



Are informal self-employment and informal employment as employee behaviorally distinct labor force states?[☆]

Luca Flabbi^{a,b,*}, Mauricio M. Tejada^{c,d}

^a Department of Economics, University of North Carolina, Chapel Hill, USA

^b IZA, Bonn, Germany

^c Department of Economics, Universidad Diego Portales, Santiago, Chile

^d Millennium Nucleus for the Study of Labor Market Mismatch: Causes and Consequences, Santiago, Chile

ARTICLE INFO

Article history:

Received 8 June 2023

Received in revised form 24 July 2023

Accepted 27 July 2023

Available online 2 August 2023

Keywords:

Labor market frictions

Search and matching

Informality

Self-employment

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 an employee be considered two different labor market states? Both analyses strongly reject equality between the two states, cautioning against aggregating them in a common “informality state”. The parametric model identifies that the variation in informal self-employment income and the short duration of informal employee jobs are the primary factors that contribute to the observed differences between these labor market states.

© 2023 Elsevier B.V. All rights reserved.

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. 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 the debate by addressing its most controversial question: Should we differentiate between informal workers that are hired as employees and those that are working as self-employed? Aggregating or differentiating these two labor

market states is relevant for estimation and policy, but 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 unregistered employees and 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 as to consider self-employed informality as a searching state in alternative to unemployment. Several other contributions take one approach or the other.¹

We use data from Colombia to test whether the informal self-employed and the informal employees should belong to two different labor market states. We follow Flinn and Heckman (1983) in providing two types of analysis. First, we conduct non-parametric tests; then, we develop and estimate a search

[☆] We would like to thank Matteo Bobba and Santiago Levy, companions of many inspiring discussions on labor markets and informality. Tejada gratefully acknowledges financial support from FONDECYT, Chile Iniciación grant project No. 11196296 and ANID – MILENIO – NCS, Chile grant project No. 2022_045.

* Corresponding author at: Department of Economics, University of North Carolina, Chapel Hill, USA.

E-mail addresses: luca.flabbi@unc.edu (L. Flabbi), mauricio.tejada@udp.cl (M.M. Tejada).

URLs: <https://sites.google.com/site/lucaflabbi/> (L. Flabbi), <https://mauriciotejada.com> (M.M. Tejada).

¹ Among the contributions that do *not* differentiate between informal employees and self-employed are early contributions in the theoretical search literature (Albrecht et al., 2009; Charlot et al., 2013); and in the macro search literature (Bosch and Esteban-Prete, 2012). More recently, Haanwinckel and Soares (2021) develop a search model with intra-firm bargaining and exclude the self-employed. 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-Prete and Kitao (2021), which allows for only one informal labor market state, excluding the self-employed; 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 entrepreneurs and informal workers; Almeida and Carneiro (2012), which differentiates between informal wage earners and self-employed.

Table 1
Kolmogorov–Smirnov test of equality of distributions.

Statistic	P-value
H_0 : Informal E duration distribution = Informal SE duration distribution:	
Values	0.3869
H_0 : U duration distribution before Informal E = U duration distribution before Informal SE	
Values	0.0326
H_0 : Informal E labor income distribution = Informal SE labor income distribution	
Values	0.1106

E denotes employees, SE self-employed, and U unemployed.

model where we directly impose the same behavior for informal self-employed and informal employees.

The paper is organized as follows. Section 2 presents the data. Sections 3 and 4 provide the non-parametric and parametric analysis, respectively. 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 and collects labor market outcomes. It also allows for a precise definition of labor market informality. We define any employed worker as *informal* if they do not contribute to social security, a definition consistent with the recommendations of the International Labor Organization (ILO) 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 employees*; if they are occupied in an activity with more independence and autonomy so that they declare themselves self-employed,² we classify them as *informal self-employed*.

To be consistent with the theoretical model, we extract a relatively homogeneous estimation sample: 25–55-year-old men, living in urban areas, who have completed at most secondary education and work full-time when employed. We focus on male unskilled workers because they are the group in which labor market informality is the most relevant and studied. We pool together all the surveys conducted from January to December 2016.

The final estimation sample includes 88,123 observations, of which about 9% are unemployed, 51% informal – 39% self-employed and 12% employees –, and 40% formal employees. A full set of descriptive statistics is available in Flabbi and Tejada (2023). Two important differences between informal employees and informal self-employed emerge from simple descriptive statistics. First, the labor income distribution for the self-employed is much more disperse than the one for informal employees, with comparable means. Second, informal self-employed jobs last much longer than informal employee jobs.³

3. Non-parametric tests

To assess whether informal self-employment and informal employment as employees are behaviorally indistinguishable states, we follow Flinn and Heckman (1983) and test whether the

distributions describing the two states are equal. If they are, they should not be considered separate states; if they are not, they should. Given the data at our disposal, we can non-parametrically estimate at least two distributions pertaining to each state. The first is the distribution of durations in each of the two states. The second is the distribution of labor incomes in each of the two states. In addition to these two, we also compute the distribution of unemployment durations before transitioning to either the informal employee state or the informal self-employed state.

We non-parametrically estimate the duration distributions using the Kaplan–Meier survival function estimator (Kiefer, 1988; Kaplan and Meier, 1958) and the labor income distributions using the empirical cumulative distribution function.

Once estimated, we use the Kolmogorov–Smirnov (K–S) non-parametric test to assess equality of distributions. The null hypothesis is whether the data draws composing the two observed samples come from the same underlying population distribution (Dodge, 2008). The K–S test is based on the maximum difference between the empirical cumulative distribution functions of the two samples and therefore does not provide information on how the two distributions differ, but only on whether they differ or not.

The test results are presented in Table 1. The null hypothesis of equal duration distributions between informal employees and informal self-employed is largely rejected, as is the test of equal labor income distributions. The equality of the unemployment duration distributions before transitioning to either one or the other informal state is also rejected, but with a higher p -value, which corresponds to a 2% confidence level.

4. Parametric tests based on search model

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 non-parametric tests cannot distinguish whether the sources of the separation are the frictions and shocks affecting the labor market dynamics or the wage offers distributions affecting the agent's decisions to accept a job. To make progress in this understanding, we follow again 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 informal self-employed and informal employees 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 infinitely lived individuals with a discount rate ρ . Individuals can be in one of the following four states: unemployment u , informal self-employment s , formal employment f , and informal employment i . We denote by $v = s, f, i$ the different types of jobs in which an agent can be employed. 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 exogenous offer distributions $G(x|v)$. Jobs terminate exogenously at rate $\eta(v)$. Formal employees are different from informal workers because they pay a proportional payroll contribution τ . In exchange, they receive benefits that are valued at a flow utility that we denote with θ . Only unemployed individuals search for a job.

We denote with U and $E(x, v)$ the steady-state values of unemployment and employment, respectively, leading to the following

² *Trabajador por cuenta propia* in the original Spanish.

³ We should mention that while it is possible to be a *formal* self-employed worker, no one in our sample is.

Table 2
Search model estimated parameters.

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

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

Bellman equation representation:

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

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

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

The optimal decision for accepting a job offer possesses a reservation values property. The reservation labor income $x^*(v)$ for job v satisfies $E(x^*(v), v) = U$, leading to:

$$x^*(v) = \frac{\rho U - \theta \iota_{v=f}}{1 - \tau \iota_{v=f}} \quad (3)$$

which also implies $\theta = \tau \rho U$. In steady-state equilibrium, flows are balanced, implying $u\lambda(v)[1 - G(x^*(v)|v)] = \eta(v)u$.

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:

$$\begin{cases} \lambda(s) = \lambda(i) \\ \eta(s) = \eta(i) \\ G(x|s) = G(x|i) \end{cases} \quad (4)$$

4.2. Estimation

We estimate the model by maximum likelihood, obtaining contributions on durations $\{t_j\}_{j \in U,S,F,I}$ and labor incomes $\{x_j\}_{j \in S,F,I}$. The full derivation of the likelihood function is available in Flabbi and Tejada (2023). 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 (4).

The identification of the model is relatively straightforward. Durations provide direct information to identify hazard rates. Labor incomes identify wage offers distributions as long as they belong to a recoverable parametric distribution (Flinn and Heckman, 1982). Following previous literature, we assume that $G(x|v)$ are log-normal with parameters $\{\mu(v), \sigma(v)\}$. The flow utility of unemployment (b) and the discount rate (ρ) are jointly identified by the equilibrium equation (1). Therefore, we re-parameterize

the likelihood to estimate ρU directly and then set $\rho = 0.12$ – the recommended discount rate for Latin America (Moore et al., 2020) – to recover b . Finally, we set $\tau = 0.16$ to match the 2016 Colombian payroll contribution (Fernández and Villar, 2017) and recover θ from condition $\theta = \tau \rho U$.

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 parameters are allowed to be different across labor market states. The column *Restricted* presents a specification where we impose the set of constraints (4): this is a model where the parameters for the informal self-employed labor market state and the informal employee labor market state 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 an employee being behaviorally indistinguishable labor market states is strongly rejected. We have also estimated and tested two intermediate models, one in which we impose equality only on the mobility parameters ($\lambda(s) = \lambda(i)$; $\eta(s) = \eta(i)$) and one in which we impose equality only on the offer distributions parameters ($\mu(s) = \mu(i)$; $\sigma(s) = \sigma(i)$). Both models are clearly rejected against the Unrestricted model.⁴

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 standard deviation being more than double the employee standard deviation. 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 employees' one.

⁴ Results on the intermediate models are available in the Appendix of Flabbi and Tejada (2023).

Table 3
Search model implied values.

Values	Unrestricted			Restricted		
	Formal $v = f$	Informal E $v = i$	Informal SE $v = s$	Formal $v = f$	Informal E $v = i$	Informal SE $v = s$
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.

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 an employee be considered two different labor market states? Both analyses 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. The parametric model indicates that the sources of the difference are the high dispersion of informal self-employment income offers and the low duration of informal employee jobs.

Data availability

Data will be made available on request.

References

- Albrecht, J., Navarro, L., Vroman, S., 2009. The effects of labour market policies in an economy with an informal sector. *Econom. J.* 119 (2009), 1105–1129.
- Almeida, R., Carneiro, P., 2012. Enforcement of labor regulation and informality. *Am. Econ. J. Appl. Econ.* 4 (3), 64–89.
- Bobba, M., Flabbi, L., Levy, S., 2022. Labor market search, informality, and schooling investments. *Internat. Econom. Rev.* 63 (1), 211–261.
- Bobba, M., Flabbi, L., Levy, S., Tejada, M., 2021. Labor market search, informality, and on-the-job human capital accumulation. *Annals Issue: Implementation of Structural Dynamic Models, J. Econometrics Annals Issue: Implementation of Structural Dynamic Models*, 223 (2), 433–453.
- Bosch, M., Esteban-Pretel, J., 2012. Job creation and job destruction in the presence of informal markets. *J. Dev. Econ.* 98 (2), 270–286.
- Charlot, O., Malherbet, F., Ulus, M., 2013. Efficiency in a search and matching economy with a competitive informal sector. *Econom. Lett.* 118 (1), 192–194.
- Dodge, Y. (Ed.), 2008. Kolmogorov-Smirnov Test in the Concise Encyclopedia of Statistics. Springer New York.
- Esteban-Pretel, J., Kitao, S., 2021. Labor market policies in a dual economy. *Labour Econ.* 68, 101956.
- Fernández, C., Villar, L., 2017. The impact of lowering the payroll tax on informality in Colombia. *Economía* 18 (1), 125–155.
- Flabbi, L., Tejada, M., 2023. Are Informal Self-Employment and Informal Employment as Employee Behaviorally Distinct Labor Force States? IZA Discussion Paper Series 16225, pp. 1–35.
- Flinn, C.J., Heckman, J.J., 1982. New methods for analyzing structural models of labor force dynamics. *J. Econometrics* 18 (1), 115–168.
- Flinn, C.J., Heckman, J.J., 1983. Are unemployment and out of the labor force behaviorally distinct labor force states? *J. Labor Econ.* 1 (1), 28–42.
- Granda, C., Hamann, F., 2015. Informality, Saving and Wealth Inequality in Colombia. IDB Working Papers 575, pp. 1–43.
- Haanwinckel, D., Soares, R.R., 2021. Workforce composition, productivity, and labour regulations in a compensating differentials theory of informality. *Rev. Econom. Stud.* 88 (6), 2970–3010.
- Kanbur, R., 2009. Conceptualising Informality: Regulation and Enforcement. IZA Discussion Papers 4186, pp. 1–18.
- Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53 (282), 457–481.
- Kiefer, N.M., 1988. Economic duration data and hazard functions. *J. Econ. Lit.* 26, 646–679.
- Meghir, C., Narita, R., Robin, J.M., 2015. Wages and informality in developing countries. *Amer. Econ. Rev.* (ISSN: 0002-8282) 105 (4), 1509–1546.
- Moore, M.A., Boardman, A.E., Vining, A.R., 2020. Social discount rates for seventeen Latin American countries: theory and parameter estimation. *Public Finance Rev.* 48 (1), 43–71.
- Narita, R., 2020. Self-employment in developing countries: A search-equilibrium approach. *Rev. Econ. Dyn.* 35, 1–34.
- Perry, G.E., Maloney, W.F., Arias, O.S., Fajnzylber, P., Mason, A.D., Saavedra-Chanduvi, J., 2007. Informality: Exit and Exclusion. In: World Bank Publications, (6730), The World Bank.
- Ulyssea, G., 2018. Firms, informality, and development: Theory and evidence from Brazil. *Amer. Econ. Rev.* 108 (8), 2015–2047.