Informality and segmentation: evidence from a self-selection model with entry barriers to formal employment in Peru

Alfredo M. Alvarado Enciso

Pontifical Catholic University of Peru (PUCP) and Inter-American Development Bank (IDB)

The Economics of Informality Conference 2018
Bogotá, Colombia

May 28-29, 2018
Outline

1. Introduction
2. Model of segmentation and self-selection
3. Informality in Peru and sources
4. Results
5. Concluding remarks and research agenda
The problem of informality

- In Peru, 3 out of 4 workers are informal (INEI, 2014)
The problem of informality

- In Peru, 3 out of 4 workers are informal (INEI, 2014)
- Informality reduces growth (Loayza, 2008)
The problem of informality

- In Peru, 3 out of 4 workers are informal (INEI, 2014)
- Informality reduces growth (Loayza, 2008)
- Informal employment is related to low income, poverty and exclusion (Perry et. al, 2007)
The problem of informality

- In Peru, 3 out of 4 workers are informal (INEI, 2014)
- Informality reduces growth (Loayza, 2008)
- Informal employment is related to low income, poverty and exclusion (Perry et. al, 2007)
- Labor informality has been a topic of debate in public policies in Peru: min. wages, special tax regimes, special labor regimes, etc
The problem of informality

- In Peru, 3 out of 4 workers are informal (INEI, 2014)
- Informality reduces growth (Loayza, 2008)
- Informal employment is related to low income, poverty and exclusion (Perry et. al, 2007)
- Labor informality has been a topic of debate in public policies in Peru: min. wages, special tax regimes, special labor regimes, etc
- Is informality a problem?
Debate about the nature of labor informality

- The academic debate has revolved around whether informality indicates:
  - Segmentation in the labor market (Fei & Ranis, 1961; Lewis, 1963; Harris & Todaro, 1970; Doeringer & Piore, 1971)
  - A competitive and voluntary option (De Soto, 1986; Maloney, 2004; Perry, y otros, 2007)
  - A mixed of both (Fields, 2005; Chen, y otros, 2005; Bacchetta & Ernst, 2009)
Debate about the nature of labor informality

The academic debate has revolved around whether informality indicates:

- Segmentation in the labor market (Fei & Ranis, 1961; Lewis, 1963; Harris & Todaro, 1970; Doeringer & Piore, 1971)
- A competitive and voluntary option (De Soto, 1986; Maloney, 2004; Perry, y otros, 2007)
- A mixed of both (Fields, 2005; Chen, y otros, 2005; Bacchetta & Ernst, 2009)
Debate about the nature of labor informality

The academic debate has revolved around whether informality indicates:

- Segmentation in the labor market (Fei & Ranis, 1961; Lewis, 1963; Harris & Todaro, 1970; Doeringer & Piore, 1971)
- A competitive and voluntary option (De Soto, 1986; Maloney, 2004; Perry, y otros, 2007)
- A mixed of both (Fields, 2005; Chen, y otros, 2005; Bacchetta & Ernst, 2009)
Debate about the nature of labor informality

The academic debate has revolved around whether informality indicates:

» Segmentation in the labor market (Fei & Ranis, 1961; Lewis, 1963; Harris & Todaro, 1970; Doeringer & Piore, 1971)
» A competitive and voluntary option (De Soto, 1986; Maloney, 2004; Perry, y otros, 2007)
» A mixed of both (Fields, 2005; Chen, y otros, 2005; Bacchetta & Ernst, 2009)
Methodological background

- Dickens y Lang (1985) use a switching regression model with unknown regimes: a wage equation with human capital variables and another without them.
- Magnac (1991) uses a multivariate generalized Tobit with three dependent variables: labor force participation, choice of sectors (self-employed or employee/blue collar with entry cost) and wage.
- Yamada (1994) uses structural probits and mincer earning equations corrected by selection bias of the employment sector (self-employed, informal and formal salaried).
- Guindling (1991) tests whether there are differences of human capital returns among self-employed/salaried formal/informal workers.
Methodological background

- Dickens y Lang (1985) use a switching regression model with unknown regimes: a wage equation with human capital variables and another without them.
- Magnac (1991) uses a multivariate generalized Tobit with three dependent variables: labor force participation, choice of sectors (self-employed or employee/blue collar with entry cost) and wage.
Methodological background

Dickens y Lang (1985) use a switching regression model with unknown regimes: a wage equation with human capital variables and another without them.

Magnac (1991) uses a multivariate generalized Tobit with three dependent variables: labor force participation, choice of sectors (self-employed or employee/blue collar with entry cost) and wage.

Yamada (1994) uses structural probits and mincer earning equations corrected by selection bias of the employment sector (self-employed, informal and formal salaried).
Methodological background

- Dickens y Lang (1985) use a switching regression model with unknown regimes: a wage equation with human capital variables and another without them.
- Magnac (1991) uses a multivariate generalized Tobit with three dependent variables: labor force participation, choice of sectors (self-employed or employee/blue collar with entry cost) and wage.
- Yamada (1994) uses structural probits and mincer earning equations corrected by selection bias of the employment sector (self-employed, informal and formal salaried).
- Guindling (1991) tests whether there are differences of human capital returns among self-employed/salaried formal/informal workers.
Heckman y Hotz (1986)

- The incentive structure of the workers focuses on maximizing the utility function rather than the income function.
  \[ \Rightarrow \] Workers also choose their sector based on a comparison of nonpecuniary concerns.
The incentive structure of the workers focuses on maximizing the utility function rather than the income function. 

\[ \implies \text{Workers also choose their sector based on a comparison of nonpecuniary concerns.} \]

Previous studies assume that competitive and non-competitive segments are exogenous.

\[ \implies \text{Self-employed sector is not always competitive, and salaried sector is not always segmented.} \]
Methodological background II

- Cunningham y Maloney (2001), through factor and cluster analysis, determines the proportion of small informal enterprises was competitive and segmented.

- Pratap y Quitin (2006), through non-parametric techniques (PSM and D&D), assesses the segmentation hypothesis by comparing informal workers who earn the same as formal workers.

- Günther y Launov (2012), through finite mixture model with sample selection, determine the number of unobservable segments of informal sector endogenously and calculates the % of involuntary informal workers.

- Alcaraz, Chiquiar y Salcedo (2015), through a utility maximization model with entry barriers and self-selection, determine the % of involuntary informal workers.
Methodological background II

- Cunningham y Maloney (2001), through factor and cluster analysis, determines the proportion of small informal enterprises was competitive and segmented.
- Pratap y Quitin (2006), through non-parametric techniques (PSM and D&D), assesses the segmentation hypothesis by comparing informal workers who earn the same as formal workers.
- Günter y Launov (2012), through finite mixture model with sample selection, determine the number of unobservable segments of informal sector endogenously and calculates the % of involuntary informal workers.

Alcaraz, Chiquiar y Salcedo (2015), through a utility maximization model with entry barriers and self-selection, determine the % of involuntary informal workers.
Methodological background II

- Cunningham y Maloney (2001), through factor and cluster analysis, determines the proportion of small informal enterprises was competitive and segmented.

- Pratap y Quitin (2006), through non-parametric techniques (PSM and D&D), assesses the segmentation hypothesis by comparing informal workers who earns the same as formal workers.

- Günther y Launov (2012), through finite mixture model with sample selection, determine the number of unobservable segments of informal sector endogenously and calculates the % of involuntary informal workers.
Methodological background II

- Cunningham y Maloney (2001), through factor and cluster analysis, determines the proportion of small informal enterprises was competitive and segmented.
- Pratap y Quitin (2006), through non-parametric techniques (PSM and D&D), assesses the segmentation hypothesis by comparing informal workers who earns the same as formal workers.
- Günther y Launov (2012), through finite mixture model with sample selection, determine the number of unobservable segments of informal sector endogenously and calculates the % of involuntary informal workers.
- Alcaraz, Chiquiar y Salcedo (2015), through a utility maximization model with entry barriers and self-selection, determine the % of involuntary informal workers.
Pratap y Quitin (2006) do not consider the heterogenous nature of informal labor market and Cunningham y Maloney (2001) does not take into account the sample selection.

The finite mixture model (Günther y Launov, 2012) does correct the sample selection of labor participation but does not address the issue of formal/informal selection.
Evidence in Peru

- Yamada (1994) self-employed workers have competitive earnings and represent a voluntary option while informal salaried are segmented in Lima in 1985, 1986 and 1990.
- Baldárrago (2015) replicated Guindling (1991)’s methodology. According to her, self-employed represent a competitive labor market while salaried are segmented in the south of Peru in 2013.
Objetives

- To adapt Alcaraz, Chiquiar y Salcedo (2015)’s methodology to the Peruvian case
- To extend the model allowing correction for sample selection of labor participation
- To test whether there are segmented, competitive or a mixed of both labor market
- To calculate the proportion of involuntary informal workers
- To propose a research agenda
Outline

1. Introduction
2. Model of segmentation and self-selection
3. Informality in Peru and sources
4. Results
5. Concluding remarks and research agenda
Assumptions of the model

- The individual maximizes his utility and not the income function (non-monetary valuations)
Assumptions of the model

- The individual maximizes his utility and not the income function (non-monetary valuations)
- The decision to join the labor force is considered (we include those who do not work)
Assumptions of the model

- The individual maximizes his utility and not the income function (non-monetary valuations)
- The decision to join the labor force is considered (we include those who do not work)
- Self-selection in the informal/formal labor market is considered
Assumptions of the model

- The individual maximizes his utility and not the income function (non-monetary valuations)
- The decision to join the labor force is considered (we include those who do not work)
- Self-selection in the informal/formal labor market is considered
- There are entry barriers to formal employment
Model

Labor force participation equation:

\[ L^* = Z_i \alpha + \epsilon_{i,1} \quad (1) \]

Choice of the segment formal/informal:

\[ F^* = X_i \beta + \epsilon_{i,2} \quad (2) \]

\[
\begin{bmatrix}
\epsilon_{i,1} \\
\epsilon_{i,2}
\end{bmatrix} \mid Z, X \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & p \\ p & 1 \end{bmatrix} \right)
\]

Choice of the segments

- do not work if \( L^* < 0 \)
- formal if \( L^* > 0 \) and \( F^* > 0 \) and is hired
- informal if \( L^* > 0 \) and \( F^* < 0 \) or \( F^* > 0 \) and is not hired
Model II

We define the hiring parameter $\delta$ as the probability of being hired or access to formal employment. Therefore, the probability of choosing an employment is:

Probability of not working

$$ p(\text{do not work}) = P(L^* > 0) = \Theta(Z_{i,1}\alpha) $$  \hspace{1cm} (3)
Model II

We define the hiring parameter $\delta$ as the probability of being hired or access to formal employment. Therefore, the probability of choosing an employment is:

**Probability of not working**

$$p(\text{do not work}) = P(L^* > 0) = \Theta(Z_i, \alpha)$$  \hspace{1cm} (3)

**Probability of working in a formal employment**

$$p(\text{formal}) = P(F^* > 0|L^* > 0)P(L^* > 0)\delta$$

$$= \delta P(X_i, \beta > 0|Z_i, \alpha > 0)P(Z_i, \alpha > 0)$$ \hspace{1cm} (4)
**Model II**

We define the hiring parameter $\delta$ as the probability of being hired or access to formal employment. Therefore, the probability of choosing an employment is:

<table>
<thead>
<tr>
<th>Probability of not working</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(\text{do not work}) = P(L^* &gt; 0) = \Theta(Z_{i,1}\alpha)$ (3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probability of working in a formal employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(\text{formal}) = P(F^* &gt; 0</td>
</tr>
<tr>
<td>$= \delta P(X_{i,2}\beta &gt; 0</td>
</tr>
<tr>
<td>$= \delta \text{BivariateNormal}(X_{i,2}\beta, Z_{i,1}\alpha, p)$</td>
</tr>
</tbody>
</table>
Model III

Probability of working in an informal employment

\[ p(\text{informal}) = P(F^* < 0 | L^* > 0)P(L^* > 0) + (1 - \delta)P(F^* > 0 | L^* > 0)P(L^* > 0) \] 

(5)
Model III

Probability of working in an informal employment

\[
p(\text{informal}) = P(F^* < 0|L^* > 0)P(L^* > 0) \\
+ (1 - \delta)P(F^* > 0|L^* > 0)P(L^* > 0) \\
= P(X_{i,2}\beta < 0|Z_{i,1}\alpha > 0)P(Z_{i,1}\alpha > 0) \\
+ (1 - \delta)P(X_{i,2}\beta > 0|Z_{i,1}\alpha > 0)P(Z_{i,1}\alpha > 0)
\] (5)
Model III

Probability of working in an informal employment

\[
p(\text{informal}) = P(F^* < 0|L^* > 0)P(L^* > 0) \\
+ (1 - \delta)P(F^* > 0|L^* > 0)P(L^* > 0) \\
= P(X_i,2\beta < 0|Z_i,1\beta > 0)P(Z_i,1\beta > 0) \\
+ (1 - \delta)P(X_i,2\beta > 0|Z_i,1\beta > 0)P(Z_i,1\beta > 0) \\
= BivariateNormal(-X_i,2\beta, Z_i,1\alpha, -p) \\
+ (1 - \delta)BivariateNormal(X_i,2\beta, Z_i,1\alpha, p)
\]

(5)
Model IV

So log-likelihood function and the parameter constrain is:

\[
\sum_{\text{formal}} \left[ \ln(\delta BVN(X_{i,2\beta}, Z_{i,1\alpha}, p)) \right]
\]

\[
+ \sum_{\text{informal}} \left[ \ln(BivariateNormal(-X_{i,2\beta}, Z_{i,1\alpha}, -p)) \right]
\]

\[
+ (1 - \delta) BivariateNormal(X_{i,2\beta}, Z_{i,1\alpha}, p))]
\]

\[
+ \sum_{\text{do not work}} \left[ \ln(\Theta(Z_{i,1\alpha})) \right], \text{where} 0 < \delta < 1
\]
We use the hiring parameter $\delta$ to calculate the proportion of involuntary informal workers. Let $FO$ be the number of formal workers, $I$ the number of informal workers and $M$ the number of workers that would prefer to be formal ($FO = \delta M$). Therefore,

\[
\text{Proportion of involuntary informal workers} = \frac{(1 - \delta)M}{I} = \frac{(1 - \delta)FO}{\delta I}
\]
Outline

1. Introduction
2. Model of segmentation and self-selection
3. Informality in Peru and sources
4. Results
5. Concluding remarks and research agenda
Definition of informality

- A Satellite Account of the Informal Economy in Peru (INEI, 2014)
- Identification of salaried informal workers with ENAHO (official from 2012):
  - Employees without health insurance granted by their employers or in a unregistered firm
- Identification of salaried informal workers with ENAHO (ad hoc narrow definition):
  - Employees without health insurance granted by their employers, without payment to pension insurance, without contract, in an unregistered firm, the firm do not have books of account
Stylized facts

- During 2012-2016, the informality rate (of urban salaried from private sector) has a slight negative tendency.

- Informality stopped shrinking in the narrow def. because of the reduction in the growth in 2014 (Cespedes, 2015).

**Evolution in the informality rate and growth of GDP, 2012-2016**

<table>
<thead>
<tr>
<th>Years</th>
<th>Informal (official)</th>
<th>Informal (narrow def.)</th>
<th>Economic growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>63%</td>
<td>84%</td>
<td>6.1%</td>
</tr>
<tr>
<td>2013</td>
<td>63%</td>
<td>85%</td>
<td>5.9%</td>
</tr>
<tr>
<td>2014</td>
<td>62%</td>
<td>84%</td>
<td>2.4%</td>
</tr>
<tr>
<td>2015</td>
<td>60%</td>
<td>85%</td>
<td>3.3%</td>
</tr>
<tr>
<td>2016</td>
<td>59%</td>
<td>85%</td>
<td>3.9%</td>
</tr>
</tbody>
</table>
Descriptive statistics of the model variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Informal (official)</th>
<th>Informal (narrow def.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Or Less</td>
<td>85%</td>
<td>96%</td>
</tr>
<tr>
<td>Secondary</td>
<td>70%</td>
<td>89%</td>
</tr>
<tr>
<td>Non-University Higher Education</td>
<td>48%</td>
<td>72%</td>
</tr>
<tr>
<td>University Higher Education</td>
<td>42%</td>
<td>72%</td>
</tr>
<tr>
<td><strong>Years old (mean)</strong></td>
<td>32.51</td>
<td>35.44</td>
</tr>
<tr>
<td><strong>Married</strong></td>
<td>41%</td>
<td>39%</td>
</tr>
<tr>
<td><strong>Head of HH</strong></td>
<td>29%</td>
<td>22%</td>
</tr>
<tr>
<td><strong>Has any insurance different from job's</strong></td>
<td>43%</td>
<td>52%</td>
</tr>
<tr>
<td><strong>White collars</strong></td>
<td>38%</td>
<td>16%</td>
</tr>
<tr>
<td><strong>Work more than 35 hours</strong></td>
<td>63%</td>
<td>84%</td>
</tr>
<tr>
<td><strong>Geographic area</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costa Norte</td>
<td>70%</td>
<td>89%</td>
</tr>
<tr>
<td>Costa Centro</td>
<td>62%</td>
<td>84%</td>
</tr>
<tr>
<td>Costa Sur</td>
<td>71%</td>
<td>88%</td>
</tr>
<tr>
<td>Sierra Norte</td>
<td>76%</td>
<td>92%</td>
</tr>
<tr>
<td>Sierra Centro</td>
<td>75%</td>
<td>90%</td>
</tr>
<tr>
<td>Sierra Sur</td>
<td>70%</td>
<td>88%</td>
</tr>
<tr>
<td>Selva</td>
<td>82%</td>
<td>93%</td>
</tr>
<tr>
<td>Lima Metropolitana</td>
<td>51%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Outline

1. Introduction

2. Model of segmentation and self-selection

3. Informality in Peru and sources

4. Results

5. Concluding remarks and research agenda
Results

- The % of involuntary informal workers is 10 % or 5 % depending of the definition.
- The low levels are coherents. In the Mexican case, the % is 15 (Alcaraz, Chiquiar, & Salcedo, 2012).
- Estimates from Tello (2015) are between 11 % and 73 % depending of the definition.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Informal (official)</th>
<th>Informal (narrow)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Years old</td>
<td>-0.133*** (-41.73)</td>
<td>-0.112*** (-31.35)</td>
</tr>
<tr>
<td>Years old^2</td>
<td>0.00140*** (38.83)</td>
<td>0.00115*** (27.65)</td>
</tr>
<tr>
<td>Primary or less</td>
<td>-1.98e-13</td>
<td>-3.67e-14</td>
</tr>
<tr>
<td>Secondary</td>
<td>-0.610*** (-28.06)</td>
<td>-0.707*** (-26.69)</td>
</tr>
<tr>
<td>Non-university higher education</td>
<td>-1.216*** (-41.16)</td>
<td>-1.386*** (-41.15)</td>
</tr>
<tr>
<td>University higher education</td>
<td>-1.392*** (-43.80)</td>
<td>-1.623*** (-45.13)</td>
</tr>
<tr>
<td>No married</td>
<td>-2.39e-16</td>
<td>-1.67e-16</td>
</tr>
<tr>
<td>Married</td>
<td>-0.414*** (-23.73)</td>
<td>-0.413*** (-23.15)</td>
</tr>
<tr>
<td>No head of HH</td>
<td>-3.05e-16</td>
<td>2.43e-16</td>
</tr>
<tr>
<td>Head of HH</td>
<td>-0.134*** (-7.48)</td>
<td>-0.146*** (-7.26)</td>
</tr>
<tr>
<td>Has health insurance diff. of job's</td>
<td>2.005*** (67.22)</td>
<td>1.884*** (55.44)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.367*** (45.49)</td>
<td>3.564*** (45.22)</td>
</tr>
<tr>
<td>p</td>
<td>0.516*** (26.37)</td>
<td>0.059 (1.82)</td>
</tr>
<tr>
<td>β</td>
<td>1.826*** (14.25)</td>
<td>2.092*** (21.79)</td>
</tr>
<tr>
<td>%involuntary informal</td>
<td>8.94*** (-0.012)</td>
<td>6.85*** (-0.007)</td>
</tr>
</tbody>
</table>

Geo. areas and years fixed eff. | Yes | Yes | Yes | Yes |
Observations | 158520 | 158520 | 158520 | 158520
Robustness check

- There are more involuntary informal white collars. Unlike blue collars, white collars really care about being formal.
- There are less involuntary informal full-time workers. Part-time workers really care about being formal.

<table>
<thead>
<tr>
<th></th>
<th>Informal (official)</th>
<th>Informal (narrow)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bechmark</td>
<td>8.94***</td>
<td>6.85***</td>
<td>4.41***</td>
<td>4.36***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.012)</td>
<td>(-0.007)</td>
<td>(-0.005)</td>
<td>(-0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White collars</td>
<td>10.46***</td>
<td>8.67***</td>
<td>4.42***</td>
<td>4.27***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.013)</td>
<td>(-0.008)</td>
<td>(-0.004)</td>
<td>(-0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work more than 35 hours</td>
<td>5.8***</td>
<td>7.04***</td>
<td>1.99***</td>
<td>2.69***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.016)</td>
<td>(-0.008)</td>
<td>(-0.004)</td>
<td>(-0.002)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Model of segmentation and self-selection
3. Informality in Peru and sources
4. Results
5. Concluding remarks and research agenda
Concluding remarks

- The hypothesis of the existence of multi-segmented labor markets is confirmed
- Around 10% of informal workers are involuntary
- Limitations of the study:
  - The model has strong assumptions about the distributions
  - There might be endogeneity with the variables correlated with earning
Research agenda

- Correct the possible endogeneity of the variables correlated with the income (IV, control function, etc).
- Explore the public policy effects over the hiring parameter within a model of self-selection of informal employment.
Thanks