Modelling the supply of informal work

a microeconometric approach

David Rodríguez Guerrero

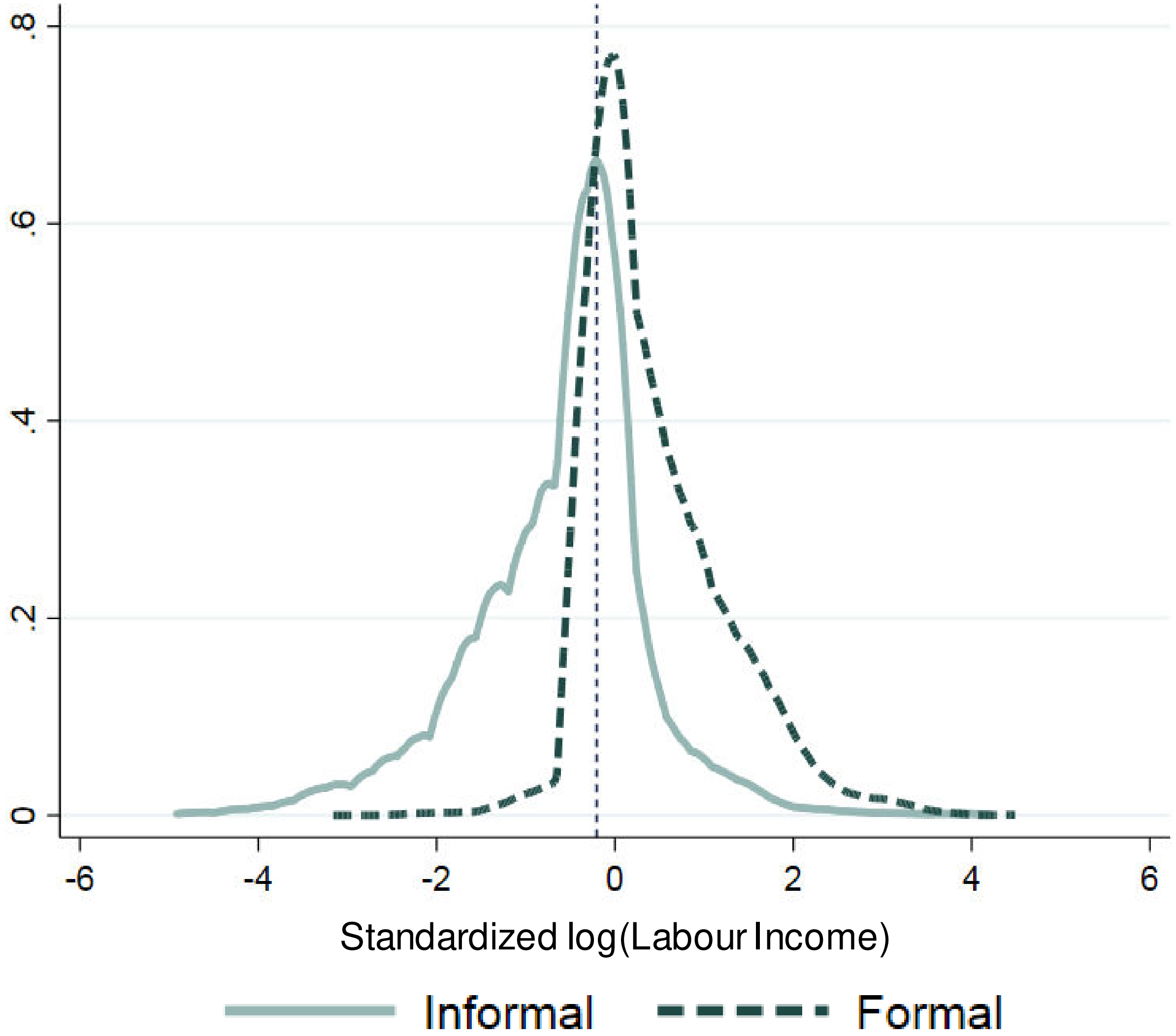
ISER
University of Essex

david.rodriguez@essex.ac.uk

Motivation

- With more than half the workforce in the informal sector in Latin-American economies, understanding the nature of informality is critical
- If markets are segmented (i.e there are significant barriers to become formal) distortions in the formal sector are indeed large and the need for reform is compelling
- If informal workers' dynamics are similar to the formal sector, the focus shifts towards the cost-benefit analysis agents undertake choosing among sectors

Earnings distribution of formal and informal employees (2014)



*minimum wage vertical dotted line Source: Author's calculations

Research Questions

- To what extent barriers to enter the formal sector affect labour supply and sector choice?
- What would be the sectoral choice responses of informal workers to exogenous changes in tax-benefit system rules or education level?

Literature Review

- Informality Choice
  - Macroeconomic and Growth models with two sectors
    - Raush (1991), Loayza (1996), Amaral and Quintin (2006), Loayza and Rigolini (2011), De Paula and Scheinkman (2011), Galiani and Weinschelbaum (2012)
  - Search and Matching Models
    - Zenou (2008), Albrecht, Navarro and Vroman (2009), Bosch and Maloney (2010)
- Labour supply with informality
  - Too few microeconometric studies
  - Pradhan and Van Soest (1997) Labour supply and sector choice of two spouses households in Bolivia (ordered vs multinomial models)
  - Gong, Van Soest and Villagomez (2004) study formality, informality or non-participation choice in Mexico with a dynamic multinomial logit panel data model

Current labour supply models

Following Aaberge and Colombino (2014) modern labour supply models can be classified in two groups:

- Discrete Choice Models (DC)
- Random Utility-Random Opportunity (RURO)

Discrete Choice Model (DC) aka Van Soest model (VS)

- Choice set of hours of worked consists of only k discrete alternatives $L_i \in [0, T] \; \forall \; i = 1,2, \ldots k$
- Stochastic utility function over only one Consumption good (disposable income) and Leisure

$$U_i = V(C_i \, , T - L_i | X) + \varepsilon_i$$

- X Individual measured characteristics such as age, gender, education or number of kids etc., that could directly affect preferences or indirectly affect disposable income (through tax and benefit rules)

- Function $\mathrm{f}(.)$ translates each element in the discrete set of hours worked into disposable income for that choice, by adding benefits and subtracting taxes and social insurance (TB) from labour $(\mathrm{wL_i})$ and non-labour $(\overline{\mathrm{Y}}_\mathrm{i})$ original income

$$\mathrm{Y_i} = \mathrm{C_i} = \mathrm{f}(\mathrm{L_i}|\mathrm{X}) = \mathrm{wL_i} + \overline{\mathrm{Y}}_\mathrm{i} + \mathrm{TB}(\mathrm{wL_i} + \overline{\mathrm{Y}}_\mathrm{i}\,|\mathrm{X})$$

- Calculated with tax-benefit microsimulation models

- Assuming ε_i is an error term with an Extreme value distribution the individual likelihood takes the multinomial logit form (L_i observed choice):

$$p_i(L_i|X) = \frac{e^{V(L_i|X)}}{\sum_{j=1}^{k} e^{V(L_j|X)}}$$

- For non-workers wages are imputed
- For couples a unitary decision household is usually assumed, utility depends on individual labour supply but household disposable income
- The model may require some fixed utility costs of working to improve fit (Van Soest,1995)

Random Utility-Random Opportunity Models (RURO)

- Intuition 1: Wages are not fixed but part of the job offer alongside the time regime.
- Intuition 2: Utilities are weighted with the intensity with which job offers are made available to each agent

- Each job offer is a bundle consisting of working time to be supplied (L_i) and a wage to be paid by the employer (w_i).

- Each non-market activity is assumed to offer a wage $w_i = 0$ and to require no hours of work $L_i = 0$

- The arrival of job offers depends on personal characteristics, labour demand conditions, the wage and time regime the job offer stipulates.

- Arrival modelled by a Poisson process with intensity parameter given by

$$\lambda_1(\varepsilon_i, q)g_1(w_i)g_2(L_i)$$

  - The opportunities function λ_1 captures labour demand conditions given by $q = \exp(\beta_q X_q)$
  - g_1 is the density (lognormal) of jobs paying wages w_i which is assumed to depend on some covariates.
  - g_2 is the density (assumed piecemeal uniform with peaks in the most frequently observed time regimes) of jobs requiring L_i hours of work

- The resulting probability that the chosen time regime L_i and wage w_i are the observed for individual i observed working is given by

$$p_i(w_i, L_i) = \frac{q\, g_1(w_i)\, g_2(L_i)\, e^{V(w_i,L_i)}}{e^{V(0,0)} + \int_{w_j \in \mathbb{W}} \int_{L_k \in \mathbb{H}} q\, g_1(w_j)\, g_2(L_k)\, e^{V(w_j,L_k)}\, dL_k\, dw_j}$$

For non participation

$$p_i(0,0) = \frac{e^{V(0,0)}}{e^{V(0,0)} + \int_{w_j \in \mathbb{W}} \int_{L_k \in \mathbb{H}} q\, g_1(w_j)\, g_2(L_k)\, e^{V(w_j,L_k)}\, dL_k\, dw_j}$$

Where $\mathbb{W}$ and $\mathbb{H}$ are the sets of wage offers and hours supplied respectively which are not observed (Decoster et al., 2016)

Labour supply of informal work

- We extend RURO to take into account sectoral choice assuming that job offer intensities, and wage and hour densities are different for informal and formal workers.
- The resulting individual likelihood is given by

$$p_i(w_i, L_i, z_i) = \frac{q_z g_{1k}(w_i)\, g_{2k}(L_i)\, e^{V(w_i, L_i, z_i)}}{e^{V(0,0,0)} + \sum_{k=0}^{1} \int_{w_j \in \mathbb{W}} \int_{L_k \in \mathbb{H}} q_z\, g_{1z}(w_j)\, g_{2z}(L_k)\, e^{V(w_j, L_k, z_i)}\, dL_k\, dw_j}$$

For non participation

$$p_i(0,0,0) = \frac{e^{V(0,0,0)}}{e^{V(0,0,0)} + \sum_{k=0}^{1} \int_{w_j \in \mathbb{W}} \int_{L_k \in \mathbb{H}} q_z\, g_{1z}(w_j)\, g_{2z}(L_k)\, e^{V(w_j, L_k, z_i)}\, dL_k\, dw_j}$$

To estimate the model we drawn from a priory density functions. The simulated likelihood for the observed time regime L_i and wage w_i is

$$p_i(w_i, L_i, z_i) = \frac{\dfrac{\mathbb{P}(0,0,0)}{\mathbb{P}(w_i, L_i, z_i)} q_z\, g_{1z}(w_i)\, g_{2z}(L_i)\, e^{V(w_i, L_i, z_i)}}{e^{V(0,0,0)} + \sum_{w_j, L_k, z_n \in \mathbb{D}} \dfrac{\mathbb{P}(0,0,0)}{\mathbb{P}(w_j, L_k, z_n)} q_n\, g_{1n}(w_j)\, g_{2n}(L_k)\, e^{V(w_j, L_k, z_n)}}$$

For non participation

$$p_i(w_i, L_i, z_i) = \frac{e^{V(0,0,0)}}{e^{V(0,0,0)} + \sum_{w_j, L_k, z_n \in \mathbb{D}} \dfrac{\mathbb{P}(0,0,0)}{\mathbb{P}(w_j, L_k, z_n)} q_n\, g_{1n}(w_j)\, g_{2n}(L_k)\, e^{V(w_j, L_k, z_n)}}$$

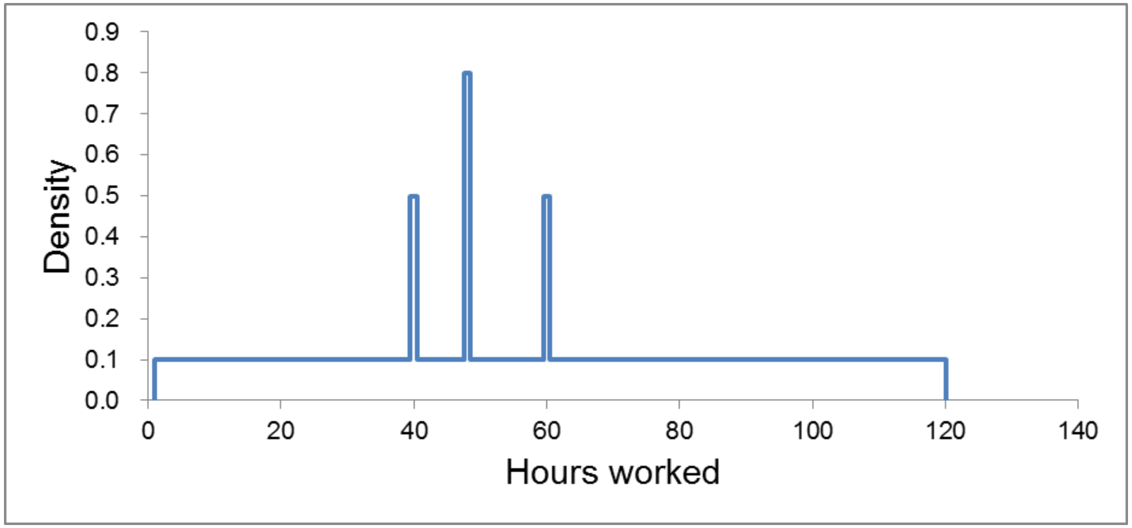
The probability of a job offer being drawn is included in the model $\left(\frac{\mathbb{P}(0,0,0)}{\mathbb{P}(w_j, L_k, z_i)}\right)$ and the observed choice must be included in the $\mathbb{D}$ subset (Train, 2009).

- For the three models the systematic part of the utility function is of the Box-Cox type

$$V(C_i, T - L_i|X) = V(w_i, L_i) = (\beta_H' X_H)\left(\frac{\left(\frac{T-L_i}{T}\right)^{\alpha_H} - 1}{\alpha_H}\right) + (\beta_Y' X_Y)\left(\frac{{Y_i}^{\alpha_Y} - 1}{\alpha_Y}\right)$$

- X_H and X_Y are a vector of parameters that shift the intensity of preference for leisure and income
- α_H and $\alpha_Y < 1$ determine the curvature of the indifference curves
  - lower values imply less substitutability between leisure and income.

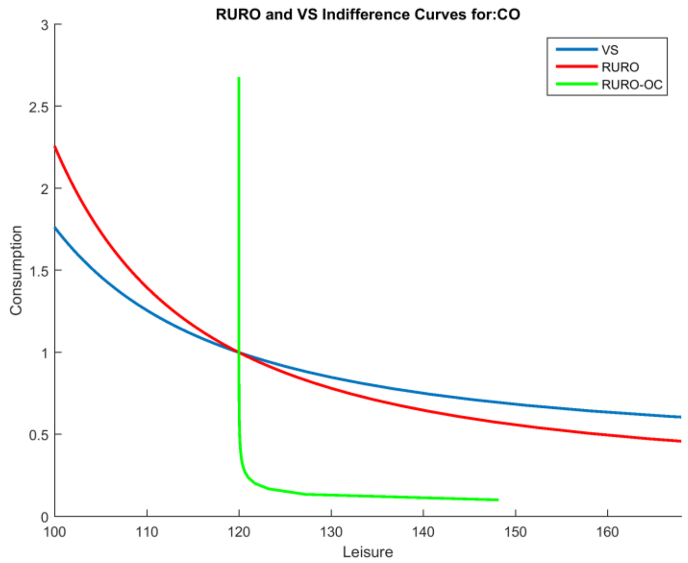
- Hours density piecemeal uniform (Only for RURO type)



Data

- Quality of Life National Survey (ENCV)
  - Income and Expenditure Household Survey for Colombia 2014
  - Cross Section
  - 67.332 observations
- We focus on singles aged 18-60 years, not in education or disabled and in urban areas
  - They must be living without other working age family members
  - They are either working as employees (formal or informal) or not working i.e self-employed are excluded
  - If working, they report only one job
  - 697 observations

Preferences



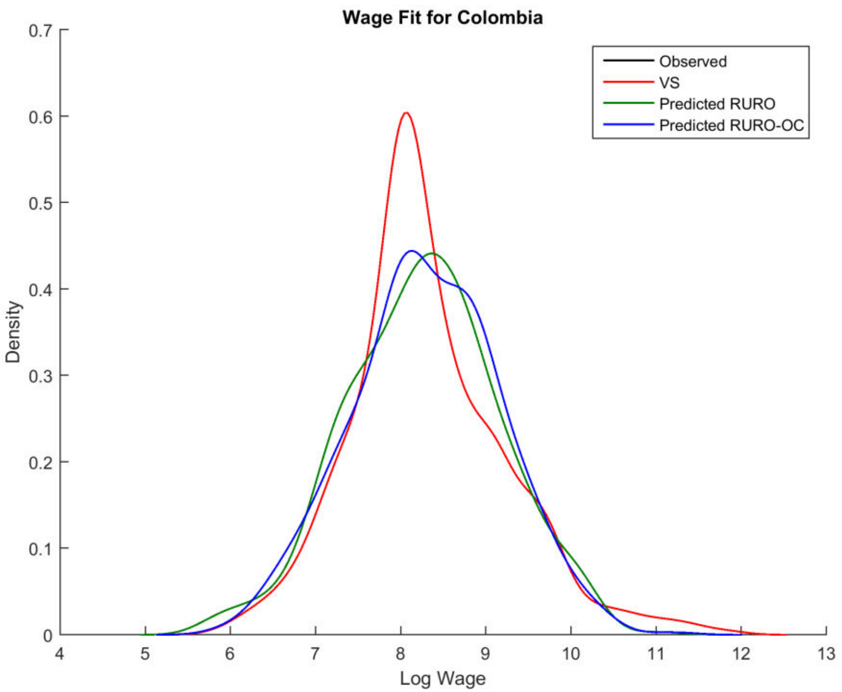
	Variable	VS coef.	VS t-value	RURO coef.	RURO t-value	RURO_OC coef.	RURO_OC t-value
All	leisureXconst	5.49	0.38	4.48	0.27	0.21	0.18
	leisure_male	-0.50	-1.94	-0.68	-1.76	-0.01	-0.47
	leisure_age	-2.56	-0.31	-2.28	-0.24	-0.16	-0.23
	leisure_age2	0.48	0.42	0.46	0.34	0.03	0.33
	leisure_child02	0.23	0.43	0.64	1.03	0.13	0.92
	leisure_child34	-0.31	-0.70	-0.45	-0.91	0.01	0.18
	leisure_child512	0.55	1.67	0.25	0.66	0.09	1.29
	alfa_leisure	-2.17	-6.48	-1.67	-2.92	-5.41	-5.04
	incomeXconst	2.30	10.96	1.07	5.48	0.00	0.21
	alfa_income	-0.36	-6.04	-0.29	-3.36	-3.19	-1.85
	fcXconst	42.94	1.97				
	fcXchild02	0.69	1.43				
	fcXage	-23.33	-1.93				
	fcXage2	3.29	1.96				

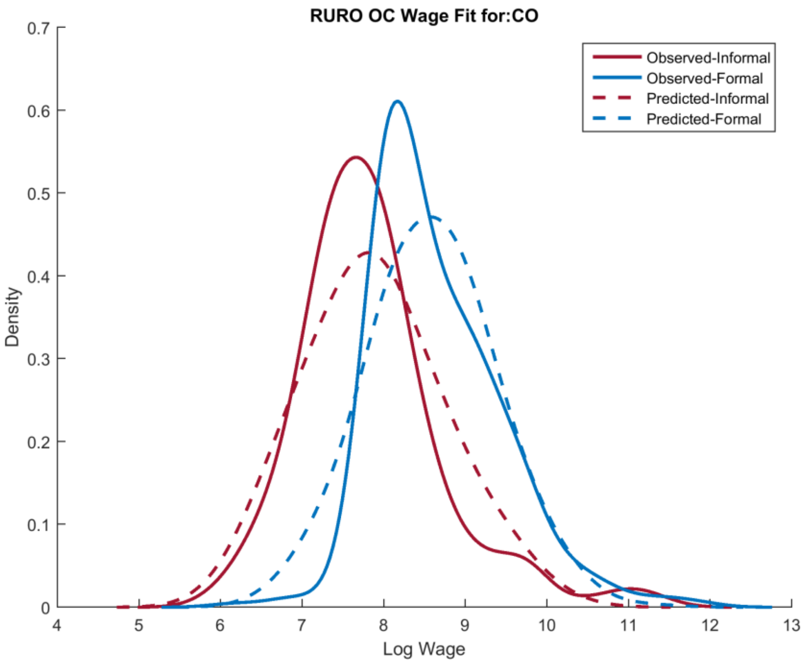
Source: Author’s calculations

Wages

	Variable	VS coef.	VS t-value	RURO coef.	RURO t-value	RURO_OC coef.	RURO_OC t-value
All	w_const			6.71	40.41		
	w_exp			3.17	2.54		
	w_exp2			-5.50	-2.20		
	w_eduhigh			1.01	11.44		
	w_male			-0.02	-0.21		
	w_capital			0.37	3.44		
	w_socsec			0.52	6.14		
	rmse			0.74	26.56		
Informal	w0_const					7.31	36.77
	w0_exp					2.06	1.17
	w0_exp2					-3.44	-1.03
	w0_eduhigh					1.18	6.65
	w0_male					0.25	2.23
	w0_capital					0.57	2.89
	rmse0					0.77	19.52
Formal	w1_const					7.82	45.85
	w1_exp					4.14	2.69
	w1_exp2					-7.51	-2.30
	w1_eduhigh					0.77	9.30
	w1_male					-0.17	-2.08
	w1_capital					0.16	1.52
	rmse1					0.65	23.56

Source: Author’s calculations





Source: Author's calculations

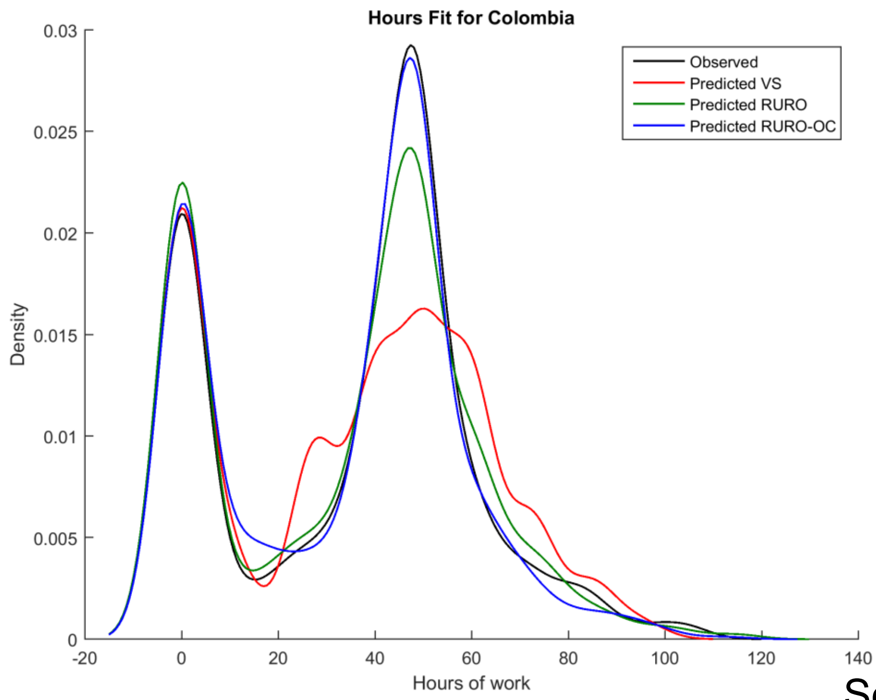
Opportunities

	Variable	VS coef.	VS t-value	RURO coef.	RURO t-value	RURO_OC coef.	RURO_OC t-value
All	opp_const			-24.24	-1.13		
	opp_age			10.98	0.91		
	opp_age2			-1.64	-0.98		
	opp_eduhigh			0.53	0.92		
	opp_formrate			0.89	0.86		
	opp_male			-0.61	-1.56		
Informal	opp0_const					-41.04	-2.48
	opp0_age					21.57	2.32
	opp0_age2					-3.21	-2.47
	opp0_eduhigh					-0.22	-0.71
	opp0_male					0.69	2.56
Formal	opp1_const					-84.70	-4.78
	opp1_age					43.43	4.40
	opp1_age2					-6.06	-4.43
	opp1_eduhigh					0.17	0.47
	opp1_form					4.68	6.88
	opp1_male					1.02	3.95

Source: Author's calculations

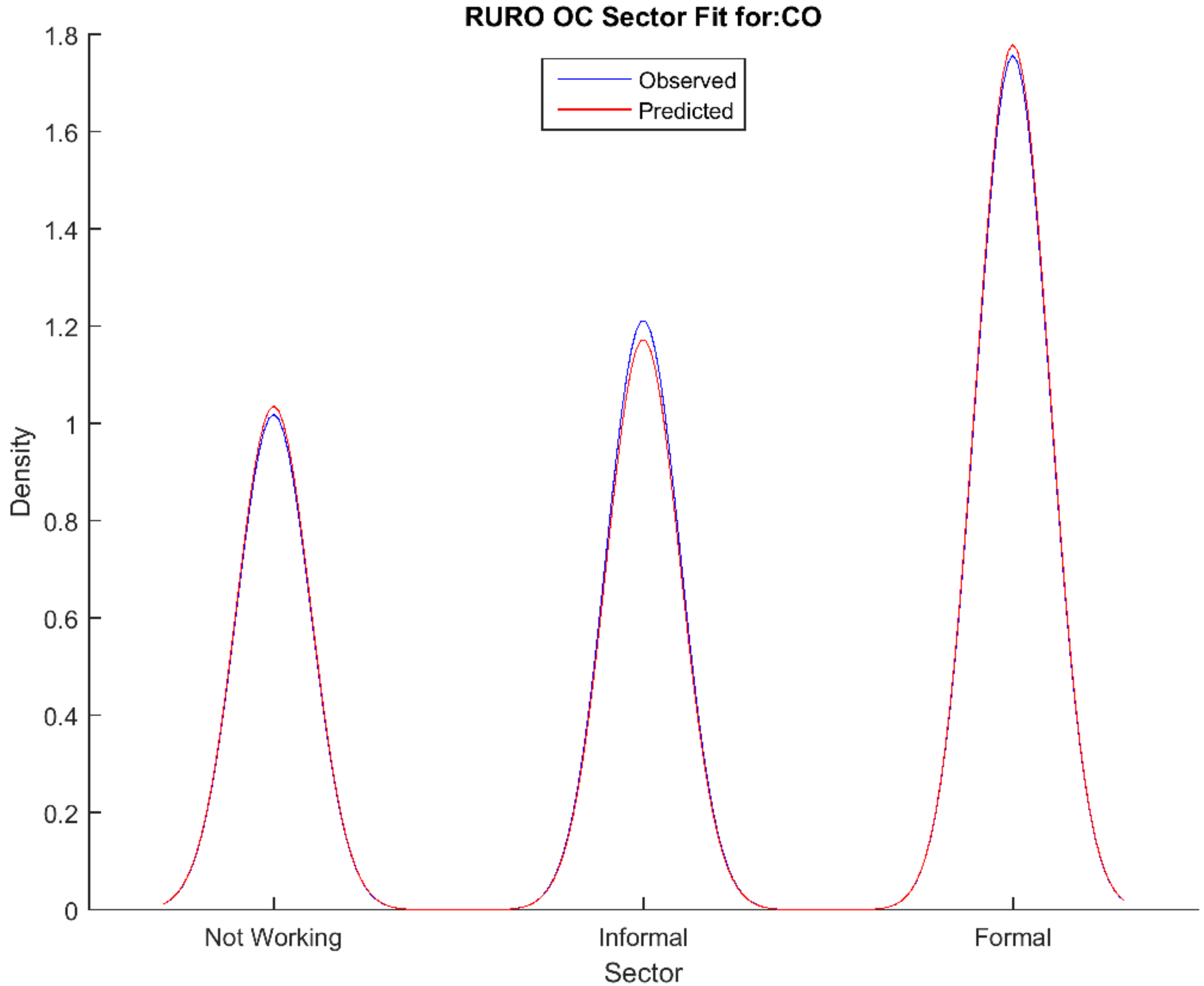
Hours

	Variable	VS		RURO		RURO_OC	
		coef.	t-value	coef.	t-value	coef.	t-value
All	hours_peak1 (40 h/w)			2.64	16.65		
	hours_peak2 (48 h/w)			3.65	31.72		
	hours_peak3 (60 h/w)			2.01	9.48		
Informal	hours0_peak1 (40 h/w)					2.16	7.96
	hours0_peak2 (48 h/w)					3.15	16.23
	hours0_peak3 (60 h/w)					2.02	6.66
Formal	hours1_peak1 (40 h/w)					3.33	17.01
	hours1_peak2 (48 h/w)					4.60	31.93
	hours1_peak3 (60 h/w)					2.48	8.76



Source: Author’s calculations

Sector Choice

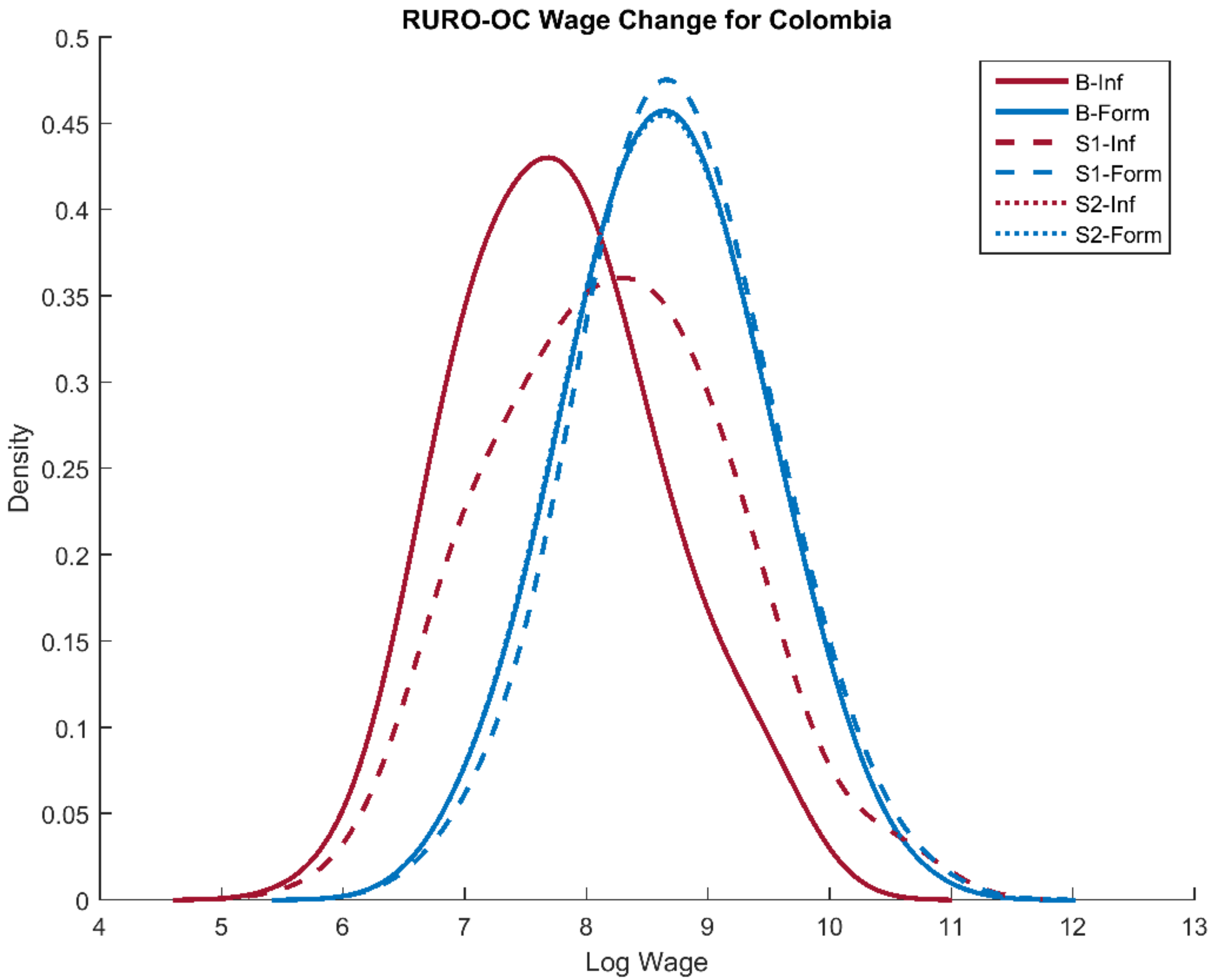


Source: Author's calculations

Simulation

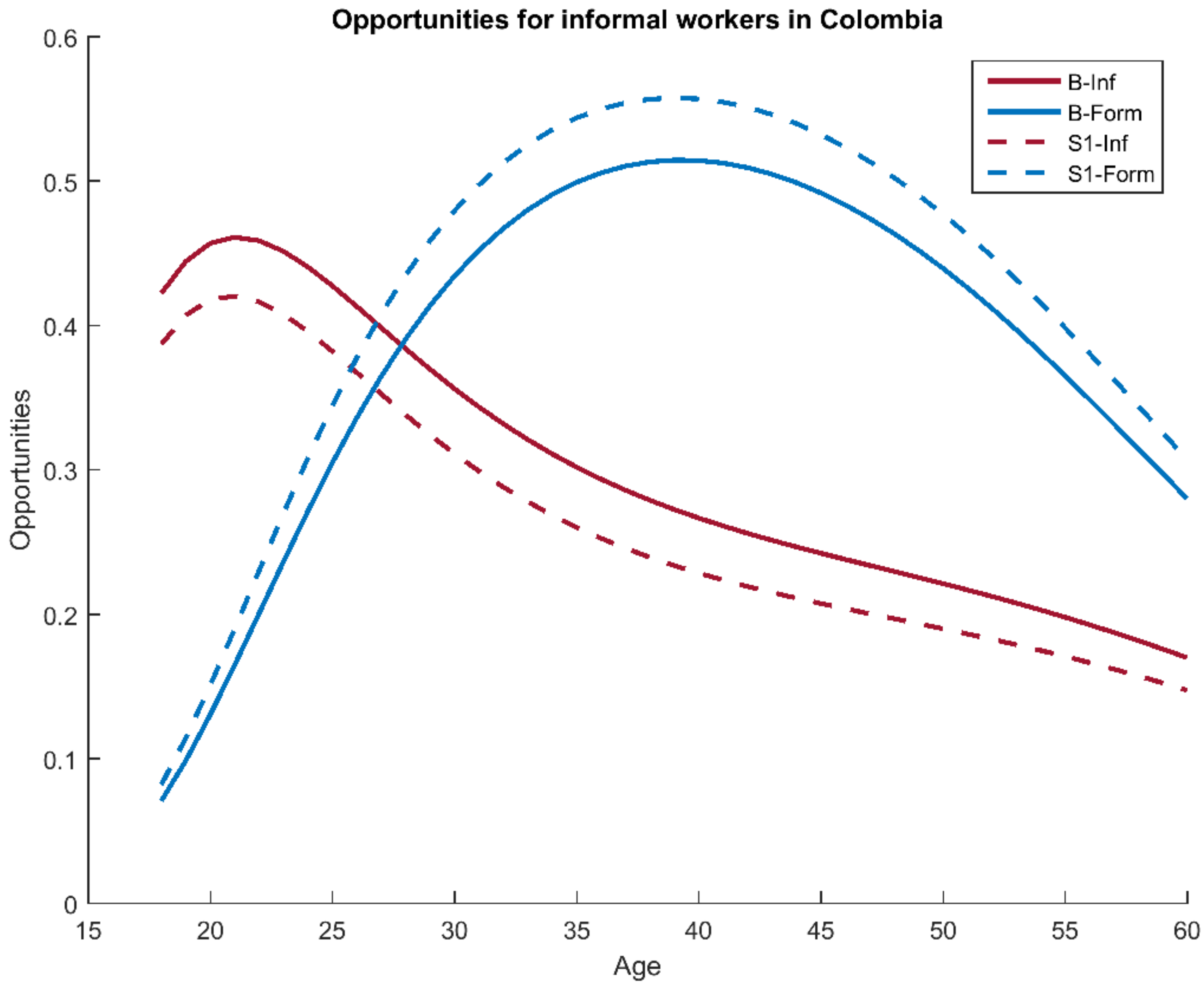
- Simulation1: We increase the observed education of informal workers aged 18-40 years to the highest level resulting in informal workers with HE 10% →70%
  - Increases opportunities, wages and incomes

- Simulation2: We eliminate social insurance contributions for formal workers earning less than 3 monthly minimum wages
  - Increases disposable incomes in the formal sector

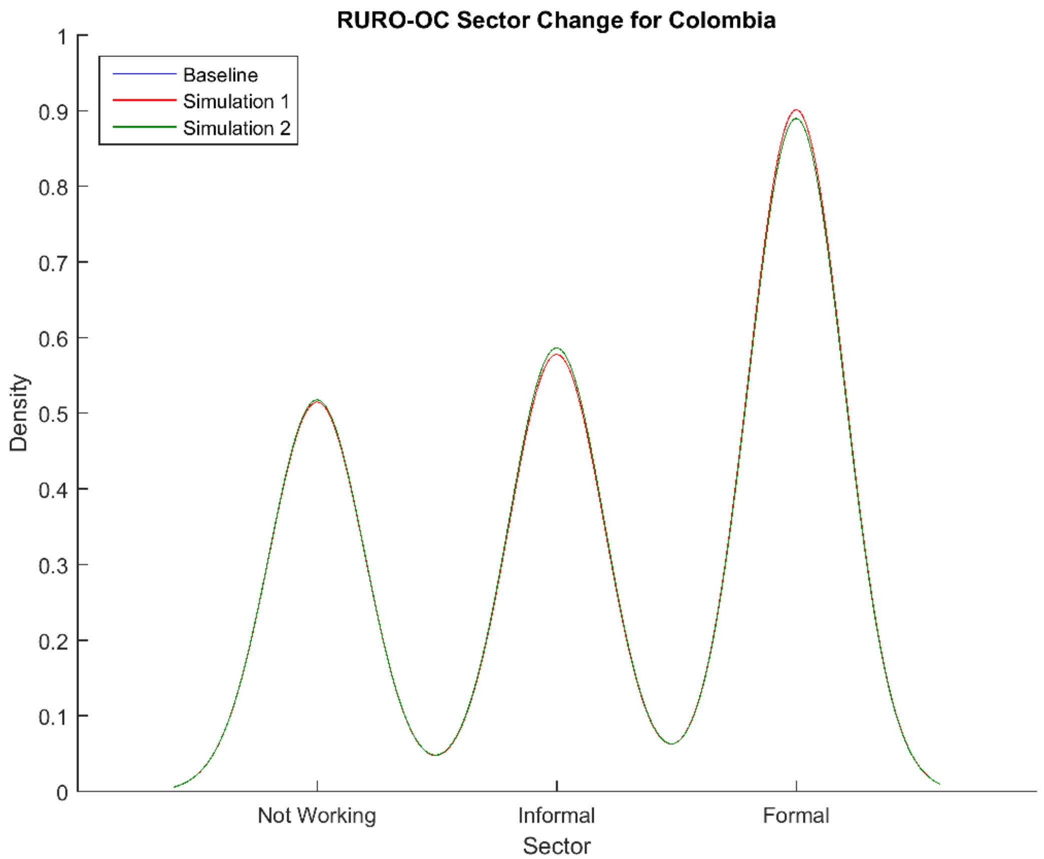


Source: Author's calculations

- Opportunities B and S1



Source: Author's calculations



	Simulation 1				Simulation 2			
		Not Working	Informal	Formal		Not Working	Informal	Formal
Baseline	Not Working	0.994	0.000	0.006	OLS	1.000	0.000	0.000
	Informal	0.000	0.961	0.039	Informal	0.000	1.000	0.000
	Formal	0.000	0.016	0.984	Formal	0.000	0.000	1.000

Source: Author’s calculations

Conclusions

- Very limited sector movement after important exogenous changes. This result favours the hypothesis of labour market segmentation
- Unobservables determine a great deal of opportunities for informal workers in the formal sector

Next steps

- Explore alternative determinants of opportunities
- Explore alternative dataset with more observations
- Calibrate a micro-founded macroeconomic model to better capture the interactions between supply and demand of formal and informal labour