

**Assessing the loss due to working
in the informal sector in Venezuela**

Josefa Ramoni and Giampaolo Orlandoni

Josefa Ramoni-Perazzi and Giampaolo Orlandoni Merli

Assessing the loss due to working in the informal sector in Venezuela

Abstract: *In Venezuela, 40% of the workers are employed in the informal sector. This sector is known for being underproductive, meaning that the income received by its workers is less than what they could earn working in formal sector jobs. This paper uses data from the Household Sample Survey (2012-2013) to estimate difference-in-differences linear and quantile regression models, controlling for some demographic characteristics, to quantify the loss associated with working in this market, as an indirect way to quantify the size of the informal sector. The parallel trend assumption is satisfied through propensity score matching, exception made for the highest quartile. The results suggest that informal sector workers lose about 34% of their potential income, loss that is larger for women and with an ambiguous behavior across levels of education. The study also indicates that the average difference in wages between the two sectors tends to narrow over time.*

Keywords: *employment in the informal sector, Venezuelan labor market, DID regression models, quantile regression, propensity score matching*

JEL classification: J46, J31, J24

Evaluación de las pérdidas asociadas al trabajo informal en Venezuela

Resumen: *En Venezuela, el 40% de los trabajadores está empleado en el sector informal. Este sector se caracteriza por baja productividad, lo que implica que el ingreso de los trabajadores que en él se emplean es inferior al que podrían obtener en empleos formales. Este trabajo utiliza datos de la Encuesta de Hogares por Muestreo (2012-2013) para estimar modelos de regresión lineal y por cuantil de diferencia en diferencias, utilizando algunas características demográficas como covariables. Se busca con ello cuantificar la pérdida asociada al trabajo en el sector informal, como una manera indirecta de estimar el tamaño del sector. Para satisfacer el supuesto de igual tendencia en los grupos se recurre a procesos de emparejamiento basados en el propensity score; este supuesto no se cumple para el cuantil superior de salarios. Los resultados sugieren que el trabajador del sector informal pierde, en promedio, cerca del 34% de su ingreso potencial, pérdida que es superior en las mujeres que en los hombres, pero sin un claro comportamiento con respecto al nivel educativo. El estudio indica que la diferencia promedio en los sueldos de ambos sectores tiende a reducirse.*

Palabras clave: *empleo en el sector informal, mercado laboral venezolano, modelos de regresión DID, regresión cuantílica, emparejamiento basado en propensity score*

Clasificación JEL: J46, J31, J24

L'évaluation des pertes liées à l'emploi informel au Venezuela

Résumé: *Au Venezuela, 40% des travailleurs sont employés dans le secteur informel. Ce secteur est caractérisé par une faible productivité, ce qui signifie que le revenu des travailleurs qui y sont occupés est plus petit par rapport à celui qu'ils pouvaient recevoir dans le secteur formel. Notre étude utilise les données de l'Enquête des Ménages (2012-2013) à fin d'estimer à la fois des modèles de régression linéaire que des modèles par quantile en différences, en utilisant certaines caractéristiques démographiques en tant que covariables. Nous voulons quantifier la perte associée aux emplois dans le secteur informel en tant que moyen indirect d'estimer la taille du secteur. Pour répondre à l'hypothèse de groupes d'égale tendance, nous utilisons un processus de matching, lequel est dérivé de la méthode propensity score. Nous montrons que cette hypothèse ne se tient pas pour le quintile supérieur des salaires. Les résultats suggèrent que les travailleurs du secteur informel perdent, en moyenne, 34% de leur revenu potentiel, une perte qui est plus grande chez les femmes que chez les hommes. Ce résultat n'est pas clair lorsqu'on tient en compte le niveau éducatif des travailleurs. L'étude montre que la différence des salaires, en moyenne, tend à diminuer dans les deux secteurs.*

Mots-clés: *emploi dans le secteur informel, marché du travail au Venezuela, modèles DID de régression, régression par quantile, matching basé sur le propensity score.*

Classification JEL : J46, J31, J24.

Assessing the loss due to working in the informal sector in Venezuela

Josefa Ramoni and Giampaolo Orlandoni*

**–Introduction. –I. Background. –II. Literature review. –III. Methodology.
–IV. Results. –Conclusions. –References.**

doi: 10.17533/udea.le.n84a02

Primera versión recibida el 7 de mayo de 2015; versión final aceptada el 20 de agosto de 2015

Introduction

Informality is a common problem in Latin American countries and has been related to poverty and to inefficient allocation of resources. A high incidence of employment in the informal sector not only points out the lack of capacity of the labor market to absorb the increasing labor force, usually as the result of economic recession, but also underestimates the unemployment problem and, therefore, limits the effect of policies designed in this regard. Cimoli, Primi and Pugno (2006) describe the persistence of informality in the region as a structural barrier to sustained growth and poverty reduction resulting from export-led policies that prevent the informal sector from participating in the economic dynamics, in spite of employing half of the urban

* *Josefa Ramoni-Perazzi*: Full professor, Universidad de Santander. Address: Calle 70 #55-210. Bucaramanga, Colombia. Email: j.ramoni@udes.edu.co

Giampaolo Orlandoni Merli: Full professor, Universidad de Santander. Address: Calle 70, #55-210. Bucaramanga, Colombia. Email: gi.orlandoni@mail.udes.edu.co.

The authors would like to thank anonymous reviewers for the insightful and very constructive comments on the original manuscript. This work could not have being done without the cooperation of the University of Los Andes (Mérida, Venezuela).

labor force, with a productivity of just 33%. A report conducted by the Inter-American Development Bank refers to poverty, informality and inequality as the three main structural problems in Latin America, where low-productivity jobs, such as those in the informal sector, lead to low incomes and poverty (IDB, 2008). Particularly, in Venezuela, the informal sector employs more than 40% of the workers, in a scenario of about 30% of households living under poverty, according to the Economic Commission for Latin America and the Caribbean (ECLAC, 2014).

The concern for the loss undergone by economies whose individuals are forced by the circumstances to work in the informal sector has generated a great deal of information about how to estimate the magnitude of this sector, without reaching any agreement. A direct approach of the problem surveys workers in this sector, collecting information on income, without controlling for their productivity-related characteristics. Indirect methods, including the monetary and global indicator approaches,¹ have proven to be difficult to be implemented due to the lack of the required information or the difficulties in satisfying some of their assumptions (Chapa, Flores and Valero, 2007). This paper estimates difference-in-differences linear and quantile regression models in an attempt to quantify the impact of having part of the working labor force employed in the informal sector in Venezuela, by measuring the earnings lost due to working in this sector.

Difference-in-differences (DID) methods have become very common to evaluate programs or estimate the impact of policies. In its simplest representation, the method compares the change in the outcome of sample units exposed to a given treatment (policy, medical treatment, training program, social or environmental issues), before and after it is introduced, to the change in the outcome of sample units not exposed to the treatment. By doing so, it is possible to generate a robust estimation of the effect of the treatment and remove possible biases due to trend and between-group differences over

1 The monetary methods (cash-deposit ratio or cash demand) assume that all transactions in the informal sector are made in cash, while the global indicator methods (physical input consumption or commodity flows) attribute to informal activity the unexplained consumption of a given input, especially electricity.

time. In this case, we are assuming that the decision about working in the informal sector results from a combination of social, political and economic factors that do not evenly affect all individuals, but just the most vulnerable ones in terms of their particular circumstances.

The paper uses information from the Households Sample Survey (HSS) provided by Venezuela's National Institute of Statistics (INE) for 2012:1-2013:2.² This period was selected mostly based on the availability of and access to the most recent information by the time the study was conducted. In addition, this was a critical period of time in Venezuela when the country faced two presidential elections, both of them won by the official candidate, after very expensive campaigns in which government was accused of distributing money through cash and in kind, especially among the most vulnerable population. Unlike previous elections, the ruling party did not use wage adjustments as political strategy. This information is complemented with official reports about informality and poverty in the country over time. By doing so, we can measure the impact of working in the informal sector and show the trend followed by these two variables in the last thirty years.

The original data set suggest some distortions in the behavior of earnings, with average wages in the informal and formal sector apparently following different trends: average nominal wages seem to increase over time in the informal sector, but tend to fall in the formal one, in spite of the high inflation. To overcome the potential violation of the parallel trend assumption, the DID regression models are estimated based on a propensity score matched subsample. The results indicate that workers employed in this sector earn, on average, 66% of what they would earn in the formal sector. This loss of about 34% of their income varies based on educational level and, especially, on gender. The loss is larger at the lowest quartile. No estimation was obtained for the third quartile, which suggests the convenience of exploring other approaches such as density-based methods.

2 The HSS is not an open access data set and can only be bought after special request from an institution. The authors thank the University of The Andes in Venezuela for helping them to obtain the information.

I. Background

The concept of informality has changed many times since its introduction by Keith Hart in 1972 to refer to individuals subsisting by working in marginal activities in an unstructured market. Over time, the definition has included terms such as poverty, underemployment and even illegal, undeclared or unregistered activities not included in National Accounting (Mincer, 1976; Feige, 1990; Blunch, Canagarajah and Raju, 2001). In the 1990s, the International Labor Organization (ILO) linked informality to the number of employees in a firm and used the term to refer to all unregistered enterprises with employees below a given number, generally micro-enterprises. Eventually, the concept expanded to include employers in informal enterprises, own-account workers in informal enterprises, unpaid family workers, and members of informal producers' cooperatives. Following the official concept of informality established in Venezuela, the data provided by the INE define informal sector workers as individuals working for money on household chores, non-professional self-employees, employers and employees at firms with less than five workers and family workers, approach that is used in this study. Generally, the sector is characterized by having small production units where the owner is directly involved in the activity; low human and physical capital investments; low levels of productivity, associated with low wages and poverty, and partial or absolute lack of control by the government.

A. Informality in Latin America

Employment in the informal sector (EIS) seems to be present in most third-world economies. Even though recent information is not often available, statistics from the ILO indicate that the highest numbers for Asia are in Nepal (73.3% in 1999), Thailand (72.9% in 2002) and Indonesia (68.2% in 1999). In Africa, the magnitude of the problem and the lack of information are even worse: 80.7% and 72.4% in Zambia and Gambia in 1993, respectively.

As for Latin America, the information provided by the ILO is relatively more abundant and updated. In the region as a whole, EIS averaged 46.8%

in 2013, with the highest incidence being observed in Guatemala (73.6%), Honduras (72.8%) and Bolivia (70%), followed by Peru and Paraguay with 64% each and Colombia (54.1%). However, it must be said that comparisons across countries may be misleading since the definition of EIS differ among them. For example, Argentina (46.1%) considers as informal any enterprise with up to five workers; Bolivia limits this number to four, six in Brazil (36.5%), and ten in Colombia. According to the Program to Promote Formalization in Latin America and the Caribbean (FORLAC-ILO, 2014) most of these countries exhibit a substantial reduction in EIS since 2003, especially Uruguay (15.1 percentage points-pp), Argentina (14.5 pp) and Brazil (10.8 pp).

B. Informality in Venezuela

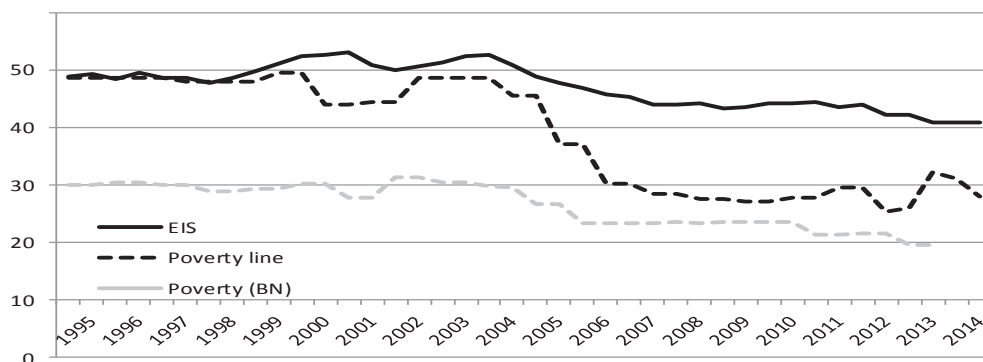
In Venezuela, the informal sector employs almost half of the workers although, according to the INE, this rate has been declining in the past decades, moving from 49.5% in 1995, to 40.8% in 2014 (see figure 1).³ The highest rate is observed in 2003, year in which, due to the 2002-2003 general strike, the government fired 40% of the state-owned oil company's workforce, raising the employment in the informal sector up to 53%. Since then, many companies have been forced to shut down operations and fire their workers, due to the increasing economic restrictions imposed on the private sector.

As said before, several studies relate informality and poverty (Tokman, 1994; Cimoli et al., 2006; IDB, 2008, and Devicienti, Groisman and Poggi, 2009). The direction of the causal effect can go either way, since informality generates poverty mainly through low wages, and poverty pushes workers into the informal sector as a result of the many constraints it poses, including malnourishment and lack of formal education. Based on the data shown in

3 It is important to highlight that several studies on the topic conducted by the ILO in the past years do not include recent information for Venezuela or do not consider this country at all. For example, the *Statistical Update on Employment in the Informal Economy* (ILO, 2012) shows information for 2009 only, while neither FORLAC-ILO (2014) nor *Panorama Temático Laboral: Transición a la Formalidad en América Latina y el Caribe* (ILO, 2014) include Venezuela among the countries analyzed.

Figure 1, poverty in Venezuela as measured by either poverty line or basic needs moves along with informality, with an estimated correlation of 0.88 and 0.90, respectively.⁴ Particularly, in the period 2012-2013, the proportion of households under poverty increased from 25.4% to 32.1% or decreased from 21.6% to 19.6% depending on whether the measure is based on the poverty line or the basic needs approach, respectively. These two measures correspond to two different methods, generated by two different institutions, which may explain the observed discrepancies. However, the increase in poverty as measured by the poverty line is more consistent with the decline in oil prices that forced the government to put some social programs aside.

Figure 1. *Poverty and employment in the informal sector in Venezuela (%), 1995-2014*



Source: authors' calculations using data from INE and ECLAC.

Because of the aforementioned relationship between poverty and informality, any policy intended to generate economic growth and poverty reduction must consider not only the reduction of the unemployment rate, but also the formalization of the EIS. However, such formalization is a necessary but not sufficient condition as long as other forms of social exclusion persist, denying some groups access to opportunities to live productive lives.

⁴ The sources of this information for each variable are: poverty line (ECLAC), basic-needs (INE), and EIS (INE).

As employment in this sector represents an inefficient allocation of resources, usually caused by rigidities in the labor market, we propose that the magnitude of the economic loss due to such an inefficient allocation can be approached by the difference between the income received by workers in the informal sector and the wages they would receive in the formal sector. Estimating such a difference is the purpose of this paper. To do that, we run difference-in-differences regression models on a matched subsample using data obtained from the HSS during the first semester of 2012 and the second semester of 2013.

II. Literature review

Several studies have been devoted to identify the causes of the high incidence of informality in Venezuela. A recent study conducted by Ramoni, et al. (2014) for Venezuela's Central Bank identifies the increasing incidence of EIS, as well as the movement out of the labor force by the unemployed, as the main reasons why the unemployment rate in that country has been declining in the middle of an economic recession. In a previous work Ramoni (2012) points out that, in the last decades, the Venezuelan labor market has moved from low-educated salaried employees working in the agricultural and manufacturing sectors to highly-educated informal and self-employed workers involved in commercial activities.

Marquez and Portela (1991) attribute this problem to lack of adequacy of the Venezuelan economic system, while Vivancos (1988) blames insufficient economic activity. Unemployment resulting from this lack of activity pushes workers into the informal sector (Garnica, 1991; Boza, 2004; Osta, 2007; Díaz and Corredor, 2008) trying to escape from poverty (Marquez and Portela, 1991; Orlando, 2000; Kolev and Morales, 2006; Guerra, 2006). Prieto, Zepa and Martinez (2008) propose the promotion of small- and medium-size firms, based on franchises, as a way to solve it.

As said before, EIS is characterized by being highly unproductive. Ramoni, Orlandoni and Castillo (2010) compare different methods for quantifying the size of the informal economy in Venezuela and conclude that this sector

employs almost half of the total working population to produce one fourth of the non-oil real GDP.

With some exceptions, in general these studies are based on descriptive statistical analysis. In contrast, DID regression models have been widely used to evaluate programs and policies. Examples of their application in the labor market field include quantification of the impact of minimum wage policies on the unemployment rate in the USA (Card and Krueger, 1994), the effect of unemployment insurance payroll taxes on wages and the unemployment rate in that same country (Anderson and Meyer, 2000), and even the effect of the Mariel boatlift migration on the labor market in Miami (Card, 1990).

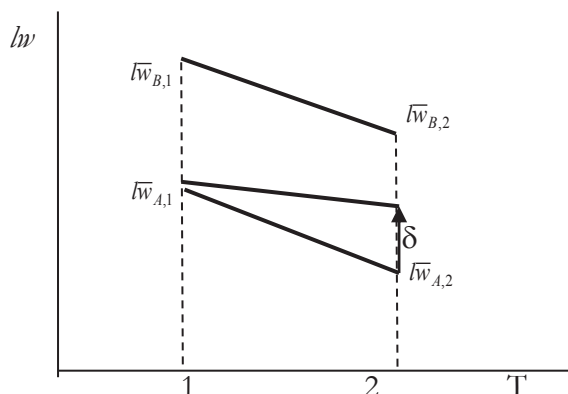
III. Methodology

How much more could an informal sector worker make in a formal sector job, according to his capabilities? Clearly, trying to answer this question by calculating the average wage earned by workers in both sectors is not enough, since we may be comparing non-comparable units. A possible answer could be comparing wages of workers who switched sectors. However, this difference would be influenced by the experience earned by the worker while waiting for a formal job. Another problem we face in these cases is that the data might exhibit selection bias, since the assignment to sectors is not at random, so that some systematic factors may affect the outcomes of an individual's decision. Given the fact that workers cannot be observed under both situations simultaneously, one simple way to overcome these problems and estimate the average treatment effect of working in the informal sector is to compare similar workers in both groups at different points in time.

DID regression models are an improved version of fixed effects regression models in which the possible bias resulting from the correlation between the decision variable and the error term is corrected by differencing twice, in an attempt to replicate an experimental research design using observatio-

nal data.⁵ They can be used to compare the outcome of individuals in two different regimes over time. Let A denote the treatment group formed by workers employed in the informal sector, and B the control group (formal sector workers), so that a variable named *sector* takes value 1 if individual is in the treatment group A, and 0 otherwise. The outcome of interest, in this case the logarithm of hourly nominal wages ($\ln w$), can be observed for individuals in both sectors at two different periods of time T , so that the movement from $\bar{\ln w}_{B,1}$ to $\bar{\ln w}_{B,2}$ in Figure 2 illustrates the trend followed by wages in the formal sector, which is assumed to be the same trend followed by wages in the informal one.

Figure 2. *Difference-in-differences estimation*



Source: authors' elaboration.

The vector δ represents the average treatment effect whose estimator is given by

5 Panel data regression models are used to control for unobservable variables that might affect wages, such as intelligence, capabilities and creativity. Among them, fixed effects regression models work with variables in deviations from the mean. That is why DID regression models are sometimes compared to fixed effects, since they both work with differences although applied in different ways. In addition, as explained in the following, this study attempts to control for unobservables by using propensity score matching methods.

$$\hat{\delta} = (\bar{lw}_{A,2} - \bar{lw}_{B,2}) - (\bar{lw}_{A,1} - \bar{lw}_{B,1}), \quad (1)$$

known as the DID estimator. This can be easily obtained from the regression model given by

$$lw = \beta_1 + \beta_2 T + \beta_3 \text{sector} + \delta \text{sector} \times T + \epsilon. \quad (2)$$

That is, the difference in output between the two groups in period 1 is subtracted from the difference in output in period 2, thus removing biases that could be due to both permanent differences between groups and the trend shown over time by the treatment group.⁶

All the ordinary least squares (OLS) assumptions apply to DID; in addition, as said in previous lines, the parallel trend assumption must be satisfied. According to Becker and Hvide (2013), matching methods may be used to prevent or correct violations of the latter assumption. Particularly, propensity score (PS) matching methods limit the comparison to paired individuals, so that it is a potent way to overcome the violation of the common trend assumption and correct selection bias (Rubin, 1973; Rosenbaum and Rubin, 1983). The PS represents the probability of assignment to treatment (informal) conditional on pre-treatment characteristics X , which can be expressed as

$$PS(X) = \Pr\{\text{sector}=1 \mid X\} = E\{D \mid X\}. \quad (3)$$

Originally, PS matching methods were proposed to reduce bias generated by unobservable confounding factors in studies with observational data where the assignment to treatment is not at random, by limiting the comparison to similar units.⁷ As Becker and Ichino (2002) indicate, the magnitude

6 Similarly, the average loss (gain) in the control group in period 2 with respect to period 1 is subtracted from the average loss (gain) in the treatment group between the two periods.

7 Even though Heckman, Ichimura and Todd (1998) show that the PS matching methods do not necessarily yield more consistent estimators compared to other matching methods, they are commonly used following two-step procedures to estimate treatment effects because of their simplicity. In fact, one of the main advantages of the PS matching is that it reduces the dimension of the process to one single variable.

at which the bias is reduced depends on the quality of the variables used to estimate the PS and the matching method applied. Particularly, we opt for a straightforward matching approach, the nearest neighbor, which compares individuals in both groups with the closest PS. Being aware of the limitations of this method, we allow for replacement to improve the bias reduction even though at the cost of higher variance (Smith and Todd, 2005). As Caliendo and Kopeinig (2005) indicate, bad matches can be avoided by allowing replacement, which is equivalent to imposing a tolerance level of the PS matching as done in radius or caliper methods.⁸

The PS is usually estimated from logit or probit models on the basis of a set of conditioning variables that affect the decision to be modeled, satisfying the balancing assumption [$\text{sector} \perp X \mid \text{PS}(X)$] in order for the PS to provide all the necessary information regarding the determinants of the treatment. The idea behind PS matching is that for each worker in the informal sector, we use the PS to identify a similar one in the formal sector at the starting point.⁹ Then, we run DID regression models comparing the wages of treated and matched control workers.

The DID estimator (1) has proven to be more efficient than the between-subjects estimate of the treatment effect ($\bar{w}_{A,2} - \bar{w}_{B,2}$), or the within-subjects estimate ($\bar{w}_{A,2} - \bar{w}_{A,1}$). In the case that other covariates are added to the model, $\hat{\delta}$ is no longer represented as before, but its interpretation remains the same. Still, the DID estimator based on OLS limits the analysis to one single point of the wage distribution, providing an estimate of the conditional average effect of switching sectors. That is why we also run a DID conditional quantile

8 Other methods such as interval matching did not fit our data. We limited our study to parametric techniques, so that non-parametric matching methods were not considered in this case. The PS matching process followed the algorithm proposed by Becker and Ichino (2002), available in Stata version 12.0.

9 As stated by Heckman (1979), selection bias may arise when considering only observed wages. Like ours, many studies comparing wages between groups (formal-informal, public-private, unionized-nonunionized) control for selection bias by either using Heckman's Mills ratio or PS matching, thus limiting the comparison to the market segments considered (see Azevedo, 2004; Ramoni, 2008, and Huesca and Camberos, 2009).

regression model (QR), which aims at estimating the conditional median and other quantiles of the wage distribution. Unlike OLS, which bases the estimation on minimizing the sum of squared residuals, QR minimizes the sum of absolute deviations. This method has some advantages over OLS: it provides a more complete statistical analysis of how covariates affect the outcome at different points of the distribution, so that new and more accurate relationships can be estimated. Additionally, QR generates robust estimates in the presence of heteroskedasticity (Koenker and Hallock, 2001).

A complete different alternative to the methods discussed above allows data to shape the functional form of the outcome. These are non-parametric techniques, which include both artificial neuronal network models and kernel estimations. While the former rely on flexible functional forms, the latter ones use none. In kernel estimations, the density of each point of the outcome (y_i) for a given value of a covariate (x_i), (y_i, x_i), is estimated based on the proportion of observations that are close to it. These nearby observations are weighted by a kernel function. Once the joint distribution is estimated, the height of the conditional density of y given x can be obtained.

Dinardo, Fortin and Lemieux (1996) implement a semi-parametric kernel procedure (DFL) to analyze the wage distribution in the USA in two different periods of time. The DFL approach works like an Oaxaca decomposition extended to the whole wage distribution. By doing so, they were able to measure the impact of some specific factors on wages. To generate counterfactual density functions of wages to compare with, DFL weights the original density function by the probabilities obtained from probit models. Later on, Huesca and Camberos (2008) adapt the DFL method to compare wages between formal and informal workers in Mexico. In this case, the weights are generated from a PS estimated based on a multinomial logit model. Then, the whole density function of wages in the formal sector is compared to the density wages that would be obtained if a worker in the informal sector were paid as a formal one. In spite of the evident goodness of kernel estimations, this first approach of wages in the informal sector in Venezuela relies on the aforementioned parametric techniques.

In this study, the information is obtained from the HSS for two different periods of time. The HSS is a biannual nationwide sample survey of 45,000 urban households, conducted by the INE to characterize the labor force in Venezuela. From one period to another, 70% of the sample is replaced; among the remaining 30%, each sample unit might be included for no longer than 6 periods, so that cohorts exhibit high attrition bias with not enough workers moving between sectors for any pattern of behavior to be observed. The sample is further reduced by the absence of data on some key variables. Because of that, the data were treated as cross-sectional random samples. The data set used in this study contains a total of 27,735 individuals aged 15 to 65 from two different periods (2012:1 and 2013:2). The data provide information about hourly wages, sector of employment, sex, age, educational level, marital status and geographical region. Estimates are obtained through both a linear regression model and a quantile regression model.

V. Results

A. Description of the original sample

The data indicate an increasing incidence of workers in the informal sector, regardless their demographic characteristics (see Table 1). Among them, participation of women as well as workers with a high-school or college diploma is higher in the second period compared to the first one. The average worker in the informal sector is about five years older. Non-single workers are more likely to work in this sector, probably because they have to shorten the job-search due to family obligations; however, the incidence of single workers in the informal sector tends to increase. Ramoni et al. (2014) find that, on average, Venezuelan workers wait almost a year for a job, search that is about two weeks longer for single workers.

Table 1. *Demographic characteristics by sector and period before matching, 2012-2013*

Variable	Values	Period 1 – 2012:1		Period 2 – 2013:2	
		Formal	Informal	Formal	Informal
Sector (%)	-	59.09	40.91	59.23	41.77
Gender (%)	Female	45.94	41.83	50.20	47.40
	Male	54.06	58.17	49.80	52.60
Education (%)	Elementary	34.16	54.23	45.27	46.65
	High school	29.57	27.43	14.96	33.95
	Technical	6.76	8.39	7.19	8.32
	University	29.51	9.94	32.58	11.08
Marital Status (%)	Married	41.84	54.10	42.08	52.04
	Div/wid	8.26	12.72	8.16	12.29
	Single	49.89	33.18	49.76	35.67
Age (years)	Mean	31.09	36.86	32.21	37.31
	SD	12.62	12.07	12.68	12.34
Hourly wage (log)	Mean	4.56	4.07	4.48	4.11
	SD	0.44	0.75	0.45	0.61

Source: authors' calculations using HSS data.

As for wages, workers in the formal sector show relatively higher average wages, but the wage gap between the two sectors shrinks as wages in the formal sector decline and informal sector workers earn more. This behavior may be explained by rigidities in a formal wage setting system that is highly competed. This fact points out that wages in Venezuela are not only not being adjusted according to the inflation rate, which was 27.1% in 2011, 20.1% in 2012 and 56.2% in 2013 according to Venezuela's Central Bank (BCV), but even pushed down since the data shown correspond to nominal wages (in logarithms). We use nominal rather than real wages to highlight the impoverishment of workers in either sector, since all wage variations are much lower than the inflation rate. This behavior is not common in an election period in Venezuela, when working conditions usually improve and wages are adjusted by the government, with a spillover effect to all of the economy. As for the

methodology, this behavior also implies a violation of the assumption of equal trend in the outcome of treatment and control group in the absence of treatment that, if ignored, yields biased estimates. All further analysis proceeds from this matched, more homogeneous subsample.

B. Propensity score estimates

Not all individuals are likely to work in the informal sector. We limit the analysis to those individuals in the formal sector that are comparable to workers in the informal sector, in terms of their probability of working in informal sector jobs. These individuals are selected based on PS matching using the nearest neighbor approach. The final sample includes 12,559 informal sector workers and 15,776 formal sector workers, matched at period 1.

To estimate the PS, we run a probit model of sector of employment on some demographic characteristics at the base line. The results are:

$$\text{sector} = -0.553 + 0.209 \cdot \text{age} - 0.308 \cdot \text{educ} + 0.349 \cdot \text{reg2} + 0.248 \cdot \text{reg3} - 0.004 \cdot \text{reg4} \quad (4)$$

(0.001) (0.008) (0.016) (0.019) (0.019)

Pseudo R2 = 10.8% LRChi2 = 4456.2*

Note: * denotes $p < 0.01$

Based on these results, the probability of working in the informal sector increases with age; workers with education at college or university level (educ) are less likely to work in this sector. Also, since job opportunities are not equally spread around the country, workers living at west (reg2) or east (reg3) states are more likely to work in this sector compared to north-central workers, where more job opportunities are available (reg4, rest of the country, is not statistically significant).

A simple indicator of differences between informal and formal sector wages before matching is the median absolute standardized variance (MSV), as used by Becker and Hvide (2013) based on Rosenbaum and Rubin (1985), which compares standardized means of the two groups at period 1:

$$\text{MSV} = [100 \cdot (\bar{w}_{A,1} - \bar{w}_{B,1}) / \bar{w}_{A,2} \sqrt{0.5(S_{IwA}^2 + S_{IwB}^2)}], \quad (5)$$

where S_{lw}^2 measures the variance of wages in sectors A and B. The comparison of the MSV indicator before and after matching shows a substantial reduction of the median standardized absolute bias from 66.28 to 14.15.

C. DID estimates

The estimates are based on the model given by $lw = \beta_1 + \beta_2 T + \beta_3 sector + \delta(sector)(T) + \gamma X + \varepsilon$, where lw is the logarithm of hourly nominal wages¹⁰ and X includes a set of traditional demographic variables such as gender, level of education, marital condition and region. The error term ε is assumed to be normally distributed $(0, \sigma^2)$.

The results of the DID estimates are shown in Table 2. As expected, the differences in wages are negative in both periods in all the cases considered, since wages in the formal sector are higher; however, this gap tends to get shorter in the second period as wages in the formal sector deteriorate. The difference is larger for women, especially at low levels of education and at lower quartiles. Notice that these differences do not get larger with the educational level and are substantially higher at the inferior quartile.

The values shown in the last three columns of Table 2 can be interpreted as the gain from formalizing jobs. This gain averages 46% for women and 25% for men and are statistically significant in all the cases considered. If wages increase with the level of education, the fact that low-skilled and low-wage workers gain relatively more (76.6%) when moving to the formal sector could indicate low and diminishing returns to human capital in the Venezuelan formal labor market. No estimation was possible for the third quartile due to the persistent violation of the same trend assumption and the impossibility to find enough appropriate matches. This fact speaks in favor of using the DFL technique in a future study.

¹⁰ The reason for using a logarithmic transformation is to facilitate interpretation of the results, as well as to reduce the variability of the data, which helps to overcome heteroskedasticity problems. Wages are considered on an hourly basis since that is their natural unit of measure. Also, by doing so, we expect to account for the effect of part-time and full-time jobs.

Table 2. *Difference in difference estimates, 2012-2013*

Regression model	Difference at period 1 D1= $\overline{lw}_{A,1}$ - $\overline{lw}_{B,1}$		Difference at period 2 D2= $\overline{lw}_{A,2}$ - $\overline{lw}_{B,2}$		DID = D2-D1		
	Men	Women	Men	Women	Men	Women	General
Lineal:							
Total	-0.593	-1.065	-0.340	-0.604	0.252	0.462	0.338
By education:							
Basic	-0.608	-1.302	-0.366	-0.759	0.242	0.542	0.326
High School	-0.578	-1.060	-0.316	-0.583	0.262	0.477	0.347
Tech & Univ	-0.495	-0.705	-0.306	-0.417	0.189	0.288	0.250
Quartile:							
0.25	-1.447	-2.177	-0.795	-1.204	0.652	0.973	0.767
0.50	-0.854	-1.326	-0.490	-0.754	0.364	0.572	0.413
0.75	-	-	-	-	-	-	-

Note: $p < 0.01$ in all cases.

Source: authors' calculations using HSS data.

Conversely, the results may be interpreted as the extension of the loss suffered by individuals forced by the market to work in informal sector activities due to the inability of the economy to provide jobs and, therefore, it may be seen as an indicator of the magnitude of the loss suffered by an economy with a high incidence of informality. That magnitude can be assessed by multiplying the proportion of workers in the informal sector times their average loss. In our particular case for Venezuela, the fact that 42% of the workers lose 33.8% of their income due to working at the informal sector represents a loss of 14.2% of the national income.

To our knowledge, the literature comparing earnings between formal and informal sectors in Latin America based on counterfactuals is quite scarce. Azevedo (2004) analyzes labor market segmentation for workers living in slum areas in Rio de Janeiro (Brazil) by testing for structural market segmentation, considering formal and informal wage earners and entrepreneurs

in 1998 and 2000. The author compares the estimates obtained from OLS applied to Mincerian earnings equations for each category of workers to those from wage equations corrected for selection bias based on a multinomial logit. In addition, the paper uses the DFL semi-parametric approach to generate counterfactuals. His results indicate relatively higher returns to education for women, with entrepreneurs earning more than salaried workers in the formal sector but less in the informal one.

Huesca and Camberos (2008) analyze the Mexican labor market in 1992 and 2002 considering four categories of employment given by wage earners and self-employed in both formal and informal sectors. To generate counterfactuals, the study relies on the DFL technique, so that the non-parametric kernel density distribution of wages in the informal sector is compared to the density of wages these workers would have been paid in the formal sector. According to their results, there exist substantial benefits from formalization at least in the last year considered, especially among wage earners. As in the case of Venezuela, women gain the most if formalized. Also for Mexico, Moreno (2007) compares self-employed and formal and informal salaried workers considering panel data for the period 2000-2003. No counterfactuals are generated in this study; instead, the gains from formalization are estimated based on the coefficients of Average Treatment, Treatment on the Treated and Treatment on the Untreated Effects, corrected from sample selection. This paper indicates that workers are better off when formalized, no matter their gender, as long as they have high levels of education; switching sectors has a negative impact among workers with low educational levels, a result that contradicts the findings for Venezuela.

All in all, even though most of the evidence speaks in favor of formalizing jobs, the gains from switching sectors, as well as the effect of gender and level of education, may vary between groups and across countries. The reader must keep in mind that these studies assume different definitions of informality and different methodological approaches.

Conclusions

In Venezuela, in 2013 the informal labor market accounted for more than 40% of the workforce, situation that could partly explain the high incidence of poverty (32%). This sector is known for its relatively low productivity, which represents an inefficient use of human capital resources.

This study measures the economic loss due to working in the informal sector by estimating difference-in-differences regression models based on a propensity score matched subsample, in an attempt to reduce biases and ensure the equal-trend assumption. The study uses data from the HSS in 2012 and 2013.

Several results are important to point out. First, during the period analyzed, nominal wages increased at rates below the inflation rate, especially in the formal sector, thus indicating a generalized purchasing power loss. Such a loss takes place in the midst of two presidential elections, processes traditionally characterized by wage adjustments used as a political tool.

Second, workers in the informal sector receive on average 66% of the income they would receive in the formal sector, which represents an economic loss of more than 14% of national income. This last result is obtained by multiplying the forgone earnings times the size of the EIS. Women gain the most from formalization. The difference in wages between the two sectors seems to decrease with the educational level, signaling low returns to education, but this behavior is ambiguous. The gains from switching to formal jobs are larger at the lower the quartile; however, density-based methodologies should be used to observe the behavior of this gain at the upper tail of the wage distribution.

The fact that the difference in average wages between the two sectors in the original data gets shorter in the second year, compared to the first, points out an unhealthy labor market in which wages in formal jobs tend to go down, while those in the informal sector tend to improve. This finding, by itself, should be a big concern for the government, who must introduce the necessary correctives to prevent the ‘informalization’ of the labor market and to encourage investment in human capital, a necessary condition for poverty reduction.

References

- ANDERSON, Patricia & MEYER, Bruce (2000). "The Effects of the Unemployment Insurance Payroll Tax on Wages, Employment, Claims, and Denials", *Journal of Public Economics*, Vol. 78, No. 1, pp. 81-106.
- AZEVEDO, Joao (2004, september). "An investigation of the labour market earnings in deprived areas: evaluating the sources of wage inequality in the slums", In: *EGDI and UNU-WIDER Conference*, Helsinki, Finlandia.
- BECKER, Sascha & ICHINO, Andrea (2002). "Estimation of average treatment effects based on propensity scores", *The Stata Journal*, Vol. 2, No. 4, pp. 358-377.
- BECKER, Sascha & HVIDE, Hans (2013). "Do entrepreneurs matter?", *IZA Discussion Paper*, No. 9295. Institute for the Study of Labor, IZA.
- BLUNCH, Niels; CANAGARAJAH, Sudharshan & RAJU, Dhushyanth (2001). "The Informal Sector Revisited: A Synthesis Across Space and Time", *Social Protection Discussion Paper Series*, No. 0119. Social Protection Unit, The World Bank.
- BOZA, María (2004). "¿Cómo es el Mercado venezolano?", *Debates IESA*, Vol. 9, No. 3, pp. 30-37.
- CALIENDO, Marco & KOPEINIG, Sabine (2005). "Some particular guidance for the implementation of propensity score matching", *Discussion Paper Series*, No. 1588. Institute for the Study of Labor, IZA.
- CARD, David (1990). "The Impact of the Mariel Boatlift on the Miami Labor Market", *Industrial and Labor Relations Review*, Vol. 23, No. 2, pp. 245-257.
- CARD, David & KRUEGER, Alan (1994). "Minimum wages and unemployment: a case study of the fast food industry in New Jersey and Pennsylvania", *The American Economic Review*, Vol. 84, pp. 316-361.
- CHAPA, Joana; FLORES, Daniel & VALERO, Jorge (2007). *La economía informal. Estimaciones, comportamiento y potencial recaudatorio*. México: Trillas.

- CIMOLI, Mario; PRIMI, Annalisa & PUGNO, Maurizio (2006). "A low-growth model: informality as a structural constraint", *Cepal Review*, No. 88, pp. 85-102.
- DEVICIENTI, Francesco; GROISMAN, Fernando & POGGI, Ambra (2009). "Informality and poverty: Are these processes dynamically interrelated? Evidence from Argentina", *Working Paper*, No. 146. Society for the Study of Economic Inequality.
- DÍAZ, William & CORREDOR, Melvy (2008). "Evolución de la economía informal en Venezuela", *Observatorio de la Economía Latinoamericana*, No. 105. pp. 1-5.
- DINARDO, John; FORTIN, Nicole & LEMIEUX, Thomas (1996). "Labor market institutions and the distribution of wages, 1973-1992: A semi-parametric approach", *Econometrica*, Vol. 64, No. 5, pp. 1001-1044.
- ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN (ECLAC) (2014). *Panorama social de América Latina, 2014*. Santiago de Chile: Author.
- FEIGE, Edgar (1990). "Defining and estimating underground and informal economies: The new institutional economics approach", *World Development*, Vol. 18, No. 7, pp. 989-1002.
- FORLAC-ILO (2014). "Recent experiences of formalization in Latin America and the Caribbean". *Notes on Formalization*. Available at: http://www.ilo.org/wcmsp5/groups/public/---americas/---ro-lima/documents/publication/wcms_245882.pdf
- GARNICA, Elizabeth (1991). *La fuerza de trabajo en la agricultura venezolana* Caracas: Fundación Polar-Universidad de Los Andes.
- GUERRA, Alexei (2006). "Ciudadanía, informalidad y Estado en Venezuela: una aproximación a través de las redes", *Espacio Abierto*, Vol. 15, No. 4, pp. 759-786.

- HECKMAN, James (1979). "Sample selection bias as a specification error", *Econometrica*, Vol. 47, No. 1, pp. 153-162.
- HECKMAN, James; ICHIMURA, Hideiko & TODD, Petra (1998). "Matching as an econometric evaluation estimator", *Review of Economic Studies*, Vol. 65, pp. 261-294.
- HUESCA, Luis & CAMBEROS, Mario (2009). "El mercado laboral mexicano 1992 y 2002: Un análisis contrafactual de los cambios en la informalidad", *Economía Mexicana Nueva Época*, Vol. 18, No. 1, pp. 5-43.
- ILO(1972). "Employment, incomes and equality: A strategy for increasing productive employment in Kenya". Available at: http://www.ilo.org/public/libdoc/ilo/1972/72B09_608_engl.pdf
- ILO (2012). The Statistical Update on Employment in the Informal Economy. Available at: http://laborsta.ilo.org/informal_economy_E.html
- ILO (2014). Panorama Temático Laboral: Transición a la Formalidad en América Latina y el Caribe. Available at: http://ilo.org/wcmsp5/groups/public/--americas/---ro-lima/documents/publication/wcms_315054.pdf
- INTER-AMERICAN DEVELOPMENT BANK (2008). *Outsiders? The changing pattern of exclusion in Latin America and The Caribbean*. Washington. Available at: http://www.iadb.org/en/research-and-data/publication-details,3169.html?pub_id=B-2008.
- KOENKER, Roger & HALLOCK, Kevin (2001). "Quantile Regression", *Journal of Economic Perspectives*, Vol. 15, No. 4: pp.143–156.
- KOLEV, Atanas & MORALES, Jesus (2006). "La política monetaria y el sector informal", *Serie Documentos de Trabajo*, N° 75. Banco Central de Venezuela.
- MÁRQUEZ, Gustavo & PORTELA, Carmen (compiladores) (1991). *Economía informal*. Caracas: Ediciones IESA.

- MINCER, Jacob (1976). "Unemployment Effects of Minimum Wages", *Journal of Political Economy*, Vol. 84, No.4, pp. 87-104.
- MORENO, Jorge (2007). "Los salarios del sector formal e informal en México: Análisis de ganancias y pérdidas por formalización", *Ensayos*, Vol. 26, No.1, pp. 1-44.
- PRIETO, Ana; ZERPA, Leiniz & MARTÍNEZ, Marle (2008). "La actividad económica informal y el negocio de franquicia en Venezuela", *Revista de Ciencias Sociales*, Vol. 14, No. 2, pp. 333-345.
- ORLANDO, María (2000). "El sector informal en Venezuela ¿Plataforma o barrera para la reducción de la pobreza?", *Borradores de Trabajo*, No. 15, Universidad Católica Andrés Bello.
- OSTA, Karelys (2007). "Desempleo e informalidad en América Latina: Definiendo políticas públicas para Venezuela", *Revista Venezolana de Gerencia*, Vol. 12, No. 38, pp. 262-278.
- RAMONI, Josefa (2004). *Compensation Comparisons Between the Public and Private Sectors in the United States: A Propensity Score Matching Approach*. A dissertation submitted for the degree of Doctor of Philosophy, Department of Economics, College of Business Administration, University of South Florida.
- RAMONI Josefa (2012). *Descripción y Análisis de los principales indicadores laborales*. Mérida: Consejo de Publicaciones ULA y Banco Central de Venezuela.
- RAMONI Josefa; ORLANDONI, Giampaolo & CASTILLO, Laura (2010). "The Size of the Informal Economy in Venezuela", *El Norte-Finnish Journal of Latin American Studies*, No. 5, pp. 1-12.
- RAMONI Josefa; ORLANDONI, Giampaolo; SINHA, Surendra; TORRES, Elizabeth & ZAMBRANO, Ángel (2014). "Análisis de la duración del desempleo en Venezuela", *Serie Documentos de Trabajo S/N*. Banco Central de Venezuela.

- ROSENBAUM, Paul & RUBIN, Donald (1983). "The central role of the propensity score in observational studies for causal effects", *Biometrika*, Vol. 70, pp. 41-55.
- ROSENBAUM, Paul & RUBIN, Donald (1985). "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score", *The American Statistician*, Vol. 39, No. 1, pp. 33-38.
- RUBIN, Donald (1973). "Matching to Remove Bias in Observational Studies", *Biometrics*, No. 29, pp. 159-183.
- SMITH, Jeffrey & TODD, Petra (2005). "Does matching overcome LaLonde's critique on nonexperimental estimators?", *Journal of Econometrics*, Vol. 125, No. 1-2, pp. 305-353.
- TOKMAN, Victor (1994). "Informalidad y pobreza: progreso social y modernización productiva", *El trimestre económico*, Vol. 61, No.1241, pp. 16-31.
- VIVANCOS, Francisco (1988). *La economía informal en Venezuela: Una aproximación econométrica*. Caracas: IIES-UCAB.