Informal employment in Chile: unobserved heterogeneity and gender-based differences

Tesis para optar al grado de
Magíster en Análisis Económico

Alumno: Ignacio Orueta Bennett
Profesor Guía: Esteban Puentes

Santiago, Junio de 2021

Informal employment in Chile: unobserved heterogeneity and gender-based differences

Ignacio Orueta Bennett
iorueta@fen.uchile.cl

June 2021

Abstract: This paper studies the presence of different groups of workers in the informal labour sector in Chile. To deal with this unobserved heterogeneity, a Finite Mixture Model (FMM) is developed, finding the optimal number of segments in the informal sector. The model also generates different wage equations for each segment, allowing to estimate in which segment workers would maximize their wages. With this, the estimated sizes of each segment and the theoretical sizes that would result from a free mobility context are compared to test the existence of segmentation. Results show statistical support for 2 informal segments in all the specifications estimated. Then, the segmentation test showed that formal workers are over-represented and that voluntary informal employment seems to be the most desirable choice. However, in a hypothetical free mobility scenario, there is also some mobility between sectors, especially for informal workers. This is evidence that informal workers may be facing some restrictions in their wage-maximizing process too. Nevertheless, voluntary informality seems to be the most prevalent situation in the Chilean labour market.

Keywords: Informal labour market, segmentation, finite mixture model.
JEL Codes: J46, O17, C14, N36.

Contents

1 Introduction

Informal labour markets have been a focus of study in economic research for a long time. This is particularly true for developing countries, where the duality of the labour market has been a great topic of interest. However, there is no consensus about the reasons that workers have to be part of the informal sector. While traditional theories describe the informal sector as a low productivity sector, where workers enter after being segmented from the formal one in the first place (Lewis, 1954), an alternative view suggests that some workers may choose the informal sector voluntarily. This second theory proposes that despite the average wage tending to be lower for informal workers, the labour market could still be competitive if people can move freely between sectors. Hence, there is competition between sectors as long as workers are free to choose the sector where they want to work. (Heckman & Hotz, 1986).

As both theories contradict each other, recent research has combined these approaches, stating that there is a coexistence between voluntary and involuntary (segmented) informal segments. This lead to a heterogeneous structure of the informal sector, where we can find a ”*lower tier*” with involuntary informal workers and a ”*upper tier*” with voluntary informal workers. (Fields, 1990).

In this paper, these hypotheses are addressed by developing a model where I can test two critical dissents between theories. First, I determine whether the Chilean informal labour market has a homogeneous or heterogeneous structure without imposing any subjective division. This is done by allowing the data itself to determine the number of segments that make the best fit. Secondly, after finding evidence of two segments, I test if that division corresponds to a competitive or a segmented labour market by calculating the number of workers that would maximize their wages in each segment to see if it is similar to the actual distribution of the data. This is done by running separate models for men and women to determine if there are gender-based differences in the structure and mobility of workers.

Despite that the structure of the informal sector is not fully understood yet, its research is crucial for policy-makers. Tackling informality needs a good understanding of the opportunities and motivations of workers who are in that situation. Therefore, a starting point is to find out differences inside this sector, quantifying them. By doing this, public policy should be able to determine how big and what type of incentives need to be implemented to formalize effectively these workers. For example, if we are in the presence of a heterogeneous informal sector with a big share of involuntary formal workers, we should rationalize that the formal sector has some entry barriers and policies need to be focused on making it more flexible to allow them to formalize. On the other hand, a heterogeneous informal sector with a big share of voluntary informal could indicate that maybe there are high labour costs or taxes that discourage workers from moving to the formal sector.

This research contributes to the literature on the Chilean informal labour market in three different ways. First, it complements the existing works about informality by adding a different methodology to test the heterogeneity of the informal sector. This leads to a better understanding of the motivations and characteristics of informal workers. Secondly, it also quantifies the relative size of voluntarily and involuntarily workers, which is something that no one has done for the Chilean case. Finally, it follows a gender approach by estimating the model separately for males and females. This allows determining if segmentation and competition in the labour market have differences based on gender.

The paper is structured as follows. Section 2 has a literature review related to informal labour markets. Section 3 specifies the econometric model and the chosen tests for market segmentation. Then, in Section 4 the data of the Chilean labour market is presented, while Section 5 proceeds with the empirical application. In section 6 there is a discussion about the results and, finally, summarize and conclusions are found in Section 7.

2 Literature Review

The nature of informal employment and the driving mechanisms of its expansion and reduction have been studied for decades. However, the reasons to be part of that sector are not fully understood yet. On one hand, the traditional dual market theory, which was first formulated by Lewis (1954), stated that the labour market is composed of a modern capitalist sector and a subsistence sector. Here, the subsistence sector is a segmented part of the labour market, where people escape from unemployment once they are segregated from the formal one. This point of view sees informal jobs as a strategy of last resort as workers are excluded from their most desirable choice, entering the informal sector only to avoid unemployment. Therefore, formal workers earn more than their identical counterparts who are in the informal sector because there are some barriers that they are not able to surpass to enter the formal sector. (Harris & Todaro, 1970; Stiglitz, 1976).

On the other hand, studies from Hart (1973) and Rosenzweig (1988) suggested that, due to observable and unobservable characteristics, workers may voluntarily choose the informal sector through individual cost-benefit analysis. Hence, there is a competitive situation between sectors, where workers are maximizing their utility rather than their earnings when they self-select the informal sector. A clear example of this could be that the flexibility in working hours of an informal job might be very valuable for some individuals and is worth taking a wage cut for more time available to fulfil other duties, like child care. Additionally, workers may have comparative advantages in the informal sector that allow them to work more efficiently. This implies that there are informal workers who are better at not taking a formal job for which they would be selected.

As we see, these theories about segmentation and competition contradict each other. Therefore, researchers have been trying to test which and where these ideas hold empirically. At first, in support of the segmented theory, Heintz & Posel (2008) showed persistent earnings differentials for formal and informal workers in South Africa, which is an indicator of segmentation. However, this is not conclusive evidence in favour of the segmented markets theory because a labour market with different returns to observable characteristics could still be a competitive one as long as individuals can move between sectors (Heckman & Hotz, 1986). Additionally, Dickens & Lang (1985) discover some evidence that supports the segmented markets theory by finding the existence of non-economic barriers that prevent non-whites from entering the high-wage sector in the United States. This was done by estimating a switching regression model, where the observable characteristics determine the sector where a worker should be selected in the absence of barriers. Then, as this model assigned more non-whites in the formal sector than the real data, they concluded that non-whites were discriminated against and faced entry barriers.

On the other hand, supporting the theory of the competitive markets, Magnac (1991) did not find any significant costs or barriers to entry for individuals who want to join the formal sector by developing a labour supply model for female Colombian workers. Therefore, despite the existence of wage differences between sectors, women in Colombia were not segmented from working formally.

Moreover, Pratap & Quintin (2006) did not find any significant differences in measures of job satisfaction between the Argentinian formal and informal sectors when estimating wage differences. This is evidence that workers do not always prefer a formal job. However, job satisfaction was measured as the fraction of workers that were looking for another job in each sector, controlling for observable characteristics. Hence, this evidence is not sufficient to support the idea of a competitive labour market as workers' motives to change jobs could not be related to voluntary or involuntary sector selection. Not wanting to change to another job does not mean that an informal worker did not try to enter the formal sector in the past, and looking for a job does not mean that informal workers will change their sector in the future.

As evidence is mixed, none of these theories sufficiently explains the structure of informal employment. With this, recent studies have combined these two opposite views and emphasized a more complex structure in the informal labour market. Fields (1990) was the first one to state that the informal sector has its internal duality, where both models -the exclusion and self-selection model- are operating simultaneously. Here, informal jobs are quite diverse, where some activities with easy entry have neither fixed hours of operation nor high set-up costs, while others have minimum working hours and high set-up costs.

Fields (1990) conceptualized these findings into what is known as the "*lower tier*" and "*upper tier*" of the informal sector. The "*lower tier*" is the easy entry segment, where there are no fixed working hours and is often represented by lower wages and unskilled self-employed workers since these activities are supposed to be oriented to subsistence. Then, the "*upper tier*" is the segment where we can find some entry barriers, higher capital or skill requirements, more regular labour relations arrangements and, hence, higher wages. Fields' framework is complemented by Maloney's (2004) vision of the "voluntary" and "involuntary" entry structure. Here, the heterogeneity of informal employment is driven by workers excluded from the formal sector (involuntary informal) and self-selected works that are a result of individual choice (voluntary informal).

There is some empirical evidence of this coexistence between these "*lower tier*" and "*upper tier*" segments in the informal sector. First, Cunningham and Maloney (1999) followed micro-enterprise workers that were originally from formal salaried work in Mexico. Doing a cluster analysis with demographic and firm variables, they found that roughly 70% of the entrepreneurs in all clusters entered voluntarily, while the other 30% did it involuntarily. They also found that workers involuntarily moving into the sector earned 25%–35% less than those who moved voluntarily, which is in line with Fields' (1990) framework. The authors said that some of the reasons that can explain the presence of voluntary informal workers could be the desire for greater independence and flexibility that offsets the instability of earnings. However, this approach was limited since it considered only small businesses, so going into formal salaried employment was never an option here.

Some Chilean studies also showed this voluntary/involuntary framework. For example, Contreras, Gillmore & Puentes (2017) found statistical support for partial queues for wage work between self-employed workers. This means that some individuals are self-employed because they were not selected into wage work, while others self-select this work in the first place. This was done by modelling the worker's decision to work in two stages. In the first stage, workers decided whether they want to become salaried workers, which is known as being in the queue (IQ). Then, the second stage model the probability of finding a job in the wage sector, which is referred as being chosen from the queue (CFQ). Finally, as the estimations support that only a fraction of workers decided to be in the queue (IQ), the rest of the workers self-select them into self-employment. This led to having two different self-employed workers: one group that chose to be part of the queue but was not selected for wage work and another group that was not in the queue and chose self-employment in the first place.

Then, the studies of Contreras, Mello & Puentes (2008) and Contreras & Puentes (2009) found partial queues for formal employment, meaning that a double selection model fits better the data than a simple selection model. Here, they measured informality as no contribution to social security. As before, double selection models consist of a first stage where people choose between formal and informal jobs and a second stage, where employers select some workers that were in the queue for the formal sector, meaning that some workers are not chosen and they are excluded from the formal sector. Hence, heterogeneity here was revealed as they found self-selection in informal jobs and some exclusion from formal employers.

Finally, Contreras, Mello & Puentes (2008) concluded that labour informality appears to be more a self-selection phenomenon than a result of rigidities in the Chilean labour market, where low educational levels and poor work experience are the major determinants of its existence. This was supported by the high probability of obtaining a formal job for workers who sought one.

As these papers show, testing the hypothesis about the structure of the informal labour market is not straightforward. Usually, the affiliation of individuals in each segment of the informal sector is unobservable since there is no information reporting the causes of informality. Hence, researchers do not know if the worker is informal because of self-selection (or not being in the queue as the Chilean studies) or because it was excluded from the formal sector (i.e. not chosen from the queue). This can lead to an arbitrary division of the informal sector based on observed characteristics or earnings and, therefore, can bias the results. (Günther & Launov, 2006). This bias comes from the fact that there are no observable characteristics that can effectively divide individuals between voluntary and involuntary segments, making partial observability a major problem. (Contreras & Puentes, 2009).

Taking into account these issues, Günther & Launov (2006, 2012) applied a Finite Mixture Model (FMM) with sample selection to estimate the unobserved heterogeneous structure of the informal labour market with only distributional assumptions. This methodology let the data determine the number of segments by separating informal workers with different wage returns to observable characteristics. They also used Heckman's correction (1979) to account for the selection bias caused by the employment decision of individuals. With this approach, the authors found 2 informal segments in the urban labour market of the Ivory Coast, each of them representing 37,7% and 28,4% of the labour force. They also found that workers in the informal segment with lower average earnings (i.e. the "*lower tier*") would have earned a higher wage in the formal sector, but, as they were not working there, they concluded that entry barriers existed. However, this did not happen for the informal segment with higher earnings (i.e. the ”*upper tier*”), so the authors argued that in that sector there were some comparative advantages that made these workers voluntary informal.

This evidence of both competitive and segmented employment supports Fields’ (1990) framework as they found some individuals that maximized their earnings in the formal sector, while others did it in the informal sector. The methodology was then replicated by Salem & Bensidoun (2012) with the same results as Günther & Launov (2006, 2012) for the labour market of Turkey. Then, applying the same methodology, Harati (2013) found different results for Egypt. This research showed that formal workers were the ones that would maximize their wages in the informal sector because of high costs of labour in the formal sector. However, the author argued that formal workers did not want to move to the informal sector because of the non-pecuniary benefits that were offsetting the potential wage-gains.

In this paper, I follow the work line of Günther & Launov (2006, 2012), applying a Finite Mixture Model (FMM) with sample selection in the Chilean labour market. As it was said before, this approach can model the unobserved heterogeneous structure of the labour market in this country and, at the same time, correct the sample selection bias in the earnings equations. (Günther & Launov, 2012). Additionally, the FMM specification is estimated separately for men and women to test differences between genders. If the results for men and women differ, then another source of heterogeneity arises, and it would not be accurate to estimate the FMM with a pooled sample. Hence, in a scenario with gender-based differences, we would improve the understandings of the informal labour market by doing the exercise with separated samples.

3 Estimation Strategy

As it was said before, testing the presence of heterogeneity in the informal sector has some complications. Despite common household surveys have information on whether workers are formal or informal, individual’s willingness to affiliate to each sector remains unobservable. Therefore, even assuming a heterogeneous structure for the informal sector, setting a determined number of segments based on observable characteristics need some assumptions that could bias the results. For example, it would be needed to artificially impose the number of segments and select the observable characteristics that would determine on which segment a worker belongs.

Following Salem & Bensidoun (2012), a labour market with J segments, where there is one homogeneous formal sector and $J-1$ heterogeneous informal sectors, can be written as [1]:

$$w_{i,F} = x_i'\beta_F + u_F, \quad \pi_F = \frac{N_f}{N} \tag{1}$$

$$w_{i,I} = \begin{cases} x_i'\beta_{I,1} + u_{I,1}, & \pi_1 = ? \\ \vdots & \\ x_i'\beta_{I,J-1} + u_{I,J-1}, & \pi_{J-1} = ? \end{cases} \tag{2}$$

Where $w_{i,F}$ and $w_{i,I}$ are the standard Mincer log-wages for formal and informal workers determined by observable characteristics, such as years of schooling and work experience. π_F is the share of formal workers and $\{\pi_1, \pi_2, ..., \pi_{J-1}\}$ represent the unknown share of each of the $J-1$ segments in the informal sector.

As an arbitrary segment division could bias the results, the estimation of the size of the informal segments (i.e. $\pi_1, \pi_2, ..., \pi_{J-1}$) needs to be done without any major assumption. Considering this, I apply a Finite Mixture Model that allows the data to determine the number of partitions that make the best fit. FMM models are a flexible semiparametric approach that is used to model unobserved heterogeneity without major distributional assumptions. (McLachlan, et al., 2019). This is useful in our case because it allows classifying informal workers into separate groups only if they have statistically different wage determinants. If this is the case, then an FMM will divide the informal sector into the J segments, each of them with a different wage equation.

3.1 Finite Mixture Model

Following the specification of Günther & Launov (2006), I assume that the labour market Y consists of J different segments, where each wage specific return β_j is identical for every worker in segment Y_j, but different across segments. Thus, each segment has its own wage equation, where errors are not correlated between segments, meaning that wage distributions are different and independent of each other.

I also assume that the affiliation of a given individual with wage w_i to any segment j is not observable. However, the probability of any individual earning w_j belonging to the segment j is given by $P(w_i \in Y_j) = \pi_j$. With these, the density of individual log-wages w_i is:

[1]Naturally, the formal sector could also have several segments. However, for the objective of the paper, this specification will be omitted to avoid unnecessary complexity.

$$f(w_i) = \sum_{j=1}^{J} \pi_j f(w_i|x_i\beta_j) \tag{3}$$

The mixture model uses the multinomial distribution to compute the probabilities for the j segments. Therefore, $\hat{\pi}_j$ is calculated with the linear predictions of the vector coefficients:

$$\hat{\pi}_j = \frac{exp(x_i\beta_j)}{\sum_{j=1}^{J} exp(x_i\beta_j)} \tag{4}$$

Then, any segment Y_j of the labour market has a segment-specific log-wage given by worker's individual characteristics:

$$w_i = x_i\beta_j + u_i, \quad u_i \sim N(0, \sigma_j^2|x_i, w_i \in Y_j), \tag{5}$$

Where x_i represent individual and observable characteristics that determine individual wages w_i. Using (3) and (5) it is easy to show that the expected wages of any individual are given by $E(w_i) = \pi_j \sum_{j=1}^{J} x_i\beta_j$.

3.2 Sample Selection

As it is well known, regression (5) might be misspecified since w_i is only observed in individuals where it exceeds their reservation wage (Heckman 1979). Therefore, the observed earnings may not be necessarily representative of the whole population, leading to a sample bias.

Following Heckman (1979), let the employment decision depend on personal characteristics z_i. Then, a probit is elaborated to model the selection equation:

$$e_i = z_i\gamma + v_i, \quad v_i \sim N(0,1) \tag{6}$$

Where e_i is the probability to participate in the labour market and being employed and $z_i\gamma$ reflects the individual decision to work. Thus, w_i is observed only if the selection variable $e_i > 0$. Additionally, the selection equation needs at least one variable that works as an exclusion variable. This means that this variable is excluded from the equation and, at the same time, it does not influence the dependent variable in the wage equation. (Heckman, 1979)

At next, if the residuals from the wage equation (u_i) and the selection equation (v_i) are correlated, the estimation of β_j will be biased. It can be assumed that the error terms follow a bivariate normal distribution and their correlation is equal to ρ_j:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \rho_j\sigma_j \\ \rho_j\sigma_j & 1 \end{pmatrix}\right) \tag{7}$$

Since this distribution is conditional on observing $e_i = 1$, we will have the following distribution of observed wages in the j^{th} segment:

$$f(w_{ij}|e_i > 0, x_i, z_i, \gamma, \beta, \sigma, \rho) = \frac{1}{\sigma_j \phi(z_i'\gamma)} \varphi\left(\frac{lw_i - x_i'\beta_j}{\sigma_j}\right) \Phi\left(\frac{z_i'\gamma + \frac{\rho_j}{\sigma_j}[w_i - x_i'\beta_j]}{\sqrt{1-\rho_j^2}}\right) \tag{8}$$

Where φ and Φ are the standard normal density and cumulative functions.[2] Additionally, we can derive the conditional distribution of wages in the whole labour market by estimating π_j as the probability that the individual i belongs to segment j. This will be given by:

$$f(w_i) = \sum_{j=1}^{J} \pi_j f(w_i|e_i > 0, \rho, \sigma_j, \beta_j) \tag{9}$$

With this, we can represent the sample selection bias as an omitted variable in (5):

$$E(w_i|e_i > 0) = E(w_i) + \sum_{j=1}^{J} E(u_i|v_i > -z_i\gamma, x_i, \beta_j, \sigma_j)\pi_j \tag{10}$$

Hence, if $\rho_j \neq 0$, we know that $E(u_{ij}|v_i > -z_i\gamma) \neq 0$. Additionally, since $\sum_{j=1}^{J} E(u_i|v_i > -z_i\gamma, x_i, \theta_j)\pi_j$ is in general not zero, the expected error term v_i in (5), is not equal to zero.

Here, I estimate the probability that an individual i belongs to the segment j assuming that each worker in that segment has the same returns to observable characteristics (i.e. the same wage equation). The advantage here is that normality is assumed only at the segment level, while the whole distribution has a flexible form. This is an advantageous assumption for the model. While the log-normality assumption in the aggregated labour market has been widely criticized (Lee, 1982; Polachek & Yoon, 1987; Baffoe-Bonnie & Gyapong, 2017), there are no arguments in favour of a particular distribution in the segment-specific wages. This allows choosing a convenient parametric form for each segment.[3] (Günther & Launov, 2012).

With this, if we take the extreme case where $J = 1$, the mixture model will be turned into the classic

[2]For this derivation, see Appendix 1 of Günther and Launov (2006)

[3]Despite this, it would be interesting for future research to study the distribution of segment-specific wages without distributional assumptions.

Heckman's (1979) model. In contrast, setting $J = N$ will lead to the case where we will have the same number of groups and observations. Ultimately, every case between $J = 1$ and $J = N$ will lead to a semi-parametric model where workers will have the same wage returns if they belong to the same segment. This will be the case only if the data find enough heterogeneity in wages.

Finally, the model still needs to be identifiable. A non-identified mixture model give the same probability as another mixture model of observing an outcome w_i for any individual. (McLachlan, et al., 2019). Hence, it is needed to restrict the possibility that this occurs. One solution is to set that the process of self-selection affects all segments (formal or informal) in the same way (i.e. $\rho_j = \rho$, $\forall j = 1....J$). This restriction meets the sufficient condition for identifiability finite mixtures models given by Teicher (1963). (For more details, see Günther and Launov, 2012).

In other words, this restriction means that the correlation between errors of the wage equation and those on the selection equation must be equal for all the segments. This restriction is not too strong. As Günther & Launov (2012) explain, a positive ρ term means that a worker with a higher probability of being employed has a higher probability to be found at the upper tail of the wages distribution, which is reasonable to happen. However, these probabilities do not need to be the same across segments because variances in each of them are still different.

3.3 Segmentation test

After determining the number of segments for the whole labour market, I need to test whether these segments exist because workers self-select into them due to earnings maximization (competitive markets theory) or because we are in presence of segmentation, where people have some restrictions that do not allow them to work in the segment where they would maximize these earnings, given their observable characteristics.

Assuming that all the returns of the sectors are well-known for the individuals, the competitive markets theory imply that every worker of the labour market, conditional on their personal characteristics, choose the segment that maximizes their wages. As Salem & Bensidoun (2012), the binary variable $I(w_i^j)$ is defined:

$$I(w_i^j) = \begin{cases} 1 & if \quad E(ln\, w_i | e_i > 0, x_i) = \max\limits_{k=1...J} E(ln\, w_i^k | e_i > 0, x_i) \\ 0 & Otherwise \end{cases} \tag{11}$$

In other words, if each individual i chooses the segment j that maximizes her wage, $I(w_i^j)$ will be equal to 1, while in any other case, will be 0. Thus, the share of individuals for whom the segment j is optimal is described for:

$$P(w_i \in Y_j|x_i) = \frac{1}{N}\sum_{j=1}^{N} I(w_i^j) = \tilde{\pi}_j \tag{12}$$

Where $\tilde{\pi}_j$ refers to the segment j that would result from a competitive market. Finally, the sector-specific expected wage for every individual is estimated by:

$$E[w_i^j|e_i > 0, x_i] = x_i\hat{\beta}_j + \hat{\rho}\hat{\sigma}_j \frac{\varphi(-z_i\hat{\gamma})}{1-\Phi(-z_i\hat{\gamma})} \tag{13}$$

As equation (12) assumes free sector mobility, $\tilde{\pi}_j$ can be defined as the expected distribution of individuals across sectors if the labour market were competitive. On the contrary, the estimated distribution of individuals across sectors in the mixture model is given by $\{\hat{\pi}_j\}_{j=1}^{J}$ in (4). Then, I can test if the probabilities $\{\hat{\pi}_j\}_{j=1}^{J}$ are not significantly different from the probabilities of $\{\tilde{\pi}_j\}_{j=1}^{J}$, which would be evidence of a competitive market with free sector mobility. On the other side, if the similarity is rejected, it can be argued that we are in presence of entry barriers between sectors.

3.4 Estimation

The estimation of the model follows a two-step procedure to consistently determine the unknown parameters γ, β_j, π_j, σ_j and ρ. First, I run a probit for the selection equation (6) separately for men and women to estimate their probability to participate in the labour market and to calculate the selection parameter γ. Then, I use $z_i\hat{\gamma}$ as consistent estimates to run the mixture model.

The estimation of the second part is done by maximum likelihood, where the likelihood for the J segments is:

$$ln\mathcal{L} = \sum_{i=1}^{N} ln\left(\sum_{j=1}^{J} \pi_j f(w_{ij}|e_i > 0, x_i, z_i, \hat{\gamma}, \beta, \sigma, \rho)\right) \tag{14}$$

Then, as the survey has information on whether the worker belongs to the formal sector, I need to maximize the log-likelihood of the formal sector, plus the log-likelihood referred to the mixture of the informal sector. Hence, the log-likelihood will be as the one in Salem & Bensidoun (2012):

$$ln\mathcal{L} = \sum_{i \in Y_F} ln(\pi_F f(w_{ij}|e_i > 0, x_i, z_i, \hat{\gamma}, \beta, \sigma, \rho)) + \sum_{i \in Y_I} ln(\sum_{j=1}^{J-1} \pi_j f(w_{ij}|e_i > 0, x_i, z_i, \hat{\gamma}, \beta, \sigma, \rho)) \tag{15}$$

Then, the likelihood is maximized using Rios-Avila (2021) module for Stata, where the already known share of formal workers can be directly computed. Here, it is not needed to correct the variancecovariance matrix from the second stage as in Günther & Launov (2006, 2012) and Salem & Bensidoun (2012) since the likelihood is already accounting for the fact that the model comes from a two-step strategy. Finally, this log-likelihood is estimated for different numbers of informal segments, where the model with the minimum information criteria (AIC and BIC) score is chosen.

4 Data

In order to develop the model described above, I need a *cross-sectional* household survey that contains individual characteristics and information on employment and wages. Thus, in this paper I use the latest round of the Chilean *Social Protection Survey* (or Encuesta de Proteción Social, *EPS* in Spanish), which was elaborated in 2015. This survey is representative at a national and regional level for people over 18 years and has detailed information about work, education, health, social security contributions and household residents.

This survey has also rounds for the years 2004, 2006, 2009 and 2012. However, the 2012 edition is not considered a finished product since it has some deficiencies in the representation of the population over 18 years old. For that reason, it is not recommended to use it for statistical purposes due to the level of error that could entail. In the round that is used for the estimations, (the 2015 edition), the sample is limited to individuals between 25 and 59 years to capture only observations of people who are not college students and are not retired yet. With this, the data has 8,996 observations.

Regarding the classification of individuals, they are divided into workers and non-workers. In the non-workers group, people without a job include both inactive individuals and those who are currently looking for a job. Then, those who work are divided into formal and informal according to the main employment and following the definition of the National Statistics Institute of Chile, which was elaborated in collaboration with the International Labour Organization (ILO, 2013). Here, an informal worker can be either a dependent worker who does not have any social security contributions or an independent worker who works in an informal economic unit. The informal economy does consider the set of companies and businesses that do not have registration in the Internal Tax Service (SII in Spanish) and do not carry an accounting system that allows separating expenses from home business. (INE, 2019). Therefore, it is considered as informal independent workers any individual who is self-employed and does not give any invoice or bill. With this, formal workers will consider dependent workers with social security contributions and independent workers who give invoice or bill.

Table 1 shows the percentage of informal workers among employed individuals by year and gender.

As we can see, men have higher informality rates than women in all years of the survey. The rates are on average 30,4% for male and 28,1% for female workers. Comparing to other developing countries, Chile has low informality rates. For example, Gunther & Launov (2006) found that 66% of workers were informal in the Ivory Coast's labour market and Salem & Bensidoun (2012) found 37% of workers were informal in Turkey. Although the definition of informality was not exactly the same in these papers, both considered social security contributions as the basic division between sectors. With this, it is reasonable to think that these informality rates are pretty comparable to the ones found in this work.

Table 1: Informality rates by year and gender

Year	Full sample (%)	Men (%)	Women (%)
2004	29.7	31.6	26.6
2006	28.6	28.8	28.4
2009	27.8	28.8	26.3
2015	31.6	32.2	30.9

Source: Own elaboration, based on EPS data.

In Table 2 we can see the wage ratios between formal and informal workers. The ratios were calculated based on hourly wages. The hourly wage is computed as the monthly wage divided by the hours worked in the month in that job. Hourly wages are used instead of monthly wages because informal workers may work fewer hours than formal workers, so this specification makes wages more comparable between sectors. This will be further discussed in Section 5 for the model estimation.

On average, formal workers always have a higher wage than informal workers. The ratios in women were also higher than men in 2006 and 2009 and were almost the same in 2004 and 2015. Additionally, women gaps were above 50% in all years, except in 2015, where formal wages were only 14% higher. On the other hand, wage ratios for men were considerably above one too. However, the difference in men was on average 35,5%, a much lower average gap than the estimated for women (46,8%). This could be an argument in favour of the segmented markets theory because the existence of persistent wage differences could be a signal that informal workers are facing entry barriers, and therefore they have to conform with a lower wage in the informal sector. However, the existence of lower wages in the informal sector not necessarily imply segmentation as long as informal workers can move freely between sectors. (Dickens and Lang, 1985). In fact, the lower wage could be a consequence of lower endowments in observable characteristics (i.e. education attainment, experience and others).

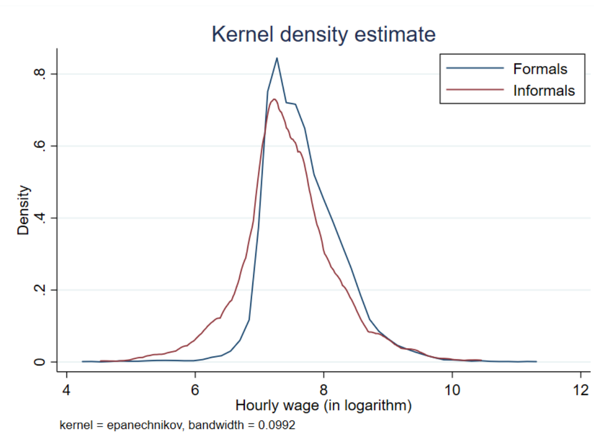
Table 2: Wage ratios between formal and informal workers, by gender

Year	Full sample	Men	Women
2004	1.59	1.60	1.59
2006	1.39	1.31	1.55
2009	1.41	1.33	1.59
2015	1.16	1.18	1.14

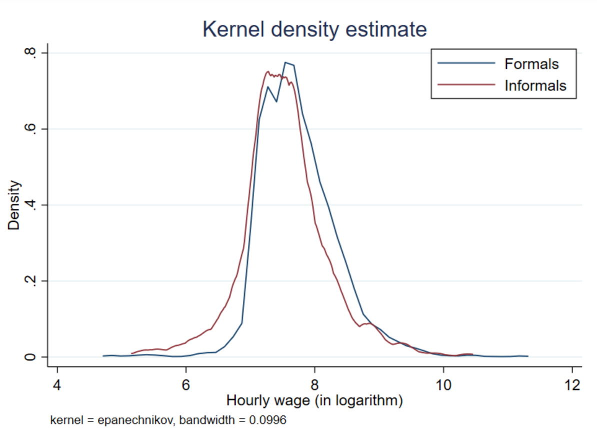
Source: Own elaboration, based on EPS data.

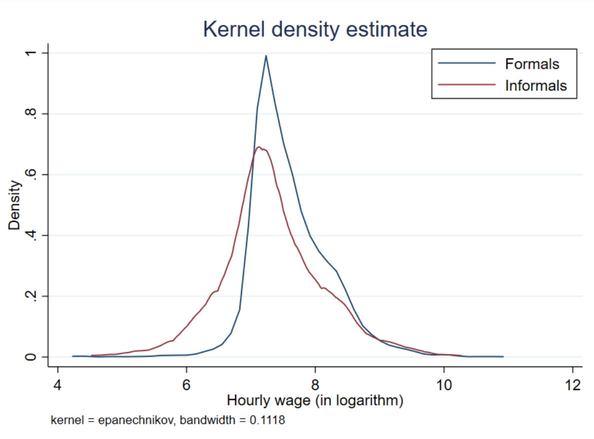
Figure 1 shows the kernel densities of informal and formal hourly log-wages for both sexes and the full sample, in 2015. Here, informal workers of all figures have lower mean wages than formal workers. Nevertheless, there is a substantial part of the densities of wages that overlaps each other, showing that not all informal workers earn lower wages than the formal workers.

Figure 1: Density of hourly wages, by worker's sector and gender



(a) Full sample





(b) Men

(c) Women

Source: Own elaboration, based on EPS 2015 data.

In Table 3 there are some descriptive statistics of the sample, divided into not working, formal and informal men and women. Here, formal men represent the 26,9% of men's sample and formal women the 22,8%. About 30% of the sample represents individuals without a job, where their majority are women (22,2%). Finally, informal workers are about 20% of the sample, where men and women are represented in similar quantities (11,4% and 8,7%, respectively). Formal men and women earn on average a higher hourly wage than their informal counterparts. This is true for monthly wages too. In terms of weekly hours worked, informal women work fewer hours on average (33,8) than formal women (42,7). However, this difference is much lower when we see informal (44,0) and formal men (46,4).

Another difference between sectors lies in the educational levels. As expected, formal men and women are more educated than informal ones of their same gender. The percentage of individuals who only reached primary education is lower both for formal men and women than for their informal counterparts. Furthermore, the percentage of men with at least some college education is 9,8% higher for formal than informal ones. On women, the percentage who have some college education is 18,5% higher on formal women than in informal ones. However, something surprising is that inactive men have a higher percentage than formal men in the level with at least some college education. However, the first group have more preponderance in the segments with no education and only primary education, which leads to conclude that this group have less educational attainment than the second one.

Concerning geographical variables, the EPS survey only provides information about the region where people work. Hence, there is no data about the geographical zone of people without a job. For employed individuals, we can see that, in both genders, formal workers tend to be more concentrated in the middle of the country (Centre and Metropolitan region) than informal ones. In the case of males, 30,9% and 34,2% of informal workers had their job in the centre and metropolitan regions, while the percentages of formal male workers in that places were 37,0% and 36,3%, respectively. On the other hand, 30,4% of informal females worked in the centre and metropolitan region, while this number for formal female was 34,4%. Finally, the metropolitan region seems to be the exception as it concentrates more on informal (44,7%) than formal women (40,0%). In general, these numbers are expected since the biggest cities of Chile are concentrated in this geographical zone (Valparaíso, Viña del Mar y Santiago, the capital) and people in big cities are more likely to work formally. (O'Clery, et al., 2016).

There also differences between the industry affiliation of workers. For example, for men, we can see that construction is a sector where there is a greater preponderance of informal (24,4%) rather than formal (16,8%) workers. The opposite happens in the manufacturing sector, where formal workers have higher participation (17,0%) than informal ones (9,9%). On the other hand, for women, there is a big difference in the commerce sector. While the vast majority of informal women work in that sector (67,1%), formal women have a much smaller preponderance there (41,5%).

Regarding family variables, there are interesting findings. At first, being married is much more common for working males (37,8% for informal and 43,2% for formal workers) than for not employed men, which may be a reflection of the former (37,2 years old) being older than the latter (41,1 years old on average). Then, for women is quite different. Around 46% of females without a job are married, which is a much higher number than the 28,4% and 27,7% of informal and formal married women. This reflects the idea that, in general, Chilean families have defined gender roles that affect female labour participation. (Contreras & Plaza, 2010).

With respect to child care, formal men are on average the group that have more presence of children in the household with less than six years among men. However, women without a job are the most preponderant group on having children under six years (34,3% of the female sample) . Additionally, the presence of people over 70 years in the household is bigger for men without a job (20,3% vs 15,2% and 17,0% for formal and informal working men, respectively). However, in the case of women, female workers have more elderly people in their households (19,1% vs 14,5% and 16,7% of not employed and informal women, respectively). This can reflect that women with elderly people in their homes have more facilities to work formally since older people can take care of their kids while they are working. With the above, it is clear that there are systematic differences across unemployed and working groups. Therefore, this could bias the results if the employment decision in the model is not considered.

Table 3: Descriptive statistics of the sample

Variables	Men			Women		
	Not working	Informal	Formal	Not working	Informal	Formal
	Mean	Mean	Mean	Mean	Mean	Mean
N	716	1026	2424	1998	779	2053
Share of sample	7,9%	11,4%	26,9%	22,2%	8,7%	22,8%
Monthly Wage ($)	-	$ 424.471	$ 550.555	-	$ 267.781	$ 435.474
Hourly Wage ($)	-	$ 2.435	$ 2.876	-	$ 2.156	$ 2.486
Hours Worked	-	44,0	46,4	-	33,8	42,7
Age	37,2	41,5	40,7	41,4	41,5	39,8
Ethnicity	10,1%	10,8%	8,2%	8,3%	9,7%	8,2%
Education						
No Education	17,0%	14,8%	8,3%	15,3%	12,2%	4,9%
Primary	10,3%	12,9%	9,8%	13,1%	10,4%	7,0%
Incomp. Secondary	12,4%	18,6%	15,8%	16,9%	18,4%	10,3%
Secondary	23,7%	32,5%	35,1%	34,0%	35,9%	36,2%
College	36,5%	21,2%	31,0%	20,8%	23,1%	41,6%
Region						
North	-	19,2%	13,4%	-	14,9%	11,1%
Centre	-	30,9%	37,0%	-	30,4%	34,4%
South	-	15,6%	13,4%	-	10,0%	14,5%
Metropolitan region	-	34,2%	36,3%	-	44,7%	40,0%
Industry						
Agriculture	-	13,8%	10,4%	-	3,8%	5,3%
Mining	-	1,7%	5,3%	-	0,3%	0,4%
Manufacturing	-	9,9%	17,0%	-	9,4%	7,9%
Construction	-	24,4%	16,8%	-	0,9%	1,5%
Commerce	-	29,0%	28,7%	-	67,1%	40,2%
Trans. & Comm.	-	14,5%	10,4%	-	1,8%	3,2%
Other	-	6,8%	11,4%	-	16,8%	41,5%
Family variables						
Married	20,7%	37,8%	43,2%	46,0%	28,4%	27,7%
Children under 6	14,9%	21,1%	22,6%	34,3%	25,0%	25,3%
Elderly	20,3%	17,0%	15,2%	14,5%	16,7%	19,1%
Family Size	4.12	4.38	4.39	4.70	4.50	4.36

*Wages are in Chilean pesos (CLP). One Euro (€) was equal to 720 CLP in 2015 aproximately.

**Children refers to the existence of one or more people with less than 6 years in the household.

***Elderly refers to the presence of people with more than 70 years in the household.

Source: Own elaboration, based on EPS 2015 data.

5 Empirical application

In this section, the econometric model that was proposed in section 3 is applied. To compute the model, some of the variables that were presented in table 3 are selected. At first, for the selection equation, at least one exclusion restriction is needed. This means that at least one variable of the selection equation is excluded and do not influence the dependent variable in the wage equation. In this case, the following variables are chosen: the married dummy variable that takes the value of 1 if the individual is married and 0 if not, family size, the number of children under 6 years in the household and the number of elderly (people over 70 years) in the household. The reasons to choose these variables are that, at first, they represent family and environment aspects that determine the personal decision to work and, at the same time, they theoretically not affect productivity once individuals decide to work since they are not observed by the employer. (Puentes, González & Jaar, 2016).

On the other hand, for the wage equation, I consider the variables of educational attainment, age, squared age (as a proxy of potential experience) and ethnicity. I also include the geographical variable that identifies the region where the individual works in the country (North, Centre, South and Metropolitan Region, where is located the capital city, Santiago) and an industry variable that is divided into 7 groups: agriculture, mining, manufacturing, commerce, construction, transport & communication and others. I use the log of hourly wages as the dependent variable. Although monthly wages reflect better the restrictions that workers face in terms of constrained hours to work, hours worked is an endogenous variable as people decide how much time they want to work. Thus, this problem is avoided by using hourly wages instead of monthly wages. (Salem & Bensidoun, 2012).

5.1 Number of segments

Before the mixture model and the sizes of the informal segment are estimated, I need to determine the number of divisions that best fit the data. To achieve this, several mixture models are calculated with a different number of *a priori* imposed segments to see which has the best fit. This is done using Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC). I perform the models by separating men and women and doing it for the full sample too to compare with other similar studies.

In Table 4 we can see the model estimation of the informal sector for one, two and three segments divided as stated before. For the full sample, both AIC and BIC models with two segments perform better as they have the smallest value among the three models. This is also true for the sample with only men, so in these two cases, the chosen models are the ones with two informal segments.

Then, the results for women are quite different. Here, the model with three segments is also worse than the other options. However, it is not clear which of the remaining alternatives is better. A model with two segments performs better in AIC, but in BIC the results are better for the model with only one informal segment. This is in line with the tendency of AIC of preferring larger models as this test is less strict with the penalization of additional segments. (Brame, et al, 2006).

With this dilemma, I choose the model with 2 segments for two reasons. At first, having chosen the model with 2 segments for the estimations with only men and the full sample makes reasonable to think that the model with only women has the same number of segments. Secondly, when estimating the model with 2 segments in the informal sector, both informal groups are considerably big (15,2% and 12,1% of the women's sample). Therefore, both segments are almost equally important and I consider crucial to take them into account separately.

Table 4: Information criteria for model selection

	N	AIC	BIC
Full sample			
1 segment	8811	27401.2	27769.6
2 segments	8811	27200.1	27710.1
3 segments	8811	27218.2	27862.8
Men			
1 segment	4060	13213.0	13522.1
2 segments	4060	13078.3	13507.3
3 segments	4060	13115.2	13600.9
Women			
1 segment	4751	13878.3	14195.2
2 segments	4751	13835.8	14275.5
3 segments	4751	13869.2	14412.4

Source: Own elaboration, based on EPS 2015 data.

5.2 Estimation results

With the chosen models, the selection and wage equations are estimated for both genders and the full sample. In appendix 1 I show the results for the selection equations. At first, educational levels are positive and significant at the 1% in all three specifications, while college education has the highest effect on participating in the labour market, except for men. In that case, secondary has the highest impact in terms of labour participation.

Then, considering third-party care issues, there are differences between genders. Concerning the presence of children under six years in the household, women have a negative and significant coefficient. For men, this coefficient is also significant, but it has the opposite sign (positive). This is not surprising knowing the extensive evidence that shows that Chile is a sexist country with defined gender roles, so women tend to be more responsible for taking care of children. (Contreras & Plaza, 2010). On the other hand, having people over 70 years in the household has a negative and significant impact on male participation. For women, however, this effect is positive and not significant, which can reflect that women may relegate childcare to the elderly and that could be offsetting the negative effect of care issues of the elderly.

Then, family size has a negative but insignificant effect on men, while for women is also negative and only significant at the 10% level. Being married has a negative and strong significant impact on the estimation for the full sample. However, there are opposite results in the other two specifications. While being married for men has a positive and significant effect on work participation, in women this effect is negative and significant. Finally, as expected, the dummy variable denoting the individual is women has a negative and significant effect on labour participation. All these results are in line with studies related to labour participation, especially those that study labour force participation separately for women. (Contreras Plaza, 2007; Contreras, Mello & Puentes, 2008; Contreras, Mello & Puentes, 2011).

Appendix 2 shows the estimation results for the mixture models chosen in the previous section. These consist of models with two informal segments for men, women and the full sample. The first thing to notice here is that the selection term (ρ) is significant for all the three models estimated. While this term is negative for men, for women is positive. This enhances the relevance of accounting for sample selection in the wage equations since the unobserved characteristics of individuals in and out of the labour market are not equal in Chile. Therefore, a model without this correction would show biased returns in the segment-specific wage equations.

As usual, almost all levels of education are positive and significant in explaining wages, especially secondary and college education, where returns are higher in the three specifications. This is true for workers in formal, informal 1 and informal 2 segments. However, comparing between segments, we see differences. For example, in the estimation for men, individuals with secondary or college education have higher returns in the informal 1 segment rather than in the formal sector and the informal

2 segment has the lowest returns to schooling. Exactly the same happens for the sample with only women. Here, female workers who are in the informal 1 sector have the highest returns to secondary and college education and the ones in the informal 2 segment have the lowest. These results contradict the ones of Günther & Launov (2012), where they found that returns to schooling were higher in the formal sector. These differences will be discussed later.

Additionally, experience (measured as age and squared age) seems to have different effects on men and women. In men, the effect is positive and significant in the formal sector, while in informal 1 and 2 are also positive but not significant at any level. On the other hand, experience for women is positive and significant in the formal and informal 1 segments (the effect is higher in informal 1), but not significant in the informal 2 segment.

Being part of an ethnic group have different effects on formal and informal workers. Among men, this variable has a negative and significant effect on wages only in the case of formal workers (in informal 1 and informal 2 segments is also negative but not significant at any level). The same happens to the women's estimation: female wages are negatively affected by ethnicity, but only in the formal sector this effect is significant. This could reflect that the formal sector can have some type of discrimination against indigenous people that does not exist in the informal sector.

The geographic area where people work is also indicative of wages, especially for women. First, for both sexes, the base value (northern region) have a significant and positive effect on wages comparing to the central area because in the north is settled in its great majority the mining sector and its services derivatives. Then, the southern and Metropolitan regions have not significant difference in wages for men in formal and informal 1 segments, while for workers in the informal 2 segment it has a negative and significant effect. This is also true for women, where the southern and metropolitan regions have a negative and significant effect on wages only in the informal 1 segment, while in the formal one this effect is not significant at the 5% level.

For comparative purposes, the model with the full sample was also estimated. Here, being women has a significant negative effect on wages. Additionally, this negative effect is lower for women in the formal sector compared to informal segments, where in informal 2 the effect is the strongest one. This could happen because high-skilled women self-select into the formal sector or maybe because the formal sector has anti-wage discrimination mechanisms that do not exist in the informal sector. Other studies that investigated the heterogeneity of the informal labour sector found the same pattern, as Khalid (2016).

Concerning the sizes of informal segments, the majority of workers are in the formal sector. For men, the 70,3% of the regression sample is in the formal sector, while the informal 1 and 2 segments represent 9,4% and 20,2%, respectively. On the other hand, formal women represent 72,7% of the sample, while the informal 1 and informal 2 segments represent 15,2% and 12,1% respectively. For the full sample, the model estimated that formal workers represent 71,4% of the active people, while the sizes of the two informal segments represent 15,5% and 1,1%.

As expected, formal workers have bigger average wages than informal workers in all estimations. In men, formal workers have an expected hourly wage of \$2.302 Chilean pesos, while informal 1 members their wage is \$1.979 CLP. Finally, in the informal 2 segment, the average wage is \$ 1.806 CLP. These differences represent a 14% gap between formal and informal 1 and a 22% gap between formal and informal 2. Then, for women, the average hourly wage is \$2.006, \$1.735 and \$1.377 CLP in the formal, informal 1 and informal 2 segments, respectively. Therefore, formal women have a 16% higher average wage than informal 1 women, and a 46% higher mean wage than informal 2 female workers. These lower average wages for informal workers are in line with what is expected for any labour market.

These big differences in wages could be fitted into Fields' (1990) framework of a "*lower tier*" and "*upper tier*". However, other results contradict this theory. For example, Fields predicted that individuals in the "*lower tier*" (informal 2) would have very low returns to schooling and experience since these activities are supposed to be oriented to subsistence activities only. This clearly is not the case here as returns to education are higher for workers who are in the informal 2 segment in the estimation for the full sample. Despite not being in line with Fields theory, strong differences were found between sectors and, more interestingly, between informal 1 and informal 2 segments. This emphasizes the importance of treating these two groups as two separate groups rather than unifying them as a unique sector.

5.3 Segmentation or Competition across sectors?

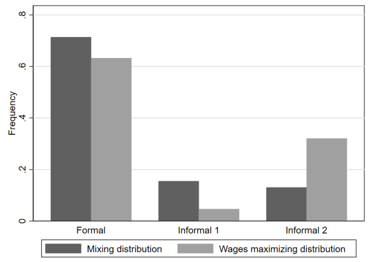
As it was discussed in Section 3.3, it is assumed that workers are wage maximizers by choosing the sector with the highest expected wages for them, given their observable characteristics. Therefore, the actual distribution of the mixing model and the distribution that would result from a completely competitive market, where workers maximize their wages, can be compared. In other words, I want to test the differences between $\{\hat{\pi}_j\}_{j=1}^J$, which are the estimated probabilities of the J segments given by the mixing model, and $\{\tilde{\pi}_j\}_{j=1}^J$ which are the estimated probabilities of the wage maximizing distribution given by (12).

Here, if $\{\hat{\pi}_j\}_{j=1}^{J}$ and $\{\tilde{\pi}_j\}_{j=1}^{J}$ are not significantly different, it can be concluded that there are no entry barriers across sectors because workers are already in the sector where they want to be maximizing their wages. However, if $\{\hat{\pi}_j\}_{j=1}^{J}$ and $\{\tilde{\pi}_j\}_{j=1}^{J}$ differs significantly, then it means that not all workers can work in the sector where they want to be. Therefore, we are in presence of entry barriers as undesired sectors are over-represented in the mixing distribution.

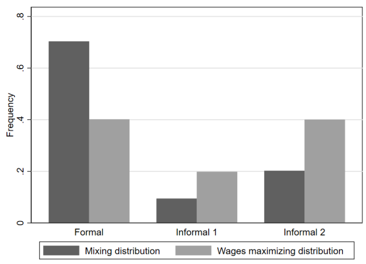
In Figure 2 and Table 5 we can see the distribution of workers across sectors in all three specifications: for men, women and the full sample. For the models with separate men and women, probabilities of each segment appear to be significantly different between the ones that were estimated by the mixing model ($\{\hat{\pi}_j\}_{j=1}^{J}$) and the ones that were estimated with the maximizing distribution ($\{\tilde{\pi}_j\}_{j=1}^{J}$). At first, in the estimation for men, the fraction of workers who would maximize their wages in the formal sector, conditional on their personal characteristics, is 40,1%. That is 30,2% lower than the actual fraction estimated by the mixing model (70,3%). The opposite happens for the informal 1 and informal 2 segments, where the fraction of workers that would prefer to be in those segments (19,8% and 40,0%) almost doubles the current fraction of workers estimated by the mixture model (9,4% and 20,2%, respectively).

There is a similar situation for the estimation with the sample with only women. Here, the share of formal workers that would maximize their wages in that sector is again lower than the estimated fraction of workers in the mixture model (72,7% vs 51,0%). Hence, some female workers who are currently working in the formal sector would earn higher wages in the informal sector, predominantly in the informal 1 segment. This means that the formal sector is less desired by wage-maximizers workers as the returns to observable characteristics are lower. Finally, the results for the full sample are less pronounced, but they follow the same pattern: the fraction of workers that would maximize their wages in the formal sector is lower than the current fraction of workers in that sector (71,4% vs 63,2%) and the preferred segment seems is the informal 2 segment.

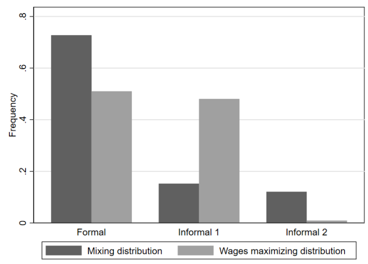
Figure 2: Distributions of workers across sectors



(a) Full sample



(b) Men



(c) Women

Source: Own elaboration, based on EPS 2015 data.

Table 5 also shows the p-values of the test of equality between the fraction of workers in the wage maximizing distribution and the fraction of workers in the mixing distribution (i.e. $\frac{\tilde{\pi}_j}{\hat{\pi}_j} = 1$). Here, we can see that equality between $\hat{\pi}$ and $\tilde{\pi}$ is rejected for all the segments in the three specifications. This means that the actual distributions of workers is significantly different from a labour market with complete free mobility and wage maximizers.

These results are opposite to those of Günther & Launov (2012), where the formal sector was the place where the majority of workers maximize their earnings. Despite these results seem contradictory, they can reflect a situation where formalization carries excessive labour costs that prevent workers to reach their maximum wage in the formal sector. Between these costs for salaried workers, we can find the contributions to unemployment insurance, social security and health insurance. Then, for self-employed workers, the cost associated with formalization can be the payment of the valued added tax (IVA, in Spanish) that reaches 19% of all of their sales in Chile.

Furthermore, the informal markets of the Ivory Coast and Chile are very different. At first, the percentage of formal workers in the Ivory Coast that Günther & Launov (2012) calculated were 33,9%, less than half of formal workers in Chile (71,4%). With this, is reasonable to think that entry barriers are easier to find in the Ivory Coast rather than in a country like Chile, with much lower informal rates. Additionally, Günther & Launov (2012) stated that the biggest formal employer in Ivory Coast is the public sector (i.e. the Government). However, in Chile, the biggest employer in the formal sector is the private sector, where 74,1% of formal workers have a job there, according to EPS survey. Therefore, the formal jobs in each country are distinct, so it seems reasonable to reach different results in each case.

Also, the results in this research are similar to the findings of Harati (2013) for the Egyptian labour market. In that paper, 17% of formal workers would be with a higher wage if they were in the informal sector. Additionally, the informal 2 segment was the most preferred choice in a context with no entry barriers. Harati (2013) explained that these results demonstrate that, despite the majority of the population would earn more in the informal sector, they may prefer the formal sector for more stability and non-financial benefits as work stability, reliable contracts and more labour rights. This could also be happening in Chile where workers may accept a lower wage in a formal job that allow them access to social protection benefits and health insurance.

Table 5: Distribution of workers across sectors and confidence intervals

	Formal	Informal 1	Informal 2
Full sample			
$\hat{\pi}$	0.714	0.155	0.131
$\tilde{\pi}$	0.632	0.047	0.321
$\frac{\tilde{\pi}}{\hat{\pi}}$	0.885	0.301	2.455
p-value*	0.000	0.000	0.000
Men			
$\hat{\pi}$	0.703	0.094	0.202
$\tilde{\pi}$	0.401	0.198	0.400
$\frac{\tilde{\pi}}{\hat{\pi}}$	0.570	2.099	1.981
p-value*	0.000	0.004	0.000
Women			
$\hat{\pi}$	0.727	0.152	0.121
$\tilde{\pi}$	0.510	0.481	0.009
$\frac{\tilde{\pi}}{\hat{\pi}}$	0.701	3.161	0.076
p-value*	0.000	0.000	0.000

Note: The p-values of the equality test are obtained from 200 replications of the sample via bootstrap.

Source: Own elaboration, based on EPS 2015 data.

5.4 Sector mobility

The last section showed that some workers are not maximizing their wages in the segment where they are currently working. This means that each segment may contain both voluntary and involuntary employment. (Khalid, 2016). Nevertheless, until now, it is unknown which would be the origin and destination of workers in a context of free mobility. Knowing this would be useful to determine possible barriers that workers may be facing. Therefore, to understand the sector mobility of workers in a fully competitive context, I estimate the number of workers who would move between sectors. This is done by taking into account the sector where they were originally working.

To do this, I follow the approach of Salem & Bensidoun (2012). In Table 6 we can see that there is substantial mobility between segments. On men, for example, formal workers who would maximize their wages into the same sector only represent a 39,2%, meaning that the other 60,8% of formal men would prefer to go from the formal to the informal 1 or informal 2 segment. Something similar happens to informal male workers. Here 57,3% of informal men would prefer to be in the same sector, while the other 42,7% would change to the formal sector in a free mobility context. Female workers also show significant mobility. At first, only 44,4% of formal women would maximize their wages in that sector, while the remaining 54,6% of them would prefer to work informally. Finally, 71,1% of informal women would also prefer to change their sector, going from the informal to the formal sector.

Table 6: Workers across sectors where their wage would be maximized.

Observed segment →	**Formal**		**Informal**	
Better paid segment ↓	$N°$ of workers	% of workers	$N°$ of workers	% of workers
Full sample				
Formal	1977	0.596	842	0.739
Informal 1	142	0.043	66	0.058
Informal 2	1200	0.362	231	0.203
Total	3319	1.000	1139	1.000
Men				
Formal	663	0.392	256	0.427
Informal 1	362	0.214	92	0.154
Informal 2	666	0.394	251	0.419
Total	1691	1.000	599	1.000
Women				
Formal	722	0.444	384	0.711
Informal 1	889	0.546	153	0.283
Informal 2	17	0.010	3	0.006
Total	1628	1.000	540	1.000

Source: Own elaboration, based on EPS 2015 data.

These results represent a higher percentage of formal workers emigrating their sector than the ones that Salem & Bensidoun (2012) found for formal workers in Turkey. In that paper, 86,5% of formal workers maximized their wages in that same sector, in contrast with the 59,6% that this research found for the full sample estimation. For informal workers, however, mobility was higher in the Turkish labour market. Here, 95,7% of informal workers would be better paid if they were formal workers, while in Chile this percentage is 73,9% for the full sample. These differences are in line with the results of the respective segmentation tests, where Salem & Bensidoun (2012) found that the formal sector would have more wage-maximizer workers than the actual distribution, while in our case the formal Chilean labour market is over-represented (the ratio between the number of workers in the wage-maximizing and mixing distribution is less than one in the formal sector).

With this, two conclusions arise. At first, the results for formal workers challenge the assumption that workers are only wages maximizers. Here, workers may be maximizing utility rather than wages, meaning that the employment benefits associated with formal jobs would be an important component in the utility maximization problem of workers. This is derived from the fact that despite formal workers would earn more being informal in the Chilean labour market, there are no reasons to think that they have entry barriers into informal employment. Thus, the most plausible explanation is that social security, health contributions and job security benefits related to formal jobs could be offsetting these wage gains, preventing formal workers from going to the informal sector.

Secondly, as a considerable share of informal workers also wanted to emigrate from their sector, it is reasonable to think that entry barriers may exist to some extent. However, given the results of Section 5.3 and the different mobility patterns with the study of Salem & Bensidoun (2012), it is difficult to conclude that segmentation is the main force driving informality in Chile. Hence, despite entry barriers cannot be discarded, the theory of competitive markets seems to be more representative of the Chilean labour market.

These findings are in line with the results of the Chilean study of Contreras, Mello & Puentes (2008). In that paper, the estimations showed that labour informality was essentially a phenomenon of self-selection rather than a segmentation process from the employers because workers that were looking for a formal job had a high probability of obtaining it.

6 Conclusions

This paper addresses whether the Chilean informal sector has a heterogeneous structure. To do this, I follow the econometric model originally proposed by Günther & Launov (2006). This model allows having multiple segments when sector affiliation is not observable, correcting at the same time for the selection bias caused by the employment decision of individuals. The advantage of this Finite Mixture Model (FMM) with sample selection is that, after determining the number of segments in the informal sector and estimating the wage equations for every segment, it allows to do a pretty simple calculation to test whether informal workers are segmented from the formal labour market or not.

As a way of determining gender-based differences, the estimations were done by separating men and women. At first, the mixture model found that the structure with 2 informal segments was better for the three specifications: only men, only women and the full sample. After selecting the models, the estimations showed that the sizes for the informal segments represented 20,2% and 9,4% of male workers and 15,2% and 12,1% of female workers.

Additionally, the segmentation test showed that a context of completely free mobility in the labour market is different to the real division that the original model estimated for both male and female workers. Here, the results showed that the formal sector is over-represented and a big fraction of formal workers would maximize their wage into the informal sector and not in their current sector. These results are contrary to the findings of Günther & Launov (2012) and Salem & Bensidoun (2012), where they found that a big majority of informal workers would have a higher wage by entering the formal sector in the Ivory Coast and Turkey, respectively. However, with the same methodology, Harati (2013) found similar results to this paper in Egypt. This may be explained by excessive high labour costs in the formal sector (social security, health and tax contributions) that discourage informal self-employed workers from formalizing.

At next, the sector mobility analysis showed that in a context of free mobility we would have substantial mobility between the formal and informal sector. At first, a big share of formal workers (60,8% in men and 55,6% in women) would prefer to go to the informal sector. These results challenge the assumption that workers are only wage maximizers. This is derived from the fact that despite formal workers would earn more being informal, there are no reasons to think that they have entry barriers into informal employment. Thus, the most plausible explanation is that non-pecuniary benefits related to formal jobs are offsetting these wage gains of the informal sector.

In the other hand, a big fraction of informal workers also showed the desire to move from the informal to the formal sector. However, the desire for mobility was lower than the one showed by Salem & Bensidoun (2012). That is in line with the main conclusion of that paper, where the most preferred sector is the formal sector and some informal workers were segmented from it. With these differences, it can be concluded that, although entry barriers may be preventing some informal workers to enter the formal sector, there is more prevalence of voluntary informal workers in the Chilean labour market. This is in line with Contreras, Mello & Puentes (2008), where they found that informality was mainly driven by a self-selection process of individuals who decided to work in that terms.

Knowing this, policy recommendations should focus on not increasing the opportunity cost of informality and on generating more monetary incentives to promote formalization. Regarding the opportunity cost of informality, increasing social protection policies such as the level of solidarity pensions could discourage individual savings and hence make formality less attractive compared with informality. Therefore, a balance must be found to avoid making these perverse incentives more attractive. (Contreras, Mello & Puentes, 2008). An alternative way to promote formalization can be done by complementing formal earnings with public subsidies. The recently implemented Guaranteed Minimum Income (”*Ingreso Mínimo Garantizado*”), which is a subsidy that aims to guarantee a minimum liquid salary of $319.600 CLP for formal workers, generates the correct incentives to promote formalization.

Finally, as the decision of selection into employment could be different in the formal and informal sector, further research should try to model different selection equations for each segment and to allow error terms to be correlated across segments. In the Chilean study of Contreras & Puentes (2009) they allowed the correlation between segment-specific errors by using Roy’s double selection model. However, they also suffered the identification problem related to the partial observability of sector affiliation. Therefore, a database with at least some information regarding voluntary or involuntary affiliation would be very useful to future research in this area. In fact, with that information, the division of segments would be clearer and, hence, returns to observable characteristics could change, leading to possible different conclusions.

Bibliography

Cameron, A. C. & Trivedi, P. K. (2005). Microeconometrics: methods and applications. Cambridge University Press. Pages 624-625.

Baffoe-Bonnie, J. & Gyapong, A. O. (2017). Definition of full-time and part-time employment and distributional assumptions: the implications for the estimated full-time and part-time wage equations. International Review of Applied Economics, 32(2), 259–279.

Contreras, D., L. De Mello & Puentes, E. (2008), Tackling Business and Labour Informality in Chile, OECD Economics Department Working Papers, No. 607, OECD Publishing, Paris.

Contreras, D., De Mello, L., & Puentes, E. (2011). The determinants of labour force participation and employment in Chile. Journal of Applied Economics, 43(21), 2765–2776.

Contreras, D., & Plaza, G. (2007). Participación laboral femenina en Chile. ¿Cuánto importan los factores culturales?.

Contreras, D. & Puentes, E. (2009). Informal jobs and contribution to social security: Evidence from a double selection model. Serie Documentos de Trabajo (307), Universidad de Chile.

Contreras, D., Gillmore, R. & Puentes, E. (2017). Self-Employment and Queues for Wage Work: Evidence from Chile. Journal of International Development, 29(4), 473-499.

Cunningham, W. V. & Maloney, W. F. (2001). Heterogeneity among Mexico's microenterprises: An application of factor and cluster analysis. Economic Development and Cultural Change, 50(1), 131-156.

Dickens, W.T. & Lang, K. (1985). A test of dual labor market theory. The American Economic Review 67 (2), 792–805.

Fields, G. S. (1990). Labour market modelling and the urban informal sector: Theory and evidence, University of Cornell.

Günther, I. & Launov, A. (2006). Competitive and segmented informal labor markets (No. 2349). IZA discussion papers.

Günther, I. & Launov, A. (2012). Informal employment in developing countries: Opportunity or last resort?. Journal of development economics, 97(1), 88-98.

Harati, R. (2013). Heterogeneity in the Egyptian informal labour market: choice or obligation?. Revue d'économie politique, 123(4), 623-639.

Harris, J. R. & Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. The American Economic Review, 60(1), 126-142.

Hart, K. (1973). Informal Income Opportunities and Urban Employment in Ghana. The Journal of Modern African Studies, 11(1), 61-89.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: Journal of the Econometric Society, 153-161.

Heckman, J. & Hotz, V., (1986). An investigation of the labor market earnings of Panamian males. The Journal of Human Resources 21 (4), 507–542.

Heintz, J. & Posel, D. (2008). Revisiting informal employment and segmentation in the South African labour market. South African Journal of Economics, 76(1), 26-44.

International Labour Organization (2013). Measuring informality: A statistical manual on the informal sector and informal employment. Geneva, International Labour Office.

Khalid, M. (2016). Heterogeneity in Informal Sector Employment in Ghana: Competitive Advantage or Last Resort. University of Manitoba, Winnipeg.

Lee, L. F, (1982). Some Approaches to the Correction of Selectivity Bias." Review of Economic Studies 49 (3): 355–372.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. Manchester School, 22, 139–191. Economic Review, 60, 126–142.

Magnac, T. (1991). Segmented or competitive labor markets. Econometrica 59 (1), 165–187.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. The World Bank Economic Review, 13(2), 275-302.

Maloney, W.F., (2004). Informality revisited. World Development 32 (7), 1159–1178.

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite Mixture Models. Annual Review of Statistics and its Application, 6, 355-378.

O'Clery, N., Gomez-Lievano, A. & Lora, E. (2016). The Path to Labor Formality: Urban Agglomeration and the Emergence of Complex Industries. CID Research Fellow and Graduate Student Working Paper Series 2016.78, Harvard University, Cambridge, MA.

Polachek, S. W. & B. J. Yoon. (1987). "Two-tier Earnings Frontier Estimation of Employer and Employee Information in the Labor Market." Review of Economics and Statistics 69 (2): 296–302.

Pratap, S. & Quintin, E. (2006). Are labor markets segmented in developing countries? A semiparametric approach. European Economic Review 50 (7), 1817–1841.

Puentes, E., González, A. & Jaar Michea, D. (2016). Underemployment and Queues for Full Time Jobs: Evidence from Chile. University of Chile, Department of Economics.

Rios-Avila, (2021) "Module for the estimation of FMM Heckman Selection model".

Rosenzweig, M. R. (1988). Labor markets in low-income countries. Handbook of Development Economics, 1, 713-762.

Salem, M. B., & Bensidoun, I. (2012). The heterogeneity of informal employment and segmentation in the Turkish labour market. Journal of the Asia Pacific Economy, 17(4), 578-592.

Stiglitz, J.E. (1976). The efficiency wage hypothesis, surplus labor, and the distribution of labour in LDCs. Oxford Economic Papers 28 (2), 185–207.

Teicher, H. (1963). Identifiability of Finite Mixtures.The Annals of Mathematical Statistics, vol. 34, no. 4, 1963, pp. 1265–69.

Appendices

Appendix 1

Table 7: Selection equation for the mixture models

Variables	Full sample	Men	Women
Ref. No education			
Primary	0.255***	0.313***	0.215***
	(0.00)	(0.00)	(0.01)
Incomplete Secondary	0.397***	0.475***	0.342***
	(0.00)	(0.00)	(0.00)
Secondary	0.571***	0.582***	0.563***
	(0.00)	(0.00)	(0.00)
College	0.700***	0.411***	0.870***
	(0.00)	(0.00)	(0.00)
Age	0.150***	0.166***	0.138***
	(0.00)	(0.00)	(0.00)
Age^2	-0.002***	-0.002***	-0.002***
	(0.00)	(0.00)	(0.00)
Children < 6 in HH	-0.018	0.186***	-0.160***
	(0.52)	(0.00)	(0.00)
Old in HH	-0.047	-0.150***	0.043
	(0.12)	(0.00)	(0.29)
Family Size	-0.013	-0.005	-0.018*
	(0.11)	(0.71)	(0.10)
Married	-0.193***	0.332***	-0.504***
	(0.00)	(0.00)	(0.00)
Women	-0.771***	-	-
	(0.00)		
Constant	-2.477***	-2.963***	-2.759***
	(0.00)	(0.00)	(0.00)
Observations	8969	4151	4818

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: Own elaboration, based on EPS 2015 data.

Appendix 2

Table 8: Mixture model with 2 informal segments for the full sample

	Full sample		
Variables	Formal	Informal 1	Informal 2
Ref. No education			
Primary	0.111***	0.048	0.013
	(0.041)	(0.065)	(0.136)
Incomplete Secondary	0.227***	0.177***	0.346***
	(0.039)	(0.060)	(0.126)
Secondary	0.348***	0.226***	0.513***
	(0.035)	(0.057)	(0.121)
College	0.803***	0.662***	1.017***
	(0.037)	(0.075)	(0.141)
Age	0.054***	0.014	0.136***
	(0.007)	(0.015)	(0.032)
Age^2	-0.001***	-0.000	-0.002***
	(0.000)	(0.000)	(0.000)
Women	-0.350***	-0.378***	-0.496***
	(0.020)	(0.042)	(0.086)
Ethnicity	-0.103***	-0.092*	-0.039
	(0.029)	(0.055)	(0.117)
Ref. North			
Centre	-0.133***	-0.163***	-0.553***
	(0.027)	(0.047)	(0.116)
South	0.011	-0.120**	-0.414***
	(0.032)	(0.060)	(0.136)
Metropolitan region	0.053**	-0.052	-0.308***
	(0.027)	(0.049)	(0.113)
Control for industry	Yes	Yes	Yes
Constant	5.968***	6.851***	4.243***
	(0.148)	(0.315)	(0.640)
σ_j	0.601	0.384	1.064
ρ	0.750***	0.750***	0.750***
	(0.045)	(0.045)	(0.045)
$\hat{\pi}$	0.714	0.155	0.131
Expected log-wage	7.678	7.393	7.507
Expected wage (CLP)	2161	1624	1820
Observations	8,811	8,811	8,811

Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, $^{***}p < 0.05$.

Source: Own elaboration, based on EPS 2015 data.

Table 9: Mixture model with 2 informal segments for men

	Men		
Variables	Formal	Informal 1	Informal 2
Ref. No education			
Primary	0.073	0.241	-0.076
	(0.052)	(0.234)	(0.077)
Incomplete Secondary	0.173***	0.400*	0.071
	(0.049)	(0.226)	(0.066)
Secondary	0.258***	0.484**	0.123**
	(0.045)	(0.226)	(0.063)
College	0.660***	0.904***	0.486***
	(0.048)	(0.237)	(0.082)
Age	0.020**	0.073	0.017
	(0.010)	(0.054)	(0.017)
Age^2	-0.000	-0.001	-0.000
	(0.000)	(0.001)	(0.000)
Ethnicity	-0.097**	-0.005	-0.075
	(0.041)	(0.219)	(0.064)
Ref. North			
Centre	-0.144***	-0.512***	-0.121**
	(0.038)	(0.192)	(0.052)
South	-0.003	-0.333	-0.151**
	(0.045)	(0.225)	(0.066)
Metropolitan region	0.043	-0.058	-0.101*
	(0.038)	(0.184)	(0.056)
Control for industry	Yes	Yes	Yes
Constant	6.925***	5.986***	6.974***
	(0.207)	(1.101)	(0.346)
σ_j	0.536	0.943	0.354
ρ	-0.252***	-0.252***	-0.252***
	(0.087)	(0.087)	(0.087)
$\hat{\pi}_j$	0.703	0.094	0.202
Expected log-wage	7.742	7.591	7.499
Expected wage (CLP)	2302	1979	1806
Observations	4,060	4,060	4,060

Standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, $^{***}p < 0.01$.

Source: Own elaboration, based on EPS 2015 data.

Table 10: Mixture model with 1 informal segment for women

	Women		
Variables	Formal	Informal 1	Informal 2
Ref. No education			
Primary	0.060	-0.341	0.199
	(0.066)	(0.224)	(0.158)
Incomplete Secondary	0.162***	0.167	0.274*
	(0.062)	(0.185)	(0.158)
Secondary	0.318***	0.347*	0.321*
	(0.056)	(0.181)	(0.171)
College	0.877***	1.044***	0.760***
	(0.058)	(0.210)	(0.207)
Age	0.052***	0.117***	-0.010
	(0.011)	(0.045)	(0.031)
Age^2	-0.001***	-0.001**	0.000
	(0.000)	(0.001)	(0.000)
Ethnicity	-0.114***	-0.096	-0.144
	(0.041)	(0.160)	(0.105)
Ref. North			
Centre	-0.128***	-0.612***	-0.277***
	(0.039)	(0.166)	(0.092)
South	0.020	-0.489**	-0.132
	(0.045)	(0.213)	(0.142)
Metropolitan Region	0.070*	-0.470***	-0.029
	(0.038)	(0.180)	(0.118)
Control for industry	Yes	Yes	Yes
Constant	5.702***	4.182***	6.972***
	(0.222)	(0.935)	(0.723)
σ_j	0.605	1.099	0.399
ρ	0.762***	0.762***	0.762***
	(0.071)	(0.071)	(0.071)
$\hat{\pi}_j$	0.727	0.152	0.121
Expected log-wage	7.604	7.459	7.228
Expected wage (CLP)	2006	1735	1377
Observations	4,751	4,751	4,751

Standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, $^{***}p < 0.05$.

Source: Own elaboration, based on EPS 2015 data.