



WIDER Working Paper 2021/50

## **Heterogeneous informality in Costa Rica and Nicaragua**

Enrique Alaniz,<sup>1</sup> T.H. Gindling,<sup>2</sup> Catherine Mata,<sup>3</sup> and Diego  
Rojas<sup>3</sup>

March 2021

**Abstract:** Informal work is often considered a place 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 that budding entrepreneurs and those looking for flexible working conditions enter voluntarily. We use the methodology developed in Günther and Launov (2012) to test for the voluntary and involuntary nature of informal work in Nicaragua and Costa Rica, without making ad hoc assumptions about labour market segmentation and self-selection. We find evidence of heterogeneous informality in both Nicaragua and Costa Rica, with one informal sub-segment where most workers are voluntarily informal and another informal sub-segment where most workers are involuntarily informal. In Nicaragua, our results suggest that 44 per cent of wage employees are involuntarily informal, while 30 per cent of self-employed workers are involuntarily informal. In Costa Rica, our results suggest that 10 per cent of wage employees are involuntarily informal, and that 66 per cent of the self-employed are involuntarily informal.

**Key words:** involuntary informality, developing economies, finite mixture model, Costa Rica, Nicaragua

**JEL classification:** F63, J8, J32, J46

**Acknowledgements:** We would like to thank Isabel Günther and Andrey Launov for providing us with the MATLAB code to replicate their analysis using Costa Rican and Nicaraguan data. We are also grateful for extensive comments from Kunal Sen, Simone Schotte, Ira Gang, and Gary Fields on drafts of this paper. We would also like to thank participants in the UNU-WIDER workshop ‘Transforming Informal Work and Livelihoods’, especially Robert Duval Hernández, for helpful comments. Section 2.2 reproduces the framework we developed in Alaniz et al. (2020).

---

<sup>1</sup> Fundación Internacional para el Desafío Económico Global (FIDEG) and Universidad Thomas More, Managua, Nicaragua; <sup>2</sup> UMBC Department of Economics, Baltimore, USA, UNU-WIDER, Helsinki, Finland, and IZA Institute of Labor Economics, Bonn, Germany, corresponding author: [gindling@umbc.edu](mailto:gindling@umbc.edu); <sup>3</sup> UMBC School of Public Policy, Baltimore, USA

This study has been prepared within the UNU-WIDER project [Transforming informal work and livelihoods](#).

Copyright © UNU-WIDER 2021

UNU-WIDER employs a fair use policy for reasonable reproduction of UNU-WIDER copyrighted content—such as the reproduction of a table or a figure, and/or text not exceeding 400 words—with due acknowledgement of the original source, without requiring explicit permission from the copyright holder.

Information and requests: [publications@wider.unu.edu](mailto:publications@wider.unu.edu)

ISSN 1798-7237 ISBN 978-92-9256-988-4

<https://doi.org/10.35188/UNU-WIDER/2021/988-4>

Typescript prepared by Joseph Laredo.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

## 1 Introduction

By a variety of measures, a large proportion of workers in developing economies are informal. Informal work is often considered a place 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 that budding entrepreneurs and those looking for flexible working conditions enter voluntarily. Analysis of informality must recognize this heterogeneity, differentiating between workers who are informal because of lack of formal employment opportunities and those who are self-employed or working in small firms voluntarily because of comparative advantage or preferences. In this paper we use the methodology developed in Günther and Launov (2012), which allows us to empirically identify different sub-segments within informality and to test for the voluntary and involuntary nature of workers in each sub-sector.

Costa Rica and Nicaragua are neighbouring countries in Central America that present contrasting economic structures. Costa Rica is classified by the World Bank as an upper-middle income country, while Nicaragua is classified as a lower-middle income country. GDP per capita in Costa Rica (US\$19,762) is more than three times that in Nicaragua (US\$5,834) and poverty is half (1.4 per cent vs 3.2 per cent below \$1.90/day, and 22 per cent vs. 41 per cent at national poverty lines).<sup>1</sup> Costa Rica has one of the most formal labour markets in Latin America, while Nicaragua has among the least formal. We estimate that more than 80 per cent of workers in Nicaragua are informal, while 40 per cent of workers in Costa Rica are informal.

In the labour market segmentation theory of dualistic labour markets, formal wages are institutionally set at higher than equilibrium (market) levels. Institutional mechanisms for maintaining above-market levels of formal wages include minimum wage enforcement and labour protection only in the formal sector, 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 also 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 are able to obtain informal employment, although at low wages. According to this theory, therefore, most informal workers are informal involuntarily because they have not been able to obtain one of the limited high-wage formal jobs.

Others argue that labour market segmentation is not the reason that workers are informal (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. Workers choose informality voluntarily because of comparative advantage or preference. 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 formal work for some workers (for example, those with a comparative advantage in informal work or those who are

---

<sup>1</sup> World Bank World Development Indicators 2018 for most recent year. All dollar amounts are in current PPP dollars.

informal to avoid taxes or costly regulations), while earnings for others may be lower in informal work (for example, for those who are willing to accept lower earnings in exchange for the flexible hours and conditions in informality), but all workers are informal voluntarily.

More recent views of dualistic labour markets recognize that voluntary and involuntary informality coexist, and distinguish between those informal workers who are voluntarily informal vs. others who are involuntarily informal because they are limited in their access to the formal employment which they would prefer (Fields 1990; Günther and Launov 2012; Maloney 2004). Fields (1990) refers to the former as upper-tier informal workers and the latter as lower-tier informal workers. While both sub-segments are informal, substantial human capital and financial capital requirements in upper-tier informality imply that there is not free mobility from lower-tier informal work into upper-tier informal work.

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 of earnings may differ between self-employment and wage employment, and because the definition of ‘earnings’ may differ between self-employment and wage employment, we examine self-employment and wage employment separately.

To provide evidence of whether workers are voluntarily or involuntarily informal in Costa Rica and Nicaragua we use data sets collected by the Costa Rican Statistics and Census agency and the Nicaraguan Fundación Internacional para el Desafío Económico Global (FIDEG) to implement the methodology developed in Günther and Launov (2012). This methodology allows us to empirically identify different sub-segments within informality and to test for the voluntary and involuntary nature of workers in each sub-segment, identifying involuntarily informal workers as those whose wages would be higher in formal work or in another sub-segment. This methodology may mistakenly identify some workers as involuntarily informal who are in fact voluntary because they are willing to accept lower wages for the flexibility of informality. For instance, women bear a greater burden from the uneven division of unpaid family care work, and therefore, they usually have fewer uninterrupted hours available to be part of the paid labour force (Antonopoulos 2009). Women from poor households are particularly concentrated in informal employment, in part due to the possibility this sector offers to combine unpaid and paid work (Kucera and Roncolato 2008; Roncolato and Radchenko 2016). To partially correct for this possibility, in the body of this paper we use data for men only.

We find evidence of heterogeneous informality among wage employees in both Costa Rica and Nicaragua and among the self-employed in Nicaragua, with one sub-segment where most workers are voluntarily informal and a second sub-segment where most workers are involuntarily informal. In Nicaragua, our results suggest that 44 per cent of wage employees (representing 63 per cent of informal wage employees) are involuntarily informal, while 30 per cent of the self-employed (representing 29 per cent of the informal self-employed) are involuntarily informal. In Costa Rica, our results suggest that 10 per cent of wage employees (representing 43 per cent of informal wage employees) are involuntarily informal. Our evidence suggests only one informal segment among the self-employed in Costa Rica, where we estimate that 93 per cent of informal self-employed are involuntarily informal (this represents 66 per cent of all self-employed).

## **2 Data and definitions**

In this section we describe the data and how we identify formal and informal workers in Costa Rica and Nicaragua.

## 2.1 Data

### *Costa Rica*

The data used in this analysis are from the 2018 Costa Rican National Household Survey (Encuesta Nacional de Hogares, ENAHO).<sup>2</sup> Household surveys are conducted annually by the Costa Rican National Statistics and Census Institute. ENAHO provides information on employment, income, poverty, education, dwelling conditions, and access to public services, among other things. The data are nationally representative and based on the sampling frame from the 2011 national census. The sampling method is probabilistic, stratified, and carried out in two steps. In the first step, the *segmentos censales* or primary sampling units are selected, while in the second step the dwellings or secondary sampling units are chosen from the primary units. About 12 dwellings per segment are selected, so the sampling comprises 1,120 segments, 13,440 dwellings, and around 40,000 people, although the final number of observations varies each year. This survey is statistically representative of the six geographical regions of Costa Rica, and of both rural and urban areas. We limit our sample to the male working-age population (15–65) who report earnings. The resulting sample includes 8,406 observations of formal and informal workers.

### *Nicaragua*

For Nicaragua, we use a panel data set collected by FIDEG that follows households and household members from 2009 to 2017. FIDEG's survey is designed to measure poverty annually using household aggregate consumption as a welfare indicator, and also includes data on family and individual characteristics, wages and other employment characteristics, access to some public services, and housing structures. It is a shorter version of an LSMS and the sample is a nationally representative panel of 1,700 households located in urban and rural areas throughout the country. The data are collected in 93 municipalities, distributed across the Pacific, Central, and Atlantic regions of Nicaragua. The sample was designed using as its sampling frame the cartography of the Population and Dwellings Census conducted in 2005 by the National Institute of Statistics (INEC) and it is representative at national, urban, and rural levels; it is probabilistic and stratified. The primary sampling units were *segmentos censales* and the second-stage units were households within each segment. Eight households were selected in each segment using systematic sampling with random start. In this paper we utilize pooled data from all years. As in Costa Rica, we limit our sample to the male working-age population (15–65) who report earnings. The resulting sample includes 14,669 observations on formal and informal workers over the entire panel.

## 2.2 Identification of formal and informal work among wage employees and self-employed

In the literature, informality may be defined relative to the employer or the worker. In this paper we focus on workers and follow the International Labor Organization's Thesaurus (ILO 2021) definition of informal work as comprising 'all remunerative work (i.e. both self-employment and wage employment) that is not registered, regulated or protected by the existing legal or regulatory framework, as well as non-remunerative work undertaken in an income-producing enterprise'. Our framework for identifying formal and informal workers is therefore based on whether or not regulations and mandatory labour protections are complied with. Formal employers and workers are those that comply with all registration requirements and labour protections, while informal

---

<sup>2</sup> We also conduct the analysis for Costa Rica using data from 2016, 2017, and pooled 2016–17. The results from each of these analyses are similar to those presented in this paper, both in the identification of informal sub-segments and in the proportion of workers who are involuntarily informal.

workers are all others. The common operationalization of ‘not protected by the existing legal or regulatory framework’ is whether or not the employer contributes to social security (through payroll taxes) for the employee. We follow this convention and identify formal wage employees as those whose employers contribute to social security for the worker.<sup>3</sup> In addition, we distinguish between wage employees and the self-employed.

### *Costa Rica*

In the ENAHO household surveys, wage employees are self-identified as ‘wage employees, unpaid assistants or private household workers’ (including domestic servants). For private household wage employees, the household for whom they work is considered the employer. Self-employed workers are those who self-identify as own-account workers or owners of firms (employers).

The common operationalization of identifying formal wage employees as those whose employers contribute to social security for the worker makes sense in Costa Rica, as social security (which provides both health care and pensions) is the most widespread and sought-after social protection, and payment of social security contributions is the most rigorously enforced tax. In fact, social security is mandated for all workers, including wage employees and the self-employed.

Self-employed workers in Costa Rica are legally required to be registered with both the Costa Rica Social Security Agency (CCSS in Spanish) and the Ministry of Finance (known as *Hacienda* in Costa Rica). Moreover, every private contractor is required to verify that a self-employed worker who is offering goods or services to the business is registered with both institutions before hiring any of their services. Each entity, the CCSS and the Ministry of Finance, is in charge of enforcing its own law and taxes, so it is possible for a self-employed worker to be registered with the CCSS but not with the Ministry of Finance (and vice-versa). However, the self-employed are subject to a ‘special regime’. This is needed because the ‘normal’ way workers contribute to social security is through their employer, which self-employed workers do not have. So the self-employed must pay both the employer’s and employee’s social security contributions.<sup>4</sup>

For self-employed workers to fully comply with the law in Costa Rica they 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 institution<sup>5</sup>

---

<sup>3</sup> Public-sector workers are also automatically included as formal-sector employees. Most public-sector workers are registered with the social security system. However, some public-sector workers, such as teachers, are affiliated with an alternative pension system but have to contribute to the social security system for health insurance and pay other mandatory payroll taxes.

<sup>4</sup> However, low-wage self-employed workers are subsidized by the government and therefore pay lower social security taxes. For example, workers earning less than 78.85 per cent of the minimum wage pay 6.43 per cent of their income into the social security system under the special regime; workers earning between 78.85 and 200 per cent of the minimum wage pay 9.3 per cent; workers earning between two and four times the minimum wage pay 12.28 per cent. The maximum social security tax for those in the special regime is 17.62 per cent of earnings (as of September 2017). This graduated payment scale by income is designed to encourage all self-employed workers, including the poor, to register with the social security system.

<sup>5</sup> Although ENAHO does not enquire directly about registration with the Ministry of Finance, it does enquire about registration with national records—which is more common for bigger and or employers—and any other public entity, which includes the Ministry of Finance. Also, those keeping formal accounting books are likely to do so for tax purposes.

or keep formal accounts for reporting to the government. We estimate that in Costa Rica 24 per cent of male wage employees and 71 per cent of male self-employed are informal.

### Nicaragua

In the FIDEG survey, a wage employee is self-identified as an ‘employee or labourer’ (including domestic servants). As in Costa Rica, we follow the convention that identifies formal wage employees as those whose employers contribute to social security for the worker. Other wage employees are informal.

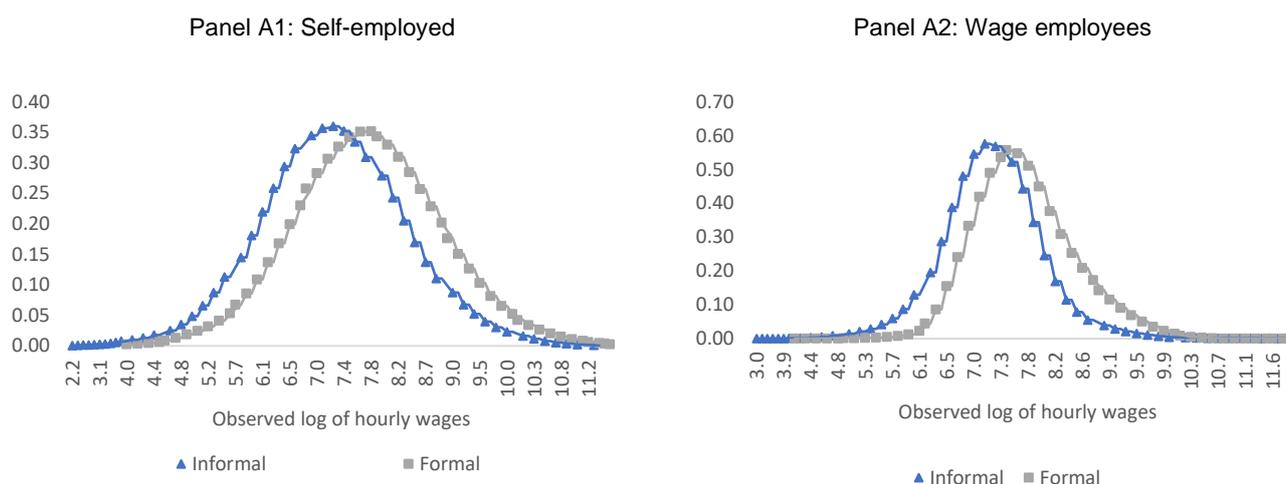
Self-employed workers are those who self-identify as own account workers or owners of firms (employers). Self-employed workers are not legally required to contribute to social security in Nicaragua. Self-employed workers can personally and voluntarily pay social security contributions through the *seguro facultativo*. However, we estimate that very few—approximately 1.5 per cent of self-employed workers—do so. Workers who do not contribute to social security still have access to local public health clinics. Formal self-employed workers are those who are registered with social security in any capacity. We estimate that in Nicaragua 70 per cent of male wage employees and over 98 per cent of male self-employed are informal.

## 2.3 Descriptive statistics

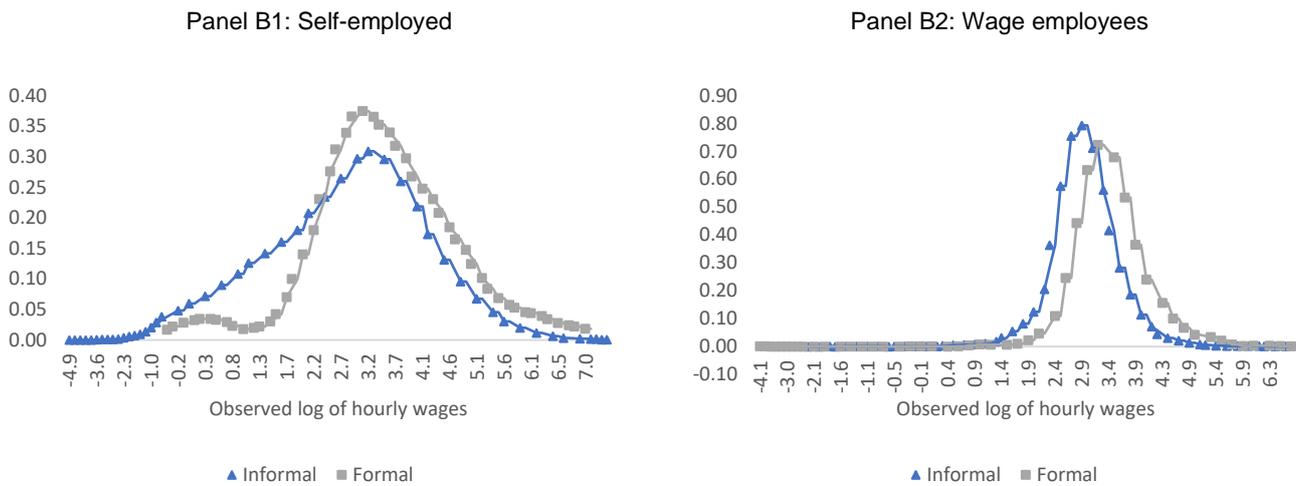
Figure 1 presents the distribution of observed wages in formality and informality for the self-employed and wage employees in Costa Rica and Nicaragua. The informal and formal distributions overlap, showing that not all informal workers earn less than formal workers, and suggesting that some workers may earn more as informal workers than they would as formal workers. Note that the distributions in Figure 1 do not control for different distributions of productivity-related characteristics of workers in the formal and informal segments and so can only be suggestive, and not direct evidence, of segmentation. We present more direct evidence later.

Figure 1: Densities of observed log of hourly wages for formal and informal workers, Costa Rica and Nicaragua

### Panel A: COSTA RICA



**Panel B: NICARAGUA**



Source: authors' calculations from 2018 ENAHO (Costa Rica); 2009–17 FIDEG panel (Nicaragua).

Table 1 presents the characteristics of formal and informal workers in Costa Rica and Nicaragua. Labour force participation and employment/population rates are over 10 percentage points higher in Nicaragua than in Costa Rica. Compared with Costa Rica, in Nicaragua there are more self-employed workers and fewer private wage employees. Public-sector wage employment is over twice as large in Costa Rica as in Nicaragua. The characteristics of formal and informal workers are, however, similar in Costa Rica and Nicaragua. Formal workers are more likely to be in mid-career (26–45 years old), while informal workers are more likely to be the oldest and youngest workers. Less educated workers are disproportionately informal, while more educated workers are disproportionately formal.

Informal work is much more prevalent in Nicaragua than in Costa Rica (Table 1). This is likely due to several differences in the economies and institutions in Nicaragua and Costa Rica. For example, enforcement of social security payroll taxes is stronger in Costa Rica than Nicaragua. Social security inspectors in Costa Rica can levy penalties and even close down firms that are not complying with regulations. Formal self-employment is particularly small in Nicaragua compared with Costa Rica. As noted previously, contributing to social security is legally required for self-employed workers in Costa Rica but not in Nicaragua. In addition, Costa Rican regulations encourage self-employed workers to contribute to social security by allowing some to contribute less than the full contributions of wage employees. Finally, large firms, export producers, international financial services providers, and foreign direct investors employ a larger proportion of wage employees in Costa Rica than in Nicaragua; these types of firms are generally more likely to comply with social security regulations than others.

Table 1: Descriptive statistics: Costa Rica and Nicaragua (men only), by formality status

	Costa Rica				Nicaragua			
	Working age male population	All male workers	Formal	Informal	Working age male population	All male workers	Formal	Informal
Total sample size	11,563	8,406	5,362	3,044	17,088	14,669	2,488	12,181
Age group (%)								
15–25	28.6	17.0	15.6	19.5	36.9	33.1	21.8	35.4
26–35	21.5	25.5	27.9	21.3	22.3	23.9	29.4	22.8
36–45	18.3	23.0	24.7	20.0	17.0	18.4	21.1	17.8
46–55	17.0	20.3	20.4	20.0	14.5	15.5	19.6	14.7
56–65	14.6	14.2	11.3	19.2	9.3	9.1	8.1	9.3
Average age	37.0	39.7	39.0	40.4	33.7	34.5	36.9	34.0
Education (%)								
none or incomplete primary	11.7	11.8	7.3	19.7	16.1	16.9	5.7	19.2
complete primary	51.4	48.9	44.4	56.9	40.9	43.4	27.3	46.6
complete secondary	17.7	18.4	21.0	13.8	33.3	30.9	41.5	28.7
some tertiary or post-graduate	19.1	20.9	27.3	9.7	9.7	8.8	25.5	5.5
Employment								
self-employed	15.8	21.8	9.7	43.0	29.5	34.4	3.2	40.8
private wage employee	48.3	66.4	71.8	56.9	51.8	60.4	65.9	59.2
public wage employee	8.5	11.7	18.4	0.0	4.5	5.2	30.8	0.0
Labour force status								
employed	72.7				85.8			
unemployed	5.9				3.0			
out of labour force	21.4				11.2			

Source: authors' calculations from 2018 ENAHO (Costa Rica); 2009–17 FIDEG panel (Nicaragua).

### 3 Methodology: involuntarily or voluntarily informal?

In this section we present the methodology for identifying the voluntary or involuntary nature of informality in Costa Rica and Nicaragua. We use the methodology developed in Günther and Launov (2012), which allows the researcher to empirically identify unobserved heterogeneity within the informal sector. This empirical model allows us to identify how many sub-segments exist within informality, and whether workers are in each informal sub-segment voluntarily or involuntarily. The technique also estimates the relative number of voluntary vs. involuntary informal workers. Using this model, Günther and Launov (2012) present evidence that in Côte d'Ivoire informal work is composed of one voluntary and one involuntary sub-segment, with the proportion of informal workers who are voluntary (55.2 per cent) slightly higher than the proportion of informal workers who are involuntary (44.8 per cent). Salem and Bensidoun (2012) apply the same methodology to data from Turkey and find that the vast majority of informal workers in Turkey (more than 89 per cent) would earn more in formal work and are therefore involuntarily informal. In contrast, Harati (2013) finds no evidence of involuntary informality in Egypt. In fact, Harati (2013) finds evidence that in Egypt 25–35 per cent of formal wage employees would maximize their earnings in one of the two informal sub-segments.

As far as we know, ours is the first study to use the Günther and Launov (2012) technique to examine whether informality is voluntary or involuntary using data from a low-income and a middle-income country in Latin America, which is the region where it is most likely for studies to find that workers are voluntarily in the informal sector (e.g. Maloney 2004).

Günther and Launov (2012) develop what they describe as a finite mixture model with self-selection. Finite mixture models, also known as latent class models, are a statistical/probabilistic technique for representing the presence of sub-populations within an overall population distribution. The model identifies the overall distribution as a weighted sum of a finite number of classes of other distributions. The weights are the estimated proportion of workers in each sub-population. The Günther and Launov (2012) technique examines the overall distribution of wages. It is assumed that the researcher can directly identify formal workers (which we do using the classification described previously). Latent wage equations describe the determinants of earnings in each sub-segment. Mincer-style wage equations are assumed to be the same for all workers in each unobserved sub-segment within the population of informal workers. Because we do not have wage data on those not working, the sample of workers is a non-random sample of the population. To address this issue, the wage equation is estimated as a self-selection model in that the Mincer wage equations are estimated using Heckman's (1979) self-selection model.

The following wage equations are estimated:<sup>6</sup>

$$\ln y_{ij} = x'_i \beta_j + \mu_{ij} \quad (1)$$

where  $i$  denotes the individual and  $j$  the segment of the labour market. The dependent variable in the Mincer wage equation is the log of hourly wages  $\ln y_{ij}$ . The independent variables ( $x_i$ ) include: age, age squared, region (rural/urban), formal education dummies<sup>7</sup>, English skills, and vocational education. Coefficients are estimated for each sub-segment  $j$ .

Furthermore, the observed distribution of earnings depends on the decision of entering (or not) the labour market. This decision is assumed to be a function of personal characteristics  $z_i$ :

$$y_{is} = z'_i \gamma + u_{is}, u_{is} \sim N(0,1) \quad (2)$$

such that earnings  $y_{ij}$  are observed only if the outcome of this last equation is positive ( $y_{is} > 0$ ). Additional instrumental variables are needed to identify the selection correction. These variables should be correlated to the choice of whether or not to enter employment but not be correlated to the error terms in the Mincer wage equations. We use the presence of a partner (i.e. spouse), number of children, and access to government-provided potable water and sanitation system as instrumental variables  $z_i$ . In Nicaragua, in both equations, we also include year-fixed effects.

---

<sup>6</sup> See Appendix B for details.

<sup>7</sup> In Costa Rica these are complete primary, complete secondary, and some tertiary, with incomplete primary as the reference category; in Nicaragua they are incomplete primary, complete primary, incomplete secondary, complete secondary and some tertiary, with no formal education as the reference category.

It is assumed that the (segment-specific) error terms of the above equations follow a bivariate normal distribution with correlation coefficient  $\rho$ , and, therefore, that the observed distribution of earnings in the  $j$ -th segment of the labour market is given by:<sup>8</sup>

$$f(y_{ij}|y_{is} > 0) = \frac{\varphi\left(\frac{\ln y_{ij} - x'_i \beta_j}{\sigma_j}\right)}{\sigma_j \Phi(z'_i \gamma)} \Phi\left(\frac{z'_i \gamma + \left(\frac{\rho}{\sigma_j}\right) [\ln y_{ij} - x'_i \beta_j]}{\sqrt{1 - \rho^2}}\right) \quad (3)$$

where  $\varphi$  and  $\Phi$  denote the standard normal probability and cumulative density functions, respectively.

Finally, the (unconditional) probability  $\pi_j$  of an individual  $i$  belonging to a segment  $Y_j$  of the informal sector, and the distribution of observed wages, are modelled by the mixture:<sup>9</sup>

$$f(y_i) = \sum_{j=2}^J \pi_j f(y_{ij}|y_{is} > 0) \quad (4)$$

In this equation  $\pi_j$  also represents the proportion of workers in each sub-segment. Equation (4) is estimated by maximum likelihood.<sup>10</sup>

In practice, the finite mixture model is estimated several times assuming a different number of informal sub-segments. We consider one homogeneous informal segment, two informal sub-segments, and three informal sub-segments. The optimal number of latent informal sub-segments is determined by minimization of information criteria such as the Bayesian Information Criterion (BIC), the Consistent Akaike Criterion (CAIC), and the Hannan-Quinn Information Criterion (HQC). The results of this exercise provide the information that allows us to estimate the proportion of workers in each informal sub-segment,  $\hat{P}_j$ , for  $j=2, \dots, J$  sectors ( $\hat{P}_j$  being our estimate of  $\pi_j$  from equation (4)).  $J-1$  is the number of informal sub-segments.

A finding of heterogeneous informality does not necessarily imply that those informal sub-segments are segmented from the formal sector. To test for segmentation the characteristics of each worker are used to estimate the worker's expected wage in a specific sub-segment. The worker maximizes wages in the sub-segment where this estimated wage is the highest. If an informal worker is in the sub-segment where they earn the highest expected wage, then they are considered voluntarily informal. If an informal worker is not in a sub-segment where they would earn the highest expected wage, then they are considered to be involuntarily in that sub-segment. Using this information, we estimate the proportion of workers who would maximize income in each sub-segment,  $\tilde{P}_j$ . The difference between  $\hat{P}_j$  and  $\tilde{P}_j$  is an estimate of the proportion of workers who are involuntarily in sub-sector  $j$ .

This methodology identifies involuntarily informal workers as those whose (expected) wages would be higher in formal work or in the other informal sub-segment. This may mistakenly identify some workers as involuntary who are in fact voluntary because they are willing to accept lower wages for the flexibility of informality.

---

<sup>8</sup> Note that the correlation coefficient  $\rho$  does not depend on the segment  $j$ . This assumption ensures that the model is identifiable. See Günther and Launov (2012) for details.

<sup>9</sup> The formal sector is denoted by  $j=1$ .

<sup>10</sup> See Appendix B for details.

## 4 Results

Table 2 presents the CAIC, BIC, and HQC for models with different numbers of informal sub-segments in Costa Rica and Nicaragua. All three information criteria are minimized for two informal sub-segments for wage employees in Costa Rica and for the self-employed in Nicaragua. For wage employees in Nicaragua, the CAIC and the BIC are minimized for the two informal sub-segments model, while the HQC is not minimized even for three informal sub-segments. For the self-employed in Costa Rica, the HQC is minimized for the model with two informal sub-segments but the CAIC and BIC are minimized for one sub-segment. In the model that assumes two informal sub-segments for the self-employed in Costa Rica we found that the distributions of expected wages for the two informal sub-segments were not significantly different using the Kolmogorov-Smirnov test (Figure 2). The Kolmogorov-Smirnov test results found that in all other sub-segments the distributions between the two informal sub-segments are significantly different. In addition, when we estimate the model for the self-employed in Costa Rica assuming two informal sub-segments, we cannot reject the hypothesis that one of the informal sub-segments has no one in it. We conclude from this evidence that informal self-employment in Costa Rica is best characterized by one informal segment, which is what we present in the rest of the paper. The other three samples are characterized by two informal sub-segments.

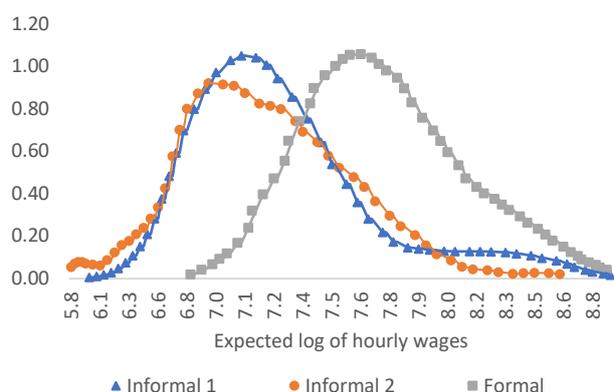
Table 2: Model selection criteria, Costa Rica and Nicaragua

<b>Panel A: Costa Rica</b>						
<b>Model</b>	<b>Self-employed</b>			<b>Wage employees</b>		
	CAIC	BIC	HQC	CAIC	BIC	HQC
One formal sector and:						
One informal	11440	11411	11288	25847	25818	25680
Two informal	11481	11441	11272	25702	25662	25472
Three informal	11546	11495	11280	25770	25719	25478
<b>Panel B: Nicaragua</b>						
<b>Model</b>	<b>Self-employed</b>			<b>Wage employees</b>		
	CAIC	BIC	HQC	CAIC	BIC	HQC
One formal sector and:						
One informal	21888	21838	21615	33399	33346	33093
Two informal	21823	21756	21458	32687	32616	32276
Three informal	21949	21865	21492	32743	32654	32228

Note: BIC = Bayesian information criterion; CAIC = Consistent Akaike information criterion; HQC = Hannan-Quinn information criterion.

Source: authors' calculations from 2018 ENAHO (Costa Rica); 2009–17 FIDEG panel (Nicaragua).

Figure 2: Densities of expected log of hourly wages for formal and informal sub-segments, Costa Rica, self-employed workers



Source: authors' calculations from the 2018 ENAHO.

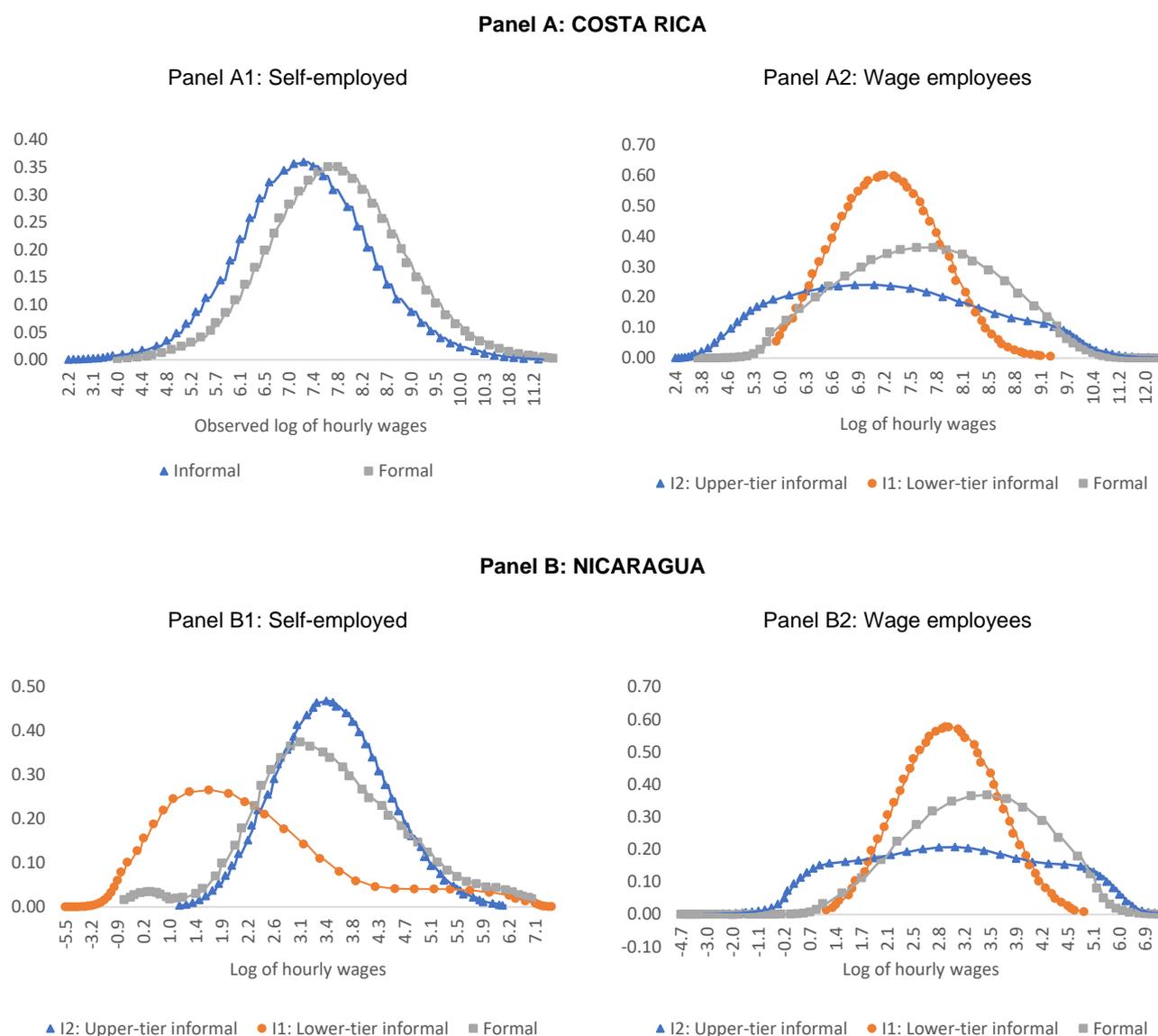
Figure 3 shows kernel density estimates of the distribution of observed wages in each sub-segment and country.<sup>11</sup> Kernel densities of the two informal sub-segments are presented for wage employees in Costa Rica (panel A2), the self-employed in Nicaragua (panel B1), and wage employees in Nicaragua (panel B2). In these three panels, there is one informal sub-segment (I1) where the distribution of wages is to the left of the distribution of wages in formal work, suggesting that wages are lower in that sub-segment than in formality for workers across the distribution. This is consistent with workers in this sub-segment being involuntarily informal. This pattern is also consistent with the I1 sub-segment as the lower-tier informal sub-segment. In the kernel density estimates for wage employees in both Costa Rica and Nicaragua, the distribution of wages in the second informal sub-segment (I2)—which is a voluntary informal sub-segment where some workers have a comparative advantage in informality while others are willing to accept lower wages than in formal work in exchange for the flexibility of informal work—has more variation, especially in the tails. Figure 2 suggests that some workers in this informal sub-segment earn the same as or more than in formality while some earn less than in formality. In particular, the larger upper tail in Figure 2 suggests that some workers in the upper-tier informal sub-segment may earn more than any worker in formal employment. This pattern is consistent with I2 as the upper-tier informal sub-segment.<sup>12</sup>

For the self-employed in Nicaragua, the distribution of wages for upper-tier informal workers is to the right of that for lower-tier informal workers and visually similar to the distribution in formal self-employment, suggesting that most workers in upper-tier informality earn more than those in lower-tier informality. For the self-employed in Costa Rica, which the information criteria evidence suggests has only one informal segment, the distribution of informal wages is to the left of that of formal wages, suggesting lower wages in informality compared with formality (panel A1).

<sup>11</sup> For this figure, informal workers are classified according to their posterior probability. Each informal worker is classified in the group in which their posterior probability is the highest. See Appendix A for details.

<sup>12</sup> Tables B1 and B2 in Appendix B present the results of the selectivity-corrected Mincer wage regressions and the selection equations for the self-employed and wage employees for Costa Rica and Nicaragua. Appendix B discusses these results.

Figure 3: Densities of log of observed hourly wages for formal and informal sub-segments, Costa Rica and Nicaragua



Source: authors' calculations from 2018 ENAHO (Costa Rica); 2009–17 FIDEG panel (Nicaragua).

Note that Figure 3 presents the distribution of observed wages in each sub-segment and as such does not control for differences in worker characteristics nor self-selection. Next we present the results of the Günther and Launov (2012) technique that does. Tables 3 and 4 compare the actual (estimated) distribution of workers in each sub-segment with the distribution of workers in each sub-segment if each worker were in the sub-segment with the highest predicted wage for that worker (the actual and maximizing distributions, respectively).<sup>13</sup> Table 3 presents the results for Costa Rica and Table 4 the results for Nicaragua.

<sup>13</sup> The number of workers in each sub-segment is slightly different from that reported in Table 1. This is because the sample used to construct Tables 3 and 4 includes only those for whom we have data on all the variables in the wage equations.

Table 3: Distribution of workers across sectors vs. maximizing distribution, Costa Rica

<b>Panel A: Self-employed</b>	<b>Actual estimated distribution</b>		<b>Maximizing distribution</b>		<b>Ratio maximizing/actual<sup>1</sup></b>	
Formal	0.293		0.953		3.248	
(P-value)	(0.000)		(0.000)		(0.000)	
95% conf. interval	0.273	0.313	0.858	1.049	2.861	3.635
Informal	0.707		0.047		0.066	
(P-value)	(0.000)		(0.336)		(0.000)	
95% conf. interval	0.687	0.727	-0.049	0.142	-0.069	0.202
<b>Panel B: Wage employees</b>	<b>Actual estimated distribution</b>		<b>Maximizing distribution</b>		<b>Ratio maximizing/actual<sup>1</sup></b>	
Formal	0.761		0.865		1.137	
(P-value)	(0.000)		(0.000)		(0.072)	
95% conf. interval	0.751	0.771	0.752	0.978	0.988	1.286
I1: lower-tier informal	0.151		0.057		0.376	
(P-value)	(0.000)		(0.172)		(0.059)	
95% conf. interval	0.119	0.184	-0.025	0.139	-0.271	1.024
I2: upper-tier informal	0.088		0.078		0.890	
(P-value)	(0.000)		(0.145)		(0.881)	
95% conf. interval	0.056	0.120	-0.027	0.183	-0.554	2.333

1 The P-value refers to the null hypothesis that the ratio equals 1.

Source: authors' calculations from the 2018 ENAHO.

Table 4: Distribution of workers across sectors vs. maximizing distribution, Nicaragua

<b>Panel A: Self-employed</b>	<b>Actual estimated distribution</b>		<b>Maximizing distribution</b>		<b>Ratio maximizing/actual<sup>1</sup></b>	
Formal	0.015		0.311		20.385	
(P-value)	(0.000)		(0.021)		(0.049)	
95% conf. interval	0.012	0.019	0.047	0.575	1.076	39.693
I1: lower-tier informal	0.568		0.002		0.004	
(P-value)	(0.000)		(0.853)		(0.000)	
95% conf. interval	0.398	0.738	-0.020	0.024	-0.036	0.043
I2: upper-tier informal	0.417		0.687		1.649	
(P-value)	(0.000)		(0.000)		(0.241)	
95% conf. interval	0.247	0.586	0.424	0.950	0.563	2.735
<b>Panel B: Wage employees</b>	<b>Actual estimated distribution</b>		<b>Maximizing distribution</b>		<b>Ratio maximizing/actual<sup>1</sup></b>	
Formal	0.305		0.750		2.459	
(P-value)	(0.000)		(0.000)		(0.000)	
95% conf. interval	0.295	0.315	0.617	0.883	2.015	2.902
I1: lower-tier informal	0.544		0.058		0.106	
(P-value)	(0.000)		(0.047)		(0.000)	
95% conf. interval	0.497	0.590	0.001	0.115	-0.001	0.214
I2: upper-tier informal	0.151		0.192		1.269	
(P-value)	(0.000)		(0.003)		(0.623)	
95% conf. interval	0.105	0.198	0.066	0.318	0.196	2.342

1 The P-value refers to the null hypothesis that the ratio equals 1.

Source: authors' calculations from the 2009–17 FIDEG panel.

The final column of Tables 3 and 4 presents the ratio of the proportion of workers in each sub-segment whose wages are maximized in that sub-segment ( $\tilde{P}_j$ ) divided by the actual proportion of workers in each sub-segment ( $\hat{P}_j$ ). If this ratio is significantly less than 1, this is evidence that the actual and maximizing distributions are different, and that workers are in informal work involuntarily. For example, the ratio of the proportion of workers in the maximizing distribution compared with the actual distribution for informal self-employed in Costa Rica is statistically less than 1. The p-value is 0.000, showing that the difference between the actual and maximizing proportions in that segment is statistically significant at any reasonable significance level. As a further illustration we also present the 95 per cent confidence intervals, and 1 is outside the 95 per cent confidence interval for the ratio, suggesting that the difference between the actual and maximizing distributions of workers in this sub-segment is statistically significant at the 5 per cent significance level.

The results presented in Tables 3 and 4 are consistent with the wage distributions shown in Figure 1. In both Nicaragua and Costa Rica, more wage employees would prefer to be formal workers than are actually formal, and the differences are statistically significant at 1 per cent in Nicaragua and 10 per cent in Costa Rica (p-value of 0.072). The overwhelming majority of wage employees in both Costa Rica and Nicaragua would earn more as formal employees than in any informal sub-segment (86.5 per cent in Costa Rica and 75 per cent in Nicaragua). This suggests that formal employment is limited, as not all of those who want to work there can, and therefore that some workers are informal involuntarily.

The results shown in Tables 3 and 4 also suggest that informality in wage employment is composed of two sub-segments, one (I1) where most informal workers earn less than they would as formal workers, and another (I2) where most workers earn the same or more as informal workers as/ than they would as formal workers. For example, in both countries there are more workers observed in lower-tier informal work (I1) than there would be if workers were found in the sub-segment where they maximized expected earnings. In Costa Rica the difference between the observed proportion and the maximizing proportion in the lower-tier informal sub-segment is approximately 10 per cent of all wage employees. This difference is statistically significant at the 10 per cent level (p-value of 0.059). In Nicaragua the difference between the actual and maximizing distributions for lower-tier informality (I1) is 49 per cent of all wage employees. This difference is statistically significant at the 1 per cent level.

On the other hand, for the upper-tier informal sub-segment (I2) in both Costa Rica and Nicaragua the difference between the actual and maximizing distributions for wage employees is not statistically significant. Indeed, for wage employees in Nicaragua more workers would maximize estimated earnings than are actually in upper-tier informality, although this difference is not statistically significant. These results suggest that most wage employees in the upper-tier informal sub-segment (I2) in both Costa Rica and Nicaragua are in that sub-segment voluntarily.

For the self-employed in Nicaragua, the proportion who maximize expected earnings in formality is 31 per cent, compared with only 1.5 per cent observed in that segment (the difference is statistically significant at 5 per cent). The results presented in panel A of Table 4 also suggest heterogeneity in informal self-employment, where almost no workers maximize wages in the lower-tier informal sub-segment (I1), while the majority of all self-employed workers (69 per cent) maximize earnings in the upper-tier informal sub-segment.

For the self-employed in Costa Rica, for whom the evidence suggests there is only one sub-segment, we estimate that the great majority of workers (over 95 per cent) would earn more in formality than in informality. In addition, the proportion of workers who maximize expected

earnings in informality (5 per cent) is not significantly different from zero. These estimates suggest that almost all of the informal self-employed in Costa Rica are in that sub-segment involuntarily.

The proportion of workers estimated to be involuntarily informal (that is, who are not in the sub-segment where they are maximizing their expected wages) is calculated by summing the difference between the actual and maximizing distributions in the informal sub-segments. Involuntary informality is more prevalent in Nicaragua than in Costa Rica. This reflects both the higher proportion of formal workers in Costa Rica and the higher proportion of informal workers who are involuntarily informal in Nicaragua. Among wage employees our results suggest that in Costa Rica 10 per cent of all wage employees (representing 43 per cent of informal wage employees) are involuntarily informal, while 66 per cent of all self-employed (representing 93 per cent of informal self-employed) are involuntarily informal. In Nicaragua, we estimate that 44 per cent of wage employees (representing 63 per cent of informal wage employees) are involuntarily informal, while 30 per cent of all self-employed (and 29 per cent of informal workers) are involuntarily informal.

## 5 Conclusions

Contrary to the traditional labour market segmentation view of dualistic labour markets, where workers are involuntarily informal, some recent literature from Latin America suggests that there is no segmentation between informal and formal workers, and that almost all informal workers are voluntarily informal. We hypothesize that this finding is because informal work is heterogeneous, with an informal sub-segment where most workers are voluntarily informal, and another sub-segment where most workers are involuntarily informal. A key purpose of this paper is to distinguish voluntary informal workers from involuntary informal workers. We use a methodology developed in Günther and Launov (2012) that explicitly allows for multiple informal sub-segments, and enables us to empirically identify different sub-sectors within the informal sector without ad hoc assumptions and to test for the voluntary and involuntary nature of workers in each sub-sector.

We find evidence of heterogeneous informality in wage employment in both Nicaragua and Costa Rica, with one informal sub-segment where most workers are voluntarily informal and another informal sub-segment where most workers are involuntarily informal. We also find evidence of a different wage structure between informal self-employment and informal wage employment. For self-employed workers we present evidence of two informal sub-sectors in Nicaragua: a lower-tier informal sub-segment where most workers would earn more in another sub-segment and are involuntarily in that sub-segment, and another where most workers are voluntary because they maximize wages in that sub-segment. On the other hand, our evidence suggests only one informal sub-segment among the self-employed in Costa Rica, where almost all workers are involuntary.

In Nicaragua, our results suggest that 44 per cent of wage employees (representing 63 per cent of informal wage employees) are involuntarily informal, while 30 per cent of self-employed (representing 29 per cent of informal self-employed) are involuntarily informal. In Costa Rica, our results suggest that 10 per cent of wage employees (representing 43 per cent of informal wage employees) are involuntarily informal, while 66 per cent of self-employed (representing 93 per cent of informal self-employed) are involuntarily informal.

Our results from an upper-middle income country and a lower-middle income country in Latin America differ from the conclusions of other recent literature that finds evidence that informality in Latin America is largely voluntary. One reason that our results differ may be that we explicitly

consider heterogeneous informality and we allow the data to determine who is voluntary and involuntary within informal employment.

## References

- Alaniz, E., T.H. Gindling, C. Mata, and D. Rojas (2020). ‘Transforming Informal Work and Livelihoods in Costa Rica and Nicaragua’. WIDER Working Paper 100/2020. Helsinki: UNU-WIDER. <https://doi.org/10.35188/UNU-WIDER/2020/857-3>
- Antonopoulos, R. (2009). ‘The Unpaid Care Work–Paid Work Connection’. Working Paper 86. Geneva: Policy Integration and Statistics Department, International Labour Office.
- De Soto, H. (1989). *The Other Path*. New York: Harper and Row. <https://doi.org/10.2139/ssrn.1176661>
- Fields, G. (1975). ‘Rural–Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDCs’. *Journal of Development Economics*, 2(2): 165–87. [https://doi.org/10.1016/0304-3878\(75\)90014-0](https://doi.org/10.1016/0304-3878(75)90014-0)
- Fields, G.S. (1990). ‘Labour Market Modelling and the Urban Informal Sector: Theory and Evidence’. In D. Thurnham, B. Salomé, and A. Schwarz (eds), *The Informal Sector Revisited*. Paris: OECD.
- 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’. *Revue d'économie politique*, 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>
- ILO (2021). *ILO Thesaurus* (February). Geneva: International Labour Organization. <https://www.ilo.org/inform/online-information-resources/terminology/thesaurus/lang-en/index.htm>
- Kucera, D., and L. Roncolato (2008). ‘Informal Employment: Two Contested Policy Issues’. *International Labour Review*, 147(4): 321–48. <https://doi.org/10.1111/j.1564-913X.2008.00039.x>
- Maloney, W. (1999). ‘Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico’. *World Bank Economic Review*, 13(3): 275–302. <https://doi.org/10.1093/wber/13.2.275>
- Maloney, W.F. (2004). ‘Informality Revisited’. *World Development*, 32: 1159–78. <https://doi.org/10.1016/j.worlddev.2004.01.008>
- Roncolato, L., and N. Radchenko (2016). ‘Women’s Labor in South Africa: Time Spent Doing Simultaneous Paid and Unpaid Work’. *International Journal of Time Use Research*, 13(1): 58–90. <https://doi.org/10.13085/eIJTUR.13.1.58-90>
- Salem, M.B., and I. Bensidoun (2012). ‘The Heterogeneity of Informal Employment and Segmentation in the Turkish Labour Market’. *Journal of the Asia Pacific Economy*, 17(4): 578–92. <https://doi.org/10.1080/13547860.2012.724546>

## Appendix 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 ML Estimator used by Günther and Launov (2012) we use Full Information Maximum Likelihood. This avoids splitting the model into two steps and obviates the necessity of computing the second-step covariance matrix correction. See Günther and Launov (2012) for additional details.

In general, 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. Each segment (the formal and the two informal sub-sectors) 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:

$$\ln y_{ij} = x'_{i} \beta_j + \mu_{ij}, i \in Y_j \quad (5)$$

where  $y_{ij}$  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. Furthermore, the observed distribution of earnings depends on the decision of entering (or not) the labour market. This decision is assumed to be a function of personal characteristics  $z$ :

$$y_{is} = z'_{i} \gamma + u_{is}, u_{is} \sim N(0, 1) \quad (2)$$

such that earnings  $y_{ij}$  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  $\rho$ , and, therefore, that the observed distribution of earnings in the  $j$ -th segment of the labour market is given by:<sup>14</sup>

$$f(y_{ij} | y_{is} > 0) = \frac{\varphi\left(\frac{\ln y_{ij} - x'_{i} \beta_j}{\sigma_j}\right)}{\sigma_j \Phi(z'_{i} \gamma)} \Phi\left(\frac{z'_{i} \gamma + \left(\frac{\rho}{\sigma_j}\right) [\ln y_{ij} - x'_{i} \beta_j]}{\sqrt{1 - \rho^2}}\right) \quad (3)$$

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

$$f(y_i) = \sum_{j=2}^J \pi_j f(y_{ij} | y_{is} > 0) \quad (4)$$

It is assumed that workers maximize earnings knowing the wage functions and, therefore, their expected earnings in each segment, given their personal characteristics. Thus, if there are no entry barriers to any segment of the labour market, each worker would be in the sector that pays the highest expected earnings given their own individual characteristics and the returns of these

---

<sup>14</sup> Note that the correlation coefficient  $\rho$  does not depend on the segment  $j$ . This assumption ensures that the model is identifiable. See Günther and Launov (2012) for details.

characteristics. This distribution is referred to as ‘maximizing distribution’, and it describes the proportion of workers that maximize expected earnings 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 entry barriers to some segments of the labour 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 of the existence of entry barriers that prevent some workers from being in the sector that pays the highest expected wage for their set of characteristics. In this case, to avoid unemployment, workers enter a sector with lower earning opportunities—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 involuntary 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,<sup>15</sup> 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:

$$\ln L = \sum_{i \notin \{Y_F, Y_I\}} \ln \Phi(-z'_i \gamma) + \sum_{i \in Y_F} \ln(\pi_F h(\theta_F, \rho | y_{iF}, y_{iS} > 0, x_i, z_i)) + \sum_{i \in Y_I} \left[ \ln \left( \sum_{j=2}^J \pi_{I_j} h(\theta_{I_j}, \rho | y_{iI_j}, y_{iS} > 0, x_i, z_i) \right) \right] \quad (6)$$

where  $h(\cdot) = \Phi(z'_i \gamma) f(\cdot)$  and  $f(\cdot)$  is given by (A1),  $\pi_F$  is the probability of belonging to the formal sector,  $\pi_{I_j}$  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 information criteria. The model is estimated using one, two, and three informal segments. The information criteria for each case are presented in Table 2.

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

$$\hat{E}[\ln y_{ij} | y_{iS} > 0, x_i] = x'_i \hat{\beta}_j + \hat{\rho} \hat{\sigma}_j \frac{\varphi(-z'_i \hat{\gamma})}{1 - \Phi(-z'_i \hat{\gamma})} \quad (7)$$

A comparison between the actual and maximizing distributions is presented in Tables 3 and 4. Finally, each informal worker can be classified in one of the sub-segments according to their posterior probability of belonging to a particular  $Y_j$ . Specifically, the mixing proportion  $\pi_j$  can be interpreted as the a priori probability that an observation belongs to the segment  $Y_j$ . Using the Bayes rule, the posterior probability that an observation  $i$  belongs to  $Y_j$  is given by  $\pi_j f(y_{ij} | y_{iS} > 0, \theta_j) / f(y_i)$ . Each informal worker is classified in the sub-segment in which this (estimated)

---

<sup>15</sup> Although not in which segment of the heterogeneous informal sector.

posterior probability is the highest. Results, presented in Figure 2, are based on this approach to classifying informal workers.

## **Appendix B: Estimated latent wage equations by sub-segment**

### **Costa Rica**

Table B1 presents the results of the selectivity-corrected Mincer wage regressions and the selection equations for the self-employed and wage employees in Costa Rica.

At least one of the identifying instrumental variables in the employment selection equation for both wage employees and the self-employed is statistically significant at 1 per cent (the presence of a partner). In the equation for the self-employed, access to public sanitation is also statistically significant.

For the self-employed, tertiary education has a statistically significant impact on wages for both formal and informal workers. Secondary education also has a statistically significant impact on informal wages. In both segments, fluency in English and living in the Central Valley have statistically significant positive impacts.

The data fit the wage equations for wage employees better than for the self-employed. For formal wage employees, all variables are statistically significant. In upper-tier informal work, secondary and tertiary education are significant, as are English skills. In the lower-tier informal sub-segment, tertiary education, vocational education, and English language skills have significant and positive impacts on wages.

### **Nicaragua**

Table B2 presents the results of the selectivity-corrected Mincer wage regressions and the selection equations for the self-employed and employees in Nicaragua. The presence of a partner is a statistically significant, identifying instrumental variables in the employment selection equation for both wage employees and the self-employed.

For all sub-segments of both informal self-employed and wage employees, the coefficients on the education variables are statistically significant and increasing for each education level. Returns to education are higher in the upper-tier informal sub-segment than in the lower-tier informal sub-segment.

Table B1: Costa Rica: Wage and selection regression results

## Panel A: Self-employed

<b>Variables</b>	<b>Formal</b>	<b>Informal</b>	<b>Selection</b>
Age	-0.108** (0.0428)	-0.0176 (0.0288)	0.263*** (0.00965)
Age squared	0.00113** (0.000470)	0.000136 (0.000315)	-0.00292*** (0.000115)
Presence of a partner			0.701*** (0.0514)
Potable water			0.00470 (0.106)
Public sanitation system			-0.181*** (0.0526)
Central Valley	0.218** (0.0851)	0.162*** (0.0541)	0.0247 (0.0458)
Complete primary education	-0.0700 (0.177)	0.0177 (0.0692)	
Complete secondary education	-0.0350 (0.193)	0.348*** (0.0912)	
Some tertiary education	0.443** (0.194)	0.931*** (0.119)	
Vocational education	-0.181* (0.0970)	-0.0880 (0.0604)	
Fluency in English	0.425*** (0.136)	0.390*** (0.123)	
Prior probability ( $\pi$ )	0.293*** (0.0108)	0.707*** (0.0108)	
(Ln of) $\sigma$	0.0210 (0.0466)	-0.0089 (0.0454)	
$\rho$			-0.5112*** (0.0928)
Constant	10.29*** (0.992)	7.788*** (0.685)	-5.723*** (0.213)
Observations	4,929	4,929	4,929

Table B1: Costa Rica: Wage and selection regression results

## Panel B: Wage employees

Variables	Formal	Informal 1	Informal 2	Selection
Age	0.0153** (0.00698)	-0.00321 (0.00886)	-0.0290 (0.0248)	0.279*** (0.00663)
Age squared	-5.34e-05 (8.64e-05)	9.57e-05 (0.000113)	0.000447 (0.000308)	-0.00348*** (8.44e-05)
Presence of a partner				0.681*** (0.0376)
Access to publicly-provided potable water				0.0512 (0.0768)
Access to publicly-provided sanitation				-0.0273 (0.0382)
Central Valley	0.0568*** (0.0150)	0.0968*** (0.0317)	0.116 (0.0903)	0.152*** (0.0320)
Complete primary education	0.162*** (0.0241)	0.0620* (0.0370)	0.0867 (0.0958)	
Complete secondary education	0.432*** (0.0279)	0.0490 (0.0677)	0.330** (0.168)	
Some tertiary education	1.011*** (0.0299)	0.688*** (0.173)	0.594*** (0.206)	
Vocational education	0.0830*** (0.0159)	0.110* (0.0614)	0.238* (0.138)	
English language skills	0.250*** (0.0249)	0.206** (0.0857)	0.834*** (0.274)	
Prior probability ( $\pi$ )	0.761*** (0.00536)	0.1513*** (0.0160)	0.0877*** (0.0158)	
(Ln of) $\sigma$	-0.678*** (0.0184)	-1.0560*** (0.0912)	-0.0446 (0.0799)	
$\rho$				-0.4046*** (0.0634)
Constant	6.753*** (0.143)	7.112*** (0.166)	7.408*** (0.475)	-4.670*** (0.131)
Observations	9,499	9,499	9,499	9,499

Note: in Costa Rica we do not consider vocational education to be part of formal education; robust standard errors in parentheses; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Source: authors' calculations from the 2018 ENAHO.

Table B2: Nicaragua: Wage and selection regression results

## Panel A: Self-employed

<b>Variables</b>	<b>Formal</b>	<b>Informal 1</b>	<b>Informal 2</b>	<b>Selection</b>
Age	0.144 (0.170)	0.0200 (0.0330)	0.0504* (0.0257)	0.177*** (0.00777)
Age squared	-0.00147 (0.00173)	-0.000134 (0.000379)	-0.000374 (0.000310)	-0.00209*** (9.73e-05)
Number of children under 12 yrs old in HH				0.0120 (0.0154)
Presence of a partner				0.956*** (0.0440)
No or incomplete primary education	-0.474 (0.353)	0.110 (0.115)	0.152 (0.120)	0.107* (0.0605)
Complete primary education	0.0682 (0.417)	0.468*** (0.116)	0.745*** (0.140)	-0.0882 (0.0679)
Incomplete secondary education	-0.126 (0.439)	0.557*** (0.127)	1.206*** (0.138)	-0.379*** (0.0611)
Complete secondary education	0.192 (0.554)	0.623*** (0.137)	1.485*** (0.174)	-0.366*** (0.0731)
Some tertiary education	1.314** (0.540)	1.249*** (0.150)	2.080*** (0.190)	-0.561*** (0.0711)
Training	-0.125 (0.331)	0.0933 (0.0898)	0.0932 (0.108)	
Year dummies	YES	YES	YES	YES
Prior probability ( $\pi$ )	0.0152*** (0.00188)	0.4166*** (0.0648)	0.5682*** (0.0648)	
(Ln of) $\sigma$	0.0981 (0.0976)	-0.1697* (0.0936)	0.4399*** (0.0276)	
$\rho$				-0.3591*** (0.0720)
Constant	-0.0745 (4.257)	2.236*** (0.732)	0.401 (0.537)	-3.009*** (0.146)
Observations	6,607	6,607	6,607	6,607

Table B2: Nicaragua: Wage and selection regression results

## Panel B: Wage employees

<b>Variables</b>	<b>Formal</b>	<b>Informal 1</b>	<b>Informal 2</b>	<b>Selection</b>
Age	0.0200** (0.00789)	0.0139*** (0.00431)	0.0544*** (0.0168)	0.143*** (0.00680)
Age squared	-0.000166 (0.000106)	-0.000154** (6.03e-05)	-0.000640*** (0.000226)	-0.00194*** (8.84e-05)
Urban	0.139*** (0.0270)	0.154*** (0.0197)	0.302*** (0.0776)	-0.601*** (0.0353)
Training	0.112*** (0.0257)	0.180*** (0.0284)	-0.0508 (0.116)	
No or incomplete primary education	-0.00217 (0.0638)	0.0411* (0.0215)	0.316*** (0.102)	0.202*** (0.0565)
Complete primary education	-0.0568 (0.0638)	0.105*** (0.0286)	0.388*** (0.116)	0.192*** (0.0622)
Incomplete secondary education	0.191*** (0.0614)	0.171*** (0.0265)	0.485*** (0.116)	-0.0340 (0.0547)
Complete secondary education	0.189*** (0.0655)	0.132*** (0.0353)	0.596*** (0.141)	0.0989 (0.0631)
Some tertiary education	0.643*** (0.0665)	0.322*** (0.0589)	1.172*** (0.185)	-0.0241 (0.0619)
Number of children under 12 yrs old in HH				0.0281** (0.0126)
Presence of a partner				0.686*** (0.0363)
Year dummies	YES	YES	YES	YES
(Ln of) $\sigma$	-0.499*** (0.0448)	-0.9377*** (0.0368)	0.0112 (0.0597)	
Prior probability ( $\pi$ )	0.3051*** (0.00518)	0.5436*** (0.0230)	0.1513*** (0.0226)	
$\rho$				-0.3441*** (0.0646)
Constant	2.505*** (0.147)	2.346*** (0.0765)	1.522*** (0.316)	-1.439*** (0.124)
Observations	10,257	10,257	10,257	10,257

Note: robust standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Source: authors' calculations from the 2009–17 FIDEG panel.