Heterogenous effects of informality: An application to labor regulation policy in Russia

Andrea Otero

UNC-Chapel Hill & Banco de la República de Colombia

Bogotá, May 28, 2018

Overview of the paper

- Estimate the heterogeneous returns of informality.
  - Model the choice of individuals of having a formal job.
- Data used: Russian Longitudinal Monitoring Survey (2009-1016).
- Application: Enforcement of the labor regulation in Russia.

Motivation

- Informality is very common phenomenon in developing countries and transition economies.
- Lack of consensus about how to measure it.
- Very hard to estimate its causal effect on earnings.
- Traditional focus measures the impact of changing social benefits such as unemployment insurance, health insurance, etc., not on enforcement.

The pros and cons of informality from the literature

- Pros: Allows workers to optimally choose the amount of social security they want to consume; more flexible schedule; potential higher earnings due to tax evasion.
- Cons: Earnings could be low and irregular; no access to safety nets such as unemployment insurance; no/limited access to credit markets.
    - Papers: Fields, 1990; Maloney, 2004; Perry et al., 2007; Slonimczyk, 2013; Narita et al., 2015; Lehmann and Zaiceva, 2013, 2015; Bobba et al., 2016.

Informality and wages

The literature linking informality and wages has found that:

- From the firm side:
  - Conditional on productivity, wages in the informal sector are higher than in the formal sector because they operate beyond the state regulation. But firms in the informal sector have lower productivity (Ulyssea, 2015).
- From the workers side:
  - Compensating differentials: Jobs that are more desirable in terms of their amenities such as fringe benefits, stability and flexibility should have lower-than-average wages (Arias and Khamis, 2008).

What other similar papers do:

- Garcia (2017, RDE)
  - The author finds evidence of both voluntary (comparative advantage) and involuntary (segmentation) informal employement. Wage gap reduces when one moves from the bottom of the wage distribution to the top. As the author claims: "Results indicate that low-paid informal workers earn less not only because they are less skilled but also because they earn lower returns on such skills, whereas highly paid informal workers earn less because formal workers have superior skills."
  - Methodology: Quantile decomposition a-la Machado-Mata controlling for selection bias caused by correlated unobserved heterogeneity.

What other similar papers do:

- Lehman and Zaiceva (2013, WP)
  - The authors find weak evidence of segmentation in the lower quantiles of the distribution for salaried workers in Russia and no wage gap in the upper quantiles. When comparing informal self-employed and entrepreneurs with formal workers, they find a negative wage gap in the lowest quartile and a a positive wage gap (but not significant) in the highest quartile that points to a segmented informal sector with a lower free entry tier and an upper rationed tie.
  - Methodology: Quantile Regression (no instruments)

The complexity of modeling informality:

1. Measurement of informality: many different measures depending on the data available.
2. Selection into formality/informality: individuals are not randomly selected into informality. Returns are likely correlated with the informality status.
3. Heterogeneity: returns to informality vary based on observable and unobservable characteristics of the individual.

Measurement of informality

How is informality measured in the literature?:

- Employees without social security or labor benefits (Arias and Khamis, 2008; Pratap and Quintin, 2006; Bobba, Flabbi, and Levy, 2016).
- Workers who work in small firms with less than 5 employees (Pianto, Tannuri-Pianto, and Arias, 2004).
- Workers who are not registered in the labor office or do not have a work card (Almeida and Carneiro, 2012; Meghir, Narita, and Robin, 2015).

For the most part, the definition of informality used by different authors is driven by the availability of data.

Job categories in the RLMS

- Formal employee: Worker at a firm who is officially registered as an employee.
- Informal employee: Worker at a firm who is not officially registered as an employee
- Works not at a firm: Worker does not work at an organization or enterprise where more than one person works. Includes self-employed and people who work for entrepreneurs. Worker may or may not be informal.
- IEA: Includes off-hour job, occasional employment against terminal contract, labor agreement, contract of work and labor, grant, or individual job under license or not.

Measurement of informality

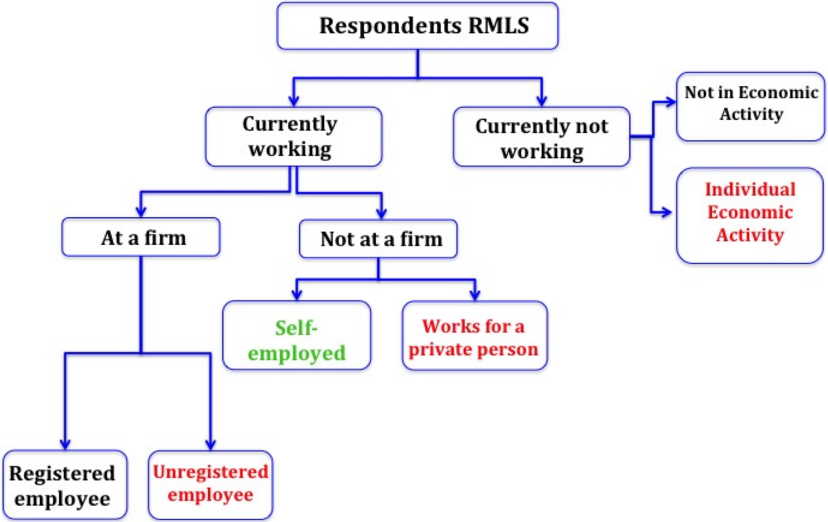
Formal at firm	82%
Informal at firm	6%
Works not at firm	10%
IEA	2%

Informality measure in my data

Main Definition

- Formal employee: Worker at a firm who is officially registered as an employee.
- Informal employee: Worker at a firm who is not officially registered as an employee

Measurement of informality



Selection into Formality

1. Inverse Probability Weighting Regression Adjustment Treatment Effects (IPWRA TE) - Doubly robust estimator
2. Heckman-type selection models - Constrained endogenous switching model (Heckman, 1976; Maddala, 1983)
3. Heckman-type selection models - Marginal Treatment Effect (Heckman and Vytlacil, 2005)

Heterogeneity

- Compute the treatment effect of formality for an individual who is indifferent between being formal and informal, but differs only in their unobservable cost of being formal.

Definition of Marginal Treatment Effect based on Heckman and Vytlacil (2005):

$$MTE(x, p) = \mathbb{E}(Y_1 - Y_0 | X = x, P(Z) = p)$$

Data

- Russian Longitudinal Monitoring Survey - HSE
  - Annual nationally representative surveys designed to monitor the effects of Russian reforms on the health and economic welfare of households and individuals in the Russian Federation.
  - I use waves 2009 to 2016.
- Federal Service for Labour and Employment (Rostrud)
  - Regional annual data about labor regulation compliance (number of inspectors, number of penalties, types of penalties, amount of money charged, etc).
- CBSD - Rosstat
  - CPI per region and population per region.

Descriptive statistics

Table: Formal and informal workers sample

Year	Formal	Informal	Total
2009	4,917	362	5,279
2010	7,542	496	8,038
2011	7,497	497	7,994
2012	7,513	542	8,055
2013	7,294	512	7,806
2014	6,038	452	6,490
2015	5,913	474	6,387
2016	5,952	466	6,418
Total	52,666	3,801	56,467

Descriptive statistics

Table: Descriptive statistics I

Variable	Formal	Informal	t-test
Female=1	0.53	0.39	-15.74
Age	39.1	36.3	-15.5
	(10.6)	(10.5)	
Married=1	0.61	0.45	-18.9
Schooling level of individual			
Secondary=1	0.33	0.50	21.4
Upper vocational=1	0.24	0.18	-8.79
Higher education=1	0.34	0.14	-25.15
Schooling level of parents			
Upper vocational=1	0.23	0.2	-4.42
Higher education=1	0.19	0.16	-4.59
Missing=1	0.15	0.21	10.17
City size			
Regional center=1	0.34	0.44	12.2
Other city=1	0.35	0.30	-5.99
Population	1,346,069	1,546,306	3.77
	(3,149,100)	(3,258,943)	
HHFormal=1	0.58	0.42	-19.06
Distance to the nearest inspection office (km)	112.86	71.39	-16.58
	(152.47)	(84.02)	
Number of inspectors per 1,000 economic entities	0.40	0.38	-4.61
	(0.269)	(0.251)	

Descriptive statistics II

Table: Descriptive statistics (in rubles)

Variable	Formal	Informal	t-test
Real wage rate	145.33	124.33	-10.04
Real labor earnings	26,572	24,442	-5.8
Before public transfers income	50,390	44,285	-5.44
Disposable income	58,105	51,304	-5.73
Expenditures	44,111	42,392	-2.51

Note: Currently RUB$1,000 are equivalent to US$17.36. The monthly minimum wage in Russia in January 2016 was RUB$7,537 (or US$130.26).

Empirical Framework

- Two types of occupations indexed by two labor market sectors: formal (treated state) and informal (untreated state).
- Y_1: potential wage of an individual with a formal job (D=1)
- Y_0: potential wage of an individual with an informal job (D=0).

The potential outcomes are:

$$Y_1 = X'\beta_1 + U_1$$
$$Y_0 = X'\beta_0 + U_0$$

Where: X contains sociodemographic and regional characteristics: age, age squared, education, parents education, marital status, urban status, and log of the size of the population at site

The decision rule:

The decision rule of an individual i, who wants to maximize his utility by choosing a formal or an informal job, is characterized by the follwing latent variable model (Willis and Rosen, 1979):

$$D = 1(D^* > 0)$$

Where:

$$D^* = Z\gamma - V$$

V represents the unobserved net cost of being formal.

- Notice that (X,Z) is observed, but (U_0, U_1, V) is not.

Assumptions

- V is a continuous random variable with a strictly increasing distribution function F_V.
- (U_0, U_1, V) is statistically independent of Z given X.
- Z is a vector that contains X and it also contains exclusion restrictions.

Rewriting the decision rule:

$$D = 1(Z\gamma_z > V)$$

Let $P(Z)$ denote the probability of choosing the formal sector $(D = 1)$ conditional on $Z = z$, such that $P(Z) = P(D = 1|Z = z) = F_V(Z'\gamma)$.

Define $U_D = F_V(V)$, which is uniformly distributed by construction. Then:

$$D = 1(P(Z) > U_D)$$

- Assume a multivariate normal distribution of (U_0, U_1, V) with mean 0 and variance Σ and (U_0, U_1, V) is independent from (X, Z).

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{10} & \sigma_{0V} \\ \sigma_{10} & \sigma_1^2 & \sigma_{1V} \\ \sigma_{0V} & \sigma_{1V} & 1 \end{pmatrix} \tag{1}$$

We can express the MTE equation as:

$$\begin{aligned} MTE(x, u_s) &= X'(\beta_1 - \beta_0) + E(U_1 - U_0 | V = \Phi^{-1}(U_S)) \\ &= X'(\beta_1 - \beta_0) + (\sigma_{1V} - \sigma_{0V})\Phi^{-1}(U_S) \end{aligned}$$

- The parameters $(\beta_1, \beta_0, \sigma_{1V}, \sigma_{0V})$ and their standard errors can be estimated by MLE.

Identification

Identification Strategy: Need exclusion restrictions that shift the probability of being formal, but do not directly affect wages of the individuals. Exclusion restrictions need to be independent conditional on X's.

- Ideal exclusion restriction: A policy that randomly and unexpectedly changes the cost of being formal for some individuals. Not available.
- Candidates used in the literature: Having another member of the household working formally, log distance to the nearest labor inspection, log ratio of inspectors per 1,000 economic entities, and an interaction term between distance and ratio of inspectors.
- Additional controls: Year and region fixed effect.

Results: No selection - OLS

Table: OLS Results

Variable	lnwage
	coef/se
formal	0.122***
	(0.009)
Female	-0.304***
	(0.004)
Age	0.040***
	(0.002)
Age squared	-0.053***
	(0.002)
Married (=1)	0.046***
	(0.004)
Schooling categories (omited: primary or less)	
High School	0.084***
	(-0.008)
Technical/Vocational	0.186***
	(-0.008)
College or more	0.458***
	(-0.008)
Log population settlement	0.045***
	(-0.002)
Constant	3.210***
	(-0.036)
N	56,467

Estimation includes education of the parents, city size, region and year fixed effects. Robust standard errors. Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Marginal Treatment Effect

Table: Selection equation: formal vs informal

Variables	Coef.	S.E.	Z
Female	0.196***	0.0177751	11.08
Age	0.021***	0.00617	3.47
Age squared	-0.014*	0.00785	-1.83
Married (=1)	0.203***	0.01806	11.27
Schooling categories (omited: primary or less)			
High School	0.148***	0.02693	5.5
Technical/Vocational	0.448***	0.03079	14.57
College or more	0.802***	0.03222	24.91
Education of the parents			
Technical/Vocational	0.012	0.02355	0.53
College or more	-0.019	0.02645	-0.74
Missing	-0.115***	0.02481	-4.64
Log population site	-0.014	0.00968	-1.5
Other formal member in HH	0.26***	0.01739	14.97
Log distance inspection	0.104***	0.01776	5.88
Ratio inspectors per 1000 entities	0.065	0.12574	0.52
DistancexRatioInspectors	0.021	0.03321	0.66
Constant	0.017	0.16285	0.1
Wald Test	2252.6		
N	56,467		

Notes: Robust standard errors in parentheses. Missing standard errors for ATE and covariances. Model includes education of the parents, log of population at settlement, city size, region and year dummies. Significance:
$***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

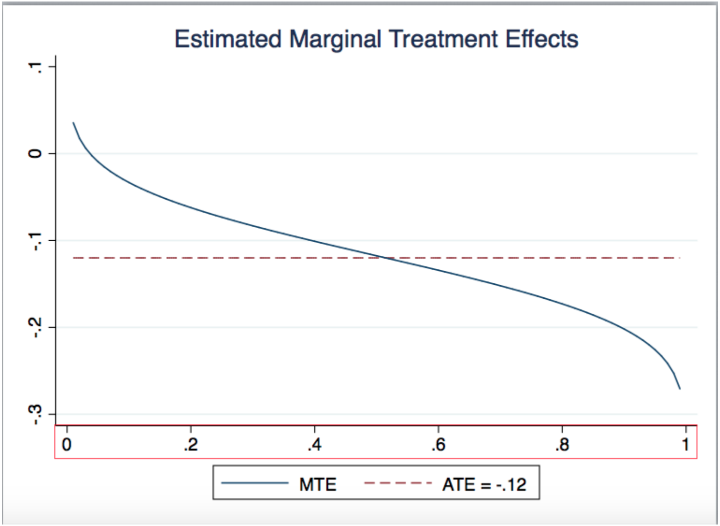
Marginal Treatment Effect

Table: Outcome equation: lnwage

Variables	**Treated [D=1]**	S.E.	Z	**Untreated [D=0]**	S.E.	Z
Female	-0.268***	0.022	-12.42	-0.298***	0.005	-65.7
Age	0.045***	0.006	7.29	0.041***	0.002	25.11
Age squared	-0.065***	0.008	-8.19	-0.053***	0.002	-26.24
Married (=1)	0.111***	0.024	4.72	0.052***	0.005	11.2
Schooling categories (omited: primary or less)						
High School	0.107***	0.027	4	0.094***	0.009	11.05
Technical/Vocational	0.271***	0.042	6.43	0.206***	0.009	22.69
College or more	0.446***	0.061	7.34	0.494***	0.009	53.81
Constant	3.49***	0.175	19.98	3.24***	0.037	88.23
Sigma	-0.187			-0.121		
Sigma1V - Sigma0V	-0.065					
ATE	-0.118					
Number of Observations	56,467					

Notes: Dependent variable: log real hourly wage rate. Robust standard errors in parentheses. Missing standard errors for ATE and covariances. Model includes education of the parents, log of population at settlement, city size, region and year dummies. Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Outcome equation: lnwage



Robustness Checks

Alternative Definition

- Formal employee: Worker at a firm who is officially registered as an employee.
- Informal employee: Worker at a firm who is not officially registered or works not at a firm but for a private person, engages in IEA, or self-employed.

Descriptive statistics

Table: Formal and informal workers sample

Year	Informal	Formal	Total
2009	1,018	4,917	5,783
2010	1,410	7,532	8,780
2011	1,550	7,497	8,829
2012	1,763	7,513	9,025
2013	1,688	7,294	8,736
2014	1,445	6,038	7,299
2015	1,277	5,913	6,986
2016	1,375	5,952	7,061
Total	11,526	52,666	64,192

Robustness Checks

Table: Selection Equation - Formal alternative definition

Variables	Coef.	S.E.	Z
Female	0.182***	(0.013)	14.5
Age	-0.000	(0.004)	-0.09
Age squared	0.010*	(0.006)	1.69
Married (=1)	0.129***	(0.013)	10
Schooling categories (omited: primary or less)			
High School	0.055***	(0.020)	2.76
Technical/Vocational	0.379***	(0.023)	16.81
College or more	0.667***	(0.023)	28.8
Log population site	-0.003	(0.006)	-0.47
Other formal member in HH	0.416***	(0.012)	33.39
Log distance inspection	0.008	(0.012)	0.7
Ratio inspectors per 1000 entities	0.290***	(0.084)	3.45
DistancexRatioInspectors	-0.049**	(0.020)	-2.51
Constant	0.254**	(0.115)	2.2
Wald Test	4,309.52		
N	64,192		

Notes: Dep. Variable: Formal alternative dummy. Robust standard errors in parentheses. Regression includes education of the parents, year and region dummies. Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

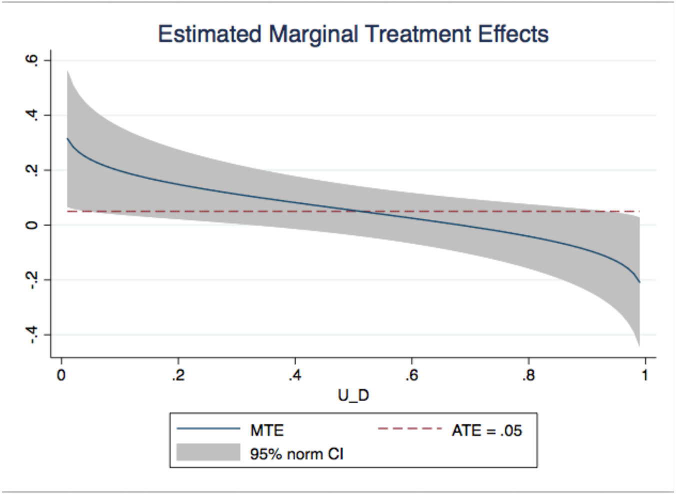
Robustness Checks

Table: Outcome Equation - Formal alternative definition

Variables	**Treated [D=1]**	S.E.	Z	**Untreated [D=0]**	S.E.	Z
Female	-0.30***	(0.005)	-63.43	-0.250***	(0.012)	-20.94334936
Age	0.040***	(0.002)	24.65	0.038***	(0.004)	9.724117789
Age squared	-0.053***	(0.002)	-25.84	-0.054***	(0.005)	-10.77127171
Married (=1)	0.048***	(0.005)	9.68	0.105***	(0.012)	8.72985316
Schooling categories (omited: primary or less)						
High School	0.089***	(0.008)	10.47	0.081***	(0.016)	4.987716416
Technical/Vocational	0.198***	(0.010)	20.33	0.166***	(0.022)	7.686923137
College or more	0.484***	(0.011)	45.00	0.370***	(0.027)	13.89948625
Log population	0.045***	(0.002)	-63.43	0.043***	(0.005)	8.175302193
Constant	3.269***	(0.038)	86.10	3.425***	(0.092)	37.12
Sigma	-0.080***	(0.025)		0.033	(0.032)	
Sigma1V - Sigma0V	-0.113***	(0.040)				
ATE	0.053	(0.046)				
Number of Observations	64,192					

Notes: Dependent variable: log real hourly wage rate. Robust standard errors in parentheses. Model includes education of the parents, log of population at settlement, city size, region and year dummies. Significance:
$***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Outcome equation: lnwage



Robustness Checks

Table: Adding Regional GDP

Variables	Treated [D=1]	Untreated [D=0]
Female=1	-0.2458617	-0.2447534
Age at the time of survey	0.0406371	0.0423161
Age squared	-0.0520392	-0.0593329
Married=1	0.0480674	0.1162418
Schooling level self		
Secondary	0.0924229	0.077824
Upper Vocational	0.2179289	0.1402234
Higher education	0.4976925	0.3366901
Log population	0.0365772	0.0372553
Log Regional GDP	0.3873677	0.2840995
Constant	-1.468756	-0.2562091
Sigma	-0.1265585	0.057045
Sigma1V-Sigma0V	-0.1836035	
ATE	0.1074744	
Number of observations	64,192	

Notes: Dependent variable: log real hourly wage rate. Robust standard errors in parentheses. Model includes education of the parents, log of population at settlement, city size, region and year dummies. Significance:
$***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Issues with Identification

- All the exclusion restrictions proposed by the literature about informality are problematic.
  - Number of labor inspectors and distance: potential double causality with informality. The government can potentially locate more inspectors in areas in which informality is higher or open new inspections office. Potential direct effect on wages: Inspected firms that are caught in a violation may comply with regulation by formalizing informal workers and adjusting wages of all workers and/or firing informal workers, which will have an impact on wages due to general equilibrium effects.
  - Having a formal family member: matching among individuals with similar characteristics.
- Other potential IVs: Bartik-like instruments using shift-share analysis, oil price shocks. Same concerns about direct effect on wages remains.

Summary of results

- Under MTE model only using employees: informal workers exhibit higher wages than formal workers. There is evidence of segmentation: $cov(U_1, V) < 0$ and $cov(U_0, V) < 0$ Potential channel: informal workers may earn higher net wages on average, but formal workers have other non-pecuniary benefits which are not monetary. Employees cannot switch easily between formality and informality due to segmentation.
- Under MTE model using broader informal definition: informal workers exhibit lower wages than formal workers. There is evidence of comparative advantage: $cov(U_1, V) < 0$ and $cov(U_0, V) > 0$ Potential channel: As informality now is very heterogeneous, individuals seem to sort into the types of jobs that they think they are going to get the highest benefits from it.

IPWRA Treatment Effects

- The RA method extends the idea of using sample means to estimate treatment effects by using a regression model to predict potential outcomes adjusted for covariates.
- In the IPW method, for subjects who did receive treatment, the weight is equal to the reciprocal of the predicted probability of treatment. For subjects who did not receive treatment, the weight is equal to the reciprocal of the predicted probability of not receiving treatment. IPW estimators view counterfactual outcomes as missing data and correct the estimators for treated and not treated sample means for the missing data.
- The IPWRA estimator is an RA estimator that uses estimated inverse-probability weights to correct the estimator when the regression function is misspecified. This estimator has a remarkable property: although it requires us to build two models, we only need to specify one of the two models correctly. If the treatment model is misspecified but we correctly specify the outcome model, we still obtain correct estimates of the treatment effect. If we correctly

Results: Accounting for selection - IPWRA

Table: Inverse Probability Weighting Regression Adjustment Treatment Effects

ATE	lnwage	lnearn	lnhhyFa	lnhhyDa	lnhhc
Formal-Informal	0.149***	0.096***	0.074**	0.064**	-0.031*
SE	0.01	0.01	0.04	0.03	0.02
t	12.12	7.74	2.04	2.48	-1.84
N	56,467	54,370	35,953	36,164	36,147

Estimation includes all X's, region and year fixed effects. Robust standard errors. Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.