197 [2.103]	3.858 [1.928]
pension_hh	0.248 [0.432]	0.189 [0.392]	0.278 [0.448]	0.195 [0.397]	0.179 [0.384]
hhs. head	0.590 [0.492]	0.522 [0.500]	0.549 [0.498]	0.425 [0.495]	0.682 [0.466]
pension_head	0.173 [0.378]	0.126 [0.332]	0.186 [0.390]	0.117 [0.322]	0.141 [0.349]
single parent	0.251 [0.434]	0.275 [0.447]	0.267 [0.443]	0.311 [0.463]	0.216 [0.412]
hhs.human capital	9.345 [5.625]	8.206 [5.271]	9.543 [5.558]	7.784 [4.908]	8.898 [5.753]
Preference for occupation					
taste	0.321 [0.221]	0.205 [0.221]	0.347 [0.217]	0.170 [0.215]	0.263 [0.218]
prefer	0.512 [0.500]	0.471 [0.499]	0.619 [0.486]	0.594 [0.491]	0.269 [0.444]
Sample Size	1924	1505	1353	934	571
Population	2508926	2028855	1738791	1258720	770135

Note: Standard deviation in brackets. Weighted averages. GBA.

Sample 2: Formal salaried workers and self-employed. Sample 3: Informal salaried workers and self-employed.

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