ON THE EMPIRICAL CONTENT OF THE FORMAL-INFORMAL LABOR MARKET SEGMENTATION HYPOTHESIS

Ricardo Paes de Barros
ON THE EMPIRICAL CONTENT OF THE FORMAL-INFORMAL LABOR MARKET SEGMENTATION HYPOTHESIS

Ricardo Paes de Barros
Federal Government of Brazil

Secretariat of Strategic Affairs of the Presidency of the Republic

Minister Roberto Mangabeira Unger

ipea Institute for Applied Economic Research

A public foundation affiliated to the Secretariat of Strategic Affairs of the Presidency of the Republic, Ipea provides technical and institutional support to government actions — enabling the formulation of numerous public policies and programs for Brazilian development — and makes research and studies conducted by its staff available to society.

President
Sergel Suarez Dillon Soares

Director of Institutional Development
Luiz Cezar Loureiro de Azeredo

Director of Studies and Policies of the State, Institutions and Democracy
Daniel Ricardo de Castro Cerqueira

Director of Macroeconomic Studies and Policies
Cláudio Hamilton Matos dos Santos

Director of Regional, Urban and Environmental Studies and Policies
Rogério Boueri Miranda

Director of Sectoral Studies and Policies, Innovation, Regulation and Infrastructure
Fernanda De Negri

Director of Social Studies and Policies, Deputy
Carlos Henrique Leite Corseuil

Director of International Studies, Political and Economic Relations
Renato Coelho Baumann das Neves

Chief of Staff
Ruy Silva Pessoa

Chief Press and Communications Officer
João Cláudio Garcia Rodrigues Lima

URL: http://www.ipea.gov.br
Ombudsman: http://www.ipea.gov.br/ouvidoria

DISCUSSION PAPER

A publication to disseminate the findings of research directly or indirectly conducted by the Institute for Applied Economic Research (Ipea). Due to their relevance, they provide information to specialists and encourage contributions.


ISSN 1415-4765


CDD 330.908

The authors are exclusively and entirely responsible for the opinions expressed in this volume. These do not necessarily reflect the views of the Institute for Applied Economic Research or of the Secretariat of Strategic Affairs of the Presidency of the Republic.

Reproduction of this text and the data it contains is allowed as long as the source is cited. Reproductions for commercial purposes are prohibited.
Este trabalho é de inteira e exclusiva responsabilidade de seu autor. As opiniões nele emitidas não exprimem, necessariamente, o ponto de vista da Secretaria de Planejamento e Coordenação da Presidência da República.
ON THE EMPIRICAL CONTENT OF THE FORMAL-INFORMAL LABOR MARKET SEGMENTATION HYPOTHESIS*

Ricardo Paes de Barros**

1. Introduction

2. Basic Concepts and Notation
   2.1 Universe, Selection Process and Potential Log-wages
   2.2 Observed Distributions and their Characteristics

3. Empirical Findings
   3.1 The Sample
   3.2 The Evidence

4. A General Framework for Wage Determination and Selection Process
   4.1 Wage Determination
   4.2 Selection Process
   4.3 The Distribution and Abilities by Sector
   4.4 Log-wage Distributions by Sector

5. Decomposing and Interpreting Observed Differences in Means, Variances and Quantiles
   5.1 The Average Impact of Formal Labor Contracts
   5.2 Decomposing and Interpreting Observed Differences in Means
   5.3 Decomposing and Interpreting Observed Differences in Variances
   5.4 Interpreting Observed Differences in Quantiles

   6.1 Homogeneous Jobs and Hedonic Functions
   6.2 Random Selection with Homogeneous Jobs
   6.3 Homogeneous Labor

Footnotes

References

* This paper is part of a larger project joint with Jaime P. Pontes and Simone Varandas on the segmentation of labor markets in Brazil. I would like to thank all of the participants of the INPES workshop on Labor Economics where most the ideas in this paper ware generated. In particular, I am grateful to José Marcio Camargo, Ricardo Cicchelli, Jaime Pontes, Guilherme Sedlacek, Octávio Tourinho, Pedro Valls, and Simone Varandas for valuable discussions. I also would like to thank Luiz Araújo, Stephen Cameron, James Heckman, and Joseph Hotz for extremely valuable comments in this version of the paper. Vandeli Guerra helped me to better understand the details of the PNAD data set. Ana Isabel Alvarenga provided excellent computational support. Finally, I would like to acknowledge the William and Flora Hewlett Foundation for partial financial support.

** Do INPES/IPEA
ABSTRACT

In this paper we pursue three objectives. First, we compare the wage-distributions in the informal and formal sectors for a group of workers employed in the Brazilian Construction Sector. The empirical regularities we encounter are not, however, specific to this particular group of workers. Indeed, similar results are also observed for several other homogeneous groups. Second, we investigate how observed differences in means, variances, and quantiles should be interpreted. Finally, we describe three models for the formal-informal segmentation of the labor market, analyse their consistency with the observed regularities, and discuss how these regularities should be interpreted in the context of each individual model. We conclude that the observed regularities are consistent with a wide range of models, although their interpretation varies remarkably depending on the model we are considering.

SINOPSE

Este artigo possui três objetivos: primeiro procura-se comparar a distribuição de salários entre os empregados com e sem carteira de trabalho. Embora tenha-se considerado apenas aqueles que trabalham na construção civil, as regularidades empíricas encontradas não são específicas a este grupo. De fato, Barros e Varandas (1987) mostraram que resultados similares são também observados para diversos outros grupos de trabalhadores. Em segundo lugar, investiga-se como as diferenças observadas tanto com respeito à média como quanto às variações e aos quantis devem ser interpretadas. Finalmente, três modelos para a segmentação do mercado de trabalho são desenvolvidos e a consistência destes modelos com as regularidades empíricas observadas é avaliada. Conclui-se que as regularidades observadas são consistentes com uma grande variedade de modelos, embora a interpretação destas regularidades varie substancialmente de acordo com o modelo observado.
INTRODUCTION

In some of our recent work (Barros and Varandas(1987a,b)) we compare the wage distributions of observably similar workers in the formal and informal sectors. In these studies we contrast not only the averages but also several indices of inequality and the quantiles of the wage distributions in the two sectors. In this paper we focus on how these observed differences should be interpreted.

As largely emphasized in the literature, if (1) workers are heterogeneous and perceived as such by firms, and (2) the selection of workers to sectors is nonrandom, then observed differences between the wage distributions in these two sectors don't have an immediate causal interpretation. This is the "selection bias" question. However, there are other sources of difficulties involved in the comparison of wage distributions which have not received as much attention. For instance, if it is possible to find equally able workers earning unequal wages within sectors, then as we are going to show, even if the selection of workers to sectors were random, it would still be nontrivial to interpret, for example, observed differences in inequality between the wage-distributions in the two sectors. In this case, it would be difficult to separate out worker heterogeneity from job heterogeneity.

In this paper we pursue three objectives. First, we compare the wage-distributions in the informal and formal sectors for a group of workers employed in the Brazilian Construction Sector. The empirical regularities we encounter are not, however, specific to this particular group of workers. Indeed, similar results are also observed for several other homogeneous groups. Second, we investigate how observed differences in means, variances, and quantiles should be interpreted. Finally, we describe three models for the formal-informal segmentation of the labor market, analyze their
consistency with the observed regularities, and discuss how these regularities should be interpreted in the context of each individual model. We conclude that the observed regularities are consistent with a wide range of models, although their interpretation varies remarkably depending on the model we are considering.

2-Basic Concepts and Notation

2.1-Uneiverse, Selection Process and Potential Log-Wages

Let $P$ be a population of workers. We assume that members of $P$ are homogeneous with respect to all observable attributes. A selection process is a mechanism to obtain a disjoint partition $(P_0, P_1)$ of $P$, such that members of $P_0$ work in the informal sector while members of $P_1$ work in the formal sector. For each worker $p$ in $P$, let $d(p)=0$ if he works in the informal sector and $d(p)=1$ if he works in the formal sector, i.e., $d(p)=0$ if $p \in P_0$ and $d(p)=1$ if $p \in P_1$.

For each worker $p$ in $P$, let $w_0(p)$ and $w_1(p)$ denote his potential log-wages in the informal and formal sectors, respectively. These log-wages are generated by allocating worker $p$ alternatively to the informal and formal sectors while holding constant the allocation of all other workers. We assume that $w_0$ and $w_1$ have finite variances.

Throughout the paper, cumulative distribution functions (c.d.f.) are denoted by $F$, whereas densities are denoted by $f$. Subscripts will indicate to which random variable they refer. So, for instance, $F_{w_0}$ denotes the c.d.f. for $w_0$ over the entire population $P$. Cumulative distributions constrained to the subpopulations $P_0$ and $P_1$ are identified by superscripts. In particular, $F_{w_0}^0$ denotes the c.d.f. for $w_0$ constrained to $P_0$. Therefore, $F_{w_0}^0(\cdot) - F_{w_0}^1(\cdot | d=0)$. If the selection of workers to sectors were random then, as one implication,
we would have \( F_w^0 = F_w^1 = F_w \).

It follows from the way \( w_0(p) \) and \( w_1(p) \) are generated that, in general, these log-wages depend on the allocation of all other workers in the population. In other words, \( F_w^0 \) and \( F_w^1 \) generally depend on which particular partition \((P_0', P_1')\) of \( P \) is being considered. In particular, they may depend on the relative size of the informal sector. For example, a policy towards the formalization of labor relations would probably lead to an exogenous movement of workers from \( P_0 \) to \( P_1 \). This movement should be expected to depress wages in the formal sector while perhaps increasing them in the informal sector. Throughout this paper, however, the partition \((P_0', P_1')\) is held fixed. So, it is not necessary to make explicit the dependence of these underlying log-wage distributions \((F_w^0, F_w^1)\) on the selection process.

2.2-OBSERVED DISTRIBUTIONS AND THEIR CHARACTERISTICS

For each worker \( p \) in \( P \), only one log-wage is actually realized: \( w_0(p) \) or \( w_1(p) \) depending on which sector the worker is allocated. For instance, if \( d(p)=1 \) then \( w_1(p) \) would be the observable log-wage for worker \( p \). As we have already mentioned, in this case \( w_0(p) \) represents the log-wage this worker would earn if he were reallocated to the informal sector (holding the allocation of all other workers fixed). Notice that \( w_0(p) \) is not necessarily the wage worker \( p \) would earn if the formal sector were abolished.

Since for each worker, only the wage in the sector he currently belongs to can be observed, \( F_w^0 \) and \( F_w^1 \) are the only log-wage distributions which can be directly observed. We concentrate our attention on three characteristics of these distributions: the mean \( (\mu) \), the variance \( (\sigma^2) \), and the quantiles \( (q(\alpha), 0<\alpha<1) \). The observed differences for each of these characteristics are given by.
\[ \Delta \mu = \int \mu_1 \cdot dF_1^0(\omega) - \int \mu_1 \cdot dF_0^0(\omega) - E[\omega_1|d=1] - E[\omega_0|d=0]; \]

\[ \Delta \sigma^2 = \text{Var}[\omega_1|d=1] - \text{Var}[\omega_0|d=0]; \]

\[ \Delta q(\alpha) = q_1^1(\alpha) - q_0^0(\alpha) \quad \text{for} \quad 0<\alpha<1, \]

where \( q_1^1(\alpha) \) and \( q_0^0(\alpha) \) are defined as follows:

\[ q_1^1(\alpha) = \inf\{q:F_1^1(q) \geq \alpha\} \]

\[ q_0^0(\alpha) = \inf\{q:F_0^0(q) \geq \alpha\} \]

In the following section we estimate these differences. In Section 5 we discuss how they should be interpreted.

3 - EMPIRICAL FINDINGS

3.1 - THE SAMPLE

The Data Set used is derived from the 1985 version of the Brazilian Yearly Household Survey (PNAD-Pesquisa Nacional por Amostra de Domicílios). More specifically, our sample consists of all currently employed workers who satisfy the following inclusion requirements: (i) are male heads of household, (ii) are residents of an urban area in one of the nine Brazilian metropolitan regions, (iii) are engaged in one of the twelve occupations which are considered typical for the construction sector, (iv) do not work for the government, (v) do not have any college education, and (vi) are not older than seventy. 2,080 workers in the PNAD-85 satisfy these restrictions.

INPES, 152/88
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Formal Sector</th>
<th>Informal Sector</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>37.6</td>
<td>37.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary (years)</td>
<td>3.00</td>
<td>2.50</td>
<td>0.5</td>
</tr>
<tr>
<td>High-School (%)</td>
<td>2.4</td>
<td>2.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Metropolitan Region (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belém</td>
<td>1.5</td>
<td>2.5</td>
<td>-1.0</td>
</tr>
<tr>
<td>Fortaleza</td>
<td>4.9</td>
<td>10.8</td>
<td>-5.9</td>
</tr>
<tr>
<td>Recife</td>
<td>6.0</td>
<td>6.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>Salvador</td>
<td>6.7</td>
<td>5.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Belo Horizonte</td>
<td>9.3</td>
<td>8.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Rio de Janeiro</td>
<td>27.0</td>
<td>29.9</td>
<td>-2.9</td>
</tr>
<tr>
<td>Sao Paulo</td>
<td>32.2</td>
<td>29.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Curitiba</td>
<td>5.1</td>
<td>3.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Porto Alegre</td>
<td>7.3</td>
<td>4.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Occupation (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bricklayer</td>
<td>37.0</td>
<td>34.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Hodman</td>
<td>21.6</td>
<td>43.1</td>
<td>-21.5</td>
</tr>
<tr>
<td>Plumber</td>
<td>11.0</td>
<td>2.2</td>
<td>8.8</td>
</tr>
<tr>
<td>Construction Foreman</td>
<td>8.5</td>
<td>4.1</td>
<td>4.4</td>
</tr>
<tr>
<td>Painter</td>
<td>7.5</td>
<td>8.0</td>
<td>-0.5</td>
</tr>
<tr>
<td>Machine Operator</td>
<td>5.6</td>
<td>0.5</td>
<td>5.1</td>
</tr>
<tr>
<td>Concrete Framer</td>
<td>4.6</td>
<td>2.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Tiler</td>
<td>1.5</td>
<td>3.0</td>
<td>-1.5</td>
</tr>
<tr>
<td>Glazier</td>
<td>1.1</td>
<td>0.2</td>
<td>0.9</td>
</tr>
<tr>
<td>Plasterer</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Paver Asphalt</td>
<td>0.5</td>
<td>1.0</td>
<td>-0.5</td>
</tr>
<tr>
<td>Calker</td>
<td>0.1</td>
<td>1.0</td>
<td>-0.9</td>
</tr>
</tbody>
</table>

Source: PNAD-85 - Public-use tapes.

Notes: 1) The sample consists of 2,080 workers, 1,468 belonging to the formal sector and 612 to the informal sector. The sample inclusion requirements are described in the text.

2) The figures include high-school dropouts.
and consequently form our sample. From this total 70.6%, that is, 1,468 workers, have formal labor contracts and hence belong to the formal sector.

To keep the sample size large enough, we have not attempted to control for observable individual attributes like age, education, region of residence and occupation. Nonetheless, the distribution of workers by the first three attributes ends up being remarkably similar in the formal and informal sectors, as shown in Table 1. The distribution of workers by occupation is, however, considerably different in the two sectors. In Barros, Pontes and Varandas (1988) we study the impact of controlling for these attributes in a linear regression framework. We show that age, education, and region of residence have only a marginal effect on the estimation of the average impact of formal labor contracts on log-wages. On the other hand, occupation has a dramatic effect. When this variable is introduced in the analysis, it causes a fourfold reduction in the estimated impact of formal labor contracts on log-wages. Due to a very likely endogeneity of occupation status, it is not clear, however, whether we really would like to perform the analysis conditional on this attribute. In particular, the wage impact of a formal labor contract may well arise via an occupational upgrading. That is, the reward may be in the form of giving the worker easier access to better occupations.

3.2-The Evidence

Figure 1 presents nonparametric estimates for the probability density of wages in the informal and formal sectors. The wages are measured in multiples of the legal minimal wage (MW). A uniform kernel estimator with a window of 0.4MW has been used. The estimated densities cross each other just once, around 1.7MW. Hence, Figure 1 gives a clear indication that the distribution of wages in the informal sector is stochastically dominated by
Figure 1

Wage probability densities by sector

Informal sector

Formal sector
Figure 2

Wage cumulative distributions by sector

Informal sector

Formal sector
Figure 3
likelihood ratio between the wage dist. in the formal and informal sector
the wage distribution in the formal sector. This fact is verified in Figure 2 where estimates for \( F_{w_0}^0 \) and \( F_{w_1}^1 \) are presented. As this Figure shows, the estimate for \( F_{w_1}^1 \) indeed lies entirely to the right of the estimate for \( F_{w_0}^0 \). Figure 3 portraits the likelihood ratio. The "monotonicity" displayed by the likelihood ratio in this figure implies that the distribution of wages in the informal sector is stochastically dominated by the distribution of wages in the formal sector in an even stronger sense (see Ross(1983,p.266)). From these figures we can also see that both the level and the inequality of wages are greater in the formal sector.

Table 2 summarizes the basic characteristics of the wage distributions by sector. The table clearly confirms the findings in Figures 1, 2 and 3 that the level and the inequality of wages are both higher in the formal sector. As the table also shows, the quantile absolute and relative differences are positive and increasing. The relative differences, however, tend to be more stable. As one would expect, the proportion of workers below the legal minimal wage is much smaller in the formal sector (4.1%) than in the informal sector (17.7%).

Table 3 summarizes similar characteristics for the log-wage distributions by sector. Of course, the findings are the same: both the level and the inequality are higher in the formal sector. The last column in the table also shows a moderately increasing pattern for the quantile differences. This phenomenon is further explored in Figure 4 where all log-wage quantiles are estimated for \( 0.1 \leq \alpha \leq 0.9 \). (Estimates in the tails are necessarily of poor quality and consequently are not reported.) Figure 4 demonstrate a clear tendency of the quantile differences to increase over the range considered.

In summary, there are three general findings: (1) The level of wages
Table 2
Summary Statistics for the Wage-Distributions in the Formal and Informal Sectors

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Formal Sector</th>
<th>Informal Sector</th>
<th>Absolute Difference</th>
<th>Relative Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.42</td>
<td>1.79</td>
<td>0.63</td>
<td>26</td>
</tr>
<tr>
<td>Quantiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q_{25}</td>
<td>1.35</td>
<td>1.08</td>
<td>0.27</td>
<td>20</td>
</tr>
<tr>
<td>Q_{50}</td>
<td>1.92</td>
<td>1.50</td>
<td>0.42</td>
<td>22</td>
</tr>
<tr>
<td>Q_{75}</td>
<td>2.88</td>
<td>2.16</td>
<td>0.72</td>
<td>25</td>
</tr>
<tr>
<td>P[Wage&lt;1HW]</td>
<td>4.1</td>
<td>17.7</td>
<td>-13.6</td>
<td>-332</td>
</tr>
<tr>
<td>P[Wage&lt;2HW]</td>
<td>53.1</td>
<td>69.7</td>
<td>-16.6</td>
<td>-31</td>
</tr>
<tr>
<td>P[Wage&lt;3HW]</td>
<td>76.9</td>
<td>87.3</td>
<td>-10.4</td>
<td>-14</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>1.62</td>
<td>1.04</td>
<td>0.58</td>
<td>36</td>
</tr>
<tr>
<td>Q_{75}/Q_{25}</td>
<td>2.13</td>
<td>2.00</td>
<td>0.13</td>
<td>6</td>
</tr>
<tr>
<td>Coeff. Variation</td>
<td>0.67</td>
<td>0.58</td>
<td>0.09</td>
<td>13</td>
</tr>
<tr>
<td>Theil</td>
<td>0.18</td>
<td>0.14</td>
<td>0.04</td>
<td>22</td>
</tr>
</tbody>
</table>

Source: PNAD-85 - Public-use tapes.

Notes: 1) The sample consists of 2,080 workers, 1,468 belonging to the formal sector and 612 to the informal sector. The sample inclusion requirements are described in the text.
Table 3
Summary Statistics for the Log-Wage Distributions in the
Formal and Informal Sectors

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Formal Sector</th>
<th>Informal Sector</th>
<th>Absolute Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.72</td>
<td>0.44</td>
<td>0.28</td>
</tr>
<tr>
<td>Quantiles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Q_{25} )</td>
<td>0.30</td>
<td>0.07</td>
<td>0.23</td>
</tr>
<tr>
<td>( Q_{50} )</td>
<td>0.65</td>
<td>0.41</td>
<td>0.24</td>
</tr>
<tr>
<td>( Q_{75} )</td>
<td>1.06</td>
<td>0.77</td>
<td>0.29</td>
</tr>
<tr>
<td>Variance</td>
<td>0.30</td>
<td>0.28</td>
<td>0.02</td>
</tr>
<tr>
<td>( Q_{75} - Q_{25} )</td>
<td>0.76</td>
<td>0.70</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Source: PNAD-85 - Public-use tapes.

Notes: 1) The sample consists of 2,080 workers, 1,468 belonging to the formal sector and 612 to the informal sector. The sample inclusion requirements are described in the text.
FIGURE 4

log-wage quantile differences between the formal and informal sectors
is generally higher in the formal sector; in particular our estimate of $\Delta \mu$ equals 0.28 (see Table 3). (2) The inequality of wages is higher in the formal sector; in particular our estimate of $\Delta \sigma^2$ equals 0.02 (see Table 3). (3) The quantile differences are everywhere positive and increasing, i.e., $\Delta q(\alpha) > 0$ and $\partial \Delta q(\alpha)/\partial \alpha > 0$ for all $0 < \alpha < 1$ (see Figure 4). Notice that $\Delta q(\alpha) > 0$ for all $0 < \alpha < 1$ implies that $F^0_\omega$ is stochastically dominated by $F^1_\omega$.

In the following sections we investigate possible interpretations of these three findings and study models for the formal-informal segmentation of the labor market which are compatible with them. First, however, we will specify how wages are determined and workers allocated to sectors.

4-A GENERAL FRAMEWORK FOR WAGE DETERMINATION, AND SELECTION PROCESS

In this section we introduce a general framework based on a weak set of assumptions. The main contribution of this framework is to permit a coherent study of a wide range of models for the formal-informal segmentation of the labor market.

4.1-WAGE DETERMINATION

Assume wages are determined by the value of the workers' marginal product. The productivity of a worker is assumed to be a function of his abilities as well as of the quality of the job he currently occupies. By hypothesis, workers in $F$ are observably equal, so they all must have the same level of observed abilities. They may differ, however, with respect to their unobserved abilities. We assume the unobserved abilities can be summarized by a scalar, which we denote by $A(p)$ for each $p$ in $F$.

We do not necessarily assume the selection process is based on pure comparative advantage. In other words, we do not necessarily assume that workers are free to choose the sector they currently work in, or that there

INPES, 152/88
is no mobility cost. Therefore, the average wage of workers with a given ability level may differ in the two sectors, i.e., sectors may value ability at different rates. Denote by $s_0(a)$ and $s_1(a)$ the average log-wage of workers with ability $A=a$ in the informal and formal sectors respectively. So,

$$s_0(a) = E[w_0 | A=a],$$

$$s_1(a) = E[w_1 | A=a],$$

for all $a \in \mathbb{R}$. These regression functions are henceforth referred to as hedonic functions. They are assumed to be strictly increasing and differentiable. When $s_0 = s_1$ we say the model has homogeneous hedonic functions.

Within sectors, equally able workers may have different wages. This is possible, since the ability of a worker is not assumed to completely determine his productivity. Differences in productivity among equally able workers within sectors are derived from differences in the quality of the jobs they currently occupy. In other words, if equally able workers interchange their jobs then their productivities and hence wages would be equally interchanged. Let

$$u_0 = w_0 - s_0(A),$$

$$u_1 = w_1 - s_1(A).$$

Under these assumptions, the variability of wages in each sector due to the components $u_0$ or $u_1$ is a measure of the degree of job heterogeneity in the sector. Accordingly, whenever $\text{Var}[u_1] > \text{Var}[u_0]$ we say that jobs in the formal sector are more heterogeneous than jobs in the informal sector. Using this notation, $w_0$ and $w_1$ can be expressed as follows:
\[ w_0 = s_0(A) + u_0 \]  \hspace{1cm} \text{(informal sector)},
\[ w_1 = s_1(A) + u_1 \]  \hspace{1cm} \text{(formal sector)}.

For simplicity, we assume that \((u_0, u_1)\) and \(A\) are independent.\(^{11}\)

4.2-Selection Process

Given a pair of hedonic functions \((s_0, s_1)\), the potential log-wages \((w_0(p), w_1(p))\) for each worker \(p\) in \(P\) is completely determined by his ability level \(A(p)\) and the quality of his pair of potential job assignments \((u_0(p), u_1(p))\). We consider only selection models in which the assignment of workers to sectors may depend on their ability levels but are independent of the quality of their potential job assignments. More precisely, we assume

\[ P[d=1|A, u_0, u_1] = P[d=1|A] = \lambda(A). \]

This equation states that equally able workers have identical probabilities of being allocated to the formal sector independently of the quality of their potential job assignments.

This assumption excludes some interesting cases. For example, suppose a demographic group suffers market discrimination in the formal sector, such that whenever they are selected into that sector they end up being assigned to jobs of low quality. In this case one should expect to observe them being over-represented in the informal sector, even holding ability constant. In this example, selection would be based on job prospects and, consequently, the hypothesis above of pure selection on ability would be violated.

This selection hypothesis has two important implications. First it implies that the hedonic functions are not distorted by the selection
process, i.e.,

\[ E[w_0|A,d=1] = E[w_0|A,d=0] = E[w_0|A] = s_0(A), \]
\[ E[w_1|A,d=1] = E[w_1|A,d=0] = E[w_1|A] = s_1(A). \]

Second, it implies that the distribution of jobs by quality in the two sectors is also not distorted by the selection process, i.e.,

\[ F^0_{u_0} = F^1_{u_0} = F_{u_0}, \]
\[ F^0_{u_1} = F^1_{u_1} = F_{u_1}. \]

In particular,

\[ E[u_0|d=1] = E[u_0|d=0] = E[u_0] = 0, \]
\[ Var[u_0|d=1] = Var[u_0|d=0] = Var[u_0], \]

with similar expressions holding for \( u_1 \).

4.3-The Distribution of Abilities by Sector

Following our notation, let \( F^0_A \), \( F^1_A \) and \( F_A \) denote the distributions of abilities over \( F \), \( F_0 \) and \( F_1 \) respectively. Notice that, using the Bayes rule, it can be easily shown that these distributions are related to each other as follows:

\[ dF^0_A = dF_A \cdot (1-\lambda)/(1-p), \]
\[ dF^1_A = dF_A \cdot \lambda/p, \]
where

\[ p = \int R \lambda(a) \cdot dF^1_A(a), \]

i.e., \( p \) is the proportion of the population selected to the formal sector. In the case that \( \lambda = p \) the selection is said to be random and \( F^0_A = F^1_A \). So, in this special case the distribution of abilities is the same in the formal and informal sectors. Consider now the following example where the selection process is not random.

**Example 1.** Let \( \lambda(a) = 0 \) for \( a < 0 \) and \( \lambda(a) = 1 - e^{-a} \) for \( a \geq 0 \). In this case \( \frac{\partial \lambda(a)}{\partial a} = 0 \). So, the higher is the ability of a worker the higher is his probability of being assigned to the formal sector. Next, assume that abilities are exponentially distributed with unit mean; that is, let \( f^0_A(a) = 0 \) for \( a < 0 \) and \( f^1_A(a) = e^{-a} \) for \( a \geq 0 \). In this case we obtain \( p = 1/2 \), \( f^0_A(a) = f^1_A(a) = 0 \) for \( a < 0 \), and

\[ f^0_A(a) = 2 \cdot e^{-2a}, \]

\[ f^1_A(a) = 2 \cdot e^{-a} \cdot (1 - e^{-a}), \]

for \( a \geq 0 \). So, fifty percent of the population is in each sector. The distribution of abilities in the informal sector is exponential with mean equal to 1/2. In the formal sector the average ability is 3/2, but the distribution is not exponential. The variance of the distribution of abilities over \( F \) is 1, while the conditional variances are 1/4 and 5/4, respectively for the informal and formal sectors. In summary, the selection process leads to a distribution of abilities in the formal sector which has a higher mean and also a higher variance.
4.4-Log-Wage Distributions by Sector

It follows from the assumption that $A$ and $(u_0, u_1)$ are independent, that

\[ F^0_w(w) = \int_{-\infty}^{+\infty} F^0(u-r) \cdot dF^0_u(r), \]
\[ F^1_w(w) = \int_{-\infty}^{+\infty} F^1(u-r) \cdot dF^1_u(r). \]

If the distribution of abilities is absolutely continuous with density $f_A$ then

\[ F^0_{s_0}(w) = \int_{-\infty}^{s_0^{-1}(w)} f_A^0(a) \cdot da = \int_{-\infty}^{s_0^{-1}(w)} \frac{1 - \lambda(a)}{1 - p} \cdot f_A^0(a) \cdot da, \]
\[ F^1_{s_1}(w) = \int_{-\infty}^{s_1^{-1}(w)} f_A^1(a) \cdot da = \int_{-\infty}^{s_1^{-1}(w)} \frac{\lambda(a)}{p} \cdot f_A^1(a) \cdot da. \]

where we use the fact that the hedonic functions are assumed to be strictly increasing. As an immediate consequence we obtain

\[ F^0_w(w) = \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{s_0^{-1}(w-r)} \frac{1 - \lambda(a)}{1 - p} \cdot f_A^0(a) \cdot da \right] dF^0_u(r), \]
\[ F^1_w(w) = \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{s_1^{-1}(w-r)} \frac{\lambda(a)}{p} \cdot f_A(a) \cdot da \right] dF^1_u(r). \]

These expressions for $(F^0_w, F^1_w)$ demonstrate that the observed distributions differ potentially due to three factors: (1) nonrandom selection, $\lambda \neq p$; (2)
differences in the hedonic functions, $s_0 = s_1$; and (3) differences in the
distributions of jobs by quality, $F_{u_0}^0 \neq F_{u_1}^1$.

Next, we introduce three particular models. Each model aims to
highlight the role of one of these factors in generating observably distinct
sectoral log-wage distributions. The empirical content of these models as
alternative descriptions of the formal-informal segmentation of the labor
market will be investigated later in Section 6.

a) Homogeneous Jobs and Hedonic Functions ($u_0 = u_1 = 0$, $s_0 = s_1$)

\[
F_{w_0}^0(w) = \int_{-\infty}^{s_0^{-1}(w)} \frac{1 - \lambda(a)}{1 - p} f_A(a) \, da.
\]

\[
F_{w_1}^1(w) = \int_{-\infty}^{s_1^{-1}(w)} \frac{\lambda(a)}{p} f_A(a) \, da.
\]

In this model all jobs are equally good, $u_0 = u_1 = 0$, and ability is equally
valuable in both sectors, $s_0 = s_1$. Consequently, workers are completely
indifferent to which sector they should belong. The observed differences in
the log-wage distributions only reflects differences in the distribution of
abilities across sectors which are an exclusive consequence of a nonrandom
selection process. Consider, as an example, how these distributions would
look like in the case of the selection process introduced in Example 1, where
the hedonic functions are linear, i.e., for all $a \in \mathbb{R}$, $s(a) = b \cdot a$ for some $b > 0$.

Example 1-(continued). Assume that $s_0(a) = s_1(a) = b \cdot a$ for all $a \in \mathbb{R}$, $u_0 = u_1 = 0$, and
consider the specific selection process and distribution of abilities
introduced in Example 1, the observed distribution of log-wages can then be expressed as follows:

\[
F^0_0(\omega) = 1 - e^{-2\mu/b}, \quad F^1_0(\omega) = (1 - e^{-\mu/b})^2
\]

for all \(\omega > 0\). Consequently, the observed means are going to equal \(b/2\) and \(3b/2\) for the formal and informal sectors, respectively. Hence, \(\Delta \mu = b > 0\). With respect to the variances, we obtain \(b^2/4\) and \(5b^2/4\) for the informal and formal sectors. Hence, \(\Delta \sigma^2 - b^2 > 0\). In summary, although there exist no differences among jobs and ability is equally valuable in both sectors, the observed log-wage distributions are very different. In fact, the process of selection on ability was able to generate a pair of observable distributions were both the level and the inequality are higher in the formal sector.

b) Random Selection with Homogeneous Jobs \((\lambda = \rho, u_0 = u_1 = 0)\)

\[
F^0_0(\omega) = F_A(s^{-1}_0(\omega)) = \int_{-\infty}^{s^{-1}_0(\omega)} f_A(a) \cdot da
\]

\[
F^1_0(\omega) = F_A(s^{-1}_1(\omega)) = \int_{-\infty}^{s^{-1}_1(\omega)} f_A(a) \cdot da
\]

In this model, due to random selection, abilities are identically distributed across sectors. Moreover, by assumption, all jobs are of the same quality. Consequently, the observed log-wage distributions differ only to the extent the two sectors value ability at different rates. In other words, the hedonic functions are the only potential source of differences among the observed
log-wage distributions. As an additional consequence of random selection, the observed distributions equal the underlying unconditional log-wage distributions, that is, \( F^0_{w_0} = F^*_{w_0} \) and \( F^1_{w_1} = F^*_{w_1} \). This fact permits the observed differences to have simple causal interpretations.

c) Homogeneous Workers (\( A= a_0 \) where \( a_0 \) is a constant)

\[
\begin{align*}
F^0_{u_0}(\omega) &= F^0_{u_0}(\omega - s_0(\omega_0)), \\
F^1_{u_1}(\omega) &= F^1_{u_1}(\omega - s_1(\omega_0)),
\end{align*}
\]

In this model the shape of the observed log-wage distributions are determined by the distribution of jobs by quality in each sector. The value of ability in each sector determine only the location of the distributions.

Since all workers are equally able, selection is necessarily random. Hence, as in the previous model the observed log-wage distributions equal the underlying distributions and observed differences have simple causal interpretations. However, the interpretations differ sharply according to which model we consider. In the current model all workers are identical, so that differences in the distributions of jobs by quality across sectors are the main reason why the observed log-wage distributions differ. In the previous model all jobs are identical whereas workers differ with respect to their ability. Abilities are identically distributed across sectors as a result of random selection. Sectoral differences with respect to the valuation of ability are the unique cause of observed differences among the log-wage distributions. The interpretation for inequality is particularly distinct in these two models. One explains inequality by job heterogeneity, while in the other inequality is rationalized by worker heterogeneity. In
most cases both workers and jobs are going to be heterogeneous, making it very difficult to interpret differences in inequality across sectors. Notice that such difficulties arise in this model despite the selection process being random.

5-Decomposing and Interpreting Observed Differences in Means, Variances and Quantiles

In this section we investigate how the observed differences in means, variances, and quantiles can be decomposed such that the components are in a one-to-one correspondence with the three sources generating differences among the observed distributions: (1) nonrandom selection, (2) differences in the hedonic functions, and (3) differences in the distributions of jobs by quality. We also study how each of these components should be interpreted.

5.1-The Average Impact of Formal Labor Contracts

Consider the population of workers who are currently working in the formal sector, \( P_1 \). The average log-wage in this population is \( E[w_1 | d=1] \). If they are, one by one, reassigned to the informal sector their average log-wage would be \( E[w_0 | d=1]\).\(^{12}\) Consequently, the average relative wage gain due to the formalization of labor relations among those workers who currently occupy jobs in the formal sector is given by

\[
D\mu = \int w, dF^1_1(w) - \int w, dF^1_0(w) = E[w_1 | d=1] - E[w_0 | d=1].
\]

Since by hypothesis the selection is based only on ability, \( E[u_1 | d=1] = E[u_0 | d=1] = 0 \). Hence,
\[ \mathcal{D}_\mu = E[s_1(A) - s_0(A) | d=1]. \]

That is, the average gain from being in the formal sector is given by the average distance between the hedonic functions in the two sectors. To evaluate the average we should use the distribution of abilities that prevails in the formal sector.

5.2 - Decomposing and Interpreting Observed Differences in Means

Consider the observed average difference in log-wages, \( \Delta \mu \)

\[
\Delta \mu = E[w_1 | d=1] - E[w_0 | d=0] \\
= E[s_1(A) - s_0(A) | d=1] + [E[s_0(A) | d=1] - E[s_0(A) | d=0]] + E[u_1 | d=1] - E[u_0 | d=0] \\
= \mathcal{D}_\mu + \mathcal{B}_\mu,
\]

where we define

\[ \mathcal{B}_\mu = E[s_0(A) | d=1] - E[s_0(A) | d=0], \]

and use the fact that

\[ E[u_1 | d=1] - E[u_0 | d=0] = 0, \]

which follows from the hypothesis that the selection process is based only on ability.

Thus, as far as means are concerned, differences between the sectors with respect to the distribution of jobs by quality does not generate observable differences. Differences in the hedonic functions are captured in
$D\mu$, whereas $B\mu$ captures the bias due to nonrandom selection. $B\mu$ is a bias in the sense that it measures the distance between the observed difference in means, $\Delta\mu$, and the average impact of formal labor contracts, $D\mu$. This bias is usually called "selection bias". It can be rewritten as follows:

$$
B\mu = E[s_0(A)|d=1] - E[s_0(A)|d=0] \\
- \int s_0(\alpha) \cdot dF^1_\alpha(\alpha) - \int s_0(\alpha) \cdot dF^0_\alpha(\alpha) \\
- \int s_0(\alpha) \cdot \left[ dF^1_\alpha(\alpha) - dF^0_\alpha(\alpha) \right] \\
= \left( \frac{1}{p(p(1-p))} \right) \cdot \int s_0(\alpha) \cdot (\lambda(\alpha) - p) \cdot dF_\alpha(\alpha) \\
= \left( \frac{1}{p(p(1-p))} \right) \cdot \text{Cov}[s_0(A), \lambda(A)].
$$

By hypothesis, average log-wages increase with ability, $\partial s_0(\alpha)/\partial \alpha > 0$ for all $\alpha$. So, if more able workers have higher probability of being in the formal sector, $\partial \lambda(\alpha)/\partial \alpha > 0$ for all $\alpha$, then $B\mu > 0$. In this case, the observable difference, $\Delta\mu$, would overstate the true average effect of being in the formal sector, $D\mu$.

$B\mu$ vanishes whenever the selection is random, so that $\lambda=p$. However, it is also possible for $B\mu$ to vanish even though the selection is nonrandom. As an example, let $A=U(0,1)$, $s_0(\alpha)=\alpha$ for all $\alpha$, and $\lambda$ be defined as follows:

$$
\lambda(\alpha) = \begin{cases} 
1 & \text{if } \alpha \in (1/4, 3/4) \\
0 & \text{if } \alpha \in [0, 1/4] \cup [3/4, 1]
\end{cases}
$$

In this case $s_0(A)$ and $\lambda(A)$ are orthogonal. So that, $B\mu=0$.

Finally, consider the three special models we introduce in Section 4.4. In model (a) sectors look identical to workers, $D\mu=0$, so that all
observed differences are the result of a nonrandom selection process, \( \Delta \mu = \delta \mu \). In both models (b) and (c) the selection process is random so that \( \delta \mu = 0 \) and the observed difference in means is actually the average causal effect of formal labor contracts on wages, \( \Delta \mu = \theta \mu \).

5.3-DECOMPOSING AND INTERPRETING OBSERVED DIFFERENCES IN VARIANCES

Consider the observed sectoral difference in the log-wage variances, \( \Delta \sigma^2 \)

\[
\Delta \sigma^2 = \text{Var}(w_1|d=1) - \text{Var}(w_0|d=0) \\
= \text{Var}(s_1(A)|d=1) - \text{Var}(s_0(A)|d=1) + \left[ \text{Var}(s_1(A)|d=1) - \text{Var}(s_0(A)|d=0) \right] + \\
\quad + \text{Var}(u_1) - \text{Var}(u_0) \\
= D\sigma^2 + B\sigma^2 + H\sigma^2,
\]

where we have defined

\[
D\sigma^2 = \text{Var}(s_1(A)|d=1) - \text{Var}(s_0(A)|d=1),
\]
\[
B\sigma^2 = \text{Var}(s_0(A)|d=1) - \text{Var}(s_0(A)|d=0),
\]
\[
H\sigma^2 = \text{Var}(u_1) - \text{Var}(u_0).
\]

Thus, we can immediately identify the three distinct factors that account for the observed sectoral difference in log-wage variances: (1) differences between sectors with respect to the value they attach to ability, \( D\sigma^2 \); (2) differences between sectors with respect to the degree of job heterogeneity, \( H\sigma^2 \); and (3) the effect of nonrandom selection, \( B\sigma^2 \).

There exists an alternative expression for \( D\sigma^2 \) which is extremely helpful in predicting its sign. This expression can be derived as follows:
\[ \Delta \sigma^2 = \text{Var}[s_1(A) \mid d=1] - \text{Var}[s_0(A) \mid d=1] \\
= E[s_1^2(A) \cdot s_0^2(A) \mid d=1] - \{E[s_1(A) \mid d=1]^2 - E[s_0(A) \mid d=1]^2\} \\
= E[(s_1(A) - s_0(A))(s_1(A) + s_0(A)) \mid d=1] - \\
\{E[s_1(A) \cdot s_0(A) \mid d=1]\} \{E[s_1(A) + s_0(A) \mid d=1]\} \\
= \text{Cov}(s_1(A) \cdot s_0(A), (s_1(A) + s_0(A)) \mid d=1). \]

By hypothesis \( s_1 \) and \( s_0 \) are strictly increasing. Hence, \( \Delta \sigma^2 > 0 \) whenever \( s_1 - s_0 \)
is also an increasing function. But, \( s_1 - s_0 \) is increasing if and only if \( \partial s_1(a)/\partial a \geq \partial s_0(a)/\partial a \) for all \( a \). In summary, if the marginal value of ability is higher in the formal sector then \( \Delta \sigma^2 > 0 \).

With respect to the term \( \mathcal{B} \sigma^2 \), it seems that there exist no simple relationship between its sign and the behavior of \( \lambda \). (See, however, Heckman and Honoré(1987) for a characterization of \( \mathcal{B} \sigma^2 \) in some particular selection models.)

As an example where all three terms are nonzero, consider the selection model introduced in Example 1 with linear hedonic functions:

**Example 1 (Continued).** Let, for all \( a, s_0(a) = b_0 \cdot a \) and \( s_1(a) = b_1 \cdot a \) with \( b_1 > b_0 \). So that, the marginal value of ability is higher in the formal sector. The selection process and the distribution of abilities considered are the ones introduced in Example 1. Thus, as already established, \( \text{Var}[A \mid d=1] = 5/4 \) and \( \text{Var}[A \mid d=0] = 1/4 \). Hence,

\[ \mathcal{B} \sigma^2 = (b_1^2 - b_0^2) \cdot \text{Var}[A \mid d=1] = \frac{5}{4} (b_1^2 - b_0^2) > 0, \]

\[ \mathcal{B} \sigma^2 = b_0^2 (\frac{5}{4} - \frac{1}{4}) = b_0^2 > 0. \]

So,

\[ \Delta \sigma^2 = \frac{5}{4} (b_1^2 - b_0^2) + b_0^2 + [\text{Var}[u_1] - \text{Var}[u_0]] \]

In words, the observed log-wage distribution is more unequally distributed in the formal sector because (1) the ability marginal value is higher in this
sector $b_1 > b_0$, that implies $D\sigma^2 > 0$; and (2) the selection process generates a pair of distributions of abilities such that the variance is higher in formal sector, that implies $E\sigma^2 > 0$. If also jobs are more heterogeneous in this sector, i.e., $\text{Var}[u_1] > \text{Var}[u_0]$, then all three forces work in the direction of greater inequality in the formal sector.

Consider the three special models introduced in Section 4.4. They are particularly valuable in understanding observed differences in inequality among sectors. In model (a) jobs are identical, so $H\sigma^2 = 0$. Moreover, the two sectors value ability at the same rate, so $D\sigma^2 = 0$. Consequently, in this particular model $\Delta \sigma^2 = E\sigma^2$. In the other two models the selection process is random, so $E\sigma^2 = 0$. In model (b) differences in inequality are uniquely explained by differences in the hedonic functions, that is, $\Delta \sigma^2 = D\sigma^2$ and $H\sigma^2 = 0$. Finally, in model (c) differences in inequality are exclusively explained by differences in the distribution of jobs by quality, that is, $\Delta \sigma^2 = H\sigma^2$ and $D\sigma^2 = 0$.

5.4-Interpreting Observed Differences in Quantiles

Up to now we have been successful in decomposing $\Delta \mu$ and $\Delta \sigma^2$ in additive components which are in a one-to-one correspondence with the three factors shown to generate differences between the observed log-wage distributions in the formal and informal sectors. So, as far as differences in means and variances are concerned these three factors do not interact with each other. In the case of differences in quantiles we were unable to find a similar simple decomposition. Accordingly, we proceed by analyzing each factor separately. This is done by considering sequentially the three special models introduced in section 4.4.
a) Homogeneous Jobs and Hedonic Functions.

As already emphasized, in this special model differences between the observed log-wage distributions are determined only by the selection process. Conversely, as we are going to show later in Section 6.1, any observed pattern for the differences in quantiles can be rationalized by some selection process. In Section 3 we mention that in our sample the quantile differences follow a very specific pattern. These differences are always positive and tend to be increasing. To illustrate that this pattern can be generated by a appropriately chosen selection process, we consider the exponential selection process introduced in Example 1. This particular selection process indeed generates a pattern for the quantile differences qualitatively similar to the one reported in Section 3.

Example 1 (continued). As we have already shown, for all \( w > 0 \),

\[
F^0_w(w) = 1 - e^{-2w/b},
F^1_w(w) = (1 - e^{-w/b})^2.
\]

So,

\[
q^0_w(\alpha) = - \frac{b}{2} \cdot \log(1-\alpha),
q^1_w(\alpha) = - b \cdot \log(1-\sqrt{\alpha}).
\]

Hence, for all \( \alpha \in [0,1] \), \( \Delta q(\alpha) > 0 \) and \( \delta \Delta q(\alpha)/\delta \alpha > 0 \) (see figure 5). In this example positive and increasing quantile differences were obtained by specifying a selection process in which the probability of being allocated to the formal sector increases with ability. This is, however, not a general property. It is possible to construct selection processes in which the probability of being allocated to the formal sector increases with ability but the quantile differences are decreasing. How properties of the selection process are reflected in the quantile differences profile is an important but
log-wage quantile differences between the formal and informal sectors
yet unsolved question.

b) Random Selection with Homogeneous Jobs

Since the selection is random, the distribution of abilities in the two sectors is identical, $F_A^0 = F_A^1$. Because all jobs are assumed to be equally good, the log-wage distribution in each sector is given by this common underlying distribution of abilities transformed by the sector-specific hedonic function. Since the hedonic functions are assumed to be strictly increasing, we obtain, for $0 < \alpha < 1$,

\begin{align*}
q^0_w(\alpha) &= s_0(q_A(\alpha)), \\
q^1_w(\alpha) &= s_1(q_A(\alpha)).
\end{align*}

where $q_A(\alpha)$ is the $\alpha$-quantile of the distribution of abilities in the overall population. Therefore, for all $0 < \alpha < 1$,

\begin{align*}
\Delta q(\alpha) &= q^1_w(\alpha) - q^0_w(\alpha) = s_1(q_A(\alpha)) - s_0(q_A(\alpha)), \\
\delta \Delta q(\alpha)/\delta \alpha &= \left[ \delta s_1(q_A(\alpha))/\delta \alpha - \delta s_0(q_A(\alpha))/\delta \alpha \right] \cdot \delta q_A(\alpha)/\delta \alpha.
\end{align*}

Hence, in this particular model, $\Delta q(\alpha)$ is the actual gain of having a formal contract for workers with ability level $q_A(\alpha)$. As $\alpha$ ranges from zero to one, $\Delta q(\alpha)$ traces these gains for workers of all ability levels. In other words, in this model $\Delta q(\alpha)$ completely describes the difference between the hedonic functions. The fact that $\Delta q(\alpha)$ is the log-wage differential among workers which are in different sectors but have equal levels of ability is a distinctive feature of this model.\textsuperscript{14}
Notice that \( \Delta q(\alpha) > 0 \) for all \( 0 < \alpha < 1 \) if and only if \( s_1(\alpha) > s_0(\alpha) \) for all \( \alpha \). Moreover, \( \delta \Delta q(\alpha)/\delta \alpha \geq 0 \) for all \( 0 < \alpha < 1 \) if and only if \( \delta s_1(\alpha)/\delta \alpha \geq \delta s_0(\alpha)/\delta \alpha \) for all \( \alpha \), since \( \delta q_1(\alpha)/\delta \alpha \geq 0 \) for all \( 0 < \alpha < 1 \). In summary, the observed quantile differences are positive and increasing \textit{if and only if} the differences between the hedonic functions in the formal and informal sectors are also everywhere positive and increasing.

c) Homogeneous Workers

In this special model observed log-wage distributions are translations of the correspondent distributions of jobs by quality in each sector. Therefore, for all \( 0 < \alpha < 1 \),

\[
q^0_\alpha = s_0(\alpha_0) + q_\alpha^0,
\]

\[
q^1_\alpha = s_1(\alpha_0) + q_\alpha^1.
\]

Hence,

\[
\Delta q(\alpha) = q^1_\alpha - q^0_\alpha = (s_1(\alpha_0) - s_0(\alpha_0)) + q_\alpha^1 - q_\alpha^0,
\]

\[
\delta \Delta q(\alpha)/\delta \alpha = \delta q^1_\alpha/\delta \alpha - \delta q^0_\alpha/\delta \alpha.
\]

In the case that both \( u_0 \) and \( u_1 \) are normally distributed and \( \text{Var}[u_1] > \text{Var}[u_0] \), \( \delta \Delta q(\alpha)/\delta \alpha > 0 \) for all values of \( \alpha \). We could then conjecture that quantile differences are everywhere increasing whenever the formal sector is the one with more heterogeneous jobs. This is, however, false as the following counterexample shows.

\text{Example 2. Let } \epsilon_1, \epsilon_2, \text{ and } \epsilon_3 \text{ be three independent } U(-1, 1) \text{ random variables.}

\text{INPES, 152/88}
FIGURE 6

log-wage quantile differences between
the formal and informal sectors
Let $u_0 = \frac{1}{2} (\epsilon_1 + \epsilon_2)$ and $u_1 = \epsilon_3$. Notice that, in this case, $\text{Var}[u_0] = 1/6$ and $\text{Var}[u_1] = 1/3$. So $\text{Var}[u_1] > \text{Var}[u_0]$, and

$$q_u(\alpha) = \begin{cases} - \left[ 1 - \sqrt{2\alpha} \right] & 0 \leq \alpha \leq 1/2, \\ \left[ 1 - \sqrt{2(1-\alpha)} \right] & 1/2 \leq \alpha \leq 1. \end{cases}$$

$$q_{u_1}(\alpha) = 2\alpha - 1.$$

Hence, despite the fact that $\text{Var}[u_1] > \text{Var}[u_0]$, $\partial \Delta q(\alpha)/\partial \alpha < 0$ for all $\alpha \in (0, 1/6) \cup (7/8, 1)$. (see Figure 6)

This conjecture is true, however, under certain suitable assumptions. Define a symmetric spread increasing function $\Phi : \mathbb{R} \to \mathbb{R}$ as a function that satisfies, for all $x \in \mathbb{R}$, the following three conditions:

(a) $\Phi(x) = -\Phi(-x),$

(b) $\Phi(x) \geq x$ for all $x \geq 0$,

(c) $\partial \Phi(x)/\partial x \geq 1.$

A simple example of a symmetric spread increasing function is $\Phi(x) = b \cdot x$, for $b \geq 1$.

**Proposition.** If

(H1) $u_0$ is distributed symmetrically around zero,

(H2) $\Phi$ is a spread increasing function,

(H3) $\Phi(u_0)$ and $u_1$ are identically distributed.

Then

(R1) $u_1$ is also symmetrically distributed around zero,

(R2) $\text{Var}[u_1] \geq \text{Var}[u_0],$

(R3) $\partial \Delta q(\alpha)/\partial \alpha \geq 0.$

**Proof:** By the symmetry of both the distribution of $u_0$ and the function $\Phi$, it

INPES, 152/88
follows that $\phi(u_0)$ is symmetrically distributed around zero. Hence (R1) follows from (H3). By (H3), $\text{Var}(u_1) = \text{Var}(\phi(u_0))$. From (H3) and (R1), $E[\phi(u_0)]=E[u_1]=0$. Consequently, $\text{Var}(u_1) = E[\phi(u_0)^2]$. By the properties (a) and (b) of $\phi$, $E[\phi(u_0)^2] \geq E[u_0^2] = \text{Var}(u_0)$. So, (R2) holds. By the fact that $\phi$ is strictly increasing and (H3), $q_{u_0}(\alpha) = \Phi(q_{u_0}(\alpha))$. So,

$$
\frac{\partial \Delta q(\alpha)}{\partial \alpha} = \left( \frac{\partial \Phi(q_{u_0}(\alpha))}{\partial \alpha} - 1 \right) \cdot \frac{\partial q_{u_0}(\alpha)}{\partial \alpha}.
$$

Hence (R3) follows from property (c) of $\phi$.

This proposition illustrates the fact that if jobs are more heterogeneous in the formal sector then it is likely that $\Delta q(\alpha)$ will be increasing in $\alpha$.

6-EVALUATING ALTERNATIVE MODELS FOR THE FORMAL-INFORMAL SEGMENTATION OF THE LABOR MARKET

Our framework for wage determination and selection process was constructed with the objective of permitting a coherent study of the three models we introduced in Section 4. These models correspond to completely distinct views of the formal-informal segmentation of the labor market. In this section we demonstrate that all three models are perfectly compatible with our empirical findings. As we will show, however, the interpretation of the findings differ remarkably depending on the model we are considering. An interesting surprise is the fact that a traditional model for the segmentation in the labor market which postulates the existence of a protected sector is clearly incompatible with our empirical findings.

6.1-HOMOGENEOUS JOBS AND HEDONIC FUNCTIONS

In this model the wage of a worker depends only on his ability. It
is independent of the job he occupies or the sector to which he belongs. Formally, this is achieved by letting $u_0 = u_1 = 0$ and $s_0 = s_1 = s$. As a consequence, for all $p \in \mathbb{P}$

$$ w_0(p) = w_1(p) = s(A(p)). $$

Workers are, therefore, completely indifferent to which sector they belong. Differences between the observed log-wage distributions in the two sectors are pure artifacts of a nonrandom selection process.

The model has no refutable empirical implications. Any pair of observed log-wage densities, $(f_0^w, f_1^w)$, can be rationalized in this context by letting

$$ s(a) = \alpha, $$

$$ f_A(a) = (1-p) \cdot f_0^w(a) + p \cdot f_1^w(a), $$

$$ \lambda(a) = p \cdot f_1^w(a)/f_A(a). $$

Although the model cannot generate refutable implications, the observed distributions impose strong restrictions on the set of admissible selection processes. In fact, for any given observed pair $(f_0^w, f_1^w)$ and choice of $s$, there exist only one pair of functions $(f_A, \lambda)$ which is compatible with them. This unique pair $(f_A, \lambda)$ is given by

$$ f_A(a) = \left\{ (1-p) \cdot f_0^w(s(a)) + p \cdot f_1^w(s(a)) \right\} \cdot \partial s(a)/\partial a $$

and

$$ \lambda(a) = \left\{ p \cdot f_1^w(s(a)) \cdot \partial s(a)/\partial a \right\} / f_A(a). $$

INPES, 152/88
\[ r = \frac{f^1_{s(a)}}{f^0_{s(a)}} \]

where

This second expression for \( \lambda \) has a surprising and important implication. It shows that, independent of the choice of \( s \), \( \lambda \) is an increasing function if and only if the likelihood ratio between the observed distributions is itself increasing. Moreover, these expressions indicate how \((f^*_A, \lambda)\) can be estimated from \((f^0_0, f^1_1)\) once we have specified a hedonic function \( s \). Figures 7 and 8 show the estimated functions we obtain based on the assumption of a logarithmic hedonic function. Notice that since the likelihood ratio is increasing the estimate for \( \lambda \) shares this same property. Remember that this property of \( \lambda \) is independent of our assumption about the logarithmic shape of \( s \).

In this model, as already mentioned, there is no wage gain related to formal labor contracts. All observed differences are pure artifacts of differences between sectors with respect to the distribution of workers by ability which are due exclusively to a nonrandom selection process. In particular, the observed difference in means equals the selection bias, \( \Delta \mu - \bar{\mu} \). The observed fact that \( \lambda \) is increasing implies that \( \bar{\mu} > 0 \) as we have shown in section 5.2. Thus, the observation that wages are, on average, higher in the formal sector is interpreted, in the context of this model, as a consequence of higher ability workers having greater chances of being allocated to the formal sector. This allocation mechanism leads to a distribution of abilities in the formal sector which dominates the correspondent distribution in the informal sector. Since wages increase with ability, the average wage ends up being higher in the formal sector.
distribution of abilities
probability of being selected into the formal sector
The other observed facts, (1) $\Delta \sigma^2 = 2\sigma^2 > 0$, (2) $\Delta q(\alpha) > 0$, and (3) $\delta \Delta q(\alpha)/\delta \alpha > 0$, do not follow from $\lambda$ being increasing. As we have mentioned before, in sections 5.3 and 5.4a, we have not yet been able to detect which properties of the selection process induce these three interesting findings.

6.2-RANDoM SELECTION WITH HOMOGENEOUS JOBS

Since jobs are assumed to be homogeneous, within sectors equally able workers must face the same wage. This common wage, however, may vary across sectors. The sectoral wage variation only reflects differences among the sectors with respect to their hedonic functions. Formally, the log-wages of workers with ability $\alpha \in \mathbb{R}$ are

$$w_0 = s_0(\alpha),$$
$$w_1 = s_1(\alpha).$$

The hypothesis of random allocation ensures that ability is identically distributed across sectors. Consequently, any observed difference between sectors with respect to the log-wage distributions must be related to differences in their hedonic functions.

As we have already shown in Sections 5.2, 5.3 and 5.4b, in this particular model the observed differences can be expressed as

$$\Delta \mu = E[s_1(A) - s_0(A)],$$
$$\Delta \sigma^2 = \text{Var}[s_1(A)] - \text{Var}[s_0(A)],$$
$$\Delta q(\alpha) = q_1^0(\alpha) - q_0^0(\alpha) = s_1(q_\lambda^0(\alpha)) - s_0(q_\lambda^0(\alpha)),$$
$$\delta \Delta q(\alpha)/\delta \alpha = \left[\delta s_1(q_\lambda^0(\alpha))/\delta \alpha - \delta s_0(q_\lambda^0(\alpha))/\delta \alpha\right] \cdot \delta q_\lambda^0(\alpha)/\delta \alpha.$$

The fact that in the population we investigate all these quantities are
positive can now be interpreted in the realm of this specific model. Notice that \( \Delta q(a) > 0 \) for all \( 0 < a < 1 \) if and only if \( s_1(a) > s_0(a) \) for all \( a \in \mathbb{R} \). Hence, everywhere positive quantile differences means that, at all levels, ability is more valuable in the formal than in the informal sector. This immediately implies that the average relative wage gain of formal labor contracts, \( \Delta \mu \), must be positive.

In Section 5.4b we also establish that quantile differences are increasing if and only if the marginal value of ability is higher in the formal sector, i.e., \( \delta s_1(a)/\delta a \geq \delta s_0(a)/\delta a \) for all \( a \in \mathbb{R} \). In Section 5.3 we show that this fact is sufficient for the log-wage variance to be higher in the formal sector, \( \Delta \sigma^2 > 0 \).

In summary, our findings are compatible with a model in which the selection process is random and jobs are homogeneous if and only if the absolute and marginal values of ability are higher in the formal sector. In this case \( s_1 \) is everywhere above \( s_0 \) and the two hedonic functions diverge. Notice that if the hedonic functions follow a pattern like the one just described then \( \Delta \mu > 0 \) and \( \Delta \sigma^2 > 0 \). The converse is not necessarily true. What ensures that the hedonic functions must follow this particular pattern are the findings about the behavior of the quantile differences. It is though extremely important to notice the crucial role of quantile differences in our analysis. Unfortunately, up to now, econometricians have only made a very limited use of quantiles.

This model is based on three primitives \( (f_A, s_0, s_1) \) while only two objects can be observed \( (f^0_{w_0}, f^1_{w_1}) \). As it was the case in the previous model, this model also does not provide us with any refutable implications. Given any observed pair \( (f^0_{w_0}, f^1_{w_1}) \) of log-wage densities it is always possible to find a triple \( (f_A, s_0, s_1) \) which perfectly rationalizes the observed densities.
Actually, by normalizing one of the primitives at any given value, we can recover the other two uniquely from the observed distributions. For instance, given $s_0$, the other primitives $(f_A, s_1)$ can be uniquely recovered as follows:

$$f_A(a) = \left( \frac{\partial s_0(a)}{\partial a} \right) \cdot F_0^0(s_0(a)),$$

$$F_1^1(s_1(a)) = F_A(a).$$

As an example, we present in figures 9 and 10 estimates for $(f_A, s_1)$ under the assumption of a logarithmic specification for $s_0$. As expected the distance between the hedonic functions mirror the observed pattern of the quantile differences.

For any given $a \in \mathbb{R}$, the difference between the hedonic functions, $s_1(a) - s_0(a)$, gives the relative wage gain of having a labor contract for workers with ability $a$. These differences being increasing means that relative gains from labor contracts are increasing with ability. A model for the segmentation of the labor market based on the idea of a protected sector would imply the opposite, that is, that gains should be higher for low ability workers. So the protected sector model would imply that the hedonic functions should be convergent. Consequently, this model can clearly be rejected as a description of the segmentation in the Brazilian Construction Sector.

6.3 Homogeneous Labor

In this model all workers are equally able. So, differences in wages within sectors are uniquely determined by the heterogeneity of jobs. By hypothesis the selection process does not affect the distribution of jobs by quality within each sector. As a consequence, the observed log-wage
FIGURE 9

distribution of abilities
Figure 10

Hedonic functions by sector

Formal sector

Informal sector

Ability
distributions truly reveal the underlying heterogeneity of jobs in each sector.

As a matter of fact, this model is just-identified. It has four primitives \((F_{0}, F_{1}, s_{0}(a_0), s_{1}(a_0))\). Any pair of log-wage distributions \((F_{0}^{0}, F_{1}^{1})\) can be rationalized in this model by letting

\[
\begin{align*}
    s_{0}(a_0) &= \int_{w_0} w \cdot dF_{0}^{0}(w) = E[w_0 | d = 0], \\
    s_{1}(a_0) &= \int_{w_1} w \cdot dF_{1}^{1}(w) = E[w_1 | d = 1], \\
    F_{0}(u) &= F_{0}^{0}(u + s_{0}(a_0)) \\
    F_{1}(u) &= F_{1}^{1}(u + s_{1}(a_0)).
\end{align*}
\]

In this model the selection process is random, which implies \(\Delta \mu = 2\mu\) and \(\Delta \sigma^2 = D\sigma^2 + H\sigma^2\). Because workers are all homogeneous \(D\sigma^2 = 0\). Therefore,

\[
\begin{align*}
    \Delta \mu &= D\mu - s_{1}(a_0) - s_{0}(a_0), \\
    \Delta \sigma^2 &= H\sigma^2 = \text{Var}[u_1] - \text{Var}[u_0].
\end{align*}
\]

Hence, \(\Delta \mu > 0\) means that on average there actually exists a positive wage gain from having a formal labor contract. The fact that \(\Delta \sigma^2 = H\sigma^2\) shows that if \(\Delta \sigma^2 > 0\) then it must be the case that jobs are more heterogeneous in the formal sector. The interpretation and implications for the pattern of the quantile differences are not clear. However, as we have shown in Section 5.4c under some weak assumptions the quantile differences should be increasing whenever jobs are more heterogeneous in the formal sector. The sufficient assumptions are (1) \(u_0\) has a symmetric distribution and (2) \(F_{1}^{1} \circ F_{0}^{0}\) is a symmetric spread increasing function. Figures 11 and 12 present our estimates for \(F_{0}^{0}\) and \(F_{1}^{1}\).
distribution of jobs by quality in the informal sector
$F^{-1} e F$. From these figures we can see that these hypothesis are satisfied only approximately.
FOOTNOTES:

1) According to our classification, a worker belongs to the formal sector when he is an employee with a formal labor contract. When he is an employee which doesn’t have a formal contract he is a member of the informal sector. All self-employed workers as well as those who work on their own account have been excluded from the sample.

2) Observables attributes are understood here as the set of all observed variables which the analyst would like to hold constant in his analysis.

3) Unfortunately, however, this procedure may not necessarily completely describe how potential wages are generated. For instance, in a segmented labor market with queues, the order of arrival of a worker may affect his wage. Moreover, in the presence of quotas it may not make sense to reallocate one worker holding the allocation of the other workers fixed.

4) The distinction we are making here is identical to Lewis’ (1986) distinction between wage gain and wage gap.

5) For a generic c.d.f F, the α-quantile is defined by $q(\alpha) = \inf \{q: F(q) \geq \alpha \}$.

6) See Table 1 for a list of these nine metropolitan regions.

7) See Table 1 for a list of these twelve occupations.

8) These workers were excluded because their labor contracts have several special features. Only 4% of the sample were eliminated at this stage.

9) F stochastically dominates G when $F(\alpha) \leq G(\alpha)$ for all $\alpha \in \mathbb{R}$. A sufficient condition for G to be dominated by F is for f to cross g only once from below.

10) Since both $s_0$ and $s_1$ are obtained from conditional expectations, they are well-defined only almost surely. So, a more precise statement would be to require that there exist versions of $s_0$ and $s_1$ which are strictly increasing and differentiable.

11) By construction, $E[u_0 | A] = E[u_1 | A] = 0$. So, without extra assumptions, $(u_0, u_1)$ would only be mean-independent of A.

12) It is important to understand that the reassignment process is not cumulative. Before a new worker can be reallocated to the informal sector, the previous worker must return to his job in the formal sector.

13) The fact that $E_\mu$ and $\text{Cov}(s_0(A), \lambda(A))$ have the same sign has also been established in Patil and Rao (1978).

14) This association between quantiles and workers’ abilities is also used in Pettengill’s (1980) work on the impact of unions on the inequality of earnings.
REFERENCES:


"Estudos para a Reforma Tributária - Tomo 5: Federализmo Fiscal", José Roberto Afonso e Thereza Lobo, Março, 153 p.


"Surplus Labor and Industrialization", Kevin M. Murphy, Andrei Shleifer e Robert W. Vishny, Maio 1987, 19 p.
113 - "Um Modelo de Consistência Multissetorial para a Economia Brasileira", Márcio Gomes Pinto Garcia, Maio 1987, 42 p.


120 - "Modelo Multissetorial CEPAL/IPEA para o Brasil", Fábio Giambiagi, Guilherme Gomes Dias, Juan José Pereira e Márcio Gomes Pinto Garcia, Outubro 1987, 124 p.


NO 127 - "Considerações sobre a Relação entre a Dívida Pública e a Inflação", José W. Rossi, Dezembro 1987, 09 p.


NO 132 - "O (Des) controle do Endividamento de Estados e Municípios - Análise Crítica das Normas Vigentes e Propostas de Reforma", Fernando Rezende e José R. Afonso, Janeiro 1988, 75 p.


Nº 139 - "O Brasil e a Atual Rodada de Negociações do GATT", José Tavares de Araujo Jr, Maio 1988, 21 p.


O INPES edita ainda as seguintes publicações: Pesquisa e Planejamento Econômico; Literatura Econômica; Coleção Relatórios de Pesquisa; Série Monográfica; Série FNPE; Série Estudos de Política Industrial e Comércio Exterior (EPICO); Relatório Interno; Informes Conjunturais; Boletim Conjuntural; Série Estudos sobre Economia do Setor Público (ESEP); Série Fac-Símile; Informe Técnico INPES e Carta de Conjuntura.
Ipea – Institute for Applied Economic Research

PUBLISHING DEPARTMENT

Coordination
Cláudio Passos de Oliveira

Supervision
Everson da Silva Moura
Reginaldo da Silva Domingos

Typesetting
Bernar José Vieira
Cristiano Ferreira de Araújo
Daniella Silva Nogueira
Danilo Leite de Macedo Tavares
Diego André Souza Santos
Jeovah Herculano Szervinsk Junior
Leonardo Hideki Higa

Cover design
Luís Cláudio Cardoso da Silva

Graphic design
Renato Rodrigues Buenos

The manuscripts in languages other than Portuguese published herein have not been proofread.

Ipea Bookstore
SBS – Quadra 1 – Bloco J – Ed. BNDES, Téreo
70076-900 – Brasília – DF
Brazil
Tel.: + 55 (61) 3315 5336
E-mail: livraria@ipea.gov.br
Ipea’s mission
Enhance public policies that are essential to Brazilian development by producing
and disseminating knowledge and by advising the state in its strategic decisions.