DEPARTMENT OF ECONOMICS
WORKING PAPER SERIES

UMBC

AN HONORS UNIVERSITY IN MARYLAND

Working Paper 25-01, economics.umbc.edu/working-papers

Job transitions and involuntary informality in Costa Rica

T.H. Gindling
University of Maryland, Baltimore County, and IZA

Catherine Mata
Brown University

Diego Rojas*
University of Maryland, Baltimore County

April, 2025

Job transitions and involuntary informality in Costa Rica

T.H. Gindling
University of Maryland, Baltimore County, and IZA

Catherine Mata
Brown University

Diego Rojas*
University of Maryland, Baltimore County

Abstract

Informal work is often considered a form of employment for marginalized and vulnerable workers who have been rationed out of the preferred formal work. However, informality can also be seen as a dynamic sector voluntarily chosen by budding entrepreneurs and those looking for flexible working conditions. This paper tests the voluntary and involuntary nature of informal work in Costa Rica. It considers the possibility that workers value not only expected earnings but also the possibility of transitioning into other sectors or into unemployment. The results suggest that considering the probability of moving across sectors and into unemployment is relevant when measuring the share of involuntary informality. In the case of Costa Rica, when accounting for these transition probabilities, more than 90 percent of all informal workers are involuntarily informal; however, when only expected earnings are considered, the share of involuntarily informal work falls to 66 percent.

Keywords: informal work, involuntary informality, job transitions, segmentation, wage differentials, developing economies, finite mixture model.

JEL classification: J46, J60, J42, J31, J21, O17

---

*Authors in alphabetical order. **Gindling**: Department of Economics and School of Public Policy, University of Maryland, Baltimore County, and Institute for the Study of Labor (IZA) (email: gindlin@umbc.edu). **Mata**: Annenberg Institute, Brown University (email: catherine_mata@brown.edu). **Rojas**: School of Public Policy, University of Maryland, Baltimore County. PUP Bldg, 4th Floor. 1000 Hilltop Circle. Baltimore, MD 21250 (email: lrojasa1@umbc.edu). The authors would like to thank Salem Abo-Zaid, Oleg Firsin, Yusuke Kuwayama, and Guillermo Vuletin for their valuable comments and insightful suggestions on earlier drafts of this article. An earlier version of this work received funding as part of the UNU-WIDER project "Transforming Informal Work and Livelihoods". Institutional support was provided by the University of Maryland, Baltimore County, and the Annenberg Institute at Brown University.

Job transitions and involuntary informality in Costa Rica

1. Introduction

By a variety of measures, a large proportion of workers in developing economies are informal. Informal work is often considered a type of employment for marginalized and vulnerable workers who have been rationed out of preferred formal work. However, informality can also be seen as a dynamic sector voluntarily chosen by budding entrepreneurs and those looking for flexible working conditions. Any analysis of informality must recognize this heterogeneity and differentiate between workers who are informal due to a lack of formal employment opportunities and those who are working in the informal sector voluntarily because of their comparative advantages or preferences. In this paper, we build on Günther and Launov (2012) to test the voluntary and involuntary nature of informal work in Costa Rica. Expanding their work, we consider the possibility that workers value not only expected earnings but also the probabilities of transition between sectors or into unemployment.

From a policy point of view, the nature of informality is relevant when determining the levels of welfare and production. As measured here, involuntarily informal workers have greater levels of expected welfare if they are able to move freely across different sectors of the labor market. A lack of mobility results in a mismatch between workers and jobs, reducing the levels of production and welfare (see, for example, Meghir *et al*., 2015). Our methodology is well suited to finding evidence of such frictions. If such evidence is present, a series of policies aiming to facilitate mobility could be considered to increase the levels of production and welfare (see, for example, Alaimo *et al*., 2015). These policies would aim to reduce the barriers to entering the formal sector. In contrast, if most informal workers are voluntarily informal, informality might instead largely reflect a problem of taxation/evasion, and measures related to changes in the tax regime, social security contributions, and/or enforcement would be more relevant.

The nature of informality is also important when considering the effects of public policies and, in particular, when evaluating the welfare effects of unemployment benefits (UBs). UBs help workers to smooth their consumption when laid off (insurance value), but they might delay re-employment (efficiency costs). The tradeoff between the insurance value and the efficiency costs of employment determines the welfare effects of UBs.[1] According to Gerard and Gonzaga (2021), the voluntary or involuntary nature of informal work has opposing implications for the value of insurance: The value of insurance will be lower if working informally is easier and informal jobs are close substitutes for formal jobs (i.e., informal work is voluntary) and will be higher if finding a formal job is more difficult and informal jobs are a poor means of self-insurance (i.e., if informal work is involuntary). Results from Gerard and Gonzaga (2021) in Brazil and Liepmann and Pignatti (2024) for Mauritius highlight the importance of the insurance value of UBs over their efficiency costs in contexts of high informality and, therefore, the importance of identifying the nature of informal work.[2]

Informal work can be involuntary due to labor market segmentation. In the labor market segmentation theory of dualistic labor markets, formal wages are institutionally set at higher than equilibrium (market) levels. Institutional mechanisms for maintaining above-market levels of

[1] See Baily (1978) and Chetty (2006).

[2] Gerard and Gonzaga (2021) find low efficiency costs in a context of high informality in Brazil, suggesting that attention should be shifted to the insurance value of UBs and, thus, towards the nature of informal work. In concordance, Liepmann and Pignatti (2024) find that the insurance value of UBs dominates the efficiency costs (for coefficients of risk aversion exceeding two) even in the presence of high informality in Mauritius. They also find that formally displaced workers find informal employment rapidly, but they perceive lower earnings, have reduced consumption, and do not return to formal jobs even long after the end of UB eligibility. Thus, their findings suggest that informal jobs are poor substitutes for formal ones, i.e., informal work tends to be involuntary.

formal wages include minimum wage enforcement, the market power of large formal sector firms, collective bargaining (unions), and public sector wage policies. The vast majority of informal workers want high-wage formal work, but not all are able to find formal work because formal sector jobs are limited (Fields, 1975; Harris and Todaro, 1970). Informal wages and employment are not subject to the regulations and other institutional mechanisms that maintain higher-than-equilibrium wages in the formal sector. Because these institutional mechanisms do not apply, wages are set at the (market) equilibrium level, which is low due to the artificially high supply resulting from the limited ability of workers to move to formal jobs. There is no limit on the number of informal jobs. Informal workers who are not able to find formal work can obtain informal employment, although at low wages. In this case, most informal workers are informal involuntarily because they are unable to obtain one of the limited high-wage formal jobs.

Others argue that labor market segmentation is not the reason that workers choose informal work (e.g., Maloney, 1999). In this view, formal wages are not set above equilibrium, formal employment is not limited, and workers are able to freely move between formal and informal work. In this argument, workers choose informality voluntarily because of their comparative advantage or preferences. These workers value the flexibility of working conditions in informal work, can avoid the costs of formalization such as social security payroll taxes and other mandatory taxes, or are entrepreneurs who find the government regulations (or acceptance of corruption) needed to start a new business too costly (De Soto, 1989). In this view, earnings may be higher in informal work than in formal work for some workers (for example, those with a comparative advantage in informal work or those who are informal to avoid taxes or costly regulations).

More recent views of dualistic labor markets recognize that voluntary and involuntary informality coexist and distinguish between those informal workers who are voluntarily informal and those who are involuntarily informal due to limited access to the formal employment that they would prefer (Fields, 1990; Günther and Launov, 2012; La Porta and Shleifer, 2014; Maloney, 2004; Meghir et al., 2015; Ulyssea, 2018 and 2020). Fields (1990) refers to the former as upper-tier informal workers and the latter as lower-tier informal workers. While both subsegments are informal, the substantial human capital and financial capital requirements in upper-tier informality imply that there is no free mobility from lower-tier informal work into upper-tier informal work.[3]

As both voluntary and involuntary informal work coexist, our methodology aims to estimate the share of each type of informal work by building on the work of Günther and Launov (2012). Our contribution to Günther and Launov's (2012) strategy is to incorporate preferences for leisure and, more relevant, the probability of unemployment in the value function of the worker. In their paper, the share of involuntarily informal workers consists of those whose expected wage in the informal sector is below that in the formal sector. In our case, low probabilities of unemployment can be a desired feature of formal work. If wage differences are small enough, the higher odds of keeping a formal job can dominate the wage differentials.

In this paper, we apply our methodology to the case of Costa Rica, an upper-middle income country. Costa Rica has one of Latin America's most formal labor markets, making it a good candidate for finding a relatively large proportion of voluntary informality. To find evidence of whether workers are voluntarily or involuntarily informal in Costa Rica, we use data sets collected by the Costa Rican Statistics and Census agency.

[3] A second type of heterogeneity in the informal sector is between those who are informal employees in firms and those who are self-employed. Self-employed workers may also be either voluntary or involuntary. Because the structure and definition of earnings differ between self-employment and wage employment, we examine self-employment and wage employment separately. Following Günther and Launov (2012), we empirically identify different subsegments within informality and test for the voluntary and involuntary nature of workers in each.

Our results suggest that the probability of moving across sectors and into unemployment should be considered when measuring the share of involuntary informality. In the case of Costa Rica, when incorporating these transition probabilities, more than 90 percent of all informal workers are involuntarily informal, but when only expected earnings are considered, the share of involuntary informal work falls to 66 percent.

2. Involuntary or voluntary informality?

This section presents the methodology used to identify the voluntary or involuntary nature of informality. The main goal is to estimate the share of workers who are in the informal sector when they would be better off in the formal sector, whom we label as "involuntarily informal". Workers who are informal and expect to be better off in the informal sector are categorized as "voluntarily informal". We do not only consider the difference in expected earnings when estimating the share of workers in each of these categories, but we also account for preferences for leisure and the possibility of unemployment in the worker's value function.

Specifically, let V_I, V_F, and V_U represent the value a worker places on holding an informal job, a formal job, or being unemployed, respectively. For simplicity, let us assume that unemployment does not give any (instantaneous) utility. Moreover, for now, let us assume that the utility received for a job, in either the formal or informal sectors, is determined only by its wage. The possibility that the informal sector brings some extra utility due to leisure is considered below.

Let us represent the probability of transitioning from segment i to segment j of the labor market as $\pi_{i,j}$. For example, let $\pi_{I,U}$ be the probability of transitioning from an informal job into unemployment and $\pi_{F,U}$ the probability of transitioning from a formal job into unemployment. Similarly, $\pi_{I,I}$, $\pi_{F,F}$, and $\pi_{U,U}$ represent the probability of keeping an informal job, a formal job, or staying unemployed, respectively.[4] Lastly, let $\beta = 1/(1+r)$ be the discount factor, where r represents the real interest rate.

When determining what type of job is preferred, we focus on wage differences between formal and informal workers and on the transition probabilities from these sectors to unemployment, while paying less attention to short-run fluctuations in economic activity. Since we are less concerned about short-run fluctuations, we consider only the steady-state value of being employed or unemployed. More specifically, the value of an informal job is given by the following steady-state Bellman equation:

$$V_I = w_I + \beta \, \{\pi_{I,I} V_I + \pi_{I,F} V_F + \pi_{I,U} V_U\},$$

that is, the value of having an informal job V_I is given by the informal wage w_I plus the continuation value or expected value of "moving" to a different sector. The continuation value is given by the present expected value of staying in the informal job, moving to a formal job, or losing the job and being unemployed. Similarly, the steady-state value of a formal job is

$$V_F = w_F + \beta \, \{\pi_{F,I} V_I + \pi_{F,F} V_F + \pi_{F,U} V_U\},$$

whereas the value of being unemployed is given by

$$V_U = 0 + \beta \, \{\pi_{U,I} V_I + \pi_{U,F} V_F + \pi_{U,U} V_U\}.$$

The last equation assumes that the utility of unemployment is zero.[5] These three equations can be used to solve for the steady-state values of V_I, V_F, and V_U in terms of wages, transition

[4] Empirically, the probability of losing a formal job is less than that of losing an informal job, that is, $\pi_{F,U} < \pi_{I,U}$ (see Fields *et al.*, 2023). This thus increases the desirability of a formal job (for given wages).

[5] Note that $\pi_{I,I} + \pi_{I,F} + \pi_{I,U} = 1$, $\pi_{F,I} + \pi_{F,F} + \pi_{F,U} = 1$, and $\pi_{U,I} + \pi_{U,F} + \pi_{U,U} = 1$.

probabilities, and the discount factor (or interest rate). After solving for V_I and V_F, it can be shown that a formal job is preferred (i.e., $V_F > V_I$) whenever[6]

$$(1)\ \frac{w_F}{w_I} > \frac{r+1-\pi_{U,U}+\pi_{F,U}}{r+1-\pi_{U,U}+\pi_{I,U}}.$$

According to this relationship, for a given wage ratio w_F/w_I, the desirability of a formal job increases with $\pi_{I,U}$ and decreases with $\pi_{F,U}$, that is, because the probability of losing the job is greater in the informal sector, the formal sector is more desirable.[7] If the probability of losing a formal job is small and the probability of losing an informal job is large, formal jobs are more desirable: Jobs with a lower probability of leading to unemployment are preferred when wages are equal. If a job offers greater wages, some probability of unemployment can be tolerated in exchange, but if formal jobs are more stable and pay higher wages, they will be strictly preferred.[8]

In addition to high wages (w) and a low probability of unemployment, workers also value leisure. To factor leisure into the worker's value function, let us now assume that a job's utility depends on both consumption (C) and hours worked (L) according to the following utility function:

$$U = \sigma_u \left[C - \frac{L^{1+\frac{1}{\psi_L}}}{1+\frac{1}{\psi_L}} \right]^{1/\sigma_u},$$

where $\sigma_u > 0$. If we impose the restriction $C = wL$, the utility function is maximized when $L =$

---

[6] See Appendix B for details.

[7] In the case of a segmented market where informal workers can only find informal jobs (and formal workers only find formal jobs), the values for having an informal job and for being unemployed are given by $V_I = w_I + \beta\{\pi_{I,I}V_I + \pi_{I,U}V_U\}$ and $V_U = 0 + \beta\{\pi_{U,I}V_I + \pi_{U,U}V_U\}$, respectively. Note that only movements between unemployment and the informal sector are allowed for informal workers because we are assuming that the market is segmented. Under segmentation, a formal job is preferred whenever

$$\frac{w_F}{w_I} > \frac{1+\frac{\pi_{F,U}}{r+\pi_{U,F}}}{1+\frac{\pi_{I,U}}{r+\pi_{U,I}}}.$$

Note that, unlike in the non-segmented labor market, the desirability of a job depends on the probability of leaving unemployment.

[8] Unlike the segmented labor market case, the desirability of one sector over the other does not depend on the different probabilities of leaving unemployment. A formal worker who loses her job can find a formal or informal job with the same probability as an informal worker who just lost her job. In a segmented labor market, when $\pi_{U,I} > \pi_{U,F}$, informal workers are unemployed for a shorter period; therefore, losing an informal job is less painful. In this case, when a formal or informal worker loses her job, she can find a job with the same probability.

w^{ψ_L} and, thus, given by

$$U(w) = \sigma_u \left[\frac{w^{1+\psi_L}}{1+\psi_L}\right]^{1/\sigma_u}.$$

When we assume that the utility is given by this relationship, instead of being equal to w, the steady state value of being in sector $k = F, I, U$ is given by

$$V_k = U(w_k) + \beta\{\pi_{k,I} V_I + \pi_{k,F} V_F + \pi_{k,U} V_U\},$$

where we assume that $w_U = 0$. Note that this is the same system of equations presented above except w_k is changed to $U(w_k)$. Thus, by substituting w_k with $U(w_k)$ in the above value functions, the formal job is preferred whenever

(2) $(1 + \psi_L) \ln w_F - \sigma_u \ln (r + 1 - \pi_{U,U} + \pi_{F,U}) > (1 + \psi_L) \ln w_I - \sigma_u \ln (r + 1 - \pi_{U,U} + \pi_{I,U})$,

or equivalently, when

(3) $\ln w_F > \ln w_I + \frac{\sigma_u}{1+\psi_L} \ln \left(\frac{r+1-\pi_{U,U}+\pi_{F,U}}{r+1-\pi_{U,U}+\pi_{I,U}}\right)$.

Relative to Günther and Launov (2012), this formulation not only considers the difference in earnings ($\ln w_F - \ln w_I$) but also includes the probability of losing a job and the preference for leisure, represented by the Frisch elasticity of substitution ψ_L.[9]

The aim of this paper is, thus, to evaluate equation (2) empirically. To that end, the value of σ_u is taken from the literature, while panel data are used to estimate the Frisch elasticity of wages ψ_L and the transition matrices between formality, informality, and unemployment, thereby setting values for $\pi_{U,U}$, $\pi_{F,U}$, and $\pi_{I,U}$. Finally, estimates of expected wages in both the formal and informal sectors are obtained using Günther and Launov's (2012) strategy, i.e., estimating a finite mixture

[9] Importantly, condition (2) is derived assuming that the labor market is not segmented. Thus, the failure of this equation to correctly assign workers is interpreted as evidence of labor market segmentation.

model (FMM) with self-selection (see Appendix A for details).[10]

This methodology allows us to estimate the worker's expected wage in a specific segment as a function of her individual characteristics. This expected wage is then combined with values for ψ_L and the transition probabilities $\pi_{U,U}$, $\pi_{U,U}$, $\pi_{U,U}$ to evaluate condition (2). If condition (2) holds for an informal worker, i.e., she prefers a formal job, then she is considered to be involuntarily informal.

Our contribution to Günther and Launov's (2012) strategy is, thus, to consider the preferences for leisure and the transition probabilities in the worker's value function. In their paper, the share of involuntarily informal workers consists of those whose estimated wage in the informal sector is below that in the formal sector. In our case, low probabilities of unemployment can be a desired feature of formal work. If wage differences are sufficiently small, low chances of unemployment can dominate wage differentials.[11]

3. Data and definitions

The data used in this analysis are from the Costa Rica National Household Surveys (Encuesta Nacional de Hogares, ENAHO). These household surveys are conducted annually by the Costa Rican National Statistics and Census Institute. The ENAHO provides information on employment, income, poverty, education, dwelling conditions, and access to public services, among others. The data are nationally representative.

[10] If our estimates indicate that an informal worker would prefer a formal job, this worker is labeled as "involuntarily informal". Similarly, if condition (2) indicates that an informal worker prefers an informal job, that worker is labeled as "voluntarily informal".
[11] An additional difference is that we estimate the model using full-information maximum likelihood instead of two steps to gain efficiency. See Appendix A for details on the estimation strategy.

We limit our sample to people of working age (15–65), who work more than 15 hours per week, and who do not report being full-time students in any pair of consecutive years in each panel. Wages refer to hourly payments received in the main occupation after taxes and social security contributions.

Workers' transition matrices are calculated from panels of individuals constructed for every pair of years between 2011 and 2018. The ENAHO uses a rotating sample design whereby in a given year, the interviewers return to approximately 75 percent of the households interviewed in the previous year. The Institute has constructed the seven year-to-year panel data sets of households and individuals we are using here (2011–2012, 2012–2013, 2013–2014, 2014–2015, 2015–2016, 2016–2017, and 2017–2018). Each year, 25 percent of households are replaced in the sample; this implies that we will be able to follow, at most, 75 percent of households from year to year, although in practice, our sample is smaller. The year-to-year panels include 37 percent of all individuals between the ages of 15 and 65 who were interviewed in the ENAHO from 2011 to 2018.

Following the International Labor Organization (ILO) Thesaurus, our framework for identifying formal and informal workers is based on whether regulations and mandatory labor protections are complied with. Formal employers and workers are those who comply with all registration requirements and labor protections, while informal workers are all others. In addition, we distinguish between wage employees and the self-employed.[12]

Self-employed workers are those who self-identify as own-account workers or owners of firms (employers). For self-employed workers to fully comply with the law in Costa Rica, they

[12] We use the terms salaried worker, wage employee, and wage employed interchangeably.

must both pay into social security and be registered. We identify formal self-employed workers as those who follow all regulations, specifically, those who both contribute to social security and are registered. Workers are identified as registered if they are registered in the National Records or other public institutions or keep formal accounts for reporting to the government.[13]

4. *Estimating ψ_L, transition probabilities, and expected wages*

Using the household surveys described in the last section, our goal is to estimate (2). To estimate (2), we need values for σ_u, r, ψ_L, expected wages, and the probabilities of moving into unemployment. We set the value of σ_u to one-third, following the international evidence presented in Vegh (2013). The interest rate is fixed at $r = 0.02$. Aside from these two parameters, the three main components of equation (2) are the Frisch elasticity of substitution ψ_L, the transition probabilities $\pi_{U,U}$, $\pi_{F,U}$, $\pi_{I,U}$, and the expected values of (log) wages in the formal and informal sectors, $\ln w_F$ and $\ln w_I$. In this section, baseline values for each of these components are presented, and then condition (2) is evaluated. Later, in section 5, we analyze how sensitive the main results are to changes in these baseline values.

[13] Although ENAHO does not inquire directly about registration with the Ministry of Finance, it does inquire about registration to national records—which is more common for bigger employers—and any other public entity, which would include the Ministry of Finance. Also, those keeping formal accounting books are likely to do so for tax purposes.

Table 1. Frisch Elasticity of Substitution

Depend. Var: Δ ln worked hours	(1)	(2)	(3)	(4)
Δ ln wage per hour	0.158∗∗∗	0.160∗∗∗	0.309∗∗∗	0.307∗∗∗
	(0.050)	(0.051)	(0.038)	(0.038)
Constant	-0.007∗∗∗	-0.009∗	-0.010∗∗∗	-0.009
	(0.002)	(0.005)	(0.002)	(0.006)
First stage				
Age	-0.008∗∗∗	-0.008∗∗∗	-0.006∗∗∗	-0.007∗∗∗
	(0.002)	(0.002)	(0.002)	(0.002)
Age2/100	0.008∗∗∗	0.008∗∗∗	0.006∗∗	0.006∗∗
	(0.003)	(0.003)	(0.003)	(0.003)
ln firm size	0.018∗∗∗	0.018∗∗∗		
	(0.002)	(0.002)		
Δ ln firm size			0.049∗∗∗	0.050∗∗∗
			(0.003)	(0.003)
Δ years of education	0.011∗∗∗	0.010∗∗∗	0.011∗∗∗	0.011∗∗∗
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.151∗∗∗	0.131∗∗∗	0.176∗∗∗	0.157∗∗∗
	(0.046)	(0.04)	(0.046)	(0.046)
Year dummies	No	Yes	No	Yes
F- First Stage	28.24	14.65	69.44	31.45
P-val overid. restr.	0.594	0.584	0.109	0.111
Observations	29 981	29 981	29 818	29 818

Source: Author's analysis based on data from the Costa Rica National Household Surveys (ENAHO), 2011–2018.
Notes: Δ denotes the change in a variable between year *t* and *t + 1*. The instrumented variable is the change in Ln wage per hour. The model is estimated by 2SLS. Statistical significance: ∗ p<0.1, ∗∗ p<0.5, ∗∗∗ p<0.01.

4.1 Estimating ψ_L

The Frisch elasticity of substitution is estimated following a methodology similar to that in MaCurdy (1981).[14] Specifically, ψ_L is the parameter resulting from regressing the change in log hours worked against the change in log wages (or earnings). As in the literature, the change in log wages is instrumented by the change in years of education, age, and age squared. As an additional instrument, we include the log of the firm size (or its change).[15]

The estimates of ψ_L are close to 0.16 when the (log of) firm size is used as an additional instrument. When the change in (the log of) firm size is used, this coefficient is approximately 0.31. Our estimates fall within the range found in the literature, which usually varies between 0.1 and 0.6.[16] As a baseline, we use the estimates presented in regression (3) of Table 1.[17]

4.2 Estimating transition probabilities

The second main component of equation (2) is given by the transition probabilities $\pi_{U,U}$, $\pi_{F,U}$, and $\pi_{I,U}$. Table 2 presents the transition matrices between formality, informality, unemployment, and out-of-labor-force in Costa Rica using panel data from 2011 to 2018. Note that formal work is the most stable sector: Approximately 83 percent of workers keep a formal job from one year to the next, compared to 63 percent of informal workers. This stability influences the probability of being unemployed or out of the labor force. While approximately 3 percent of those with a formal job are unemployed one year later, approximately 5 percent of informal workers are.

[14] This approach is commonly used in the literature. See, for example, Altonji (1986), Keane and Rogerson (2012), Lee (2001), Keane and Neal (2023), Reynaga and Rendon (2012), among others.
[15] See MaCurdy (1981), Keane and Rogerson (2012), Lee (2001), and Keane and Neal (2023). Reynaga and Rendon (2012) propose the use of firm size as an additional instrument.
[16] See Keane and Rogerson (2012), Lee (2001), Keane and Neal (2023), Chetty (2012) and Chetty *et al*. (2013).
[17] Additionally, we also included the inverse of the mills ratio in the first stage to account for self-selection, however, we found similar results. The inverse mills ratio was estimated by a probit model that included age, age squared, gender, living in the central region (standing alone and interacted with females), being head of the household, the number of children under 12 in the household (standing alone and interacted with female), and year fixed effects.

The odds of moving out of the labor force are even more striking: While only 3 percent of those in a formal job are out of the labor force one year later, approximately 17 percent of informal workers leave the labor force from one year to the next.

Also note that a slightly greater proportion of those unemployed move to informal jobs than to formal employment (28 vs. 24 percent). However, it is most common to enter the job market through informality.[18] While 15 percent of individuals out of the labor force enter the labor force through an informal job, only 3.6 percent do it through a formal job. Thus, taking together unemployment and those out of the labor force, the informal sector absorbs more people without employment than the formal sector.

Table 2. Transition Matrix

Transitions between *t* (row) and *t* + 1 (column)				
	Formal	Informal	Unemployed	Out of Labor Force
Formal	0.8332	0.1073	0.0282	0.0313
Informal	0.1529	0.6288	0.0498	0.1685
Unemployed	0.2376	0.2835	0.2246	0.2542
Out of Labor Force	0.0366	0.1470	0.0531	0.7634

Source: Author's analysis based on data from the Costa Rica National Household Surveys (ENAHO), 2011–2018.
Notes: Rows reflect the work status in *t*, while columns indicate the work status one year later in *t* + 1.

In summary, the probability of staying in a formal job from one year to the next is greater than that of staying in an informal job. Similarly, the probability of being without employment in the following period is greater for those in the informal sector. These probabilities increase the desirability of formal jobs over informal jobs.

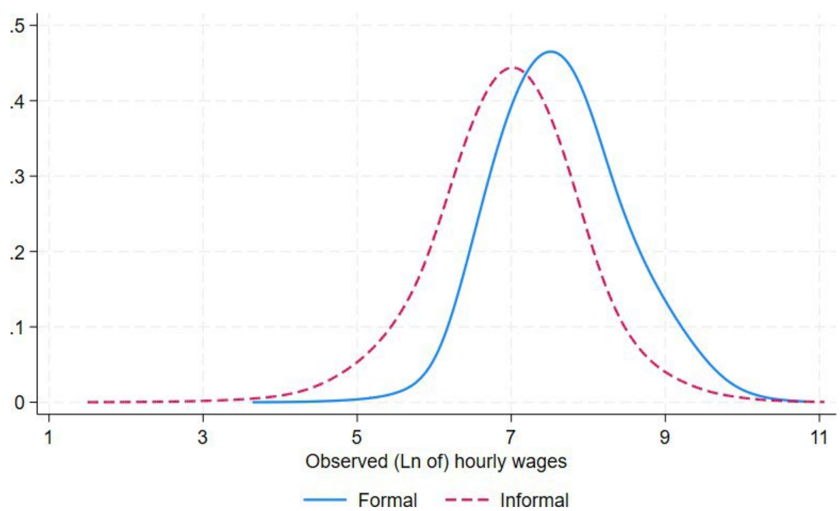
[18] Arias *et al* (2018) find similar results for Brazil and Mexico when accounting for mobility costs.

4.3 Estimating expected wages

Given the estimates presented in Tables 1 and 2 for ψ_L, $\pi_{U,U}$, $\pi_{F,U}$ and $\pi_{I,U}$, we are only missing expected log wages to calculate equation (2). Expected wages will be estimated using the FMM strategy developed by Günther and Launov (2012). In this section, only two observed segments are considered: one for formal workers and one for informal workers. This assumption is relaxed below, where we extend the model to include unobserved segments within informality (see Appendix A for details).

The average log wage across workers is illustrated in Figure 1. This figure presents the distribution of observed wages for formal and informal workers in Costa Rica. Not surprisingly, the wage distribution for formal workers tends toward the right, showing that on average, formal workers earn more than informal workers. However, the wage distributions of formal and informal workers overlap, suggesting that some workers may earn more in informal jobs than they would in formal employment.

Figure 1: Densities of Observed Ln of Hourly Wages for Formal and Informal Workers



Source: Authors' calculations using panel data from the Costa Rica National Household Surveys (ENAHO). Notes: Income in colones as of June 2015. Pooled data of all workers between 15 and 65 years old who are included in the panels from 2011–2012 to 2017–2018.

Considering together the data on transition probabilities and the above distribution of wages, we can expect the average formal job to be more desirable than the average informal job. Not only do formal jobs offer on average higher wages, but they also provide a greater likelihood of maintaining the same type of employment. However, we are interested in estimating involuntary informality across all workers and not only for the average worker. Not all workers earn more in the formal sector. Expected wages in each sector depend on different individual characteristics. To consider these characteristics, we estimate wages as a function of different explanatory variables.

As explanatory variables of (log) wages, we include age, age squared, years of education, and gender and indicators for whether the person lives in the central region of the country, whether they are fluent in English as a foreign language, whether they are professionals,[19] and whether they migrated to Costa Rica from Nicaragua. The selection equation also includes age, age squared, a dummy variable for women, an indicator of people living in the central region of the country, and an indicator of whether a person migrated from Nicaragua. Additionally, an indicator of whether a person is the head of the household and the number of children in the household under 12 years old (interacted with gender) are used as instruments.[20] Year fixed effects are included in all regressions.

In our simple model, we assume the existence of only two segments: formality and informality.

[19] Professionals or professional experience is defined as being employed as a technician, associate professional, or clerical support worker in the current or previous year.

[20] The number of children is a common instrument used in the literature for the selection into work. See, for example, Huber and Mellace (2014), Mulligan and Rubinstein (2008), Schaffner (2002), Wellington (1993), and Deolalikar (1993). The intuition is that employers do not have (reliable) information or do not consider household composition when offering a wage. Under the same logic, being the head of the household can also be considered as a variable excluded from the wage equation.

That is, we assume that the formal and informal sectors represent the only two distinctive segments of the labor market, and thus, the returns to each characteristic of the worker are potentially different in each of these two segments.[21] The results from the FMM model with one formal and one informal sector are presented in Table 3.

The regressions represented in Table 3 can now be used to estimate expected (log) wages for each worker in the formal and informal sectors based on their individual characteristics. Assuming that there is no segmentation between the formal and informal sectors, a worker would choose the sector with the highest value according to condition (2). In the next section, we use this condition to estimate the share of workers who are expected to prefer a formal to an informal job and, thus, classify informal workers into voluntarily or involuntarily informal workers.

[21] Below, we extend the model to consider the possibility of heterogeneous segments within the informal sector (as in Günther and Launov, 2012).

Table 3. Finite Mixture Model for (Ln of) Hourly Wage

	Formal	Informal	Selection
Age	0.0303***	0.0129**	0.150***
	(0.00331)	(0.00549)	(0.00301)
Age^2/100	-0.0204***	-0.0104	-0.202***
	(0.00419)	(0.00680)	(0.00368)
Years of Education	0.104***	0.0547***	
	(0.00100)	(0.00244)	
Female	0.0472**	-0.0618*	-0.942***
	(0.0188)	(0.0320)	(0.0148)
Central Region	-0.00271	0.128***	0.257***
	(0.00805)	(0.0154)	(0.0119)
English	0.0909***	0.250***	
	(0.0235)	(0.0605)	
Central × English	0.115***	0.250***	
	(0.0272)	(0.0770)	
Professionals	0.0210***	0.327***	
	(0.00791)	(0.0256)	
Nicaraguan	-0.106***	0.0342	0.0735***
	(0.0136)	(0.0229)	(0.0230)
Ln σ_j	-0.588***	-0.188***	
	(0.0111)	(0.0129)	
π_j	0.651***	0.349***	
	(0.00251)	(0.00251)	
Head of Household			0.690***
			(0.0143)
Children under 12			0.0531***
			(0.0111)
Children under 12 × Female			-0.227***
			(0.0142)
ρ			-0.387***
			(0.0460)
Constant	5.790***	6.253***	-1.968***
	(0.0671)	(0.114)	(0.0579)
Year FE	Yes	Yes	Yes
Observations	59 391	59 391	59 391

Source: Author's analysis based on data from the Costa Rica National Household Surveys (ENAHO), 2011–2018.

Notes: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, *p<0.1. The model corresponds to a finite mixture model with sample selection, assuming only two observed segments (formal and informal). See Appendix A for details.

5. Estimation of involuntary informality

Table 4 shows the results from estimating the shares of workers that are informal voluntarily and involuntarily. Here, we compare the actual distribution of workers with the maximizing distribution, which is the distribution that would occur if each worker were placed in the segment in which they have the highest predicted value.

Column (1) of Table 4 presents the "actual distribution" of workers, that is, the share of workers observed in each segment. This distribution is calculated from the parameters π_j estimated in the FMM model and shown in Table 3. In this simple model, with only two observed sectors, the point estimates of π_j coincide with the share of formal and informal workers in the sample.

Column (2) of Table 4 presents the "maximizing distribution". The maximizing distribution estimates the share of workers who would have the highest value in each segment if they were free to choose between formal and informal work. This distribution is derived from equation (2), for which we need values for ψ_L and the transition probabilities. We used the estimates of the Frisch elasticity of substitution ψ_L presented in regression (3) of Table 1. For the results presented in Panel A of Table 4, the transition probabilities of Table 2 are used. We call these probabilities unconditional, since they do not consider the characteristics of each individual. In Panel B of Table 4, the probabilities of transitioning into unemployment depend on individual characteristics as explained in the next section. Finally, log wages are estimated using the FMM model of Table 3.

With this information, for each individual i and according to (2), we calculate the value

$$W_{i,j} = (1 + \psi_L)\, lnw_{i,j} - \sigma_u ln(r + 1 - \pi_{i_{U,U}} + \pi_{i_{j,U}}),$$

where j represents the sector (formal or informal). If $\tilde{W}_{i,j0}$ is such that

$$\tilde{W}_{i,j_0} = max_j W_{i,j},$$

we say that individual i maximizes in segment j_0. The maximizing distribution is given by the share of workers who maximize in each segment j. This distribution is shown in column (2) of Table 4. Note that we are assuming free mobility or no segmentation across the formal and informal sectors.

The last column of Table 4 presents the proportion of workers who maximize in each segment divided by the actual proportion of workers in that segment, that is, the maximizing distribution (column 2) divided by the actual distribution (column 1). If this ratio is significantly less than 1, it is evidence that the actual and maximizing distributions are different and that workers are in informal work involuntarily.

For example, in Panel A of Table 4, the ratio of the maximizing distribution for informal workers compared to their actual distribution is statistically less than 1, showing that a significant difference exists between the actual and the maximizing proportion in the informal sector.[22]

[22] Standard errors are estimated by bootstrapping taking r, σ_u, and ψ_L as given.

Table 4. Distribution of Workers across Sectors vs. Maximizing Distribution

	(1) Actual Distribution	(2) Maximizing Distribution	(3) Ratio Maximizing/Actual
PANEL A. Unconditional $\pi_{.,U}$			
Formal	0.651***	0.981***	1.507***
95% Conf. Interval	[0.646 - 0.656]	[0.971 - 0.991]	[1.488 - 1.526]
Informal	0.349***	0.019***	0.054***
95% Conf. Interval	[0.344 - 0.354]	[0.009 - 0.0289]	[0.026 - 0.082]
PANEL B. Conditional $\pi_{.,U}$			
Formal	0.651***	0.965***	1.483***
95% Conf. Interval	[0.646 - 0.656]	[0.950 - 0.980]	[1.457 - 1.508]
Informal	0.349***	0.035***	0.100***
95% Conf. Interval	[0.344 - 0.354]	[0.020 - 0.050]	[0.058 - 0.143]
PANEL C. Comparing only wages			
Formal	0.651***	0.953***	1.464***
95% Conf. Interval	[0.646 - 0.656]	[0.934 - 0.972]	[1.432 - 1.495]
Informal	0.349***	0.047***	0.135***
95% Conf. Interval	[0.344 - 0.354]	[0.028 - 0.066]	[0.081 - 0.190]

Source: Authors' analysis based on Costa Rica National Household Survey (ENAHO) data.
Notes: The significance levels in column (3) refer to the null hypothesis that the ratio equals one. Differences between dividing column (2) by column (1) (i.e., maximizing over actual distribution) and the value presented in column (3) (i.e., ratio Maximizing/Actual) are due to rounding errors. Standard errors are estimated by bootstrapping taking r, σu, and ψL as given. Results are from 1200 replications. Panel A assumes that the transition probabilities into unemployment are those presented in Table 2 and that they do not depend on the characteristics of each individual. In Panel B the probabilities of transitioning into unemployment depend on the characteristics of each worker according to the probit model presented in Table 5. In panel C a sector is preferred when expected wages are higher in that sector. *** p<0.01, ** p<0.05, * p<0.1.

In sum, when considering the unconditional probabilities of transitioning into unemployment,[23] the overwhelming majority of informal workers would be better off in formal

[23] These probabilities are presented in Table 2.

employment. According to Panel A of Table 4, only 1.9 percent of all workers maximize in the informal sector, which represents 5.4 percent of all informal workers. These results suggest that formal employment is limited and that the majority of informal workers are informal involuntarily.

However, the share of involuntarily informal workers depends on our assumptions about segmentation, as well as on the inclusion of the transition probabilities into unemployment and the preferences for leisure ψ_L. Given the potential sensitivity of the results to these assumptions, in the next sections, we explore how our main results change depending on the values of ψ_L, the transition probabilities, and the inclusion of different job types (wage vs. self-employed) and segments within the informal sector.

5.1 How sensitive is the share of involuntary informality to ψ_L and the probability of unemployment?

Condition (2) can be rewritten as

$$(8)\ \ln w_F > \ln w_I + \frac{\sigma_u}{1+\psi_L} \ln\left(\frac{r+1-\pi_{U,U}+\pi_{F,U}}{r+1-\pi_{U,U}+\pi_{I,U}}\right),$$

that is, a formal job is preferred to an informal job when the log wage in the formal sector is greater than that in the informal sector plus a term that depends on 1) the Frisch elasticity ψ_L, 2) the probability of falling into unemployment from a formal job $\pi_{F,U}$, 3) the probability of falling into unemployment from an informal job $\pi_{I,U}$, and 4) the probability of leaving unemployment, $1-\pi_{U,U}$. In this section, we analyze the importance of this term in the case of Costa Rica. Note that when this term is ignored, only log wages are compared (as in Günther and Launov's 2012 model).

The main result of this section is that the last term in condition (8) plays an important role. This can be illustrated by comparing the simple model presented in Panel A of Table 4 that considers condition (8) with the results from comparing only log wages, that is, assuming that the last term of condition (8) is zero (as presented in Panel C). While in the simple model, the share of involuntary informality is 94.6%, this share falls by 8.2 points (to 86.4%) when only log wages are compared. Furthermore, this difference is mainly explained by the transition probabilities into unemployment rather than by the Frisch elasticity of substitution.

Note that as $\psi_L \to \infty$, the last term of condition (8) tends to zero, reducing the model to comparing (only log) wages, that is, as ψ_L increases, the 8.2 percentage point difference could be fully explained by this parameter. However, empirically, this does not seem to be the case. Taking values of ψ_L between 0.1 and 0.6, which are commonly found in the literature,[24] the share of involuntary informality for wage employees varies from 95.4 percent to 93.4 percent. With a value for ψ_L of one, the share of involuntary informality for wage employees would be 92.3 percent. That is, using a large value of ψ_L relative to what is found in the literature decreases the 8.2 difference by only 2.3 points. Thus, ψ_L only explains a small proportion of the change.

Since ψ_L only explains a small fraction of the difference between the simple model and the model comparing only log wages, a large fraction of this difference must be explained by the transition probabilities $\pi_{F,U}$, $\pi_{I,U}$, and $\pi_{U,U}$. When $\pi_{F,U} = \pi_{I,U}$, condition (8) reduces to a comparison of differences in log wages. From this point, as $\pi_{F,U}$ decreases and $\pi_{I,U}$ increases, formal work becomes more desirable than informal work, and the share of involuntarily informal workers increases. Indeed, our estimate of $\pi_{F,U}$ is smaller than $\pi_{I,U}$, and this explains most of the difference between the simple model and one that only compares log wages without taking

[24] See Keane and Rogerson (2012), Lee (2001), Keane and Neal (2023), Chetty (2012), and Chetty *et al.* (2013).

transition probabilities into account. In conclusion, the share of involuntarily informal workers is sensitive to our assumptions about the probability of transitioning into unemployment.

Since the results are sensitive to these probabilities, let us now extend the model to consider the case when transitions into unemployment depend on worker's characteristics. The estimates presented in Panel A of Table 4 consider only unconditional probabilities, that is, they assume that the probabilities of transitioning into unemployment are independent of the individual characteristics of each worker. To consider the possibility of different odds of unemployment for each worker, instead of calibrating the model with the aggregate transitions of Table 2, we use a probit model to estimate the transition probabilities as a function of each worker's individual characteristics.

Transition probabilities are assumed to be a function of age, age squared, gender, levels of education, additional training, fluency in English, the number of children under age 12 in the household, and living in the central region of the country. In addition, several changes between t and $t + 1$ are included. For example, completing primary, secondary, or some tertiary education from t to $t + 1$ are included as explanatory variables. Gaining some training, becoming fluent in English, and increasing the number of children under 12 years old (interacted with gender) are also included in the specification.

The results of the probit model are shown in Table 5. The dependent variable in the first column $\pi_{U,U}$ equals one when the worker is unemployed or out of the labor force in both t and $t + 1$, and includes only the sample of those who are unemployed or out of the labor force in t. Similarly, the second column $\pi_{F,U}$ includes only those who are in the formal sector in t, and the dependent variable equals one for those moving from the formal sector to unemployment or out of the labor force in the following year. The third column $\pi_{I,U}$ includes only informal workers, and the dependent variable identifies those who lose their job in the following year.

Table 5. Probit Model of Transition Probabilities

	(1)	(2)	(3)
	$\pi_{U,U}$	$\pi_{F,U}$	$\pi_{I,U}$
Age	-0.0425***	-0.149***	-0.120***
	(0.00505)	(0.00777)	(0.00607)
$Age^2/100$	0.0732***	0.181***	0.143***
	(0.00640)	(0.00960)	(0.00742)
Female	0.704***	0.330***	0.896***
	(0.0291)	(0.0358)	(0.0302)
Complete Primary	-0.149***	-0.189***	-0.0934***
	(0.0330)	(0.0530)	(0.0349)
Complete Secondary	-0.204***	-0.307***	-0.208***
	(0.0422)	(0.0609)	(0.0480)
More than Secondary	-0.379***	-0.514***	-0.245***
	(0.0469)	(0.0601)	(0.0559)
Earn Primary	-0.171**	-0.161	-0.125
	(0.0747)	(0.116)	(0.0771)
Earn Secondary	-0.0371	-0.144	-0.0323
	(0.0543)	(0.0964)	(0.0831)
Earn Some Tertiary	0.0888	-0.0944	-0.133
	(0.0945)	(0.0944)	(0.122)
Training	-0.185***	0.0195	-0.0349
	(0.0273)	(0.0312)	(0.0287)
Gain Training	-0.275***	0.00637	-0.272***
	(0.0347)	(0.0418)	(0.0432)
English	-0.286***	-0.0295	-0.0939
	(0.0555)	(0.0515)	(0.0655)
Learn English	-0.375***	0.0457	-0.112
	(0.0736)	(0.0809)	(0.113)
Number of Children under 12	-0.141***	-0.0420*	-0.0994***
	(0.0242)	(0.0221)	(0.0206)
Children under 12 × Female	0.226***	0.122***	0.192***
	(0.0277)	(0.0322)	(0.0277)
Change in Children under 12	-0.0399	-0.0606	-0.0834**
	(0.0384)	(0.0382)	(0.0346)
Change in Children under 12 × Female	0.122***	0.110*	0.160***
	(0.0462)	(0.0569)	(0.0471)

Central Region	-0.0846***	-0.0568**	-0.123***
	(0.0220)	(0.0281)	(0.0243)
Constant	1.127***	1.393***	1.423***
	(0.0967)	(0.156)	(0.121)
Year FE	Yes	Yes	Yes
Observations	21 796	22 919	14 596

Source: Author's analysis based on data from the Costa Rica National Household Surveys (ENAHO), 2011–2018.

Notes: Standard errors are in parentheses. Independent variables are measured in t, and changes are measured between t and $t+1$. The dependent variable in column $\pi_{U,U}$ equals one for those who are unemployed or out of the labor force in both t and $t+1$, and zero otherwise. The dependent variable in column $\pi_{F,U}$ equals one for those in the formal sector in t and unemployed/out-of-labor-force in $t+1$, and zero otherwise. The dependent variable in column $\pi_{I,U}$ equals one for those in the informal sector in t and unemployed/out-of-labor-force in $t+1$, and zero otherwise. Column $\pi_{U,U}$ includes only those unemployed in t; column $\pi_{F,U}$, those in formal employment in t; and column $\pi_{I,U}$, those in informal employment in t. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

As in the baseline model, our estimates of $\pi_{F,U}$ tend to be smaller than $\pi_{I,U}$, making the formal sector more desirable. Only for 21 out of the 23 409 individuals unemployed in *t+1* the probability $\pi_{F,U}$ is greater than $\pi_{I,U}$. Going back to condition (8), a greater $\pi_{I,U}$ (relative to $\pi_{F,U}$) increases the share of involuntarily informal workers relative to the case that only considers log wages.

The actual and maximizing distributions when using the probit model to estimate the transition probabilities according to individual characteristics are presented in Panel B of Table 4. We find that approximately 90 percent of all informal workers are involuntarily informal. This figure is smaller than the 94.6 percent in our baseline model (Panel A), but it is still greater than the 86.4 percent in the model that only compares log wages (Panel C).

In summary, in the case of Costa Rica, the preferences for leisure ψ_L do not seem to play a major role in determining the share of involuntary informal work when the full model is compared with a model that only considers utility from (today's) wages. In contrast, the probabilities of transitioning into unemployment play an important role in explaining the share of involuntarily informal workers.

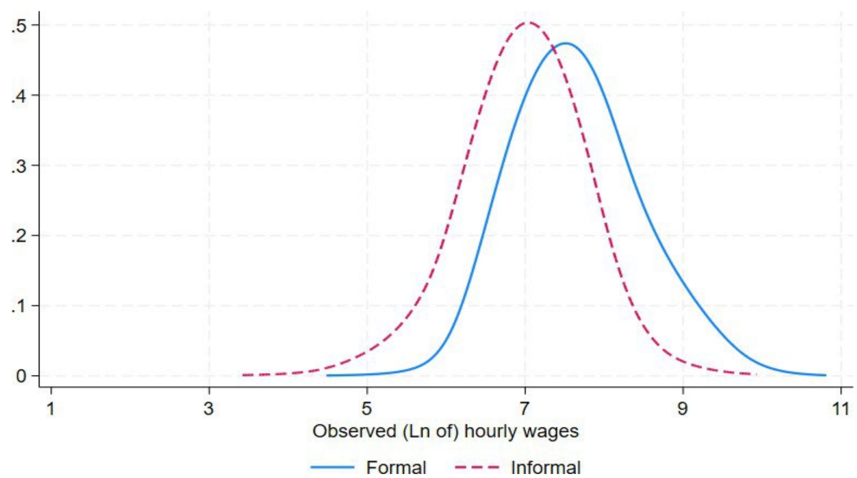
6. Informal segments and types of job

In the simple model presented in the previous sections, we assume that 1) earnings do not depend on the type of job and that 2) both the formal and informal sectors are (internally) homogeneous, i.e., each sector can be represented by one segment. In this section, we relax these two assumptions.

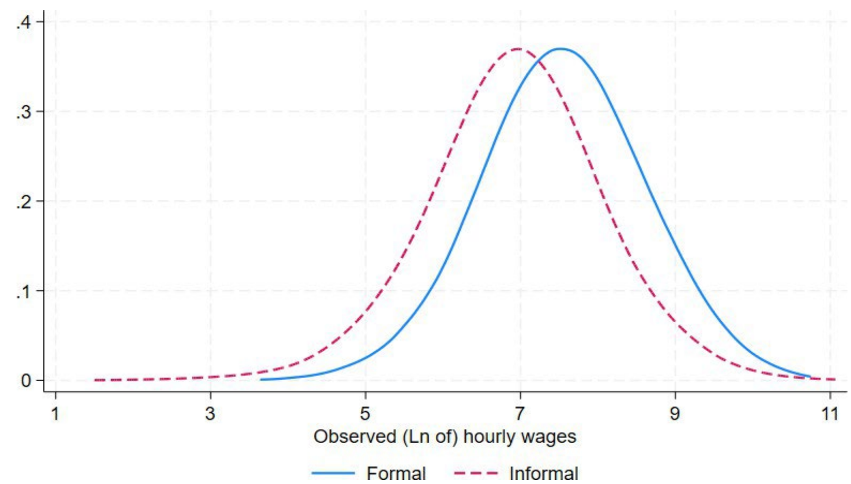
Regarding the type of job, we now divide workers into two categories: wage employees and self-employed. Panels A and B of Figure 2 show the densities of log income by the formality of the job for wage employees and self-employed workers, respectively. As shown in the figure, the densities of these two types of jobs are different, particularly for informal workers. Wages in the informal sector have a greater standard deviation for self-employed than for salaried workers. The self-employed present both lower and higher wages at the tails of the distribution relative to wage employees. The income distribution for informal wage employees seems to be more concentrated around the mean. Given these differences, and since these two types of jobs are different in nature, we applied the FMM strategy to both salaried and self-employed workers separately to estimate expected wages.

Figure 2. Densities of the Observed Ln of Hourly Wages for Formal and Informal Workers

Panel A. Wage Employees



Panel B. Self-Employed Workers



Source: Authors' analysis based on data from the Costa Rica National Household Survey (ENAHO), 2011–2018.
Notes: Income in colones as of June 2015. Pooled data of all workers between 15 and 65 years old are included in the panels from 2011–2012 to 2017–2018.

We also allow for the possibility of multiple segments within the informal sector. Up to this point, we have assumed that the formal and informal sectors are different from each other but internally homogeneous. However, empirical evidence suggests that there is heterogeneity within the informal sector (see Fields et al., 2023, and Günther and Launov, 2012; among others). Some literature suggests that there are at least two segments or tiers within the informal sector, a "lower-tier" characterized by low wages and an "upper-tier" characterized by higher wages and a higher percentage of voluntary informality (see Fields et al., 2023, for example).

However, these segments within the informal sector might not be observable. As shown by Günther and Launov (2012), an advantage of the FMM methodology is that it allows us to estimate the model assuming different unobserved segments within informality. Each of these segments is characterized by its own wage equation, such that the returns to individual characteristics vary by (unobserved) segment.[25] That is, the FMM methodology allows us to move beyond our assumption of homogeneity within the observed formal and informal sectors and assume that the informal sector might comprise different (unobserved) heterogeneous segments.

The number of segments within informality is selected using the BIC and the CAIC. Table 6 presents the CAIC and BIC for models with different numbers of informal segments for both salaried and self-employed workers. The two information criteria are minimized for two informal subsegments in each case, as shown in the table.

[25] Allowing the wage equation to vary across job types, sectors, and segments is a way to consider differences in productivity across these groups. This allows us to account for differences in productivity related to the characteristics of and (unobserved) average levels across groups (collected by the constant term and the year fixed effects).

Table 6. Finite Mixture Model Selection

	Wage Workers	
Model	CAIC	BIC
One Formal Sector and		
One Informal	128 818.29	128 767.29
Two Informal	127 931.40	127 862.40
Three Informal	127 982.94	127 895.94
Four Informal	128 104.96	127 999.96
	Self-Employed	
Model	CAIC	BIC
One Formal Sector and		
One Informal	54 732.46	54 681.46
Two Informal	54 680.97	54 611.97
Three Informal	54 785.59	54 698.59
Four Informal	54 765.16	54 663.16

Source: Authors' analysis based on data from the Costa Rica National Household Survey (ENAHO) panel 2011–2018.
Notes: Results are from estimating the finite mixture model described in the methodology and the appendix. CAIC= consistent Akaike information criterion, BIC= Bayesian information criterion.

Now we have two sectors (formal and informal) divided into two types of jobs (wage and self-employed). Additionally, for each type of job, the informal sector is divided into two segments, while we maintain the assumption that the formal sector is homogeneous (within each type of job). The results from the FMM model with two informal segments are presented in Tables 7 and 8. Table 7 presents the results for wage employees and Table 8 for the self-employed. The regressions represented in these tables can now be used to estimate six expected wages for each worker depending on the type of job (wage or self-employed), sector (formal or informal), and segment of informality (Informal 1 or Informal 2).

Finally, as before, we need the transition probabilities to estimate the share of involuntarily informal workers. We assume that the probabilities of moving between sectors and types of jobs are the same for both segments within the informal sector (i.e., Informal 1 and Informal 2). The transition matrix is presented in Panel B of Table 9.

Table 7. Finite Mixture Model for Wage Workers

	Formal	**Informal 1**	**Informal 2**	**Selection**
Age	0.0360***	0.0563***	0.00764*	0.152***
	(0.00249)	(0.0101)	(0.00448)	(0.00320)
$Age^2/100$	-0.0258***	-0.0585***	-0.00578	-0.214***
	(0.00318)	(0.0129)	(0.00589)	(0.00393)
Years of Education	0.106***	0.0654***	0.0211***	
	(0.000939)	(0.00696)	(0.00401)	
Female	-0.0121	-0.402***	0.101***	-0.856***
	(0.0115)	(0.0433)	(0.0201)	(0.0159)
Central Region	-0.0117	0.111***	0.0788***	0.303***
	(0.00730)	(0.0366)	(0.0151)	(0.0126)
English	0.0954***	0.271	0.222	
	(0.0221)	(0.185)	(0.140)	
Central × English	0.112***	0.521**	-0.233	
	(0.0256)	(0.210)	(0.157)	
Professionals	0.00141	0.232***	0.245***	
	(0.00762)	(0.0695)	(0.0403)	
Nicaraguan	-0.109***	0.0888*	0.00205	0.106***
	(0.0131)	(0.0535)	(0.0198)	(0.0239)
Ln σ_j	-0.717***	-0.215***	-1.090***	
	(0.00709)	(0.0212)	(0.0305)	
π_j	0.756***	0.102***	0.142***	
	(0.00256)	(0.00544)	(0.00554)	
Head of Household				0.706***
				(0.0148)
Children under 12				0.0669***
				(0.0119)
Child. under 12 × Female				-0.265***
				(0.0150)
ρ				-0.240***
				(0.0285)
Constant	5.643***	5.279***	6.652***	-2.070***
	(0.0494)	(0.206)	(0.0868)	(0.0606)
Year FE	Yes	Yes	Yes	Yes
Observations	51 591	51 591	51 591	51 591

Source: Author's analysis based on data from the Costa Rica National Household Survey (EBAHO), 2011–2018.
Notes: Robust standard errors are in parentheses. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Finite Mixture Model for Self-employed

	Formal	**Informal 1**	**Informal 2**	**Selection**
Age	-0.00755	0.00503	0.0189	0.209***
	(0.0183)	(0.0173)	(0.0312)	(0.00520)
$Age^2/100$	0.0130	0.00112	-0.0222	-0.241***
	(0.0201)	(0.0200)	(0.0361)	(0.00598)
Years of Education	0.0691***	0.0813***	0.0640***	
	(0.00556)	(0.00811)	(0.0212)	
Female	0.0628	0.0186	-0.272**	-1.258***
	(0.0717)	(0.0848)	(0.127)	(0.0232)
Central Region	0.165***	0.238***	0.0731	0.0992***
	(0.0414)	(0.0395)	(0.101)	(0.0186)
English	0.126	0.380***	0.0462	
	(0.116)	(0.147)	(0.211)	
Central × English	0.120	0.181	0.485	
	(0.145)	(0.189)	(0.355)	
Professionals	0.232***	0.316***	0.669***	
	(0.0553)	(0.0853)	(0.187)	
Nicaraguan	-0.0598	0.00390	-0.235*	-0.0423
	(0.110)	(0.0797)	(0.138)	(0.0414)
Ln σ_j	-0.0786***	-0.329***	0.243***	
	(0.0194)	(0.1083)	(0.08162)	
π_j	0.271***	0.470***	0.259**	
	(0.00503)	(0.01053)	(0.10524)	
Head of Household				0.683***
				(0.0207)
Children under 12				0.0723***
				(0.0156)
Child. under 12 × Female				-0.165***
				(0.0209)
ρ				-0.314***
				(0.0434)
Constant	7.006***	6.148***	6.225***	-4.434***
	(0.422)	(0.362)	(0.698)	(0.107)
Year FE	Yes	Yes	Yes	Yes
Observations	31 234	31 234	31 234	31 234

Source: Author's analysis based on data from the Costa Rica National Household Survey (EBAHO), 2011–2018.
Notes: Robust standard errors are in parentheses. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 9. Transition Matrix: Wage and Self-Employed

Panel A: Observed Odds

	Wage Formal	Wage Informal	SE Formal	SE Informal	Unemployed/ OLF
Wage Formal:	0.8636	0.0494	0.0040	0.0225	0.0605
Wage Informal:	0.1328	0.4959	0.0184	0.1168	0.2361
SE Formal:	0.0331	0.0683	0.4800	0.3690	0.0496
SE Informal:	0.0532	0.1317	0.1017	0.5162	0.1971
Unemployed/OLF:	0.0609	0.1006	0.0038	0.0654	0.7693

Panel B: Calibration

	Wage Formal	Wage Informal 1	Wage Informal 2	SE Formal	SE Informal 1	SE Informal 2	Unemployed/ OLF
Wage Formal:	0.8636	0.0206	0.0287	0.0040	0.0145	0.0080	0.0605
Wage Informal 1:	0.1328	0.2073	0.2886	0.0184	0.0753	0.0415	0.2361
Wage Informal 2:	0.1328	0.2073	0.2886	0.0184	0.0753	0.0415	0.2361
SE Formal:	0.0331	0.0286	0.0398	0.4800	0.2379	0.1311	0.0496
SE Informal 1:	0.0532	0.0550	0.0766	0.1017	0.3328	0.1834	0.1971
SE Informal 2:	0.0532	0.0550	0.0766	0.1017	0.3328	0.1834	0.1971
Unemployed/OLF:	0.0609	0.0421	0.0586	0.0038	0.0422	0.0232	0.7693

Source: Authors' analysis based on data from the Costa Rica National Household Survey (ENAHO), 2011–2018.
Notes: Rows reflect the work status in t, while columns indicate the work status one year later in t+1. SE= Self-employed, OLF=Out of the labor force. Panel A (Observed Odds) refers to pooled data observed from the ENAHOs, 2011–2018. Panel B (Calibration) is derived from Panel A assuming (for rows Wage Informal 1–2 and SE Informal 1–2) that the probabilities of moving from either segment 1 or 2 within the informal sector to other sectors are the same. Columns Wage Informal 1–2 and SE Informal 1–2 in Panel B are derived from columns Wage Informal and SE Informal of Panel A weighted by the share of each informal segment. These shares are derived from the point estimates of π_j in Tables 7 and 8.

6.1 Estimation of involuntary informality across sectors and types of job

When we have different types of jobs and segments within sectors, we must first solve for the expected value of each job and then compare them to pin down the sector in which each worker maximizes. Assume that the labor market is divided into J groups (in our case, combinations of sector, job type, and segment within informality). Let Π be the transition matrix across these groups between t and $t+1$. That is, the element (i, j) of Π represents the probability of moving from group i in t to group j in $t+1$. Defining $V = (V_1, \ldots V_J)'$ and $U = (1 + r)(U_1, \ldots, U_J)'$, the system of interest is given by

$$(1 + r)V = U + \Pi V.$$

We solve this system and compare across values of V_j to pin down the group of the labor market at which each worker maximizes.

In this section, Π is given by the values presented in Panel B of Table 9, and U is estimated using the baseline values of σ_u, ψ_L, and the estimated wages from the FMM models presented in Tables 7 and 8.

Table 10 compares the actual distribution of workers with the maximizing distribution. The latter is the distribution that would be obtained if each worker was in the group with the highest predicted value. The actual distribution is calculated from the parameters π_j estimated in the FMM model and shown in Tables 7 and 8.

For formal workers (for whom the segment is known), the value π_j matches the share of formal workers in the sample used in the estimation. Therefore, for formal wage employees, the parameter π_j shown in the second column of Table 7 represents the share of formal wage employees out of all wage employees. Similarly, for formal self-employed workers, the parameter π_j shown in the second column of Table 8 represents the share of formal self-employed workers out of all self-employed workers.

For the informal segments, Informal 1 and Informal 2 in columns 3 and 4, respectively, of Tables 7 and 8, the parameter π_j is an estimate of the share of workers in those segments out of the total number of wage and self-employed workers, respectively. Thus, these estimates and the shares of wage employees and self-employed out of all workers are used to estimate the distribution of workers across different sector–job–segment triplets.[26] We refer to this as the "actual distribution" in Table 10, and it represents the (estimated) share of workers observed in each segment.

In Panel A of Table 10, the maximizing distribution estimates the share of workers who would obtain the highest value V_j in each group if they were free to choose any of the possible six. Note that we are assuming free mobility or no segmentation, not only across sectors (formal and informal) but also across types of jobs (wage and self-employed). The maximizing distribution in Panel B of Table 10 also assumes free mobility across all six sector–job–segment triplets, but it compares only the value of log wages (instead of V_j).

The last column of Table 10 presents the proportion of workers in each group who maximize in that group divided by the actual proportion of workers in that group, that is, the maximizing distribution divided by the actual distribution. If this ratio is less than one, it is evidence that the actual and maximizing distributions are different and that workers are in informal work involuntarily.

[26] Approximately 78 percent of all workers are wage employees, and 22 percent are self-employed.

Table 10. Distribution of Workers across Sectors vs. Maximizing Distribution

	Actual Distribution	Maximizing Distribution	Ratio Maximizing/Actual
Panel A. Comparing V			
Wage Formal	0.5920***	0.8633***	1.4580***
Wage Informal 1	0.0801***	0.0024	0.0305***
Wage Informal 2	0.1109***	0.0194***	0.1748***
SE Formal	0.0588***	0.1062**	1.8048
SE Informal 1	0.1020***	0.0007	0.0071***
SE Informal 2	0.0562**	0.0080	0.1421
Panel B. Comparing Log Wages			
Wage Formal	0.5920***	0.6338***	1.0705
Wage Informal 1	0.0801***	0.0208*	0.2593***
Wage Informal 2	0.1109***	0.0666***	0.6005**
SE Formal	0.0588***	0.2477***	4.2094***
SE Informal 1	0.1019***	0.0110	0.1083***
SE Informal 2	0.0562**	0.0201	0.3579

Source: Authors' analysis based on data from the Costa Rica National Household Surveys (ENAHO).
Notes: Panel A compares the values of V from solving the system $[(1 + r)I - \Pi]V = U$ using the baseline values of ψ_L, r, and those from Table 9. Expected log wages are calculated from the FMM model presented in Tables 7 and 8. Panel B compares only log wages. Differences between the last column and the ratio of the first two columns are due to rounding errors. The significance levels in the las column refer to the null hypothesis that the ratio equals one. Standard errors are estimated by bootstrapping taking r, σ_u, ψ_L, and the probabilities from Table 9 as given. Formality and type of job are used as strata. Results are from 1200 replications. *** p<0.01, ** p<0.05, * p<0.1.

The results in Table 10 show how the share of involuntarily informal workers depends on the different assumptions on transition probabilities. In Panel A of Table 10, when it is assumed that there is free mobility and the transition probabilities are considered, the great majority of informal work is involuntary. Moreover, the preferred job is a formal wage job.

While 59.20 percent of all workers are in formal wage jobs, 86.33 percent of workers maximize in this sector. Similarly, although 5.88 percent of workers are formal self-employed, 10.62 percent maximize in this sector. Adding these two sectors together, we find that 96.95 percent of all workers maximize in the formal sector, while it represents 65.08 percent of all workers. This means that an extra 31.87 percent of workers would like to move (from the informal sector) to a formal job. Therefore, among informal workers, 8.73 percent appear to be in the informal sector voluntarily, while the remaining 91.27 percent are involuntarily informal.[27]

When the value of a job is measured only by wages, the share of involuntary informality falls to 66 percent. That is, when we do not consider the transition probabilities, the share of involuntarily informal workers falls from 91.27 to 66 percent. Moreover, formal wage jobs are not as desirable when the value of the job is given only by its wage. As shown in Panel B of Table 10, the relationship between those who are and those who maximize in the formal wage segment is close to 1, i.e., in net terms, only a small fraction of new workers want to move to formal wage jobs. This contrasts with the results shown in Panel A of Table 10. In Panel A, when the transition to unemployment is considered, approximately 86.33 percent of all workers maximize in formal wage jobs. When only wages are compared in Panel B, this percentage falls to 63.38. In other words, wage formal jobs are not as attractive as before.

We measure this decrease in terms of potential or desired absorption in the wage formal sector. In terms of "potential net absorption", when the probability of transitioning into unemployment is considered in Panel A, approximately 27.13 percent of workers who are not in a

[27] While the great majority of informal workers would prefer a formal sector job, Table 10 also presents evidence that informal workers prefer one informal sector over the other. Specifically, a much higher proportion of workers actually in Informal 2 maximize utility in that sub-sector relative to those in Informal 1 who maximize utility for Informal 1. This is true for both wage employees and self-employed, and despite more than twice as many informal workers in the less preferred Informal 1 relative to the more preferred Informal 2. These results suggest that informal workers are divided into a lower-tier informal sector (Informal 1) and a higher utility upper-tier informal sector where barriers such as substantial human capital and financial capital requirements limit mobility into an upper-tier informal sector (Fields, 1990).

formal wage job would like to move to this segment. In contrast, when only wages are compared in Panel B, only 4.18 percent of workers would like to move to a formal wage job.

In contrast to formal wage jobs, formal self-employed jobs are more desirable when only wages are considered. When only wages are considered, the formal self-employed sector has the greatest “potential net absorption” or the greatest percentage of workers willing to move. Those maximizing as formal self-employed increase from 10.62 percent of all workers in Panel A to 24.77 percent in Panel B. This means a “potential net absorption” of 4.74 percent in Panel A vs. 18.89 percent in Panel B.

The difference in “potential absorption” between Panels A and B is explained by the transition probabilities. When considering only wages (in Panel B), formal self-employed jobs are more attractive because they have high earnings, explained in part by the fact that their earnings consider some returns to capital. However, according to the transitions presented in Table 9, this type of job is less stable than a formal wage job. Only 48 percent of formal self-employed workers are in the same type of job one year later. In contrast, approximately 86.4 percent of wage formal workers are in this type of job one year later. Thus, when considering the odds of keeping a formal wage job vs. the odds of keeping a formal self-employed job, most workers prefer the formal wage job (even when they might perceive lower immediate wages).

In summary, the share of involuntarily informal workers is sensitive to the probability of moving across types of jobs and into unemployment. When the transition probabilities across types of jobs and into unemployment are considered, approximately 91 percent of all informal workers are involuntarily informal. When the value of moving is not considered, i.e., when the value of the job is only measured by earnings, the share of involuntary informality falls to 66 percent.

Conclusions

This paper provides evidence supporting the traditional labor market segmentation view of dualistic labor markets in which most workers are involuntarily employed in the informal sector. The primary objective of this paper is to distinguish voluntarily informal workers from involuntarily informal workers by expanding the methodology developed in Günther and Launov (2012) to explicitly account for preferences for leisure and the probability of transition across jobs and into unemployment.

In the case of Costa Rica, our findings indicate that preferences for leisure, as measured by the Frisch elasticity of substitution, do not explain a relevant proportion of those voluntarily engaged in informal work. In contrast, the probabilities of transitioning between segments of the labor market, particularly into unemployment, appear to play an important role in determining the desirability of formal wage jobs. When accounting for these transition probabilities, more than 90 percent of all informal workers are found to be involuntarily informal. However, when only expected earnings are considered, the proportion of involuntarily informal workers drops to 66 percent.

Our results for an upper-middle-income country in Latin America challenge the idea of other literature that finds that informality in the region is largely voluntary. One reason that our results differ is that we explicitly account for heterogeneous informality and, in particular, recognize that formal jobs offer greater stability. Specifically, wage employees face a lower probability of transitioning into unemployment or a lower-paying job compared to other types of workers.

Appendices

A. Finite mixture model with sample selection

This appendix presents a few details of the methodology. The econometric model follows Günther and Launov (2012), from which this appendix borrows heavily. The only difference in our specification is that instead of applying the limited information maximum likelihood estimator used by Günther and Launov (2012), we use full information maximum likelihood. This approach avoids splitting the model into two steps and obviates the necessity of computing the second-step covariance matrix correction. Please see Günther and Launov (2012) for additional details.

Günther and Launov's (2012) methodology consists of estimating a Finite Mixture Model (FMM) with sample selection. FMMs are a statistical technique used to represent the presence of subpopulations within an overall population. The model identifies the overall distribution as a weighted sum of a finite number of classes of other distributions. In our case, the overall distribution of wages is modeled as a mixture or weighted sum of distributions across segments or groups of workers.

It is assumed that the researcher can directly identify formal workers and that these workers form a homogeneous segment. The same assumption holds for informal workers in the simple model, but in the full model presented in Section 6, we assume that informal workers can belong to different (unobserved) segments within informality. The main advantage of the FMM methodology is that it allows us to estimate expected wages even when assuming the existence of these unobserved segments within informality. Günther and Launov (2012) expand the FMM methodology to account for sample selection.

Specifically, an individual can be employed in the formal sector, employed in the informal sector, or out of work. It is assumed that the formal sector is homogeneous, whereas the informal sector can be heterogeneous.[28] Each segment (formal and the two informal subsectors) is characterized by its own wage equation. Specifically, workers can be divided into J segments, Y_j. The log-wage in each segment Y_j is described by the equation

$$\ln w_{ij} = X_i'\beta_j + \mu_{ij},\ i \in Y_j,$$

where w_{ij} are the earnings of an individual i in segment j, and the error term $\mu_{ij} \sim N(0, \sigma_j)$ is uncorrelated across segments.[29] Furthermore, the observed distribution of earnings depends on the decision to enter (or not) the labor market. This decision is assumed to be a function of personal characteristics Z:

$$y_{is} = Z_i'\gamma + v_{is},\ v_{is} \sim N(0, 1),$$

such that earnings w_{ij} are observed only if the outcome of this last equation is positive ($y_{is} > 0$). It is assumed that the (segment-specific) error terms of the above equations follow a bivariate normal distribution with correlation coefficient ρ, and therefore, the observed distribution of earnings in the j-th segment of the labor market is given by

[28] In our simple model, it is assumed that the formal and informal sectors are homogeneous, i.e., they are formed by only one segment. In section 6 this assumption is relaxed to consider the possibility of heterogeneous segments within the informal sector. In practice, to estimate the number of heterogeneous segments, the finite mixture model is estimated several times assuming a different number of informal segments. From these repetitions, the number of segments that minimize an information criterion —such as the Bayesian information criterion (BIC) or the consistent Akaike criterion (CAIC)— is selected.

[29] The dependent variable in the Mincer wage equation is the log of hourly wages (*ln wij*). The independent variables include age, age squared, gender, region, years of education, English skills, migration from Nicaragua, and professionals. Region is a dummy that takes the value of one for those living in the central region of the country. English skill is a dummy that equals one for those fluent in English. Migration from Nicaragua is a dummy indicating if the worker was born in Nicaragua. Professionals are defined as being employed as a technician, associate professional, or clerical support worker in the current or previous year.

$$(9)\ f(w_{ij}|y_{is} > 0) = \frac{\phi\left(\frac{lnw_{ij}-X_i'\beta_j}{\sigma_j}\right)}{\sigma_j\Phi(Z_i'\gamma)}\Phi\left(\frac{Z_i'\gamma+\left(\frac{\rho}{\sigma_j}[lnw_{ij}-X_j'\beta_j]\right)}{\sqrt{1-\rho^2}}\right),$$

where ϕ and Φ denote the standard normal probability and cumulative density functions, respectively. Let $j = 1$ represent the formal sector. In the informal sector, the (unconditional) probability π_j of an individual i belonging to segment Y_j, and the distribution of observed wages are modeled by the mixture:

$$f(w_i) = \sum_{j=2}^{J} \pi_j f\left(w_{ij}\middle|y_{is} > 0\right),$$

If there are no entry barriers to any segment of the labor market, each worker will be in the sector with the highest expected value given their own individual characteristics and the returns on these characteristics. This distribution is referred to as the "maximizing distribution", and it describes the proportion of workers that maximize their expected value in each segment.

However, the actual distribution of individuals across sectors—given by $P(i \in Y_j) = \pi_j$—might differ from the maximizing distribution if there are barriers to entry for some segments of the labor market. In other words, if it cannot be rejected that the actual distribution and the hypothetical maximizing distribution are equal, then it cannot be rejected that workers select each segment according to their comparative advantage, and thus, there is no evidence of market segmentation. On the contrary, rejection of the equality of the actual and maximizing distributions is evidence that entry barriers exist that prevent some workers from being in the sector with the greatest value. In this case, to avoid unemployment, workers enter a sector with lower value—or a last resort job—which is evidence of market segmentation. Consequently, the difference between the maximizing distribution and the actual distribution can be used as an estimate of involuntarily informal employment. Thus, the goal is to estimate these two distributions.

The actual distribution can be estimated by maximum likelihood (ML). Since it is observed whether an individual works in the formal or informal sector, let us denote by Y_F the set of earnings in the formal sector and by Y_I the set of earnings in the informal sector. Thus, those individuals who are neither in Y_I nor in Y_F represent those who are out of work.

Considering this (observed) division of workers between formal and informal, the log-likelihood function can be written as:

$$(10)\ ln\mathcal{L} = \sum_{i\notin\{Y_F,Y_I\}} ln\Phi(-Z'_i\gamma) + \sum_{i\epsilon Y_F} ln\left(\pi_F h\left(\theta_F|w_{i,F}, y_{is} > 0, X_i, Z_i\right)\right) +$$

$$\sum_{i\epsilon Y_I}\left[ln\left(\sum_{j=2}^{J}\pi_{I_j}h\left(\theta_{I_j}|w_{i,I_j}, y_{is} > 0, X_i, Z_i\right)\right)\right],$$

where $h(\cdot) = \Phi(Z'_i\gamma)f(\cdot)$, and $f(\cdot)$ is given by (9), π_F is the probability of belonging to the formal sector, π_{I_j} is the probability of belonging to the j-th segment of the informal sector, and $\theta_j = \{\beta_j, \gamma, \sigma_j, \rho\}$ is the set of other parameters to be estimated. These parameters are estimated by following a full information ML approach, while the number of segments J is selected on the basis of the information criteria. The model is estimated using one, two, three, and four informal segments. The information criteria for each case are presented in Table 6.

Once the information criteria are used to select the number of segments J, and the actual distribution π_j is estimated by ML, the maximizing distribution is estimated from the share of workers who have the highest expected value in each sector. The expected (log) wage (or earnings) of individual i in segment j are calculated as

$$\widehat{\mathbb{E}}[\ln \mathrm{w_{ij}}|y_{is} > 0, X_i] = X'_i\hat{\beta}_j + \hat{\rho}\hat{\sigma}_j\frac{\phi(-Z'_i\hat{\gamma})}{1-\Phi(-Z'_i\hat{\gamma})}.$$

These expected values are used to estimate the maximizing distribution.

B. Derivation of condition $V_F > V_I$

Assume that the labor market is divided into J segments. Let Π be the transition matrix across these segments between t and $t+1$. That is, the element (i, j) of Π, let's say $\pi_{i,j}$, represents the probability of moving from segment i in t to segment j in $t+1$. Furthermore, let V_i represent the (present) value of being in segment i in t, with a discount rate of r, and let U_i represent the utility of being in segment i in t. Defining $V = (V_1, \dots V_J)$' and $U = (1+r)(U_1, \dots, U_J)$', the system of interest is given by

$$(1+r)V = U + \Pi V.$$

The solution is given by[30]

$$V = [(1+r)I - \Pi]^{-1}U,$$

where I is the identity matrix. With only three segments, let's say formal $j = 1$, informal $j = 2$, and unemployment $j = 3$, the above system reduces to

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = (1+r)\begin{pmatrix} r+1-\pi_{1,1} & -\pi_{1,2} & -\pi_{1,3} \\ -\pi_{2,1} & r+1-\pi_{2,2} & -\pi_{2,3} \\ -\pi_{3,1} & -\pi_{3,2} & r+1-\pi_{3,3} \end{pmatrix}^{-1}\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}.$$

Solving for the inverse on the right-hand side, the above system reduces to

[30] Note that if $r > 0$, then $[(1+r)I - \Pi]$ is not singular. The eigenvalues of $[(1+r)I - \Pi]$ associated with the eigenvectors s are the values α such that $[(1+r)I - \Pi]s = \alpha s$, or equivalently, $\Pi s = (r+1-\alpha)s$ That is, if s is an eigenvector of $[(1+r)I - \Pi]$ with associated eigenvalue α, then s is an eigenvector of Π with associated eigenvalue $r+1-\alpha$. Since Π is a transition matrix, all its eigenvalues are less than or equal to one. Therefore, $\alpha \geq r > 0$. We conclude that zero is not an eigenvalue of $[(1+r)I - \Pi]$, and hence it is not singular.

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \frac{(1+r)}{|D|} \begin{pmatrix} (r+1-\pi_{2,2})(r+1-\pi_{3,3}) - \pi_{3,2}\pi_{2,3} & \pi_{1,2}(r+1-\pi_{3,3}) + \pi_{3,2}\pi_{1,3} & (r+1-\pi_{2,2})\pi_{1,3} + \pi_{1,2}\pi_{2,3} \\ \pi_{2,1}(r+1-\pi_{3,3}) + \pi_{3,1}\pi_{2,3} & (r+1-\pi_{1,1})(r+1-\pi_{3,3}) - \pi_{1,3}\pi_{3,1} & (r+1-\pi_{1,1})\pi_{2,3} + \pi_{1,3}\pi_{2,1} \\ (r+1-\pi_{2,2})\pi_{3,1} + \pi_{2,1}\pi_{3,2} & (r+1-\pi_{1,1})\pi_{3,2} + \pi_{1,2}\pi_{3,1} & (r+1-\pi_{1,1})(r+1-\pi_{2,2}) - \pi_{1,2}\pi_{2,1} \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}$$

where $|D| > 0$ represents the determinant[31] of $[(1 + r)I - \Pi]$ for $J = 3$. Since each row of Π sums to one, the above solution implies that

$$V_1 > V_2 \Leftrightarrow (r + 1 - \pi_{3,3} + \pi_{2,3})\ U_1 > (r + 1 - \pi_{3,3} + \pi_{1,3})\ U_2 + (\pi_{2,3} - \pi_{1,3})\ U_3,$$

which reduces to condition (2) when $j = 1$ is the formal sector, $j = 2$ is the informal sector, $j = 3$ is unemployment, and the instantaneous utility of unemployment is zero, $U_3 = 0$.

Data Availability Statement:

This paper uses confidential data from the Costa Rican Institute of Statistics and Census (INEC). The data can be obtained by filing a request directly with the INEC.

Acknowledgments:

The authors would like to thank Salem Abo-Zaid, Oleg Firsin, Yusuke Kuwayama, and Guillermo Vuletin for their valuable comments and insightful suggestions on earlier drafts of this article. An earlier version of this work received funding as part of the UNU-WIDER project "Transforming Informal Work and Livelihoods". Institutional support was provided by the University of Maryland, Baltimore County, and the Annenberg Institute at Brown University.

[31] Note that $|D|$ must be positive since, as we argued in the previous footnote, all the eigenvalues of $[(1 + r)I - \Pi]$ are positive.

References

Altonji, J.G. (1986). "Intertemporal substitution in labor supply: Evidence from micro data." *Journal of Political Economy*, 94(3, Part 2): S176–S215.

Alaimo, V., M. Bosch, D. Kaplan, C. Pagés, and L. Ripani. (2015). *"Jobs for Growth,"* IDB Publications (Books), Inter-American Development Bank, number 7203, June. Washington, DC.

Antonopoulos, R. (2009). The Unpaid Care Work–Paid Work Connection. Working Paper

86. Geneva: Policy Integration and Statistics Department, International Labour Office.

Arias, J., E. Artuc, D. Lederman, and D. Rojas. 2018. "Trade, informal employment and labor adjustment costs". *Journal of Development Economics*, 133: 396-414.

Bailey, M. N. (1978). "Some Aspects of Optimal Unemployment Insurance." *Journal of Public Economics*, 10(3): 379-402.

Chetty, R. (2006)."A General Formula for the Optimal Level of Social Insurance." *Journal of Public Economics*, 90(10-11): 1879-1901.

Chetty, R. (2012). "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply". *Econometrica*, 80(3): 969-1018.

Chetty, R., A. Guren, D. Manoli, and A. Weber. (2013). "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-analysis of Extensive Margin Elasticities". NBER macroeconomics Annual, 27(1): 1-56.

Deolalikar, A. (1993). "Gender Differences in the Returns to Schooling and in School Enrollment Rates in Indonesia." *Journal of Human Resources*, 28(4): 899-932.

De Soto, H. (1989). The Other Path. Harper and Row. https://doi.org/10.2139/ssrn.1176661

Fields, G., T.H Gindling, K. Sen, M. Danquah, and S. Schotte (2023). "*The Job Lad- der: Transforming Informal Work and Livelihoods in Developing Countries.* UNU-WIDER Studies in Development Economics, Oxford University Press.

Fields, G. (1975). Rural–Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDCs. *Journal of Development Economics*, 2(2), 165–87.

Fields, G.S. (1990). Labour Market Modelling and the Urban Informal Sector: Theory and Evidence. In D. Thurnham, B. Salomé & A. Schwarz (Eds), *The Informal Sector Revisited*. Paris, OECD.

Gerard, F., and G. Gonzaga. (2021). "Informal Labor and the Efficiency Cost of Social Programs: Evidence from the Brazilian Unemployment Insurance Program." *American Economic Journal: Economic Policy*, 13(3): 167-206.

Günther, I., and A. Launov (2012). Informal Employment in Developing Countries: Opportunity or Last Resort? *Journal of Development Economics*, 97, 88–98. https://doi.org/10.1016/j.jdeveco.2011.01.001

Harati, R. (2013). Heterogeneity in the Egyptian Informal Labour Market. R*evue D'économie Politque*, 123(4), 623–39. https://doi.org/10.3917/redp.234.0623

Harris, J.R., and M. Todaro (1970). Migration, Unemployment and Development: A Two- Sector Analysis. *American Economic Review*, 40, 126–42.

Heckman, J.J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 153–61. https://doi.org/10.2307/1912352

Huber, M. and G. Mellace. (2014). Testing Exclusion Restrictions and Additive Separability in Sample Selection Models. *Empirical Economics*, 47: 75-92.

International Labour Organization (ILO) (2021), ILO Thesaurus, February, 2021, https://www.ilo.org/inform/online-information-resources/terminology/thesaurus/lang–en/index.htm

International Labour Organization (ILO) (2018). ILOSTAT database, ILO Modelled Estimates, November. Geneva: International Labour Office. https://www.ilo.org/ilostat/

Keane, M., and T. Neil. (2023). "Robust Inference for the Frisch Labor Supply Elasticity". Opportunity and Inclusive Growth Institute, Federal Reserve Bank of Minneapolis, Institute Working Paper No. 81. November.

Keane, M., and R. Rogerson. (2012). "Micro and macro labor supply elasticities: A reassessment of conventional wisdom." *Journal of Economic Literature*, 50(2): 464–76.