Attitude of informal workers towards formality in Peru

Erick Lahura and María Paula Vargas

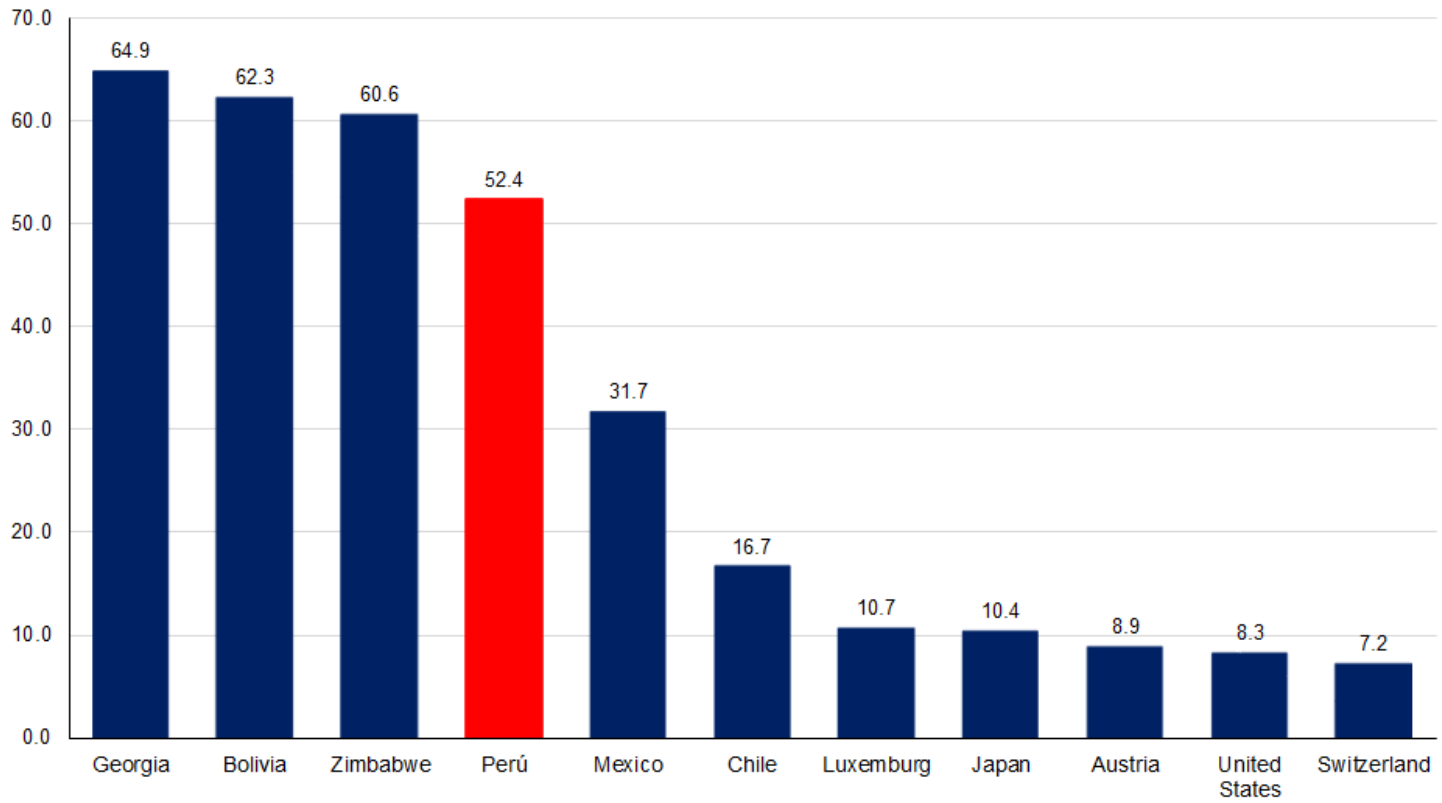
XXXVI Encuentro de Economistas del BCRP

October, 2018

Structure

Introduction

Motivation: Peru has one of the highest levels of informality in the world



Note: As percentage of GDP. **Source**: Medina and Schneider (2018).

Introduction
What do we do?

- **Research question**: **Why informal workers believe "it is not necessary" to be formal?**
  - Recent survey data provides information on the reasons for being informal in Peru.
  - We explore potential factors that lead informal workers to believe that it is not necessary to be formal in Peru.
- **Hypothesis**: informality is related to several socio−economic and demographic characteristics.

Introduction
What do we do?

- **Data**: National Household Survey (ENAHO), 2014-2016.
- **Empirical methodology**: based on standard binary choice models:
  - linear probability, logit and probit models.
  - Oaxaca/Blinder decomposition.
- **Results**: income, age, type of worker, location of business and gender, among the main factors. Men are less committed to formality.

Introduction
Literature Review

- **Peru**: Loayza (2008), Chong y Otros (2008), Del Valle (2009), Jaramillo (2013), Díaz (2014), Herrera e Hidalgo (2014), Machado (2014), Tello (2014), Chacaltana (2016).

- **Empirical methods**: There is no single empirical approach for identifying the predictors of informality.

- **Studies**: Norris et al. (2008), Lehmann and Zaiceva (2013), Dau and Cuervo (2014), Thanh Thai and Ekaterina (2014), Babbitt et al. (2015), Williams et al. (2015), Elbahnasawy et al. (2016), etc.

Introduction

Literature Review: Key determinants/predictors of informality.

Author(s)	Entrepreneur's demographic and socioeconomic	Industry and Firms characteristics	Financial market development	Tax and Regulation burden	Quality of Institutions and Government Effectiveness
Macro Level					
Dau and Cuervo (2014)			×		×
Thanh Thai and Ekaterina (2014)				×	×
Elbahnasawy et al. (2016)				×	×
Moreno and Posadab (2018)					×
Firm Level					
Norris et al. (2008)		×	×	×	×
Williams et al. (2015)	×	×		×	×
Jimenez et al. (2015)	×				
Individual Level					
Babbitt et al. (2015)	×	×			
Lehmann and Zaiceva (2013)	×	×	×		

Introduction

Literature Review: by type of data

	Macro Level	Firm Level	Individual Level
	Dau and Cuervo (2014)	**Norris et al. (2008)**	**Babbitt et al. (2015)**
Research Question	Do pro-market institutions determine informal or formal entrepreneurships?	What causes firms to hide output?	Are female entrepreneurs more likely to prefer the formal sector or the informal sector?
Informality - measure	Informal entrepreneurship: Number of new unregistered businesses as a percent of the working-age population.	• The share of sales kept informal: *what percentage of total sales would you estimate the typical firm in your area of activity keeps âĂIJoff the booksâĂİ?* • 7 answers: i) none at all, ii) 1-10%; iii) 11-20%; iv) 21-30%; v) 31-40%; vi) 41-50%; and vii) more than 50%.	• Based on the question: *âĂIJDo you want to formalize your business?âĂİ* • 55% of female entrepreneurs responded âĂIJyesâĂİ and 48% of male entrepreneurs responded âĂIJyes.âĂİ
Data and sample	51 countries, period: 2002-2009.	4000 firms in 41 countries.	Indonesia, 141 entrepreneurs individuals (85 female and 56 male), period: 2012.
Determinants	Pro market institutions: Economic Liberalization and National Governance.	Index of quality of legal institutions, tax and regulation burden, financial market development, entrepreneurial characteristics.	Socio-demographic and Industry-Firms characteristics.
Control Variables	GDP per capita, GDP growth, immigration rate, dummy of crises, unobserved country-specific factors.	GDP per capita.	
Methodology	Panel Data.	Ordered Probit Model.	Logistic Regression Model.
Main findings	• Institutional environment matters: start an entrepreneurial venture and whether be informal or formal. • Economic liberalization may facilitate their growth within the informal sector.	• Quality of legal institutions is important in determining the size of the informal sector. • Taxes, regulations, and financial constraints are not significative.	• Female entrepreneurs are less likely to be informal. • Their decision is conditional on other factors.

Introduction
Literature Review: Gender Differences

The literature provides several studies that explore the *importance of gender in determining people preferences about informality*, although no consensus appears to be reached.

- Sethuraman (1998), Chen et al (2006), Chant and Pedwell (2008), Chakrabarti (2009) and Otobe (2017): given the different structural disadvantages between men and women, women are more likely to be informal than men.

- *Peru: are women more likely to prefer the formal rather than the informal sector?*

Structure

Data and summary statistics

Source: Peruvian National Household Survey (ENAHO)

- Annual survey; national, urban, and rural coverage.
- Information on poverty, living conditions, household expenses and income, etc.
- Questionnaire ENAHO.04: "**Income of the independent worker**".
  - Question: "**What is the main reason why you have not registered in SUNAT?**" (applied for the first time in 2014).
  - Period of analysis: 2014 – 2016.

Data and summary statistics

ENAHO: "What is the main reason why you have not registered in SUNAT?"

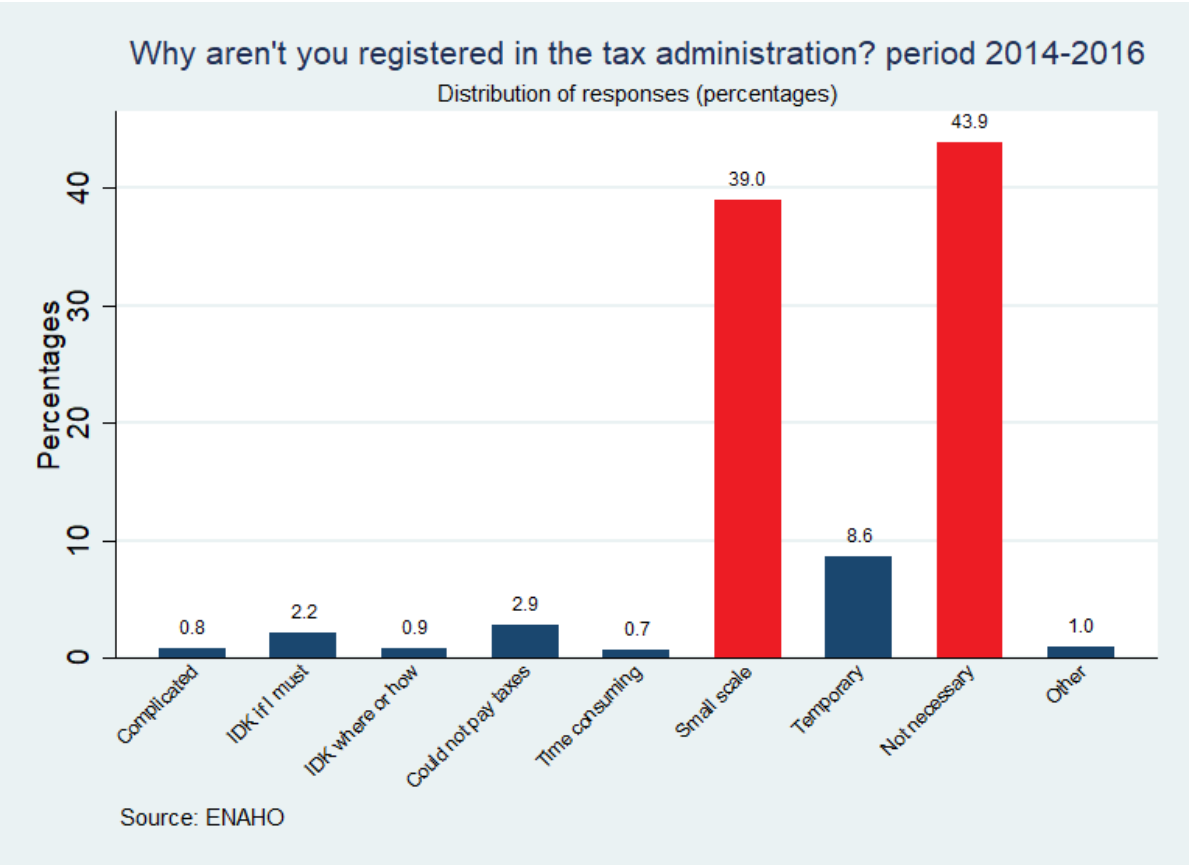
Employer and Independent worker

	Answers	Main Occupation 2014	2015	2016
1	The procedures are very complicated	0.85%	0.63%	1.09%
2	I do not know if I must register	3.32%	1.62%	1.65%
3	I do not know where or how to register	1.03%	0.79%	0.87%
4	I could not assume the tax burden if I were registered	2.90%	3.34%	2.46%
5	It is time consuming	0.89%	0.74%	0.52%
6	My business is small / I produce small quantities	41.44%	37.85%	37.34%
7	It is an eventual work	9.62%	8.41%	8.26%
8	**I do not believe it is necessary**	**38.71%**	**46.0%**	**46.74%**
9	Another	1.23%	0.82%	1.10%
	TOTAL in %	100%	100%	100%
	TOTAL	12 395	12 057	13 919

	Grouped answers	2014-2016
1	**It is not necessary** (option 8)	43.93
0	Other reasons (all except option 8)	56.07
	TOTAL	38 371

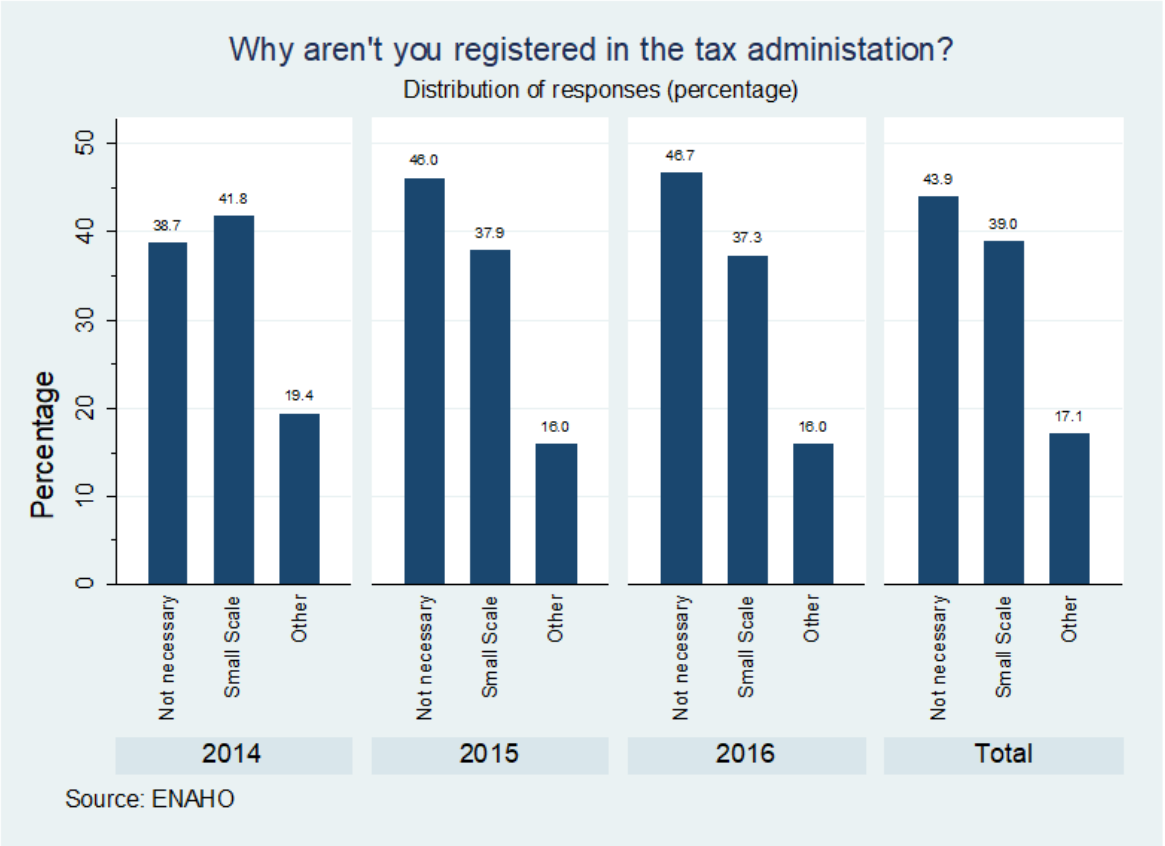
Data source

Distribution of all answers 2014-2016



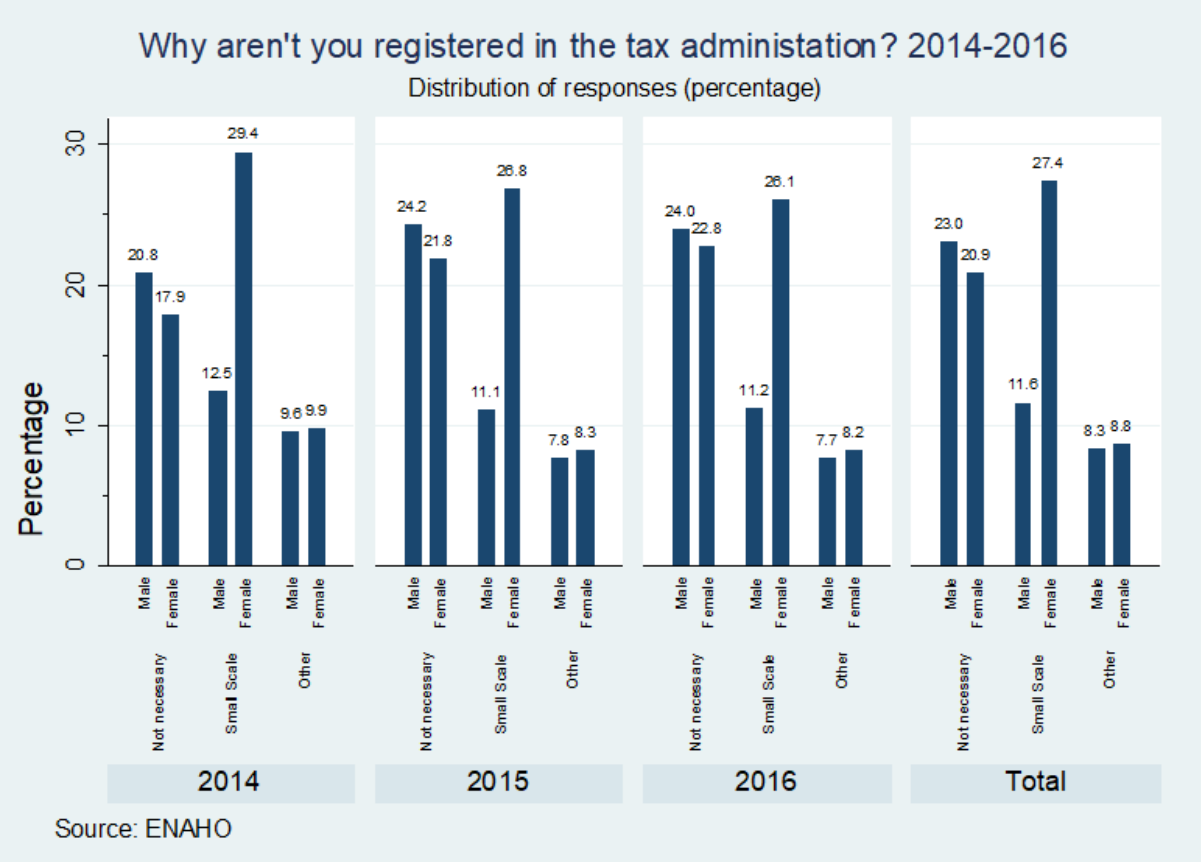
Data and summary statistics

Distribution of two most important answers



Data and summary statistics

Distribution of answers by gender



Data and summary statistics
main facts

- The evidence suggests that most informal workers do not consider it necessary to become formal.
- **Informal workers who are less committed to pay taxes** are men, between 36 and 55 years, without education or basic school education, partners or married, whose business location is in Lima and whose economic activity belongs to the transports and retail sales sector.

Structure

Methodology

- In order to identify the factors that determine the attitude of informal workers towards formality, we define our endogenous variable y_i as follows:

$$y_i = \begin{cases} 1 & , \quad \text{“}i\text{” believes it is not necessary to be formal} \\ & \quad (i.e., \text{ tax compliant}) \\ 0 & , \quad \text{otherwise} \end{cases}$$

where “i” represents an individual that is informal. In order to estimate the probability that $y_i = 1$ conditional on a $k \times 1$ vector of explanatory variables x_i

Methodology
Discrete choice models

First we assume that this probability is a linear function of x_i:

$$\begin{aligned} y_i &= Pr(y_i = 1|x_i) + u_i \\ &= x_i^{'}\beta + u_i \end{aligned}$$

for $i = 1, 2, \ldots, n$, where β is a $k \times 1$ vector of unknown parameters and $E(u_i) = 0$. The vector x_i includes factors that describe informal worker i, such as age, gender, educational attainment, income, among others.

Methodology
Discrete choice models

In order to avoid any potential biases, we also estimate two standard discrete choice models. First, we estimate a probit model, which assumes that the probability of y_i is described by:

$$Prob[y_i = 1] = \Phi(x_i' \beta)$$

where $\Phi(\cdot)$ represents a standard normal distribution. Second, we estimated a logit model, which assumes that the probability of y_i is described by :

$$Prob[y_i = 1] = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}$$

Methodology
Oaxaca/Blinder (OB) decomposition

The implementation of OB requires the estimation of separate regressions for males and females:

$$y_{m,i} = x'_{m,i}\beta_m + u_{m,i}$$
$$y_{f,i} = x'_{f,i}\beta_f + u_{f,i}$$

The OD is an expression that compares the expected value of y_i between males and females:

$$\begin{aligned} E[y_{m,i}] - E[y_{f,i}] &= x'_{m,i}\beta_m - x'_{f,i}\beta_f \\ &= [x_{m,i} - x_{f,i}]'\beta_f + x'_{m,i}[\beta_m - \beta_f] \end{aligned}$$

Methodology
Oaxaca/Blinder (OB) decomposition

The differences in the attitude towards formality can be explained by two components:

1. Differences in observed characteristics specified in x, $[x_{m,i} - x_{f,i}]$, usually called the explained component.
2. Differences related to being male and female (treatment effect), $[\beta_m - \beta_f]$.

The Oaxaca decomposition can be estimated using the OLS estimates of β_m and β_f and the sample mean values of $x_{m,i}$ and $x_{f,i}$:

$$\bar{y}_m - \bar{y}_f = [\bar{x}_m - \bar{x}_f]' \hat{\beta}_f + \bar{x}_m[\hat{\beta}_m - \hat{\beta}_f]$$

Methodology
Explanatory variables

Variable name	Description
lincome	The natural log of the annual income in the main activity of informal worker.
female	A dummy variable = 1 if the informal worker is female; = 0 otherwise.
age18_25	A dummy variable = 1 if the informal worker is aged 18−25 years; = 0 otherwise.
age26_35	A dummy variable = 1 if the informal worker is aged 26−35 years; = 0 otherwise.
age36_55	A dummy variable = 1 if the informal worker is aged 36−55 years; = 0 otherwise.
with_edu	A dummy variable = 1 if the informal worker has a level of education higher than pre−school (primary, secondary, university, postgraduate, or other high education, either complete or incomplete; = 0 otherwise.
marrpart	A dummy variable = 1 if the informal worker is married or partner; = 0 otherwise.
widivse	A dummy variable = 1 if the informal worker is divorced or separated; = 0 otherwise.
d2015	A dummy variable = 1 if the year is 2015; = 0 otherwise.
d2016	A dummy variable = 1 if the year is 2016; = 0 otherwise.
independent	A dummy variable = 1 if the informal worker is independent; = 0 otherwise.
secondary_act	A dummy variable = 1 if the economic activity of informal worker belongs to secondary sector; = 0 otherwise.
terciary_act	A dummy variable = 1 if the economic activity of informal worker belongs to terciary sector; = 0 otherwise.
department	A dummy variable = 1 for each department of Peru indicating the location of the business of informal worker. There are in total 24 region dummies; = 0 otherwise.

Structure

Summary statistics by gender group

Variable	Pooled	Female	Male
	Mean/SD	Mean/SD	Mean/SD
not_necessary	0.43930	0.36638	0.53607
	(0.4963)	(0.4818)	(0.4987)
lincome	9.16290	8.84016	9.59119
	(1.3096)	(1.3940)	(1.0445)
female	0.57026	1	
	(0.4950)	0	
age18_25	0.10139	0.07403	0.13770
	(0.3018)	(0.2618)	(0.3445)
age26_35	0.18919	0.18307	0.19731
	(0.3916)	(0.3867)	(0.3979)
age36_55	0.46774	0.49794	0.42765
	(0.4989)	(0.5001)	(0.4947)
with_edu	0.94618	0.91580	0.98651
	(0.2256)	(0.2776)	(0.1154)
marrpart	0.66627	0.66564	0.66710
	(0.4715)	(0.4718)	(0.4713)
widivse	0.18619	0.23858	0.11666
	(0.3893)	(0.4262)	(0.3210)
d2015	0.31470	0.31393	0.31573
	(0.4645)	(0.4641)	(0.4649)
d2016	0.36360	0.36384	0.36329
	(0.4810)	(0.4811)	(0.4809)
independent	0.94927	0.97304	0.91772
	(0.2194)	(0.1619)	(0.2748)
secondary_act	0.11451	0.13725	0.08433
	(0.3184)	(0.3441)	(0.2778)
terciary_act	0.84945	0.85751	0.83876
	(0.3576)	(0.3495)	(0.3677)
Observations	37 279	21 259	16 020

Standard Deviation in parentheses.

Marginal and Impact effects

Estimates from the linear probability model

	Dependent variable: not_necessary							
	Model 1: 2014−2016		Model 2: 2014		Model 3: 2015		Model 4: 2016	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
lincome	0.015***	(0.002)	0.014***	(0.004)	0.017***	(0.004)	0.014***	(0.004)
female	-0.144***	(0.005)	-0.153***	(0.010)	-0.142***	(0.010)	-0.135***	(0.009)
age18_25	0.023**	(0.010)	0.043**	(0.018)	0.021	(0.018)	0.007	(0.018)
age26_35	0.030***	(0.008)	0.047***	(0.014)	0.028**	(0.014)	0.019	(0.013)
age36_55	0.027***	(0.006)	0.026**	(0.011)	0.040***	(0.011)	0.020*	(0.011)
with_edu	-0.014	(0.011)	-0.020	(0.020)	-0.002	(0.020)	-0.022	(0.019)
marrpart	-0.009	(0.008)	-0.008	(0.014)	-0.010	(0.014)	-0.012	(0.013)
widivse	0.002	(0.010)	0.007	(0.017)	-0.007	(0.017)	-0.000	(0.016)
d2015	0.073***	(0.006)						
d2016	0.081***	(0.006)						
independent	0.063***	(0.012)	0.086***	(0.019)	0.051**	(0.021)	0.050**	(0.020)
secondary_act	-0.018	(0.015)	0.037	(0.027)	-0.039	(0.027)	-0.043*	(0.026)
terciary_act	0.020	(0.014)	0.079***	(0.024)	-0.013	(0.024)	-0.001	(0.023)
constant	0.339***	(0.033)	0.346***	(0.055)	0.457***	(0.059)	0.405***	(0.057)
R−squared	0.100		0.101		0.120		0.099	
Observations	37095		11975		11710		13410	
F stat regression	111		38		46		42	
P val F regression	0.000		0.000		0.000		0.000	

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

Marginal and Impact effects 2014-2016

LPM, Probit and logit

Dependent variable: not_necessary						
	LPM		PROBIT		LOGIT	
	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
lincome	0.015***	(0.002)	0.015***	(0.002)	0.016***	(0.002)
female	-0.144***	(0.005)	-0.143***	(0.006)	-0.142***	(0.006)
age18_25	0.023**	(0.010)	0.023**	(0.010)	0.023**	(0.010)
age26_35	0.030***	(0.008)	0.030***	(0.008)	0.029***	(0.008)
age36_55	0.027***	(0.006)	0.027***	(0.006)	0.026***	(0.006)
secondary_act	-0.018	(0.015)	-0.020	(0.015)	-0.022	(0.015)
tertiary_act	0.020	(0.014)	0.020	(0.013)	0.020	(0.013)
with_edu	-0.014	(0.011)	-0.012	(0.012)	-0.012	(0.012)
marrpart	-0.009	(0.008)	-0.009	(0.008)	-0.009	(0.008)
widivse	0.002	(0.010)	0.002	(0.010)	0.002	(0.010)
d2015	0.073***	(0.006)	0.073***	(0.006)	0.073***	(0.006)
d2016	0.081***	(0.006)	0.082***	(0.006)	0.081***	(0.006)
independent	0.063***	(0.012)	0.061***	(0.011)	0.061***	(0.011)

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

Oaxaca/Blinder (OB) decomposition

	Dependent variable: not_necessary			
	Model 1: 2014−2016	Model 2: 2014	Model 3: 2015	Model 4: 2016
total				
female	0.367***	0.313***	0.383***	0.400***
	(110.57)	(55.68)	(64.20)	(71.29)
male	0.536***	0.486***	0.561***	0.558***
	(135.49)	(69.48)	(80.10)	(84.97)
difference	-0.169***	-0.173***	-0.178***	-0.158***
	(-32.77)	(-19.28)	(-19.35)	(-18.26)
explained	-0.0250***	-0.0233**	-0.0356***	-0.0206**
	(-5.63)	(-2.97)	(-4.57)	(-2.73)
unexplained	-0.144***	-0.150***	-0.143***	-0.137***
	(-22.04)	(-13.02)	(-12.47)	(-12.49)

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

Log Income regressions

	Dependent variable: not_necessary							
	Model 1: 2014−2016		Model 2: 2014		Model 3: 2015		Model 4: 2016	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
female	-0.664***	(0.013)	-0.702***	(0.023)	-0.662***	(0.023)	-0.631***	(0.021)
age18_25	-0.052**	(0.025)	-0.057	(0.044)	-0.042	(0.043)	-0.055	(0.041)
age26_35	0.289***	(0.019)	0.301***	(0.034)	0.296***	(0.033)	0.270***	(0.031)
age36_55	0.353***	(0.015)	0.356***	(0.027)	0.350***	(0.027)	0.353***	(0.025)
with_edu	0.482***	(0.027)	0.521***	(0.049)	0.477***	(0.048)	0.447***	(0.045)
marrpart	0.334***	(0.019)	0.370***	(0.034)	0.317***	(0.034)	0.312***	(0.031)
widivse	0.336***	(0.023)	0.322***	(0.041)	0.313***	(0.041)	0.356***	(0.037)
d2015	0.057***	(0.015)						
d2016	0.077***	(0.014)						
independent	-1.157***	(0.027)	-1.158***	(0.047)	-1.134***	(0.049)	-1.174***	(0.045)
secondary_act	-0.428***	(0.037)	-0.445***	(0.067)	-0.359***	(0.066)	-0.479***	(0.060)
tertiary_act	0.457***	(0.033)	0.439***	(0.059)	0.510***	(0.058)	0.418***	(0.053)
constant	9.472***	(0.062)	9.388***	(0.108)	9.508***	(0.110)	9.666***	(0.102)
R−square	0.257		0.260		0.258		0.262	
Observations	37095		11975		11710		13410	
F stat regression	357		123		119		140	
P val F regression	0.000		0.000		0.000		0.000	

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Structure

Conclusions

- The results suggest that variables such as gender, income, age, type of worker and location of the business, help explaining why informal workers believe it is not necessary to become formal.
- Evidence of gender differences: informal women have a more positive attitude towards formality than men.
- These results are robust to the sample period and alternative estimation methods.

Thanks

Attitude of informal workers towards formality in Peru

Erick Lahura and María Paula Vargas

XXXVI Encuentro de Economistas del BCRP

October, 2018