

The Formal-Informal Labor Market Segmentation Hypothesis Revisited*

Gabriel Ulyssea**

Abstract

Despite the existence of a large empirical literature, there is no conclusive answer to whether labor markets in developing countries are segmented. Moreover, a prior and perhaps more fundamental question remains largely unanswered, namely, whether the formal-informal segmentation hypothesis actually has empirical content. This paper tackles the latter question by developing a stochastic dynamic discrete choice model to analyze workers' choices between unemployment, formal and informal jobs. This general framework is then used to assess the empirical implications of different models of labor market functioning. As a complete taxonomy exercise is beyond the scope of this paper, I focus on the two most polar cases in the present context: the fully integrated and fully segmented labor markets. The final goal is to analyze to which extent these two extreme benchmarks can generate implications that are empirically distinguishable and, more broadly, whether the findings in the empirical literature can be supported by a single model or by a broad array of models. The analysis shows that both models can support the observed regularities regarding formal-informal wage gaps, transitions and job duration. Hence, it does not seem possible to identify the existence of segmentation from the moments typically analyzed in the empirical literature.

Keywords: Segmentation, Informal Sector.

JEL Codes: O17.

*Submitted in May 2011. Revised in August 2011. I am extremely grateful to Steven Durlauf for detailed comments and suggestions. I would also like to thank James J. Heckman, one anonymous referee and the editor, Cristine Campos de Xavier Pinto, for their comments and suggestions. All remaining errors are, of course, my own. Financial support from CAPES and IPEA is gratefully acknowledged.

**University of Chicago and IPEA. E-mail: ulyssea@uchicago.edu

1. Introduction

The coexistence of large informal sectors and substantial wage differentials between observably equal workers has motivated an extensive literature on the formal-informal labor market segmentation hypothesis. Indeed, most empirical studies rely on the analysis of wage differentials and the process of wage determination in both sectors.¹ The underlying definition of segmentation in this literature is one where equally productive workers receive higher wages if they work in formal jobs. As a corollary, if one is able to identify the existence of wage differentials between equally productive workers, then one is able to empirically show the existence of segmentation [e.g. Barros et al. (1993)].

However, to distinguish the true causal effect (if it exists) of holding a formal contract from compensating differentials and self-selection is extremely difficult. Heckman and Hotz (1986), for example, argue that earning functions in different sectors/segments can differ for a variety of reasons and therefore they are not an adequate test for the existence of segmentation [see also Magnac (1991)]. This identification problem has led some authors to argue that one cannot test for the existence of segmentation by the observation of wage differentials alone, and there has been an increasing emphasis on the analysis of job-to-job transitions between sectors [e.g. Maloney (1999), Gong et al. (2004) and Curi and Menezes-Filho (2006)].² This approach is supported by the theoretical argument that labor market segmentation might manifest itself not only through different skill pricing or job quality across sectors, but also through restrictions to workers' mobility between sectors [see Fields (1975), Dickens and Lang (1985) and Dickens and Lang (1992)].

Despite the existence of an extensive literature, the question of whether labor markets are segmented and the empirical relevance of this phenomenon remain inconclusive. Moreover, a prior and perhaps more fundamental question also remains largely unanswered, namely, whether the formal-informal segmentation hypothesis actually has empirical content. In a pioneering study, Barros (1988)³ examines the empirical content of the formal-informal labor market segmentation hypothesis by analyzing to what extent the evidence regarding moments of the wage distributions is consistent with a unique model or with a wide range of models. His results are not encouraging, as he concludes that the empirical regularities relative to the wage distributions can be supported by very different views of labor market functioning.

The objective of this paper is to further analyze the empirical content of the formal-informal labor market segmentation hypothesis. For that, I build on the

¹See Ulyssea (2006) for a literature review.

²An important early exception is the work of Barros et al. (1990), which investigates the mobility of formal and informal employees in Brazil using data from the metropolitan region of Sao Paulo.

³For an updated and modified version of Barros (1988), see Barros and Ulyssea (2010).

framework presented by Barros and Ulyssea (2010), to develop a stochastic dynamic discrete choice framework to model workers' choices between unemployment, formal and informal jobs. Individuals are heterogeneous with respect to their ability, unobservable to the econometrician, which is also a source of heterogeneity in the utility they can derive from a job in either sector. The latter is a function of worker's potential wage in the given sector and of a bundle of non-pecuniary, sector-specific characteristics (e.g. flexible hours and job security). Potential log-wages have two additively separable components:

- (i) a hedonic function that describes how wages depend on worker's ability; and
- (ii) a job quality shock, which is drawn from a sector-specific distribution.

The model's novelty is the presence of a job rationing component, which allows for the possibility that workers might not be able to get the job that yields the highest total expected value. This innovation aims at capturing the fact that there might exist barriers to mobility across formal and informal jobs, which is one of the central features of the segmentation hypothesis.

With this general framework at hand, it is possible to assess the empirical implications of specific models of labor market functioning, which can be obtained as special cases of the former. Obviously, a complete taxonomy exercise is beyond the scope of this paper. For this reason, I focus on the two most polar views of labor market functioning in the present context: the fully integrated and fully segmented labor markets. One of the advantages of the general framework briefly described above is that it allows me to precisely define these two benchmarks, which is often not done in the literature. Using these definitions, I analyze what the empirical implications would be if the data generating process corresponded to these two extreme cases. The final objective is to analyze to which extent they can generate implications that are empirically distinguishable. More broadly, I ask whether the findings in the empirical literature can be supported by a single model or by a broad array of models that imply completely opposite views of labor market functioning.

The analysis shows that both views of labor market functioning – segmented *versus* integrated – can support the observed regularities regarding formal-informal wage gaps, and transition rates between sectors. In particular, both models can generate positive or negative wage gaps, positive or negative ability bias in employment sector choice and duration dependence (even after conditioning on unobserved ability). Hence, without imposing further structure there seems to be little hope for identification. More specifically, it does not seem possible to identify which model is behind the data generating process from the moments typically analyzed in the empirical literature.

The remainder of the paper is organized as follows. Section 2 presents some evidence on the moments that are discussed in the theoretical analysis. Section

3 presents the general framework, while Section 4 analyzes the different empirical implications of the segmented *versus* integrated models. Section 5 concludes.

2. Stylized Facts

The objective of this section is to provide some stylized facts regarding the moments that are analyzed in the following sections. The goal is both to motivate and contextualize the theoretical discussion, which is the core of the paper. In this section, I focus on the analysis of workers' mobility, in particular, on duration analysis.⁴ The evidence regarding the formal and informal wage differentials is discussed in Barros and Ulyssea (2010) and it is therefore omitted here.

The data used here come from the Monthly Employment Survey (*Pesquisa Mensal de Emprego – PME*) for the period from March 2002 to December 2008. The PME is representative of the six major metropolitan areas in Brazil, and it is a rotating panel with a similar structure to that of the Current Population Survey (CPS) in the United States: selected households are interviewed once a month for four months, leave the sample and return eight months later for another round of four interviews. I restrict the sample to individuals who were either formal or informal employees at the time of the survey, and aged between 18 and 65 years. Throughout the analysis, I define informal workers as those *employees* who do not have a formal labor contract (*sem carteira de trabalho assinada*). Analogously, formal workers are those employees who have a work permit (*com carteira de trabalho assinada*). Finally, to avoid inconsistencies in the job duration measure, I only use information from the first four interviews and exclude individuals holding more than one job. In both datasets, I also exclude public employees and domestic workers.

The natural starting point is to analyze a simple transition matrix between states. Table 1 shows the patterns of transitions out of four possible states (informal or formal employees, self-employed and unemployed) and into all possible destinations. The matrix displays a well-known fact about informal labor markets: Informal employees face much higher turnover rates than their formal counterparts and therefore the probability of staying in the informal sector is much lower than staying in the formal one (from the diagonal entries, we have 47.6% and 84.6%, respectively). Transitions from informal to formal employment are much more likely than the other way around (20.2% and 5.2%, respectively). This evidence has also been found for other countries, and has been interpreted as suggestive that if labor market segmentation exists, it is not severe, as informal workers do not seem to stay long in the informal sector and, moreover, they are likely to transit to the formal sector.

Taking a step further, the survival function and the cumulative hazard tell similar stories, as the probability of survival in the informal employment status

⁴The facts presented here constitute an updated subset of the evidence presented in Szerman and Ulyssea (2007).

Table 1
Transition matrix: 2002-2008

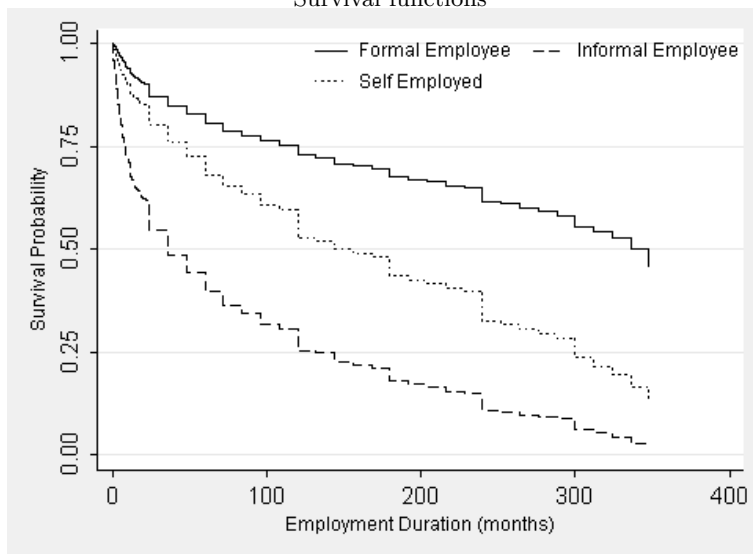
	Unemp.	Formal	Informal	Self Emp.	Other	O.L.F.†
Unemp.	59.4	11.3	14.1	14.3	1.1	–
Formal Emp.	2.2	84.6	5.2	2.4	0.7	5.0
Informal Emp.	4.9	20.2	47.6	15.8	2.9	8.5
Self Emp.	3.3	4.3	9.8	61.3	8.4	12.9

Source: Author's calculations from the PME.

† Out of the labor force.

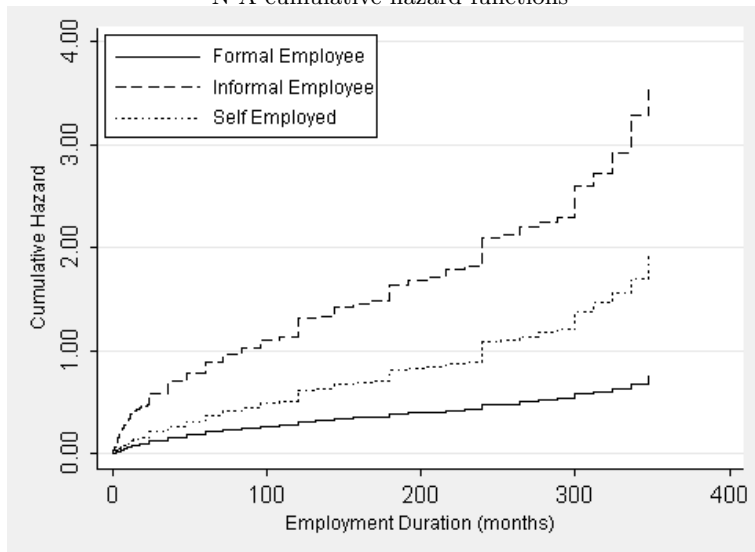
is much lower than in self-employment and in formal employment. However, the cumulative hazard for informal employees displays a concave shape for almost the entire job duration range, indicating the existence of negative duration dependence. Thus, it suggests that the longer individuals stay in the informal sector, the less likely they are to leave.

Figure 1
Survival functions



Although informative, these figures have two main shortcomings: they do not control for workers' observable characteristics and they do not account for the fact that there are multiple destinations an individual can go to once she leaves a given occupation status. To address these issues in a very simple way, I estimate the Cox

Figure 2
N-A cumulative hazard functions



Proportional Hazard model, with competing, but independent risks. This allows the analysis of duration dependence controlling for observable characteristics and looking at specific transitions.⁵ The Cox model has the attractive feature of being semi-parametric, and it is not necessary to make any functional form assumptions regarding the baseline hazard.

The competing risks analysis shows a very different picture from the one suggested by the transition matrix presented before. Once one controls for workers' characteristics, the probability of transiting out of the informal sector and into the formal sector is much smaller than in the opposite direction. This fact is depicted in Figure 3, which shows that the survival probability in the informal sector with respect to the risk of exiting to a formal job is much higher than the other way around. More interestingly, they are basically the same at the very beginning of the spell in each sector (around the first three months) and then they significantly drift apart. It is worth noting that the Brazilian labor regulation allows for a three-month period of "experience", during which the employer can dismiss a formal employee at no cost.

Figure 3 thus shows a very different scenario from the one found in simple transitions, survival and the cumulative hazard functions. According to the view that the existence of segmentation implies barriers to workers' mobility across sectors,

⁵The Cox model regressions include the usual sociodemographic controls (age, gender, race, marital status and schooling), plus workers' firm size and industry (services, manufacturing, construction and so on).

Figure 3
Cox Baseline Survival Functions

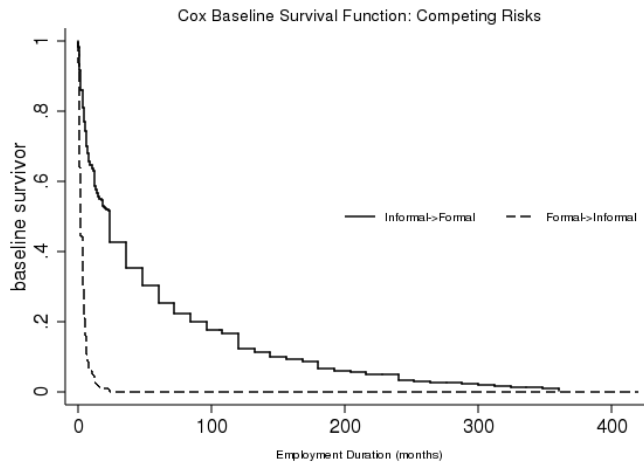


Figure 3 could be interpreted as evidence in favor of the existence of segmentation. However, as the remainder of the paper discusses, even when taken together, the evidence presented here and elsewhere in the literature is not conclusive about the labor market segmentation hypothesis.

3. The General Model

3.1 Basic structure

The total population of workers is assumed to be constant and is denoted by \mathbb{L} . All workers $i \in \mathbb{L}$ are assumed to be homogeneous with respect to their characteristics observable to the econometrician, and differ only with respect to their unobserved ability, A , which, for simplicity, is assumed to be a scalar.⁶ In any period, the population of workers consists of a disjoint partition of unemployed and employed workers, denoted by \mathbb{U}_t and \mathbb{E}_t , respectively. Let the unemployed set be denoted by 0, and the formal and informal sets by 1 and 2, respectively. The formal and informal sectors constitute a disjoint partition $(\mathbb{E}_{1,t}, \mathbb{E}_{2,t})$ of \mathbb{E}_t . Define the indicator variable for unemployed worker as $I_{0,t}(i) = 1 [i \in \mathbb{U}_t]$ and for formal worker as $I_{j,t}(i) = 1 [i \in \mathbb{E}_{j,t}]$, $j = 1, 2$. The population of workers is fixed, but the

⁶As long as unobserved skills are characterized by fixed bundles and the different elements cannot be sold separately in the market, the assumption of scalar ability does not fundamentally change the present analysis, but it makes exposition easier.

employed-unemployed and the formal-informal partitions can change from period to period. Even though the transitions in and out of unemployment are of interest, the focus of the analysis is the assignment process of workers that generates the partition $(\mathbb{E}_{1,t}, \mathbb{E}_{2,t})$ and the implications for the cross-sectional distribution of wages in both sectors and the transition rates between sectors.

In every period, employed workers must choose whether they want to remain in their current job, if they want to go to the other sector, or if they want to go to unemployment. Similarly, unemployed workers must choose between a job in either sector or remaining unemployed and waiting for a better opportunity. Even though individuals can always choose to be unemployed, it is possible that they cannot get their most preferred job choice due to a stochastic job rationing component. I introduce this potential job rationing in a very simple and stylized way. I assume that in every period workers draw a random vector $D_t = (d_{1,t}, d_{2,t})$, where $d_{j,t} = 1$ if the worker can choose to work in sector j and zero otherwise. Put differently, if $d_{j,t} = 0$ the worker is rationed out of sector j in period t . If the worker draws $D_t = (0, 0)$, she is rationed out of both sectors and becomes an unemployed worker. Thus, the discrete random vector D_t has only four possible realizations, $D_t \in \{(1, 1), (1, 0), (0, 1), (0, 0)\}$.

The wage setting mechanism in each sector has a very mechanical form in the present model. I assume that the log-wage in each sector has only two components: an ability pricing function, which is an equilibrium outcome that is left unspecified; and a “job quality” component, which from the perspective of the worker is an exogenous shock (conditional on worker’s ability). Hence, in any given period t , a worker of ability $A = a$ has two *potential* log-wages in each sector $j = 1, 2$,⁷ which are given by:

$$w_{i,j,t} = \omega_j(a_i) + \phi_{i,j,t} \quad (1)$$

where $\omega_j(\cdot)$ is the equilibrium pricing function in sector j (which is time-invariant), and $\phi_{i,j,t}$ is the job quality component drawn by individual i in period t for sector j .

Albeit important in itself, the mechanism that leads to this equilibrium pricing function is left unspecified for two reasons: (i) the central feature for the present analysis is whether the pricing function is different across sectors and not *why* it is so; and (ii) leaving it unspecified avoids introducing unnecessary structure (given the goals of this paper) that could lead to further modeling misspecification error.⁸ Nonetheless, this option is not assumption-free, as it implies assuming that the

⁷Of course, only one of these potential wages is in fact realized. If the worker is allocated to the informal sector, then $w_{1,t}$ should be interpreted as the counterfactual wage, were this worker reallocated to the formal sector.

⁸In other words, not modeling the mechanism that gives rise to the ability pricing function avoids taking a strong stand on which economic framework is more adequate to model the wage setting in each sector (e.g. wage bargain, efficiency-wage, competitive or monopsony models).

wage setting in both sectors does not depend on the outside options available to the worker in each period. More generally, it excludes any structure that would imply that the pricing equation changes over time.

Write $\Phi_j^t = (\phi_{j,0}, \dots, \phi_{j,t})$; workers draw their potential job quality from the distributions $F_{\phi_{j,t}|\Phi_j^{t-1}}$, $j = 1, 2$, which they take as given. For notational simplicity, I omit the conditioning on the past history of shocks, denoting the conditional distribution simply as $F_{\phi_{j,t}}$.⁹ These distributions are characteristics of the sectors and not of the workers. For simplicity, I assume that the job quality component is independent of workers' ability:

Assumption (A.1): The job quality shocks, Φ_j^t , are independent of workers' ability: $\Phi_j^t \perp A$, $\forall t$ and $j = 1, 2$.

Finally, workers derive utility from the labor income, $w_{j,t}$, and from a non-pecuniary component specific to sector j that is denoted by Q_j . The Q_j can be interpreted as a scalar or as a bundle of characteristics specific to each sector, such as flexible hours, job security, and so on.

3.2 Worker's problem

Before describing the value functions of each possible state, it is important to precisely describe the timing of worker's problem and her information set. From this point on, I omit the time and individual subscripts whenever possible to simplify the notation. At the beginning of each period, employed and unemployed workers face the same decision problem, which unfolds as follows:

1. The worker draws job-specific shocks in both sectors, $(\phi_{1,t}, \phi_{2,t})$.
2. Given the draws, the worker chooses the optimal sector based on the total discounted expected values of each state. If the optimal choice is the unemployment sector, then the worker certainly steps into unemployment.
3. If the optimal choice is an employment sector, then the worker draws $D_t = (d_{1,t}, d_{2,t})$. If her first-best option is available, she goes to the segment chosen in step 2. If she is rationed out of her first-best, then she goes to her second-best, which can be either the other employment sector or unemployment.

Given the timing just described, the relevant information set for the worker when she is making her decision (stage 2 above) is $\Omega_t = \{\Upsilon_t, \phi_{1,t}, \phi_{2,t}\}$, where

⁹Throughout the paper these shocks are assumed to be correlated across time and independent across sectors. A natural assumption here would be that the job quality shock follows a first-order Markov process (instead of conditioning on the entire history of shocks). Nonetheless, at this point the analysis would not gain much with this assumption so I opt to leave the notation at this higher level of generality.

$\Upsilon_t = \{\Phi^{t-1}, D^{t-1}, \mathbb{I}^{t-1}, a\}$ and $\Phi^{t-1} = \{(\phi_{1,0}, \phi_{2,0}), \dots, (\phi_{1,t-1}, \phi_{2,t-1})\}$, $D^{t-1} = \{(d_{1,0}, d_{2,0}), \dots, (d_{1,t-1}, d_{2,t-1})\}$ and $\mathbb{I}^{t-1} = \left\{ \{I_{j,0}\}_{j=0}^2, \dots, \{I_{j,t-1}\}_{j=0}^2 \right\}$. Hence, the information set in t is comprised of the entire history of job quality shocks (Φ^{t-1}), of job opportunities in both sectors (D^{t-1}) and realized sector membership (\mathbb{I}^{t-1}). Note that the set Υ_t only contains the information up to the previous period. This is true because the worker makes her decision about which sector she would like to work in before she draws the current shock $(d_{1,t}, d_{2,t})$. This is trivially true for \mathbb{I}^t , as the current-period membership, $I_{j,t}$, is the final outcome of her decision process.

The value functions of being unemployed, and employed in the formal and informal sectors, respectively, can be written as follows:

$$\begin{aligned} V_0(\Omega) &= b + \delta \sum_{j=0}^2 P_{0,j}(\Omega) E[V_j(\Omega') | \Omega] \\ V_1(\Omega) &= U(w_1, Q_1) + \delta \sum_{j=0}^2 P_{1,j}(\Omega) E[V_j(\Omega') | \Omega] \\ V_2(\Omega) &= U(w_2, Q_2) + \delta \sum_{j=0}^2 P_{2,j}(\Omega) E[V_j(\Omega') | \Omega] \end{aligned} \tag{2}$$

where the primes denote next period objects, b denotes the utility from unemployment, $P_{k,j}(\Omega)$ is the worker's probability of being in sector j next period given that she is in sector k in the current period, and Ω denotes the current-period information set.

Define $\theta_j = 1[V_j(\Omega) \geq V_s(\Omega), \forall s \neq j]$, for $j = 0, 1, 2$; in words, $\theta_j = 1$ means that segment j is the worker's first best in the current period. In the absence of job rationing, θ_j would provide all the information necessary to determine in which sector (unemployed, formal or informal employment) the worker would be in the current period. With this notation at hand, it is possible to write the expected transition rates out of the formal and unemployment states between two periods as follows (the transition rates out of the informal state are analogous to those out of the formal state and are therefore omitted):

$$\begin{aligned} P_{1,1}(\Omega) &\equiv \Pr[\theta'_1 = 1, d'_1 = 1 | \Omega] + \Pr[\theta'_2 = 1, D' = (1, 0), V'_1 \geq V'_0 | \Omega] \\ P_{1,2}(\Omega) &\equiv \Pr[\theta'_2 = 1, d'_2 = 1 | \Omega] + \Pr[\theta'_1 = 1, D' = (0, 1), V'_2 \geq V'_0 | \Omega] \\ P_{1,0}(\Omega) &\equiv \Pr[\theta'_0 = 1 | \Omega] + \Pr[D' = (0, 0), V'_0 < V'_j, j \neq 0 | \Omega] + \\ &\quad + \sum_{j=1}^2 \Pr[\theta'_j = 1, d'_j = 0, V'_0 > V'_s, s \neq j | \Omega] \end{aligned}$$

$$\begin{aligned}
 P_{0,1}(\Omega) &\equiv \Pr[\theta'_1 = 1, d'_1 = 1|\Omega] + \Pr[\theta'_2 = 1, D' = (1, 0), V'_1 \geq V'_0|\Omega] \\
 P_{0,2}(\Omega) &\equiv \Pr[\theta'_2 = 1, d'_2 = 1|\Omega] + \Pr[\theta'_1 = 1, D' = (0, 1), V'_2 \geq V'_0|\Omega] \\
 P_{0,0}(\Omega) &\equiv \Pr[\theta'_0 = 1|\Omega] + \Pr[D' = (0, 0), V'_0 < V'_j, j \neq 0|\Omega_t] + \\
 &\quad + \sum_{j=1}^2 \Pr[\theta'_j = 1, d'_j = 0, V'_0 > V'_s, s \neq j|\Omega]
 \end{aligned}$$

where $V'_j \equiv V_j(\Omega')$; the first term in the above probabilities describes the situation where the stochastic rationing is not binding, and thus reflects the probability of workers making the transition solely based on their choice. The second term describes the situation where the worker is pushed to her second-best option, or into unemployment.

4. The Empirical Content of the Segmentation Hypothesis

As mentioned in the introduction, in order to assess the extent to which the formal-informal segmentation hypothesis actually has empirical content, I analyze the empirical implications of the two most extreme benchmarks in the present context: the segmented *versus* integrated labor markets. For that, it is necessary to state these benchmarks very precisely [which is done in definitions (D.1) and (D.2)], as they shape all the analysis that follows. However, before proceeding to these definitions, I impose some additional structure to the model. First, I make the following assumption regarding the utility function, $U(w_{j,t}, Q_j)$:

Assumption (A.2):

1. The utility function is quasi-linear: $U(w_j, Q_j) = w_j + g(Q_j)$.
2. Workers are heterogenous with respect to the utility they derive from Q_j : $g(Q_j) \equiv g(Q_j, \epsilon)$, where ϵ is the taste heterogeneity parameter and $\epsilon \equiv \epsilon(A)$. Thus, the idiosyncratic preference parameter is a function of worker's ability.

The first part of Assumption (A.2) simplifies the analysis considerably and yet it seems a fairly reasonable description of individuals' utility. The second part is meant to allow for a more general structure of preferences and it also introduces an additional source of heterogeneity in the model. In addition to (A.2), I impose the following assumption regarding the information set and the relevance of past information:

Assumption (A.3): The tenure in the current job, which is denoted by $T_{j,t}$, is a sufficient statistic for the history of sector choices (Θ^t) and job rationing

draws (D^t) .¹⁰

Thus, Assumption (A.3) states that the tenure in the current job contains all the relevant information for the worker to make her choice at the beginning of period t . If the worker ended period $t - 1$ in sector j , the tenure at t can be written as $T_{j,t} = \tau$, where $\tau = \sum_{s=1}^{\tau} I_{j,t-s}$, for some $\tau \in \mathbb{N}^+$ and $j = 0, 1, 2$. It is worth emphasizing that this assumption does not imply a smaller information set in the sense that the $I_{j,t}$ summarizes the information regarding choices (Θ^t) and job rationing (D^t) . This assumption, however, has important implications for the nature of the time persistence of these shocks (both job quality and rationing). In particular, it implies that once the worker switches sectors, the history of job shocks in the previous sector is no longer relevant. This assumption does not fundamentally change any of the results discussed ahead, but it does make the analysis of job duration more straightforward, as will be clear in the following sections.

The following two definitions are central to the remainder of the analysis, as they define the two polar views regarding labor market functioning: segmented *versus* integrated labor markets. Using the notation developed so far, segmentation is defined in the following way:

Definition (D.1): The labor market is said to be *segmented* if and only if the following conditions hold:

- (i) Formal and informal sectors have different ability pricing functions, $\omega_j(\cdot)$, $j = 1, 2$.
- (ii) Formal and informal sectors have different potential job quality distributions, denoted by $F_{\phi_{j,t}}$, $j = 1, 2$.
- (iii) There is non-price rationing of jobs in the formal, but not in the informal sector: $\Pr [d_{1,t} = 1] < 1$ and $\Pr [d_{2,t} = 1] = 1$.¹¹

This is the case where both sectors differ in all of their basic elements and thus are regarded as completely segmented. The second definition describes the situation where the labor market is completely integrated. At this point it is important to emphasize that the definitions of segmented versus integrated markets regard the equilibrium outcomes observed in each sector, given the existing set of labor market institutions in the economy. In the segmented markets definition, these

¹⁰Analogously to the previous definitions, write $\Theta^t = \left\{ \left\{ \theta_0^j \right\}_{j=0}^2, \dots, \left\{ \theta_t^j \right\}_{j=0}^2 \right\}$.

¹¹I assume that only formal jobs are rationed because the formal-informal labor market segmentation literature never considers the situation where informal jobs are rationed. The reason for that is that formal jobs are typically considered to be of higher quality or “protected”, while informal jobs are not.

institutions lead to different equilibrium outcomes on the pricing functions, on potential job quality distributions and on job availability. In the integrated market definition, although both sectors have the same equilibrium outcomes, there are still sector-specific characteristics that differentiate them (which are summarized by Q_j). Thus, even in the case of fully integrated labor markets it is still meaningful to talk about two sectors.

Definition (D.2): The labor market is said to be *integrated* if and only if

- (i) Formal and informal sectors have the same pricing function: $\omega_1(\cdot) \equiv \omega_2(\cdot) \equiv \omega(\cdot)$.
- (ii) Formal and informal sectors have the same potential job quality distribution: $F_{\phi_{1,t}} \equiv F_{\phi_{2,t}} \equiv F_{\phi_t}, \forall t$.
- (iii) There is no job rationing: $\Pr[d_{1,t} = 1] = \Pr[d_{2,t} = 1] = 1, \forall t$.¹²

Finally, before proceeding to the analysis of the implications of the different views of labor market functioning, it is useful to state one last definition. Let the sector assignment processes be defined as the conditional probability (on worker's ability) of an individual to be allocated to sector j in a given period: $\pi_{j,t}(A) = \Pr[I_{j,t} = 1|A]$, $j = 0, 1, 2$, respectively. Let $p_{j,t} = \int_A \pi_{j,t}(a) dF_A(a)$ denote the unconditional probability of any given individual to be allocated to sector j , which in the data simply corresponds to the share of individuals in state $j = 0, 1, 2$ in period t . With this notation in mind, consider the following definition:

Definition (D.3): The assignment process is said to be *unbiased* if, and only if, the employment probabilities do not depend on worker's ability. If this is the case, then one would observe $\pi_{j,t}(A) = p_{j,t}, j = 1, 2$.

I now proceed to the analysis of the empirical predictions under each of the two polar views defined in (D.1) and (D.2). I discuss the empirical implications of both models regarding transition rates and job duration in the following section. The implications regarding the cross-sectional wage distribution are analyzed in subsection 4.2.

4.1 Transition rates and job duration

4.1.1 Integrated labor markets

Since formal and informal sectors have the same pricing functions and job quality distribution and there is no job rationing, the assignment process is entirely

¹²Condition (iii) is quite strong and could be replaced with a weaker form of integration: (iii)' $\Pr[d_{1,t} = 1] = \Pr[d_{2,t} = 1] = \bar{P} < 1$. That is, job rationing exists, but it is the same for both sectors. Assuming $\bar{P} = 1$ is not necessary to any of the following results, but it does simplify the analysis and hence this stronger version of the assumption is used.

based on individual's choice. If choice was solely based on current and future expected wages, selection would be unbiased (see definition D.3). This is so because the wage determination mechanism is the same in both sectors. However, because workers place a different value on each sector's bundle (Q_j), the assignment is not unbiased. I thus proceed to analyze the effect of unobserved ability on transition probabilities.

The effect of unobserved ability

Consider first the probability of a worker being in state s in the next period given that she ended the current period in state j (here again time subscripts are omitted and primes denote next-period objects). In the integrated market model, this probability is simply $\Pr [I'_s = 1 | T_j, A] = \Pr [\theta'_s = 1 | T_j, A]$. Since the focus of the formal-informal labor market segmentation literature lies on the job-to-job transition, in what follows I will analyze the individual hazard rate out of the informal sector and into the formal sector (in the integrated model the job-to-job hazards are symmetric). To simplify notation, I also assume that the persistence in the job quality shocks is high enough so that the current informal job is still preferred to unemployment.¹³ One can thus write the informal to formal hazard rate as follows:

$$\begin{aligned}
 h_{2,1}(\tau | A) &\equiv \Pr [I_1 = 1 | T_2 = \tau, A] && (3) \\
 &= \Pr [\theta_1 = 1 | \Upsilon] \\
 &= \Pr [\phi_1 - \phi_2 \geq g(Q_2, \epsilon) - g(Q_1, \epsilon) + K_2(\Upsilon) - K_1(\Upsilon) | \Upsilon] \\
 &= 1 - F_{\phi_D}(\bar{C} | \Upsilon_t) && (4)
 \end{aligned}$$

where $\bar{C} \equiv g(Q_2, \epsilon) - g(Q_1, \epsilon) + K_2(\Upsilon) - K_1(\Upsilon)$ and $K_s(\Upsilon) = \delta \sum_{j=0}^2 P_{s,j}(\Omega) E[V_j(\Omega') | \Upsilon]$, $s = 1, 2$; $\phi_D = \phi_1 - \phi_2$ and F_{ϕ_D} is the c.d.f. of the difference, which is assumed to be absolutely continuous with density f_{ϕ_D} ; and I use Assumption (A.3) to write $\Upsilon = \{T_j, A\}$.

A quick inspection of expression 3 shows that the effect of worker's ability on the probability of leaving the informal sector for the formal one is largely determined by the term $g(Q_2, \epsilon) - g(Q_1, \epsilon)$.¹⁴ If $g(Q_2, \epsilon) - g(Q_1, \epsilon)$ is decreasing in A , $\forall A$, then $\frac{h_{2,1}(\tau | A)}{\partial A} > 0$.¹⁵ This is true if and only if $\frac{\partial g(Q_1, \epsilon)}{\partial A} > \frac{\partial g(Q_2, \epsilon)}{\partial A}$,

¹³Given that the worker was currently holding an informal job, with tenure $T_2 = \tau_2$, it must have been the case that she preferred this option to unemployment for τ_2 periods. With enough persistence in the job quality shocks, this will still be the case in the following period.

¹⁴Clearly, a worker's skill affects her expectation of future rewards, as it is part of her information set. However, in this scenario, the only way skills affect the value function is through the $g = (\cdot)$ and these remain constant over time.

¹⁵To see why this is true, one can simply differentiate the hazard rate with respect to a :

$\forall A$. In words, as we move from the bottom to the top of the skill distribution, the rate of change in the utility derived from a formal job must be higher than the one from informal jobs. Following the same type of reasoning, it is easy to verify that the same holds for job duration in both sectors: if $\frac{\partial(g(Q_2,\epsilon)-g(Q_1,\epsilon))}{\partial A} < 0$, then more skilled workers are more likely to survive in a formal job.

Duration dependence

As for the the analysis of duration dependence, it is more convenient to rewrite the informal to formal hazard rate as

$$\begin{aligned}
 h_{2,1}(\tau|A) &= \Pr[\phi_1 \geq \phi_2 + \bar{C} | T_2 = \tau, A] \\
 &= 1 - \int_{-\infty}^{\infty} \int_{-\infty}^{\phi_2 + \bar{C}} dF_{\phi|T_2,A} dF_{\phi|T_2,A}
 \end{aligned}
 \tag{5}$$

where I use the fact that in the integrated model both sectors have the same job quality distribution.

As expression 5 shows, once one conditions on worker’s ability, the behavior of the individual hazard rate with respect to the duration of the spell (duration dependence) is completely determined by the time pattern of the job quality shock. In particular, if the shocks are persistent and positively correlated, then one would observe negative duration dependence, while if they are persistent, but negatively correlated, then one one would observe positive duration dependence. In the extreme case where the job quality shock is constant, one would observe $h_{1,2}(\tau|A) = h_{2,1}(\tau|A) = 0$ for all workers and all time periods.

This is important as the existence of high mobility rates between sectors is usually associated with absence of segmentation, while low informal-formal transition rates are interpreted as evidence of barriers to mobility. However, this framework shows that low informal-formal transition rates can emerge even in a completely integrated labor market. Similarly, even if the econometrician was able to confidently remove the effect of unobserved ability, the existence of negative duration dependence in the informal sector would not necessarily imply the existence of segmentation nor an “informality trap.” The negative duration dependence could emerge even in the context of perfect mobility, as long as there is enough persistence in the shocks ϕ .

Nonetheless, this model does deliver a very sharp prediction regarding the observed duration dependence pattern across sectors. From expression 5 it is possible to see that once one controls for the unobserved heterogeneity, job duration in both sectors should have the same pattern of duration dependence. In particular, transition rates from formal to informal and the other way around should be symmetric,

$$\frac{h_{2,1}(\tau|A)}{\partial A} = -f_{\phi_D}(\bar{C}_t) \left[\frac{\partial(g(Q_2,\epsilon)-g(Q_1,\epsilon))}{\partial A} \right], \text{ which is positive as long as } \frac{\partial(g(Q_2,\epsilon)-g(Q_1,\epsilon))}{\partial A} < 0.$$

for a given level of productivity and job tenure. This result is a consequence of the assumption that in the integrated model the shocks have the same distribution in both sectors, and thus display the same autocorrelation patterns.

4.1.2 Segmented labor markets

The effect of unobserved ability

In this model, selection into the different sectors clearly depends on ability. Once again, I start by analyzing the effect of ability on job-to-job transitions. As before, I assume that workers always prefer either employment statuses to unemployment. Following the analysis in Section 4.1.1, it is possible to write the job-to-job hazard rates in the formal and informal sectors, respectively, as follows:

$$h_{1,2}(\tau|A) = \Pr[\theta_2 = 1|T_1 = \tau, A] + \Pr[\theta_1 = 1|T_1 = \tau, A] \times \Pr[d_1 = 0|T_1 = \tau, A] \tag{6}$$

$$h_{2,1}(\tau|A) = \Pr[\theta_1 = 1|T_2 = \tau, A] \times \Pr[d_1 = 1|T_2 = \tau, A] \tag{7}$$

where I use the fact that, conditional on the employment history, d_1 and θ_j are independent (so one can factor out the two probabilities).

The above expressions can be rewritten as follows (I focus on the informal to formal hazard, but the formal-informal one would be analogous):

$$h_{2,1}(\tau|A) = \Pr[\theta_1 = 1|T_2 = \tau, A] \times \Pr[d_1 = 1|T_2 = \tau, A] = \left[1 - F_{\phi_D}(\tilde{C}|T_2 = \tau, A)\right] \times \Pr[d_1 = 1|T_2 = \tau, A] \tag{8}$$

where I use the expression derived in equation 3 with $\tilde{C} = \bar{C} + \omega_2(A) - \omega_1(A)$. Differentiating the the above expression with respect to worker’s ability, it is straightforward to see that

$$\frac{\partial h_{2,1}(\tau|A)}{\partial A} = \Pr[d_1 = 1|T_2 = \tau, A] \times \left(-f_{\phi_D}(\tilde{C}|T_2 = \tau, A) \underbrace{\frac{\partial \tilde{C}}{\partial A}}_{\leq 0} \right) \leq 0$$

where

$$\frac{\partial \tilde{C}}{\partial a} = \frac{\partial(\omega_2(A) - \omega_1(A))}{\partial A} + \frac{\partial(g(Q_2, \epsilon) - g(Q_1, \epsilon))}{\partial A}$$

Thus, as in the integrated labor market model, the sign of the selection bias depends on how the utility from formal and informal jobs is affected by unobserved

ability. If the total utility derived from a formal job increases in worker's ability at a higher rate than that derived from informal jobs, then there will be positive ability bias out of the informal sector and into the formal one. This will be the case if $\frac{\partial \omega_2(A)+g(Q_2,\epsilon)}{\partial A} < \frac{\partial (\omega_1(A)+g(Q_1,\epsilon))}{\partial A}$. This is very close to the condition derived in the integrated market model, except that in this case ability has a different direct effect on workers' potential earnings in both sectors (through the hedonic pricing functions ω_j).

Duration dependence

As opposed to the integrated labor market model discussed in subsection 4.1.1, in the segmented model, transitions are determined both by workers' choice and by the job rationing component. In this case, low transition rates between sectors can emerge due to high persistence of the job quality shocks or due to a high degree of formal job rationing [$\Pr (d_{1,t} = 1)$ is very low], or both. This can be seen by once again examining the job-to-job hazards:

$$\begin{aligned} h_{1,2}(\tau|A) &= \Pr [\theta_2 = 1|T_1 = \tau, A] + (1 - \Pr [\theta_2 = 1|T_1 = \tau, A]) \\ &\quad \times \Pr [d_1 = 0|T_1 = \tau, A] \\ &= \Pr [\theta_2 = 1|\Upsilon] \times (1 - \Pr [d_1 = 0|\Upsilon]) + \Pr [d_1 = 0|\Upsilon] \\ &= \left(1 - \int_{-\infty}^{\infty} \int_{-\infty}^{\phi_1 - \tilde{C}} dF_{\phi_2|\Upsilon} dF_{\phi_1|\Upsilon} \right) \times (1 - \Pr [d_1 = 0|\Upsilon]) \\ &\quad + \Pr [d_1 = 0|\Upsilon] \end{aligned}$$

$$\begin{aligned} h_{2,1}(\tau|A) &= \Pr [\theta_1 = 1|T_2 = \tau, A] \times \Pr [d_1 = 1|T_2 = \tau, A] \\ &= \left(1 - \int_{-\infty}^{\infty} \int_{-\infty}^{\phi_2 + \tilde{C}} dF_{\phi_1|\Upsilon} dF_{\phi_2|\Upsilon} \right) \times \Pr [d_1 = 1|\Upsilon] \end{aligned}$$

As it is clear from the expressions above, if the labor market is highly segmented – in the sense of high barriers to entry into the formal sector (i.e. $\Pr [d_1 = 1|\Upsilon] \rightarrow 0$) – then one would observe nearly zero informal-formal transition rates and possibly higher formal-informal transition rates for any given spell τ . Nonetheless, this is not empirically distinguishable from the situation where there is no rationing of formal jobs, but job quality shocks are highly and positively correlated across time in both sectors.

4.1.3 Discussion

The non-identification problem pointed out above is central to the discussion of empirically identifying the existence of segmentation. As demonstrated by Dickens

and Lang (1985) and Dickens and Lang (1992), the existence of barriers to workers' mobility between sectors is a crucial feature of the segmentation theory. The existence of different wage determination mechanisms in both sectors is important, but it is also present in the human capital theory, most notably in the Roy model [see Heckman and Hotz (1986) for an early example of the Roy model applied to the analysis of labor market segmentation].

Moreover, the existence of low mobility between sectors caused by high persistence of job quality shocks has a completely different welfare interpretation from low mobility that arises as a consequence of persistent job rationing. In the first case, there is no transition due to workers' choice and hence it should not raise equality or welfare concerns, as long as these shocks are random. This is not the case in the second situation, as there is a share of workers in the informal sector that would be better off in the formal sector. In this case, low mobility has a direct impact on workers' welfare.¹⁶

Nonetheless, the only hope of finding footprints of either model seems to be through the analysis of the duration dependence pattern in both sectors. As discussed in subsection 4.1.1, the integrated model implies that conditional on unobserved heterogeneity, both sectors should have the same pattern of duration dependence. This is a sharp prediction. Indeed, if one does not find the same pattern of duration dependence in both sectors, then one can reject the fully integrated model. However, one cannot reject the existence of a slightly weaker form of integration, which would be both sectors being completely integrated, but having different job quality distributions. This is arguably a completely different scenario for labor market functioning from the one with a fully segmented market. In particular, the former does not have any restriction on workers' mobility. So, there remains a significant identification problem: as one moves slightly away from the fully integrated model, it is no longer possible to empirically differentiate any of the models.

This is true even if one is willing to assume that the distribution of job quality shocks in the formal sector first-order stochastically dominates (FOSD) the informal sector one, which conforms with the existing empirical evidence that formal firms tend to be more productive than their informal counterparts.¹⁷ The FOSD, however, is a statement about how the distributions of job quality in both sectors compare in the cross-section, which will determine workers' sorting across sectors. But conditional on sector membership, the transition and duration patterns will be highly determined by the persistence of the process that generates the job quality shock. For instance, if a worker decided to sort into the informal sector at t , it must have been the case that she had a relatively better draw of job quality in

¹⁶It is important to note, however, that this model is not an equilibrium model. Thus, it is not suited to perform welfare analysis, as the general equilibrium implications of these different views of labor market functioning cannot be analyzed.

¹⁷I would like to thank an anonymous referee for suggesting this discussion.

the informal than in the formal sector (under perfect mobility). From then on, it is likely that she will still prefer the informal sector due to the high persistence of shock processes, despite the FOSD property. In the data, this would show up as low informal to formal transition, which is the same outcome that one would observe with high barriers to mobility.

This exercise points to the importance of imposing additional structure to the model in order to achieve identification. This comes from the fact that the segmented model has eight central building blocks – F_{ϕ_1} , F_{ϕ_2} , $\omega_1(\cdot)$, $\omega_2(\cdot)$, F_A , $\Pr[d_1 = 1]$, $g(Q_1, \epsilon)$ and $g(Q_2, \epsilon)$ – that determine two reduced-form relationships: $\frac{\partial h_{s,j}(\tau|A)}{\partial \tau}$, $s, j = 1, 2$. Hence, one can always manipulate these objects to obtain the same duration dependence pattern in the formal and informal sectors. The need to match the observed sign for the formal-informal wage gap and the difference in wage variance imposes additional restrictions (discussed in the following section), but there are still enough remaining degrees of freedom to obtain any intended result. Hence the need to impose additional assumptions that can pose restrictions on the behavior of these building blocks, such that the underlying model can generate empirical predictions that are both falsifiable and distinguishable from other models.

4.2 Cross-sectional distribution of log-wages

This section closely follows Barros and Ulyssea (2010), and I show that their differences in mean and variance decomposition also holds in this more general framework. In any period,¹⁸ let F_A and $F_A^j(\cdot) = F_A(\cdot|I_j = 1)$ denote the distributions of ability over \mathbb{L} and \mathbb{E}_j , $j = 1, 2$, respectively. Using Bayes rule one can write

$$dF_A^j = \frac{dF_A \pi_j(A)}{p_j}, \quad , j = 1, 2 \tag{9}$$

and hence under unbiased assignment (see D.3) one has that $dF_A^1 = dF_A^2 = dF_A$; that is, the distribution of abilities is the same in both sectors and is equal to the distribution of abilities over the entire population.

Let $F_w^j(\cdot) = F_w(\cdot|I_j = 1)$ denote the observed wage distribution in sector $j = 1, 2$ (i.e. the wage distribution over \mathbb{E}_j). From (A.1) it follows that we can write these conditional wage distributions as

$$F_w^j(x) = \int_{-\infty}^{\infty} F_{\omega_j}(x - r) dF_{\phi_j}^j(r), \quad j = 1, 2 \tag{10}$$

where $dF_{\phi_j}^j(\cdot) = dF_{\phi_j}(\cdot|I_j = 1)$, $j = 1, 2$, are the distributions of job quality in

¹⁸Since this section focuses on the cross-sectional distribution of log-wages, time subscripts are omitted everywhere.

the formal and informal sectors that are actually observed in any given period.¹⁹ Assuming that the distribution of abilities is absolutely continuous with density f_A , one can write $F_{\omega_j}(\cdot)$ as

$$F_{\omega_j}(x) = \int_{-\infty}^{\omega_j^{-1}(x)} f_A^j(y) dy = \int_{-\infty}^{\omega_j^{-1}(x)} \frac{\pi_j(A)}{p_j} f_A(y) dy, \quad j = 1, 2 \quad (11)$$

where I use the fact that the pricing functions are strictly increasing (by assumption). Substituting 11 into 10 we get a key expression:

$$F_w^j(x) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\omega_j^{-1}(x-r)} \frac{\pi_j(A)}{p_j} f_A(y) dy \right] dF_{\phi_j}^j(r), \quad j = 1, 2 \quad (12)$$

Hence, as in the static framework of Barros and Ulyssea (2010), the observed log-wage distributions in both sectors may differ due to three factors:

- (i) biased assignment: $\pi_j(A) \neq p_j$, $j = 1, 2$.
- (ii) different pricing functions, ω_j , $j = 1, 2$.
- (iii) different distributions of job quality across sectors: F_{ϕ_j} , $j = 1, 2$.

However, in Barros and Ulyssea (2010) the assignment process itself, the pricing functions and the distribution of job quality in both sectors are exogenous to the model. In the present framework, the assignment process depends on workers' sector choice, which by its turn crucially depends on current and future job prospects, as well as on the pricing function in each sector. Thus, opposed to Barros' analysis, the assignment process is endogenously determined.

¹⁹The dF_{ϕ_j} are the *potential* job quality distributions in each sector.

The formal-informal wage gap

The observed average wage gap between the formal and informal sectors can be decomposed in a very straightforward way:²⁰

$$\begin{aligned}
 \Delta\mu &= E[w_1|I_1 = 1] - E[w_2|I_2 = 1] & (13) \\
 &= \underbrace{E[\omega_1(A) - \omega_2(A)|I_1 = 1]}_{\Delta\mu_\omega} \\
 &+ \left(\underbrace{E[\omega_2(A)|I_1 = 1] - E[\omega_2(A)|I_2 = 1]}_{\Delta\mu_s} \right) \\
 &+ \left(\underbrace{E[\phi_1|I_1 = 1] - E[\phi_2|I_2 = 1]}_{\Delta\mu_\phi} \right)
 \end{aligned}$$

where $\Delta\mu_\omega$ corresponds to the average difference in the pricing of ability in both sectors (difference in the hedonic functions); $\Delta\mu_s$ is the bias due to self-selection on ability; and $\Delta\mu_\phi$ is the difference in average job quality in both sectors. The first term, $\Delta\mu_\omega$, is the equivalent of the treatment on the treated parameter (TOT). It has the causal interpretation of the average impact of holding a formal contract for those who would have been assigned to the formal sector.

4.2.1 Integrated labor markets

From expression (14), one can see that, in this case, the components relative to the average difference in the hedonic function and the difference in average job quality are equal to zero: $\Delta\mu_\omega = \Delta\mu_\phi = 0$. However, because selection depends on individuals' ability, the term relative to selection bias is not zero, $\Delta\mu_s \neq 0$. Thus, in this model, the wage differential is completely determined by the component relative to biased selection.

Nearly all of the empirical studies of formal-informal wage gap report a positive gap in favor of formal workers. As ω_j is strictly increasing in a , if more able workers have a higher probability of being in the formal sector, then $\Delta\mu_t = \Delta^s\mu_t > 0$. As discussed in the previous subsection, this will be the case if $g(Q_2, \epsilon) - g(Q_1, \epsilon)$ is decreasing in A for all A .

²⁰The difference in wage variances can be decomposed in the same way:

$$\begin{aligned}
 \Delta\sigma^2 &= Var[w_1|I_1 = 1] - Var[w_2|I_2 = 1] = \underbrace{Var[\omega_1(A)|I_1 = 1] - Var[\omega_2(A)|I_1 = 1]}_{\Delta\sigma_\omega^2} \\
 &+ \underbrace{(Var[\omega_2(A)|I_1 = 1] - Var[\omega_2(A)|I_2 = 1])}_{\Delta\sigma_s^2} + \underbrace{Var[\phi_2|I_1 = 1] - Var[\phi_2|I_2 = 1]}_{\Delta\sigma_\phi^2}
 \end{aligned}$$

where the three components have the same interpretation as above.

4.2.2 Segmented labor markets

In this model, the wage gap has no clear interpretation, as all three components in equation 14 are different from zero. Under the assumption that $\frac{\partial \bar{C}}{\partial A} < 0$, then the self-selection component is positive, $\Delta\mu_s > 0$. If one adopts the traditional view of the formal sector being the protected sector and the informal sector the disadvantaged one, one would expect to have $\omega_1(A) \geq \omega_2(A)$, $\forall A$, and $\omega_1(A) > \omega_2(A)$ for some A , which implies $\Delta\mu_\omega > 0$. This assumption is consistent with the evidence that the quantile log-wage differences are everywhere positive [see Barros and Ulyssea (2010)]. If this is the case, then one does not need to assume that the average job quality in the formal sector is higher than in the informal sector to obtain $\Delta\mu > 0$. As long as the two previous conditions hold, the average job quality in the informal sector could be higher and still the formal-informal wage gap would be positive as long as $\Delta\mu_\omega + \Delta\mu_s > |\Delta\mu_\phi|$. Of course, if the average job quality in the formal sector is higher ($\Delta\mu_\phi > 0$), then the wage gap will be unambiguously positive. The analogous reasoning is valid for the variance decomposition.

It is also worth noting that the existence of a positive formal-informal wage gap does not immediately admit a causal interpretation. If one controls for selection bias (using a control function approach, for example), there will be a remaining unobserved component due to job quality differences across sectors. If, however, one has access to matched employer-employee data, then it would be possible to control for job quality as well and one could be more confident to have identified the TOT parameter, $\Delta\mu_\omega$. Following the same logic, if after controlling for selection bias and firms' characteristics the estimated gap is close to zero, then this is indicative of no causal effect of holding a formal contract.

5. Final Remarks

The objective of this paper was to analyze the empirical content of the formal-informal labor market segmentation hypothesis. For that, I build on and extend the pioneering work of Barros (1988) into the following directions:

- (i) I consider a dynamic framework;
- (ii) workers' sector choice is explicitly modeled; and
- (iii) I introduce a stochastic job rationing component in both sectors.

The last feature is motivated by the fact that labor market segmentation might manifest itself not only through different workers' skill pricing or job quality across sectors, but also through restrictions to workers' mobility between sectors.

This general framework is specialized to two polar views of labor market functioning, namely the fully integrated and segmented labor markets. I then analyze

what the empirical implications would be if the data generating process corresponded to these two polar cases, with the final goal of assessing whether or not they can generate implications that are empirically distinguishable. Put differently, the objective is to assess to which extent the findings in the empirical literature can be supported by a single model or by a broad array of models that imply completely opposite views of labor market functioning.

The analysis indicates that there seems to be little hope of empirically identifying the existence or not of segmentation: The two polar cases considered here – fully segmented and fully integrated labor markets – can generate the observed regularities regarding formal-informal wage gaps, transition rates and job duration in the formal and informal sectors. In particular, both models can generate positive or negative wage gaps, positive or negative ability bias in employment sector choice and duration dependence (even after conditioning on unobserved ability).

A natural extension of the present analysis is to therefore investigate which plausible assumptions, if any, would ensure identification. Barros and Ulyssea (2010) and Barros et al. (1993) make some efforts in this direction, but more is needed in order to better understand what can be learned from the available empirical evidence, as well as to unveil new facts that can be informative about the existence or not of labor market segmentation.

References

- Barros, R., Reis, J. G., & Rodriguez, J. (1990). Segmentação no mercado de trabalho: A carteira de trabalho na construção civil. *Brazilian Review of Econometrics*, 10(2):313–335.
- Barros, R. P. (1988). On the empirical content of the formal-informal labor market segmentation hypothesis. *Unpublished*.
- Barros, R. P., Mello, R., & Pero, V. (1993). Informal labor contracts: A solution or a problem? Texto para Discussão 291, Instituto de Pesquisa Economica Aplicada.
- Barros, R. P. & Ulyssea, G. (2010). On the empirical content of the formal-informal labor market segmentation hypothesis. *Brazilian Review of Econometrics*, 2.
- Curi, A. T. & Menezes-Filho, N. A. (2006). O mercado de trabalho brasileiro é segmentado? Alterações no perfil da informalidade e nos diferenciais de salários nas décadas de 80 e 90. *Estudos Econômicos*, 36:867–899.
- Dickens, W. T. & Lang, K. (1985). A test of dual labor market theory. *American Economic Review*, 75(4):792–805.
- Dickens, W. T. & Lang, K. (1992). Labor market segmentation theory: Reconsidering the evidence. NBER Working Papers 4087, National Bureau of Economic Research.

- Fields, G. S. (1975). Rural-urban migration, urban unemployment and underemployment and job-search activity in LDCs. *Journal of Development Economics*, 2(2):165–187.
- Gong, X., van Soest, A., & Villagomez, E. (2004). Mobility in the urban labor market: A panel data analysis for Mexico. *Economic Development and Cultural Change*, 53(1):1–36.
- Heckman, J. J. & Hotz, V. J. (1986). An investigation of the labor market earnings of panamanian males evaluating the sources of inequality. *The Journal of Human Resources*, 21:507–542.
- 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. *World Bank Economic Review*, 13(2):275–302.
- Szerman, D. & Ulyssea, G. (2007). Job duration and the informal sector in Brazil. *Presented at the Second IZA-World Bank Conference on Employment and Development*.
- Ulyssea, G. (2006). Informalidade no mercado de trabalho brasileiro: Uma resenha da literatura. *Brazilian Review of Political Economy*, 26(3):596–618.