Are Informal Self-Employment and Informal Employment as Employee Behaviorally Distinct Labor Force States? *

Luca Flabbi[†] Mauricio Tejada[‡]

June 7, 2023

Abstract

The paper performs both a parametric and non-parametric analysis to address a fundamental question in the growing literature using search models to study labor market informality: should informal self-employment and informal employment as employee be considered two different labor market states? Both the non-parametric and the parametric tests strongly reject equality between the two states, cautioning against aggregating them in a common "informality state." The parametric model indicates the source of the difference in the high dispersion of informal self-employment income and in the low duration of informal employee jobs.

Keywords: Labor market frictions, Search and matching, Informality, Self-employment.
JEL Codes: J46, J64, O17.

*Tejada gratefully acknowledges financial support from FONDECYT, grant project No. 11196296.
†Department of Economics, University of North Carolina - Chapel Hill, USA. E-mail: luca.flabbi@unc.edu.
‡Department of Economics, Universidad Diego Portales, Santiago Chile. E-mail: mauricio.tejada@udp.cl.

1 Introduction

In an influential paper published in 1983, Flinn and Heckman asked: "Are Unemployment and out of the Labor Force Behaviorally Distinct Labor Force States?" (Flinn and Heckman, 1983). The question was relevant because labor economists had started to study labor market dynamics with richer theoretical and empirical models, forcing researchers to take a stand on which labor market states were relevant, what possibility of transitions between them existed, and what assumptions were worth imposing. The study of labor market dynamics in economies with high informality is experiencing a similar transformation. Richer labor market models have recently been developed and estimated, prompting a crucial debate on what the relevant labor market states are, and which transitions between them we should focus on.

This paper contributes to this debate by addressing its most controversial question: should we differentiate between informal workers that are hired as employee and those that are self-employed? Aggregating or differentiating these two labor market states has proven relevant both for obtaining credible estimates and for drawing policy implications. In addition, both approaches have been used by influential papers, without producing a consensus in the literature. For example, Meghir et al. (2015) is one of the first estimated search model on a market with high informality (Brazil) and aggregates all the unregistered employees and the self-employed in the same labor market state; Bobba et al. (2022) is a recent contribution on Mexico but strongly differentiates between the two states, so much so as to consider self-employed informality as a searching state in alternative to unemployment. A number of other contributions take one approach or the other, either driven by data availability or by the desire for a richer or more parsimonious specification.[1]

We use data from Colombia – the fourth economy in Latin America, a region with particularly high levels of labor market informality – to test if the informal self-employed and the informal

[1] Among the contributions that do *not* differentiate between informal employee and self-employed are early contributions in the theoretical search literature, such as Albrecht et al. (2009); Charlot et al. (2013); and in the macro search literature, such as Bosch and Esteban-Pretel (2012). More recently, Haanwinckel and Soares (2021) develop a search model with intra-firm bargaining and exclude the self-employed. On the other side, contributions developing search models of the labor market that *do* differentiate informal self-employment as a distinct state from informal employment include Bobba et al. (2021); Narita (2020).

Other examples beyond the search literature that take a stand in this debate include: Esteban-Pretel and Kitao (2021), which allows for only one informal labor market state, excluding the self-employed from the calibration sample; Ulyssea (2018), which considers informality choices of both firms and workers, separating informal workers who are employees from informal firms; Granda and Hamann (2015), which distinguishes between informal entrepreneur (informal self-employed) and informal worker (informal employee); Almeida and Carneiro (2012), which clearly differentiates between informal wage earners and self-employed.

employee should belongs to two different labor market states when modeling labor market dynamics. We follow Flinn and Heckman (1983) in providing two types of analysis. First, we conduct non-parametric tests, using transitions to compare Kaplan-Meier survival functions and empirical wage distributions to perform Kolmogorov–Smirnov tests. Then, we estimate a standard partial equilibrium search model, which allows to distinguish the effect of mobility parameters from the effect of wage offers on individuals' behavior. Under this parametric model, we can directly impose the same behavior for self-employed and informal employee and perform likelihood ratio tests to assess the validity of the restriction.

Both the non-parametric and the parametric tests strongly reject that informal self-employment and informal employment as employees are behaviorally indistinguishable labor force states. The result cautions against aggregating them when modeling either empirically or theoretically labor markets with high informality. Instead, informal self-employed and informal employees should be considered different labor market states and neither one should be considered the sole representative of the behavior of the typical informal worker. For the Colombian case, we estimate the main sources of the difference to be the dispersion in labor income offers (much higher for the informal self-employed) and the job termination rate (much higher for the informal employees).

The paper is organized as follows: Section 2 presents the data, Section 3 provides the non-parametric analysis and Section 4 the parametric one. Section 5 concludes.

2 Data

We use the Colombian *Gran Encuesta Integrada de Hogares (GEIH)* for 2016. GEIH is a nationally representative survey collected monthly by the *Administrative Department of National Statistics* (DANE). The survey contains individual characteristics – such as gender, age, and schooling – and collects labor market outcomes – such as employment status, durations, monthly labor income, weekly hours worked, and occupational characteristics. It does also allow for a precise definition of labor market informality. We define any employed workers to be *informal* if they do not contribute to social security, a definition consistent with International Labor Organization (ILO)'s recommendations and with the previous literature on LAC (Perry et al., 2007; Kanbur, 2009; Bobba et al., 2022). If these workers are in a subordinate working relationship with a well-defined employer, we classify them as *informal employee*; if they are occupied in an activity with more independence, such as selling cheap goods in a street corner, we classify them as *informal self-employed*.

To be consistent with the theoretical model, we extract an estimation sample relatively homogeneous over demographic characteristics and education. We focus on 25-55 years old men, living in urban areas, that have completed at most secondary education and work full-time when employed. We focus on male unskilled workers because this is the group on which labor market informality is the most relevant and the most studied. To gain sample size, we pool together all the surveys conducted in the year, from January to December 2016. The final estimation sample include 88,123 observations. Two important differences between informal employees and informal self-employed are already evident in the descriptive statistics of the estimation sample. First, the labor income distribution for the self-employed is much more dispersed than the one for informal employees, with standard deviations of, respectively, 0.56 and 0.36 (to be compared with means of 1.073$/h and 1.064$/h). Second, informal self-employed jobs last much longer than informal employee jobs, with an average of, respectively, 106 and 33 months. In the other two labor market states, the unemployed are searching for a job for an average of 4 months, while formal employees earn on average more than informal workers (with a mean of 1.419$/h). We should mention that formal self-employed workers are a tiny proportion of this demographic group and therefore are ignored in the analysis.

3 Non-parametric tests

We begin by analyzing non-parametric differences in the dynamics of unemployment inflows and outflows into and from the current job, as well as labor income, between informal employees and informal self-employed workers. To accomplish this, we compare the empirical survival probabilities computed using duration data in the unemployment state immediately prior to the current job, the empirical survival probability computed using ongoing duration data of the current job, and the empirical cumulative distribution function computed using hourly labor income data.[2] To compute the empirical survival probability, that is the probability of staying in given state last longer than t or $\Pr[\tau > t]$, we use the Kaplan-Meier survival function estimator.[3]

Additionally, we use the Kolmogorov-Smirnov (K-S) non-parametric test to statistically com-

[2]The empirical cumulative distribution function (ECDF) is compute as $F(x_j) = \Pr[x_i \leq x_j] = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}[x_i \leq x_j]$ where $\mathbf{1}[C]$ is an indicator variables that takes the value of 1 es the condition C is satisfied and zero otherwise and n is the total number of observations.

[3]The Kaplan-Meier survival function is estimated by calculating the product of the survival probabilities at each time point, which in turn are calculated at any given time by the proportion of individuals who have survived to that time point (Kaplan and Meier, 1958). That is $S(t_j) = \Pi_{i=1}^{j} \frac{n_i - h_i}{n_i}$ where t_j is the duration of the spell j, h_j is the number of completed spell of duration t_j, and n_j is the total number of spells not completed before t_j (see for example Kiefer, 1988).

pare the probability distributions of the duration and labor income data between informal employees and informal self-employed workers. In the K-S test, the null hypothesis is whether the data draws of the two samples we observe come from the same underlying distribution in the population Dodge, ed (2008). Formally, the K-S test is based on the maximum difference between the empirical cumulative distribution functions of the two samples,[4] in our case the survival probability distribution function on the one hand, and the cumulative labor income distribution function on the other hand. It is worth mentioning that this test does not provide information on how the two distributions differ, but rather on whether they are different or not.

Figure 1 shows the results of the Kaplan-Meier survival function estimator. As observed in panel (a), the probability of remaining in the unemployment state drops drastically (below 0.2) in the first 10 months, and this holds true for those who exit as informal employees or as self-employed. The survival probabilities seem to be very similar for an unemployed individual who exits to any of these types of jobs, except for durations close to a year of searching where there is a small but statistically significant gap. In panel (b), the comparison results are very different. Indeed, the probability of remaining in an informal job drops below 0.25 in the first 5 years and further below 0.1 in the first 10 years. For the self-employed, the probability of remaining in this employment state is above 0.6 and slightly below 0.4 for these same periods of time. Also, notice that the gap in the survival probability starts to close after 30 years on the job. When comparing the empirical cumulative distributions of hourly labor income, in Figure 2, it can be observed that the distributions cross each other and that there are noticeable differences above and below the common mean (around 1.1 $/h). Below the mean, it is more likely for the self-employed to earn a low labor income, while above the mean the opposite is true, that is it is more likely for an informal employee to earn a high labor income.

The K-S test results, presented in Table 1, confirm what we described above. The null hypothesis of equal duration distributions between informal employees and self-employed is largely rejected at any significant level for the duration of the current job, while it is rejected at a 5% confidence level for the unemployment duration in the spell immediately prior to the current job. Under the alternative that the duration distribution of the self-employed always lies below that of the informal employees, the null of equal distributions cannot be rejected. While under the alternative that the duration distribution of the self-employed always lies above that of the informal employees, the null of equal distributions can be rejected in favor of the alternative even at a 1% confidence level. Finally, the null hypothesis of equal hourly labor income distributions

[4]The K-S test uses the empirical cumulative distribution function $F(x_j)$, as defined in footnote 2, for two samples, say 1 and 2, to compute the following statistic: $D = \sup |x\,(F_1(x_j) - F_2(x_j))|$.

between informal employees and self-employed can also be rejected at any confidence level, and because the distributions cross each other, the null hypothesis is rejected in favor of all the alternatives.

4 Search model based tests

The non-parametric tests presented in Section 3 already give a strong indication that informal employees and informal self-employed should be considered two separate labor market states. But they cannot distinguish if the sources of the separation are the frictions and shocks affecting the labor market dynamics or the wage offers distributions affecting agent's decisions to accept a job. To make progress in this understanding, we follow Flinn and Heckman (1982) and develop and estimate a simple search model of the Colombian labor market. Under this parametric approach, we can directly impose the same behavior for self-employed and informal employee and perform likelihood ratio tests to assess the validity of the restriction.

4.1 Environment and equilibrium conditions

Time is continuous, the environment is stationary and the economy is populated by a large number of infinitely-lived individuals with discount rate ρ. At any point in time, each individual can be in one of the following four states: unemployment u, informal self-employment s, formal employment f, and informal employment i. We denote with $v = s, f, i$ the different job "type" in which an agent can be employed in. While unemployed, individuals receive flow utility b and search for jobs, meeting offers at a Poisson rate $\lambda(v)$. Offers are fully described by labor income x, drawn from the job-type-specific exogenous offer distributions $G(x|v)$. All jobs terminate exogenously at a rate $\eta(v)$. Formal employees are different from informal workers because they pay a proportional payroll contribution τ. But, in exchange, they receive benefits that are valued at a flow utility that we denote with θ. Only unemployed individuals search for a job.

Let U be the steady-state value of unemployment, and let $E(x, v)$ be the steady-state value of employment of type v and labor income x. The optimal behavior in this economy can be characterized by the following set of Bellman equations:

$$\rho U = b + \sum_{v=s,f,i} \lambda(v) \left[\int \max \{E(x, v) - U, 0\} \, dG(x|v) \right] \quad (1)$$

$$\rho E(x, v) = x [1 - \tau \iota_{v=f}] + \theta \iota_{v=f} + \eta(v) [U - E(x, v)] \quad (2)$$

where $\iota_{v=f}$ is an indicator variable equal to 1 if the type of job is formal and zero otherwise.

The optimal decisions for accepting a job have a reservation values property. In particular, the reservation labor income $x^*(v)$ for the type of job v satisfies $E(x^*(v), v) = U$. Using Equations (1) and (2), this condition translates to $x^*(v) = \frac{\rho U - \theta \iota_{v=f}}{1 - \tau \iota_{v=f}}$. Note that in order to make an unemployed individual indifferent between accepting any type of job at the reservation value, it has to hold that $x^*(v) = \rho U$ for all $v = s, f, i$, which implies $\theta = \tau \rho U$. Thus, the utility value of a formal job should account for the tax loss of the job's present value, which is the present value of the job search (the unemployment value).

In steady-state equilibrium, the flows out of unemployment to each type of jobs and from those jobs to unemployment are equal, which implies that $u\lambda(v)\left[1 - G(x^*(v)|v)\right] = \eta(v)v$. Finally, we normalize the total population to 1 such that $\sum_{v=s,f,i} v = 1$.

In this model environment, the *null hypothesis* of informal self-employment and informal employment as employee being behaviorally indistinguishable labor market states is represented by the following set of parametric constraints:

$$\begin{cases} \lambda(s) = \lambda(i) & \text{Job opportunities arrival rates} \\ \eta(s) = \eta(i) & \text{Termination rates} \\ G(x|s) = G(x|i) & \text{Labor income offers distributions} \end{cases} \tag{3}$$

4.2 Estimation

We estimate the model by maximum likelihood. As described in Section 2, the available data for estimation are: (1) the hourly labor income (measured in US dollar per hour) and the on-going duration in the state (measured in month) for all employment states (self-employment, formal and informal employment); and (2) the on-going unemployment or searching duration (measured again in months). For given observation j, we denote durations with $\{t_j\}_{j \in U,S,F,I}$ and labor income with $\{x_j\}_{j \in S,F,I}$.

To derive the durations' contributions to the likelihood function, we define the hazard rates for each labor market state. An unemployed individual leaves the state if he receives an acceptable job offer from any of the three types of jobs. Therefore, the hazard rate out of unemployment is given by $h_u = \sum_{v=s,f,i} \lambda(v)\left[1 - G(x^*(v)|v)\right]$. An employed individual in any type of job leaves the state if a termination shock occurs. Therefore, the hazard rate out of employment is $h_v = \eta(v)$. Since no hazard rates depend on the duration in the state, all the durations follow a negative exponential distribution with parameter equal to the corresponding hazard rate (see for example Eckstein and van den Berg, 2007).

To derive the labor incomes' contributions to the likelihood function, it is important to first consider that the labor income observed in the data only includes accepted job offers, while the primitive of the model is the overall distribution of job offers. Conditioning on the model's equilibrium, the accepted offers satisfy $x \geq x^*(v)$ for job types v. Consequently, the density of the observed labor income is $g_x(x|v) = \frac{g(x|v)}{1-G(x^*(v)|v)}$. In addition, to account for the possibility of labor income being measured with error, we assume that the observed labor income is $x^o = x \times \epsilon$, where ϵ represents the measurement error and follows a distribution $Q(\epsilon)$. We also assume that the conditional expectation of observed wages is equal to the true wages, that is $\mathbb{E}[x^o|x] = x$, which implies $\mathbb{E}[\epsilon|x] = 1$.[5] The density function of observed labor income is then given by $g_x^o(x^o|v) = \int_{x^*(v)} \frac{1}{x} q\left(\frac{x^o}{x}\right) \frac{g(x|v)}{1-G(x^*(v)|v)} dx$ for $v = s, f, i$.

In conclusion, the logarithm of the likelihood function to be maximized to find the set of parameters $\Theta = \{\lambda(v), \eta(v), G(x|v)\}_{v=s,,f,i} \times \{b, \theta, Q(\epsilon)\}$ is:

$$\begin{aligned}
\mathcal{L}(\Theta) &= \sum_{j \in U} \left[\log\left(h_u \exp(-h_u t_j) \times u\right)\right] \\
&+ \sum_{j \in S} \left[\log\left(\eta(s) \exp(-\eta(s) t_j \times \int_{x^*(s)} \frac{1}{x} q\left(\frac{x_j^o}{x}\right) \frac{g(x|s)}{1-G(x^*(s)|s)} dx \times s\right)\right] \\
&+ \sum_{j \in F} \left[\log\left(\eta(f) \exp(-\eta(f) t_j \times \int_{x^*(f)} \frac{1}{x} q\left(\frac{x_j^o}{x}\right) \frac{g(x|f)}{1-G(x^*(f)|f)} dx \times f\right)\right] \\
&+ \sum_{j \in I} \left[\log\left(\eta(i) \exp(-\eta(i) t_j \times \int_{x^*(i)} \frac{1}{x} q\left(\frac{x_j^o}{x}\right) \frac{g(x_j|i)}{1-G(x^*(i)|i)} dx \times i\right)\right] \qquad (4)
\end{aligned}$$

where each of the contributions has been weighted by the probability of being in each labor market state given that the data is observed only for those who are currently in those states.

An important advantage of using maximum likelihood estimation in our application is that we can employ the Log-likelihood Ratio Test (LR) to directly test the null hypothesis expressed by the set of constraints (3).[6]

We conclude the estimation subsection by briefly discussing the identification strategy. Our approach follows the standard strategy formulated by Flinn and Heckman (1982). First, the duration data provides direct information to identify the hazard rates for transitioning out of each labor market state, thanks to the implication of a negative exponential distribution. By combining these four hazard rates with the three steady-state conditions that ensure equality of

[5]If $Q(\epsilon)$ is a log-normal distribution with location and scale parameters μ_ϵ and σ_ϵ, this assumption implies that $\sigma_\epsilon = \sqrt{-2\mu_\epsilon}$, thus only one parameter of the measurement distribution needs to be estimated.

[6]Recall that the Log-likelihood Ratio statistic is $LR = -2\left[\mathcal{L}(\Theta_0) - \mathcal{L}(\hat{\theta})\right]$, where $\mathcal{L}(\Theta_0)$ is the value of the log-likelihood function under the null hypothesis and $\mathcal{L}(\hat{\theta})$ is the value of the log-likelihood function of the unconstrained model.

flows between unemployment and different job types, we have sufficient information to identify the mobility parameters: the arrival rates of job opportunities ($\lambda(v)$) and the termination rates ($\eta(v)$). Second, in order to identify the offered hourly income distributions from the observed labor income data, it is necessary to recover the original distribution from its truncated version at a known truncation point (known as the invertibility condition). To achieve this, we assume that labor income offers in each employment state are drawn from log-normal distributions with location and scale parameters $\mu(v)$ and $\sigma(v)$ respectively. These parametric distributions satisfy the invertibility condition. Third, the parameters for the flow utility of unemployment (b) and the discount rate (ρ) are jointly identified through the reservation labor income ρU. Therefore, we estimate ρU directly by maximizing the likelihood function in equation (4), and then we recover b using the equilibrium condition in equation (1) given ρ. We set ρ at the recommended discount rate for Latin America by multilateral development banks, which is 0.12 (Moore et al., 2020). Finally, we follow Fernández and Villar (2017) to set the institutional parameter τ to 0.16, with which we recover θ using the constraint $\theta = \tau \rho U$.

4.3 Results

Table 2 reports the estimated parameters. The column *Unrestricted* presents the unconstrained model: this is the model presented in Section 4 where all the parameters are allowed to be different across labor market states. The column *Restricted* presents a specification where we impose the set of constraints (3): this is a model where the parameters for the informal self-employed labor market state ($v = s$) and the informal employee labor market state ($v = i$) are constrained to be equal.

The Log-likelihood Ratio Test is presented at the bottom of the Table and clearly rejects the Restricted model against the Unrestricted one. Therefore, the null hypothesis of informal self-employment and informal employment as employee being behaviorally indistinguishable labor market states is rejected, and at very low P-values. We have also estimated and tested two intermediate models, one where we impose equality only on the mobility parameters ($\lambda(s) = \lambda(i); \eta(s) = \eta(i)$) and the other where we impose equality only in the offers distributions parameters ($\mu(s) = \mu(i); \sigma(s) = \sigma(i)$). Also these two models are clearly rejected against the Unrestricted model, with P-values of, respectively, .These results indicate that the sources leading to informal self-employment and informal employment being two different labor markets states are to be found both in the frictions and shocks affecting the labor market dynamics and in the wage offers distributions affecting agent's acceptance decisions do. So much so, that even one of the two would be enough.

Looking at the actual point estimates in conjunction with the implied values reported in Table 3, we observe a very large difference in the dispersion of the labor income distributions between the two informal states, with the self-employed's standard deviation being more than double the employee's one. This is the actual source of the differences coming from the wage offers distributions, not a difference in mean offers. In terms of mobility parameters, the source of the difference is in the termination rates, with the self-employed's termination rate being less than a third than the employee's one.

5 Conclusions

The paper performs both a parametric and non-parametric analysis to address a fundamental question in the growing literature using search models to study labor market informality: should informal self-employment and informal employment as employee be considered two different labor market states? Both the non-parametric and the parametric tests strongly reject the equality of the two labor market states, cautioning against aggregating them in a common "informality state" as done in important previous contributions in the literature.[7] The parametric model indicates that the source of the difference are the higher dispersion in informal self-employment income offers than in informal employee wage offers and the lower duration of informal employee jobs with respect to informal self-employment positions.

[7] Meghir et al. (2015) is an influential example. For additional references see footnote 1.

References

Albrecht, James, Lucas Navarro, and Susan Vroman, "The Effects of Labour Market Policies in an Economy with an Informal Sector," *Economic Journal*, 2009, *119* (2000), 1105–1129.

Almeida, Rita and Pedro Carneiro, "Enforcement of labor regulation and informality," *American Economic Journal: Applied Economics*, 2012, *4* (3), 64–89.

Bobba, Matteo, Luca Flabbi, and Santiago Levy, "Labor Market Search, Informality, and Schooling Investments," *International Economic Review*, 2022, *63* (1), 211–261.

_ , _ , _ , and Mauricio Tejada, "Labor Market Search, Informality, and on-the-Job Human Capital Accumulation," *Journal of Econometrics*, August 2021, *223* (2), 433–453.

Bosch, Mariano and Julen Esteban-Pretel, "Job Creation and Job Destruction in the Presence of Informal Markets," *Journal of Development Economics*, July 2012, *98* (2), 270–286.

Charlot, Olivier, Franck Malherbet, and Mustafa Ulus, "Efficiency in a search and matching economy with a competitive informal sector," *Economics Letters*, 2013, *118* (1), 192 – 194.

Dodge, Yadolah, ed., *Kolmogorov-Smirnov Test in The Concise Encyclopedia of Statistics*, Springer New York, 2008.

Eckstein, Zvi and Gerard van den Berg, "Empirical Labor Search: A Survey," *Journal of Econometrics*, February 2007, *136* (2), 531–564.

Esteban-Pretel, Julen and Sagiri Kitao, "Labor market policies in a dual economy," *Labour Economics*, 2021, *68*, 101956.

Fernández, Cristina and Leonardo Villar, "The Impact of Lowering the Payroll Tax on Informality in Colombia," *Economía*, 2017, *18* (1), 125–155.

Flinn, Christopher J and James J. Heckman, "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 1982, *18* (1), 115–168.

_ and James J Heckman, "Are Unemployment and out of the Labor Force Behaviorally Distinct Labor Force States?," *Journal of Labor Economics*, 1983, *1* (1), 28–42.

Granda, Catalina and Franz Hamann, "Informality , Saving and Wealth Inequality in Colombia," *IDB Working Papers*, 2015, *575*, 1–43.

Haanwinckel, Daniel and Rodrigo R Soares, "Workforce composition, productivity, and labour regulations in a compensating differentials theory of informality," *The Review of Economic Studies*, 2021, *88* (6), 2970–3010.

Kanbur, Ravi, "Conceptualising Informality: Regulation and Enforcement," *IZA Discussion Papers*, May 2009, *4186*, 1–18.

Kaplan, E. L. and Paul Meier, "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 1958, *53* (282), 457–481.

Kiefer, Nicholas M., "Economic Duration Data and Hazard Functions," *Journal of Economic Literature*, 1988, *26*, 646–679.

Meghir, Costas, Renata Narita, and Jean-Marc Robin, "Wages and Informality in Developing Countries," *American Economic Review*, April 2015, *105* (4), 1509–1546.

Moore, Mark A., Anthony E. Boardman, and Aidan R. Vining, "Social Discount Rates for Seventeen Latin American Countries: Theory and Parameter Estimation," *Public Finance Review*, January 2020, *48* (1), 43–71.

Narita, Renata, "Self-employment in developing countries: A search-equilibrium approach," *Review of Economic Dynamics*, 2020, *35*, 1–34.

Perry, Guillermo E., William F. Maloney, Omar S. Arias, Pablo Fajnzylber, Andrew D. Mason, and Jaime Saavedra-Chanduvi, *Informality : Exit and Exclusion* number 6730. In 'World Bank Publications.', The World Bank, 2007.

Ulyssea, Gabriel, "Firms, informality, and development: Theory and evidence from Brazil," *American Economic Review*, 2018, *108* (8), 2015–2047.

Table 1: Kolmogorov-Smirnov Test of Equality of Distributions

	Statistic	P-value	Statistic	P-value	Statistic	P-value
Alternative:	Two-Sided		Less		Greater	
H_0 Duration for Informal E and Informal SE are draw from the same distribution:						
Previous unemployment spell	0.0326	0.0177	0.0018	0.9857	0.0326	0.0089
Current employment spell	0.3869	0.0000	0.0000	1.0000	0.3869	0.0000
H_0 Labor Income for Informal E and Informal SE are draw from the same distribution:						
Current employment spell	0.1106	0.0000	0.1106	0.0000	0.0556	0.0000

Note: The "Two-Sided" alternative means that there are differences in the ECDF no matter where those differences are. The "Less" alternative means that the ECDF of the first sample always lies below that of the second sample. Finally the "Greater' alternative means that the ECDF of the first sample always lies above that of the second sample.

Table 2: Search Model Estimated Parameters

Parameters	Unrestricted			Restricted		
	Formal	Informal E	Informal SE	Formal	Informal E	Informal SE
	$v = f$	$v = i$	$v = s$	$v = f$	$v = i$	$v = s$
$\lambda(v)$	0.0897	0.0553	0.0529	0.097	0.0446	
	(0.0015)	(0.0012)	(0.0009)	(0.0015)	(0.0007)	
$\eta(v)$	0.0157	0.0317	0.0096	0.0159	0.0115	
	(0.0001)	(0.0006)	(0.0001)	(0.0001)	(0.0001)	
$\mu(v)$	0.2956	0.0071	-0.0500	0.2957	-0.0363	
	(0.0033)	(0.0055)	(0.0001)	(0.0068)	(0.0068)	
$\sigma(v)$	0.3145	0.3434	0.5048	0.3141	0.4723	
	(0.0099)	(0.0098)	(0.0058)	(0.0222)	(0.0142)	
b		-1.7205			-1.6053	
		(0.0839)			(0.0954)	
θ		0.0106			0.0160	
		(0.0048)			(0.0060)	
σ_ϵ		0.1196			0.1205	
		(0.0275)			(0.0509)	
Log-Likelihood		-470885.0			-480498.0	
LR Statistic		–			19226.0	
P-value		–			0.0000	

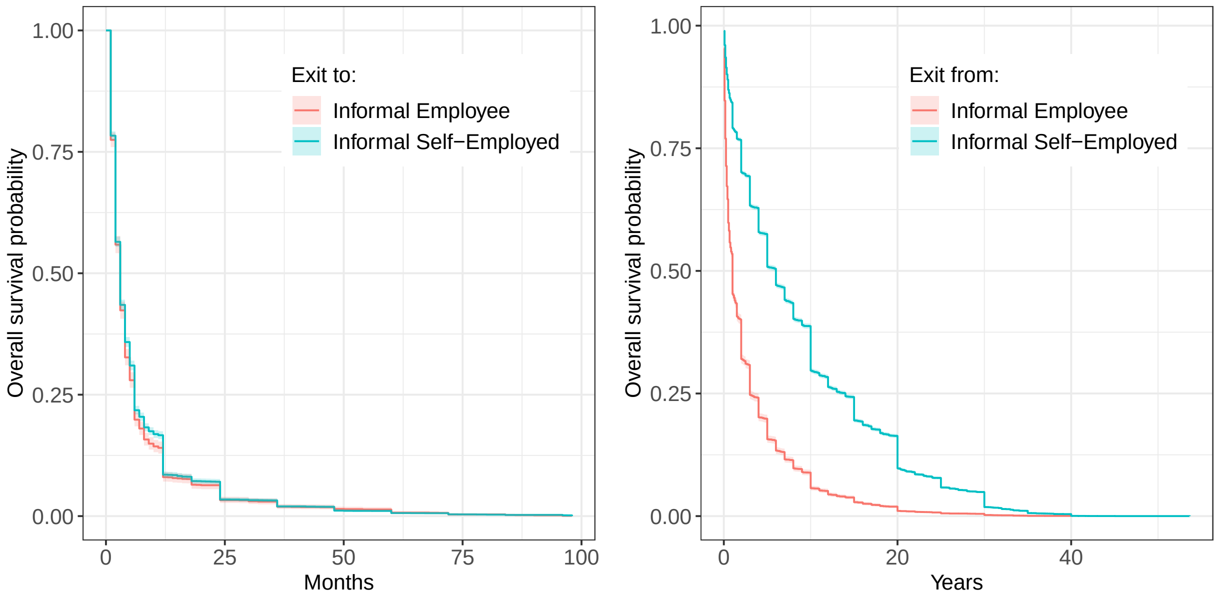
NOTE: Bootstrapped standard errors in parenthesis. The Restricted Model imposes the constraints $\lambda(s) = \lambda(i)$, $\eta(s) = \eta(i)$ and $\mu(s) = \mu(i)$, $\sigma(s) = \sigma(i)$. LR denotes the Log-likelihood Ratio Test.

Table 3: Search Model Implied Values

Values	Unrestricted			Restricted		
	Formal	Informal E	Informal SE	Formal	Informal E	Informal SE
	$v = f$	$v = i$	$v = s$	$v = f$	$v = i$	$v = s$
Employment:						
$E[t\|v]$	63.5	31.5	104.4	63.0	86.8	
$E[x\|v]$	1.412	1.068	1.080	1.412	1.078	
$SD[x\|v]$	0.207	0.143	0.339	0.207	0.291	
Unemployment:						
$E[t\|u]$		5.1			5.4	

NOTE: Values obtained from the point estimates reported in Table 2.

Figure 1: Kaplan-Meier Survival Functions



(a) Unemployment Duration conditional on Exit

(b) Employment Duration

Figure 2: Empirical CDF of accepted hourly wages