the review of income and wealth

Review of Income and Wealth
Series 57, Special Issue, May 2011

EARNINGS STRUCTURES, INFORMAL EMPLOYMENT, AND SELF-EMPLOYMENT: NEW EVIDENCE FROM BRAZIL, MEXICO, AND SOUTH AFRICA

BY OLIVIER BARGAIN*

UCD–Geary Institute, IZA, ESRI, CHILD, CEPS–INSTEAD

AND

PRUDENCE KWENDA

UCD–Geary Institute

We estimate the conditional earnings gap between formal and informal sectors, distinguishing between salary and self-employed workers. Rich panel datasets for Brazil, Mexico, and South Africa are assembled to define informality in a comparable way and to control for (time-invariant) unobserved heterogeneity. Estimations are conducted at different points of the conditional earnings distributions. Interesting results emerge. First, informal salary workers are systematically underpaid compared to their formal sector counterparts, in all countries and at almost all conditional quantiles. Yet penalties are very moderate in Brazil and Mexico while more substantial in South Africa, a country where legal advantages in formal employment are effective. Second, informal self-employment contributes to a more dispersed earnings distribution in all three countries. International comparisons reveal a continuum of situations reflecting historical and legal differences across countries, from very large self-employment penalties in South Africa to significant conditional earnings premia in Mexico.

JEL Codes: J24, J31, O17

Keywords: self-employed, informal sector, conditional earnings differential, quartile regression, fixed effects model

1. INTRODUCTION

Recent evidence indicates that the informal sector is likely a long-term feature of emerging and developing economies, particularly in Africa and Latin America (Charmes, 2000).[1] Given that informality represents a large share of the working force, understanding its workings is essential to comprehending labor markets and income distribution in these countries. The main difficulty is perhaps the huge heterogeneity characterizing this sector and therefore the difficulty to describe its functioning and its role. According to the traditional view (Fields, 1975; Dickens and Lang, 1985), salary workers enter informality to escape unemployment or

Notes: We are grateful to Derek Blades, François Bourguignon, Eliane El Badaoui, Francisco Ferreira, Bill Maloney, Blaise Melly, Martin Rama, Arthur van Soest, Elizabeth Villagomez, and participants/discussants of the 2009 RIW conference, workshops at IZA and Brunel University, and seminars at CEPS–INSTEAD and UCD for useful advice. The usual disclaimers apply.

*Correspondence to: Olivier Bargain, University College Dublin, School of Economics, Newman Building, Belfield Campus, Dublin 4, Ireland (olivier.bargain@ucd.ie).

[1]Evidence on informality is summarized in Leontaridi (1998), Perry *et al*. (2007), Jütting *et al*. (2007), and Ruffer and Knight (2007), among others.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA, 02148, USA.

because they are rationed out of the formal sector as a result of labor market regulation.[2] Yet some authors have recently questioned this paradigm, arguing that an important fraction of informal jobs may reflect the voluntary choice of workers given their preferences, skill endowments, and competing earnings prospects (e.g. Maloney, 1999). Even if recent labor market modeling suggests adopting a dual representation of the informal sector whereby a competitive/voluntary entry segment coexists with a rationed/segmented group (Funkhouser, 1997; Blunch *et al.*, 2001; Fields, 2005), empirical studies are necessary to grasp the much broader variety of situations that exist both within a country and across country-specific labor market experiences and legal contexts. Arguably, data constraints have limited the number of comprehensive empirical studies where informality can be defined and studied in a comparable way across countries and where individual heterogeneity can be exploited.

One of the important aspects is how the presence of informal sectors affects the earnings structure in these countries. According to the traditional view, informal salary workers earn less than *identical* workers in the formal sector because of legal reasons (minimum wages, higher unionization) that may push up formal sector wages above market-clearing levels.[3] More recent studies nonetheless show that informal–formal pay differentials may not be negative or may not be explained solely by regulatory environments. For instance, higher remunerations can be found for self-employed workers in some Latin American countries (e.g. Yamada, 1996; Maloney, 1999, 2004; Saavedra and Chong, 1999). Several challenges exist in estimating the existence of informal earnings penalty/premium (after controlling for workers' characteristics), and notably the necessity to incorporate unobserved heterogeneity and to depart from mean estimations that may not characterize the returns of the majority of informal sector workers if a handful of prominent entrepreneurs push up the average earnings (see Hamilton, 2000). Some authors have attempted to deal with the unobserved characteristics that affect both the selection into a particular sector and earnings levels, either by explicitly introducing selection equations (e.g. Carneiro and Henley, 2001; Gong and van Soest, 2002) or by purging estimations from unobservables by using fixed effects models (e.g. Badaoui *et al.*, 2008). Others have used quantile regressions to account for the variety of situations along the earnings distribution that may be concealed by comparisons at the mean (e.g. Tannuri-Pianto and Pianto, 2008).

The present paper suggests an attempt to capture the diversity of the informal sector by addressing these two issues in a comprehensive way and in an international perspective. We estimate the conditional earnings penalties/premia carried by informal sector workers, carefully distinguishing between informal self-employed and informal salary workers.[4] We exploit large (rotating) panels to

[2]In a similar way, informal self-employment is often seen as a means of overcoming economic: hardships in developing countries (Leibenstein, 1968).

[3]Similarly, some authors report lower average earnings in informal self-employment compared to paid (formal) employment (e.g. Aronson, 1991; Carrington *et al.*, 1996; Sullivan and Smeeding, 1997).

[4]Like ours, some studies consider in turn the comparison between formal and informal salary workers and between the latter and self-employed workers (e.g. Arias and Khamis, 2008; Bosch and Maloney, 2010). Some other studies focus exclusively on formal versus informal salaried work (e.g. Badaoui *et al.*, 2008; Bargain and Kwenda, 2009). At the other extreme, some studies approximate informality by self-employment (e.g. Yamada, 1996, for Peru).

account for workers' unobserved heterogeneity by estimating fixed effects regressions at the mean but also at different points of the conditional earnings distribution. This approach allows us to unveil more complex patterns that prove useful for characterizing the different segments of the labor market. We focus on three countries which have received a considerable amount of attention in the literature on informality, namely Brazil, South Africa, and Mexico. One of the challenges of international comparative work is to define informality in the most comparable way across countries. In our view the present contribution to the current literature thus derives from both a methodological perspective and data quality. Cross-country comparisons reveal interesting regularities concerning the relative earnings position of informal salary workers or the dispersion effect of self-employment on the earnings distribution. They also point to contrasted situations especially regarding the relative position of self-employed workers in the different countries.

We can summarize the main findings as follows. First, informal salary workers are systematically underpaid compared to their formal sector counterparts, in all countries and at almost all conditional quantiles. Penalties are larger at lower conditional quantiles—those who do badly conditional on their observed characteristics do especially poorly in informal salary work. Yet penalties obtained after controlling for unobserved heterogeneity are extremely modest in Latin America. They are more substantial in South Africa, a country where legal advantages in formal employment are effective (in particular unionization and minimum wage). Second, the comparison between self-employment and formal salary work shows a continuum of situations reflecting historical and legal differences across countries. The two polar cases are South Africa, with a largely penalized group of self-employed workers, and Mexico where very significant conditional earnings premia appear for half of the self-employed. Brazil is an intermediary case with moderate premia (penalties) at the top (bottom) of the conditional earnings distribution.

The paper is organized as follows. Section 2 briefly describes the data and discusses the nature of informality in the three countries under study. The econometric approach is detailed in Section 3. In Section 4 we discuss and interpret the results. Section 5 reports robustness checks and Section 6 concludes.

2. Measuring Individual Earnings in All Three Sectors

2.1. *Data, Selection, Definition*

For Brazil, we make use of the Monthly Employment Survey (Pesquisa Mensal de Emprego, PME) conducted by the *Instituto Brasileiro de Geografiae Estatistica* (IBGE). This is a monthly household survey on the six largest metropolitan areas of Brazil (i.e. Belo Horizonte, Porto Alegre, Recife, Rio de Janeiro, Salvador, and Sao Paulo). Households are interviewed four months in a row and re-interviewed eight months later for another four months. We create a panel with observations that are a year apart, focusing on the years 2002 to 2007. For South Africa, we use the Labor Force Survey (LFS), a bi-annual rotating panel conducted by Statistics South Africa (Stats SA) and covering all provincial areas.

Twenty percent of the sampling units are rotated out of the survey and replaced with a new sample every six months; workers are therefore observed five times at most over a two-and-a-half year period. We use the waves of September 2001 to March 2007. For Mexico, we use the Mexican National Occupation and Employment Survey (ENOE) conducted by the *Instituto Nacional de Estadistica, Geographica e Informatica* (INEGI). This is a quarterly survey where workers are observed at most five times over a five-quarter period. We use data from the first quarter 2005 to the third quarter 2008.

These surveys provide information about job characteristics, incomes, work duration, demographics, and education. Since households are identified over time but individuals are not, we construct panels of individual workers by linking persons within households over time on the basis of gender, race, and age. For the baseline estimates, we select workers that are observed at least twice in the data. The attrition resulting from this procedure corresponds to 30 percent of the initial sample for Brazil, 19 percent for South Africa, and 17 percent for Mexico. In the last section, we check for possible non-random attrition that could bias our results. We restrict our sample to urban men aged 15 to 65 years, not engaged in any form of education, and in full time employment in the private sector. We focus on men because in all three countries a large proportion of women are not active or are engaged in unpaid work—accounting for selection into the labor market is not yet standard in quantile estimations (see Albrecht *et al.*, 2009). We select only workers in the private sector, which excludes unpaid family workers (whose implicit earnings are difficult to evaluate) and public sector employees; for the latter, there are indeed important differences in institutional mechanisms regulating wages, both across countries and compared to the private sector.

We opt for the legalistic/social protectionist definition of informality which refers to the lack/avoidance of formal registration, taxation, and labor standards and the lack of social security protection. This is a broad concept of informality as it recognizes the possible presence of unregistered/unprotected workers in large firms (Perry *et al.*, 2007).[5] Specifically, the group of informal salaried workers is identified on the basis of lack of compliance with labor legislation, which is relatively straightforward to capture in the surveys in use. In Mexico employees have to contribute to the social security agency (IMSS). Similarly, employees in Brazil must hold a labor card (*carteira assinada*), the signing of which guarantees them access to formal labor protection. Therefore those wage employees not registered with the social security agency in Mexico or not holding a signed labor card in Brazil are considered as informal salaried (similar definitions are used for instance in Amuedo-Dorantes, 2004; Tannuri-Pianto and Pianto, 2008; Bosch and Maloney, 2008, 2010). For South Africa, the LFS contains several questions regarding fringe benefits and other aspects of the job that can be used to identify

[5]ILO traditionally recommends classifying informal as workers in small establishments of fewer than 5–10 employees, who tend to be informal along different dimensions. Henley *et al.* (2006), Perry *et al.* (2007), and Bosch and Maloney (2008) show there is substantial overlap in these definitions. We have checked this for the datasets in use and find that it is broadly the case, except for Brazil where we found many informal (unregistered) salaried workers in large firms—see Bargain and Kwenda (2009) for more details.

the sector, in particular questions regarding whether the firm provides medical aid and deducts unemployment insurance contributions (see also Badaoui *et al.*, 2008).

The group of self-employed workers also belongs to a large extent to the informal sector as defined above or as characterized by the LIO. For Brazil, the self-employed do have the legal obligation to pay social security contributions; Henley *et al.* (2006) report that around 95 percent do not do so. A relatively small group of self-employed in Mexico (less than 6 percent of all self-employed in our survey) satisfy the INISS registration although are under no legal obligation to do so. They would be counted as formal sector self-employment, a category which we do not include, and hence is dropped from our selection. In South Africa, although self-employed workers can make contributions to social security, we find that only 3 percent do so. The data allows us to identify those owners of a registered firm and who pay taxes (19.5 percent of all self-employed). Those who are registered/pay taxes or make social security contributions are excluded from our sample. As a further check, we find that very few self-employed workers own firms of more than five employees (15 percent in Brazil, less than 5 percent in Mexico, and 10.5 percent in South Africa). Note that we also check the validity of the self-reported employment state with relevant information for each country.[6]

This selection leaves an unbalanced panel of 22,186 men with 44,372 observations in Brazil, 9,237 men with 22,757 observations in South Africa, and 107, 465 men with 363,911 observations in Mexico. Summary statistics are reported in Table 1 and discussed below. We categorize workers in one of the three states, namely self-employed, formal, and informal paid work. We construct a measure of *hourly earnings* for all workers, using monthly gross earnings and reported work hours in the primary job. For the self-employed, information on monthly earnings does not allow distinguishing between returns to labor and to capital—we discuss this point in the concluding section. Earnings are adjusted over time using the national consumer price indices provided by the IBGE, Stats SA, and the Central Bank of Mexico.

2.2. *Informality in Brazil, South Africa, and Mexico*

Self-employment accounts for 34 percent of total employment in Brazil, 10 percent in South Africa, and 26 percent in Mexico. Informal salary work accounts for 12 percent of total employment in Brazil, 11 percent in South Africa, and 33 percent in Mexico. These proportions are in line with existing evidence, which shows that whatever the definition, informal employment is significant in all three countries (for instance, see Marcouiller *et al.*, 1997, for Mexico; or Carneiro and

[6]For instance, holding a working permit in Brazil should only apply to those in paid employment. We drop the few self-employed workers who declare having such signed labor cards (3.4 percent of them), interpreting it as an indication of misclassification. For South Africa, we use a question on whether a worker runs his/her own business. The marginal fraction declaring that they do not have their own enterprise is excluded from the sample. Data also allows distinguishing between firm owners and own account workers. Hence, we verify that the latter are consistently located in firms of size equal to one (errors are of an order less than 1 percent).

TABLE 1

Selected Samples: Descriptive Statistics

	Brazil				South Africa				Mexico		
	Formal Salaried	Informal Salaried	Self-Employed		Formal Salaried	Informal Salaried	Self-Employed		Formal Salaried	Informal Salaried	Self-Employed
Hourly gross earnings	4.79	3.79	6.39		2.54	1.06	1.79		2.78	2.02	3.82
Std dev.	(6.97)	(5.39)	(10.76)		(3.69)	(1.71)	(3.43)		(2.32)	(2.02)	(5.53)
Demographics											
Age	36.6	36.6	43.3		38.5	39.2	41.6		34.6	32.3	41.8
Household size	3.8	3.9	3.7		5.9	6.4	7.7		4.6	4.9	4.5
# children (0–10 years)	3.2	3.9	3.2		1.7	2.2	2.7		1.8	1.6	2.6
Married	0.64	0.56	0.73		0.63	0.49	0.66		0.62	0.45	0.66
Black	0.07	0.07	0.06	Black	0.74	0.86	0.94				
Brown	0.32	0.31	0.27	Colored	0.26	0.14	0.06				
White	0.61	0.61	0.67								
Education											
No schooling	0.01	0.02	0.02	No schooling	0.09	0.15	0.09	No education	0.02	0.05	0.05
1–3 years	0.04	0.06	0.06	Primary	0.31	0.40	0.33	1–3 years	0.04	0.08	0.10
4–7 years	0.24	0.31	0.29	Secondary	0.53	0.42	0.52	4–7 years	0.24	0.34	0.34
8–10 years	0.18	0.19	0.16	Vocational	0.07	0.03	0.05	8–10 years	0.45	0.40	0.32
11+ years	0.53	0.43	0.48	University	0.01	0.00	0.01	11+ years	0.25	0.13	0.19
Region											
Recife	0.06	0.05	0.05	Western Cape	0.21	0.10	0.06	>100,000 Inhab.	0.72	0.56	0.64
Salvador	0.07	0.06	0.07	Eastern Cape	0.09	0.16	0.12	15,000–99,999 Inhab.	0.11	0.16	0.15
Belo Horizonte	0.16	0.12	0.16	Northern Cape	0.08	0.05	0.02	2,500–14,999 Inhab.	0.08	0.14	0.12
Rio de Janeiro	0.27	0.32	0.31	Free State	0.11	0.07	0.09	<2,500 Inhab.	0.08	0.13	0.10
Sao Paulo	0.25	0.29	0.24	Kwazulu-Natal	0.11	0.15	0.16				
Porto Alegre	0.19	0.16	0.17	North West	0.11	0.13	0.10				
				Gauteng	0.12	0.12	0.17				
				Mpumalanga	0.10	0.10	0.15				
				Limpopo	0.05	0.12	0.14				
Economic sector											
Manufacturing	0.31	0.18	0.12		0.35	0.08	0.10		0.36	0.17	0.14
Construction	0.07	0.15	0.21		0.12	0.26	0.20		0.10	0.34	0.25
Trade & retail	0.23	0.29	0.35		0.28	0.18	0.52		0.24	0.14	0.23
Services	0.24	0.23	0.19		0.17	0.33	0.10		0.19	0.20	0.27
Transport and comm	0.13	0.11	0.11		0.07	0.14	0.09		0.09	0.15	0.11
# panel observations	24,105	5,200	15,067		18,102	2,444	2,211		151,370	118,357	94,184
Share of workers	54%	12%	34%		79%	11%	10%		42%	33%	26%

Notes: Statistics concern the selected sample of males aged 15–65, neither in education nor in the public sector. Data covers the period 2002–07 for Brazil, 2001–07 for South Africa, and 2005–08 for Mexico. Hourly earnings in 2002 PPP international $.

Henley, 2001, for Brazil). The large share of informal salary work in Brazil and a growing self-employment are often viewed to be a result of a stringent labor legislation (Barros and Corseuil, 2004), the enforcement of these regulations (Almeida and Carneiro, 2009), and a series of economic crises (Moro *et al*., 2003). The first two aspects have been particularly reinforced following the 1988 constitutional changes which increased the degree of workers' protection and hence labor costs for firms (Barros and Corseuil, 2004; Bosch *et al*., 2007). In South Africa, informality coexists with the presence of classic unemployment. According to Kingdon and Knight (2007), the overall proportions of informal employment and unemployment are estimated to be 24 and 29 percent, respectively, of the labor force in 2003. One of the reasons for the relatively smaller informal salary work sector in this country pertains to higher reservation wages compared to lower income countries. Indeed the unemployed who receive some support from within or beyond the household (social grants) may prefer to remain outside the low-tier informal sector where real income is very low (Kingdon and Knight, 2004). Informal employees alone account for 11 percent of total salary work according to Badaoui *et al*. (2008), which is similar to our findings for that category.

2.3. *Sample Description and Earnings Dispersion*

Table 1 presents the descriptive statistics for Brazil, South Africa, and Mexico and Table 2 completes the description of the samples by reporting the estimates of a multinomial logit of the workers' states, either formal salary workers (the reference group), informal salary workers, or the informal self-employed. Other things being equal, self-employed workers are substantially older in Latin America and informal paid workers tend to be younger than all other workers in all three countries, which is consistent with the view that informal salaried work acts as an entry point into the labor market. Another standard result is that highly educated workers are more concentrated in formal employment and, to a lesser extent, in informal self-employment.

In line with previous studies, we observe that self-employed workers earn more on average than wage earners in Latin America but earn less than formal salaried workers in South Africa. To depart from simple mean values, we provide a few representations of the unconditional earnings distribution in Figures 1 and 2. Figure 1 shows earnings density functions for each sector. It appears that inequality is highest among the self-employed and that this group is over-represented among top earners in Mexico and, to a lesser extent, in Brazil. We also see that informal salary workers are highly concentrated at low earnings levels. In Mexico and Brazil, the earnings distribution of formal salary workers lies somewhere in-between. Since we are interested in earnings differentials between sectors, Figure 2 represents the difference in mean earnings between each decile of the informal sector (salary workers or self-employed) and the corresponding decile of the formal sector. Wages in the informal salary sector are systematically lower than in the formal sector while self-employment earnings dominate formal sector wages for a large part of the unconditional distribution in Brazil and Mexico.

TABLE 2

Selected Samples: Multinomial Logit of Workers' Status

	Brazil					South Africa					Mexico			
	Informal Salaried		Self-Employed			Informal Salaried		Self-Employed			Informal Salaried		Self-Employed	
Demographics	*Ref: white, single*					*Ref: black, single*					*Ref: single*			
Age	−0.141	(0.012)	0.121	(0.010)		−0.072	(0.016)	−0.021	(0.021)		−0.106	(0.004)	0.138	(0.005)
Age squared	0.002	(0.000)	−0.001	(0.000)		0.001	(0.000)	0.001	(0.000)		0.001	(0.000)	−0.001	(0.000)
Married	−0.340	(0.045)	−0.044	(0.036)		−0.510	(0.071)	−0.114	(0.083)		−0.524	(0.016)	−0.260	(0.019)
Household size	0.068	(0.024)	0.008	(0.020)		−0.049	(0.014)	−0.062	(0.014)		0.035	(0.003)	−0.014	(0.004)
# of children (0–10 years)	0.063	(0.027)	−0.080	(0.022)		0.156	(0.027)	0.164	(0.029)		−0.001	(0.004)	0.027	(0.005)
Black	−0.170	(0.082)	−0.445	(0.066)	Colored	−0.291	(0.102)	−0.857	(0.142)					
Brown	−0.105	(0.049)	−0.334	(0.040)										
Education	*Ref: no schooling*					*Ref: no schooling*					*Ref: no schooling*			
1–3 years	0.142	(0.172)	0.006	(0.138)	Primary	−0.366	(0.093)	−0.168	(0.139)	1–3 Years	−0.298	(0.050)	−0.175	(0.056)
4–7 years	0.164	(0.154)	0.091	(0.122)	Secondary	−1.162	(0.101)	−0.397	(0.146)	4–7 Years	−0.574	(0.044)	−0.378	(0.049)
8–10 years	0.033	(0.158)	0.051	(0.125)	Vocational	−1.769	(0.175)	−0.683	(0.201)	8–10 Years	−0.995	(0.044)	−0.662	(0.050)
11+ years	−0.280	(0.156)	0.153	(0.122)	University	−2.325	(1.003)	−0.009	(0.693)	11+ Years	−1.330	(0.046)	−0.592	(0.051)
Economic sector	*Ref: construction*					*Ref: construction*					*Ref: construction*			
Manufacturing	−1.212	(0.068)	−2.121	(0.057)		−2.070	(0.111)	−1.782	(0.122)		−1.793	(0.020)	−1.698	(0.025)
Trade & retail	−0.626	(0.068)	−1.417	(0.055)		0.184	(0.107)	−0.260	(0.138)		−0.204	(0.025)	−0.655	(0.030)
Services	−0.771	(0.076)	−1.313	(0.061)		−0.027	(0.089)	−1.086	(0.134)		−1.438	(0.024)	−1.022	(0.029)
Transport and comm	−0.437	(0.064)	−0.674	(0.051)		−1.136	(0.093)	0.119	(0.095)		−0.495	(0.025)	0.083	(0.028)
Region	*Ref: Recife*					*Ref: Western Cape*					*Ref: >100,000 Inhab.*			
Salvador	0.007	(0.117)	0.051	(0.090)	Eastern. Cape	1.040	(0.121)	0.971	(0.171)	15,000–99,999 Inhab.	0.608	(0.021)	0.429	(0.024)
Belo Horizonte	−0.086	(0.100)	0.050	(0.078)	Northern Cape	0.147	(0.147)	−0.145	(0.257)	2,500–14,999 Inhab.	0.789	(0.023)	0.624	(0.027)
Rio de Janeiro	0.286	(0.095)	0.039	(0.074)	Free State	0.206	(0.146)	0.722	(0.189)	<2,500 Inhab.	0.550	(0.024)	0.346	(0.029)
Sao Paulo	0.428	(0.096)	0.030	(0.076)	North West	0.692	(0.136)	0.609	(0.193)					
Porto Alegre	0.033	(0.103)	−0.067	(0.081)	Gauteng	0.433	(0.135)	0.825	(0.176)					
					Mpumalanga	0.276	(0.142)	0.957	(0.181)					
					Limpopo	1.270	(0.151)	1.595	(0.196)					
					Kwazulu Natal	0.532	(0.134)	0.906	(0.175)					
Constant	1.655	(0.294)	−2.609	(0.256)		1.023	(0.363)	−1.308	(0.462)		3.704	(0.079)	−2.549	(0.100)
Pseudo-R2	0.103					0.186					0.142			

Note: Multinomial logit with base outcome = being a formal salaried worker.

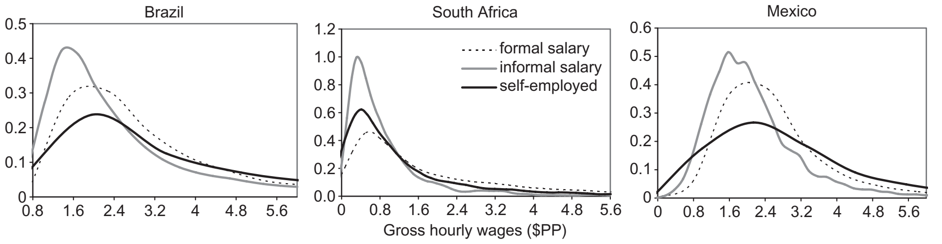

Figure 1. Earnings Distribution by Sector (Kernel-Density)

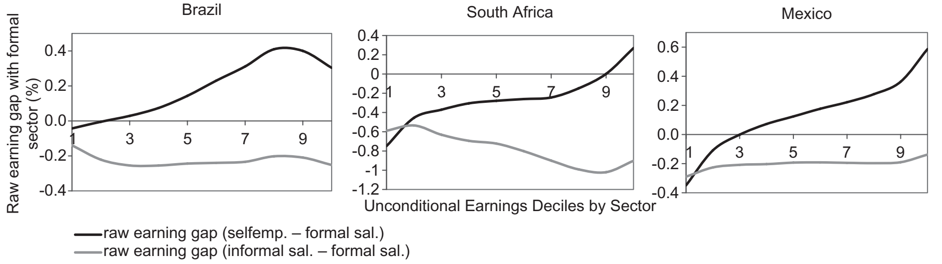

Figure 2. Raw Earnings Differences Between Self-Employed/Informal Sector and Formal Sector

While these pictures result from differences in both workers' characteristics and returns between sectors, the rest of the paper attempts to pinpoint the true differences in remuneration across sectors, i.e. the distribution of earnings gaps conditional on workers' observed and time-invariant unobserved characteristics.

3. Econometric Approach

We first estimate standard Mincer earnings equations at the mean and at various quantiles on pooled years data with clustered standard errors at the individual level. Denote y_{it} the hourly earnings of worker i at time t. Explanatory variables x_{it} comprise standard human capital information (age, age squared, education) and other individual/household characteristics as reported in Table 1 (race, number of children, marital status, region) as well as broad industry dummies to control for the possible structural differences between sectors. The model estimated by OLS is simply written as:

$$y_{it} = x_{it}\beta + \delta I_{it} + \rho S_{it} + u_i \tag{1}$$

where dummy I_{it} (resp. S_{it}) takes value one if person i observed at time t is an informal salary worker (resp. self-employed). The estimated coefficients $\hat{\delta}$ and $\hat{\rho}$ are interpreted as a measure of the conditional earnings premium/penalty experienced by informal salary workers and self-employed workers, respectively, compared to formal wage earners. In the case of quantile regressions (QR), we can write for any worker i the τ^{th} quantile of the y distribution conditionally on observables as:

$$F_{y_{it}}^{-1}(\tau \mid x_{it}) = x_{it}\beta(\tau) + \delta(\tau)I_{it} + \rho(\tau)S_{it}, \quad \forall \tau \in [0,1]. \tag{2}$$

Importantly, the effects of the covariates are permitted to depend on the quantile of interest, in particular the informal sector premia/penalties $\delta(\tau)$ and $\rho(\tau)$.

While the datasets in use are arguably relatively rich in individual- and worker-level characteristics, there may still be a considerable probability that there are other unobserved factors that determine both selection into a particular sector and earnings. One obvious example would be unobserved productivity that is not correlated with the educational level. The failure to account for such factors could then lead to biased estimates of the conditional earnings penalty/premium. A way to purge estimations from such unobservables, if they are time invariant, is to use the rotating panel nature of the data to introduce fixed effects in the model. We first compare the results of a simple fixed effects model (FE) for each country to the results based on standard OLS. We also extend this approach to the whole distribution by estimating fixed effects quantile regression (FE-QR), to be compared to the results of the standard QR.

With x_{it} a set of controls (same as above except that time-invariant characteristics are dropped), α_i the time-invariant heterogeneity (the individual fixed effect), and ε_{it} an i.i.d. normally distributed stochastic term accounting for possible measurement error, the FE model is simply written:

$$y_{it} = \alpha_i + \gamma_t + x_{it}\beta + \delta I_{it} + \rho S_{it} + \varepsilon_{it} \tag{3}$$

where $E[\varepsilon_{it}, \alpha_i, x_{it}, I_{it}, S_{it}] = 0$ for all individuals i and periods t. The FE estimator is consistent even if unobserved characteristics are correlated with both selection and earnings, as long as those characteristics are constant over time. The conditional earnings $\hat{\delta}$ and $\hat{\rho}$ gaps are derived from the comparison between those who move between employment states and the "stayers." Denote $C = 1, 2, 3$ the three different states, respectively self-employed, informal salary workers, and formal salary workers. Let us illustrate the identification by a simple two-period example and three of the possible transitions:

$$E[y_{i2} - y_{i1} | C_{i1} = k, C_{i2} = k] = \Delta \text{ for } k = 1, 2, 3$$

$$E[y_{i2} - y_{i1} | C_{i1} = 1, C_{i2} = 3] = \Delta - \rho$$

$$E[y_{i2} - y_{i1} | C_{i1} = 2, S_{i2} = 3] = \Delta - \delta$$

$$\text{with } \Delta = \gamma_2 - \gamma_1 + (x_{i2} - x_{i1})\beta.$$

The changes in earnings for those moving into formal employment and coming from self-employment (second line) or from informal salary work (third line) contribute to identify the conditional earnings gaps ρ and δ, together with the stayers in any state (first line) which capture the change in time-varying observables. Beyond this example, actual identification is naturally completed by the movers for all other possible permutations between states, that is, from self-employment or informal salary work into formal salary work or between self-employment and informal salary work. Also, the example above is based on two periods only but we explore in practice all of the possible transitions of an individual over the several periods observed in the data (some workers may move several times or be accounted as both stayers and movers if observed more than twice). Note that at this stage, we do not account for possible differences in the conditional earnings gaps whether identified on workers moving from formal to informal sectors or on those moving in the other direction, but allow for asymmetrical effects in the robustness checks section.

We extend the standard QR model to longitudinal data as follows. For any worker i, the τ^{th} quartile of the y distribution conditionally on observables can be written as:

$$(4) \qquad F_{y_{it}}^{-1}(\tau \mid x_{it}) = \alpha_i + \gamma_t(\tau) + x_{it}\beta(\tau) + \delta(\tau)I_{it} + \rho(\tau)S_{it}, \quad \forall \tau \in [0,1].$$

Fixed effects α's have a pure *location* shift effect on the conditional quantiles of the response (i.e. they affect all quantiles in the same way).[7] We can use Koenker's (2004) approach to estimate this model or the alternative and simpler approach suggested by Canay (2010). The latter exploits the assumption that α terms are pure location shifters, so that they can be estimated in a first step by traditional mean estimations (for instance by OLS estimator in first differences). Then it is possible to use the estimated $\hat{\alpha}_i$ in order to regress corrected earnings $\hat{y}_i = y_i - \hat{\alpha}_i$ on the other covariates by traditional QR.

4. Empirical Results

4.1. *Main Results*

Our main results are represented in Figures 3, 4, and 5 and commented on below. For each country, we report the estimated coefficients $\hat{\delta}$ and $\hat{\rho}$, i.e. the conditional earnings penalties/premia from informal salary work and informal self-employment compared to the formal sector. The left panel of each graph shows the estimates from OLS and QR on pooled years while the right panel depicts the estimates of the FE and FE-QR on panel data. Dashed lines and empty diamonds represent the bootstrapped 95% confidence intervals. For each country, we can see that QR (resp. FE-QR) estimates are not all contained in the interval surrounding the OLS (resp. FE) coefficient and there are important differences along the conditional earnings distribution. In Table 3, we also report the conditional earnings gaps at the mean, the median, and two extreme quantiles as well as the bootstrapped standard errors.[8]

[7]As explained by Koenker (2004), it is unrealistic to attempt to estimate *distributional* shift $\alpha_i(\tau)$ for a worker i if the number of periods of observations is too small. This is the case in the present study, and we can only estimate an individual specific location-shift effect.

[8]The full estimation tables, not reported because of space limitation, are available from the authors.

We first discuss the results concerning the conditional earnings gap between informal and formal salary workers. According to OLS estimations reported in Table 3, informal salary workers face an average earnings penalty of around 9 percent in Brazil, 63 percent in South Africa, and 15 percent in Mexico when controlling for workers' observed characteristics. We also see that these magnitudes are close to the penalties estimated at the median of the conditional earnings distribution. Grey lines in the left panels of Figures 3, 4, and 5 confirm that informal salary workers are systematically underpaid in all countries and at all conditional quantiles.

When accounting for unobserved heterogeneity (right panels of the same figures), the penalties faced by informal salary workers decrease. A similar pattern is observed in all three countries, with the largest penalties to be found at the bottom of the conditional earnings distribution. Penalties remain relatively large in South Africa, between 10 and 30 percent, but are now very small in other countries (10–15 percent at most, and insignificant for upper conditional quantiles).

We now consider the conditional earnings gap between informal self-employed workers and formal wage earners. Results reveal a more contrasted situation across countries than in the case of informal versus formal salary workers. OLS estimates indicate that on average self-employment confers an earnings premium of 11 percent in Brazil and 13 percent in Mexico but a penalty of 30 percent in South Africa. Again, these are close to QR estimates at the median of the conditional earnings distribution. QR at other levels, as represented by the black lines in the left panels of Figures 3, 4 and 5, show a more complex pattern with penalties at the bottom and premia at the top of the conditional distribution in all countries, indicating that self-employment contributes to a more dispersed earnings distribution.

Accounting for unobserved heterogeneity tends to increase premia in Mexico and decrease penalties in South Africa. In Brazil, fixed effects are also important but act in the opposite direction (in particular FE-QR yield lower premia in the upper part of the conditional distribution). We observe contrasted situations, with penalties at almost all conditional quantiles in South Africa (except the very top) and premia for almost all self-employed workers in Mexico (except the very bottom)—which is in line with Cunningham and Maloney (2001) who show that a minority (13 percent) of the self-employed in this country are in the lower-tier informal sector. Again self-employment premia increase with conditional quantiles in Mexico and Brazil (and penalties decrease with conditional quantiles in South Africa) so that the dispersion effect of self-employment is maintained.

4.2. *Interpretations and Extensions*

Conditional versus Unconditional Earnings Caps

First, it is important to note that our study is limited to results concerning *conditional* quantiles. That is, Figures 3, 4, and 5 show conditional penalties/premia, i.e. earnings differences after adjusting for differences in workers' characteristics. In particular, higher premia at higher conditional quantiles for the self-employed indicate that individuals who receive higher remunerations, given their human capital characteristics, do especially well in informal self-employment.

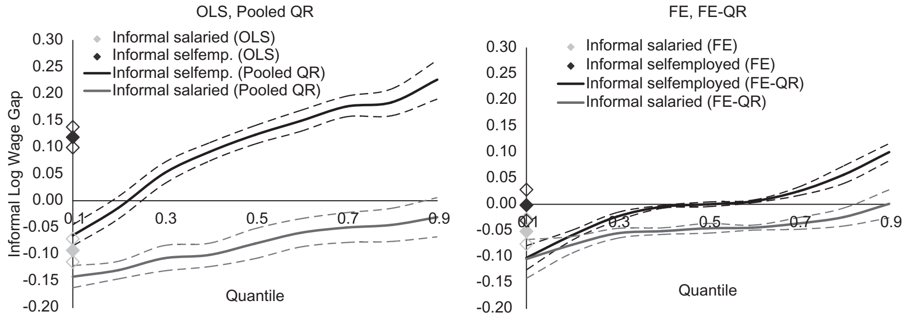

Figure 3. Estimated Earnings Gaps at the Mean and at Conditional Quantiles (Brazil)

Left panel (estimations on the pooled years sample): full diamonds indicate the conditional earnings gap between self-employment (dark) or informal salary work (light) and formal employment using OLS. Lines indicate the earnings gap at different conditional quantiles using QR. Bootstrapped 95% confidence intervals are represented by empty diamonds (OLS) and dashed lines (QR).

Right panel (estimations using panel information): full diamonds indicate the conditional earnings gap between self-employment (dark) or informal salary work (light) and formal employment using FE. Lines indicate the earnings gap at different conditional quantiles using FE-QR. Bootstrapped 95% confidence intervals are represented by empty diamonds (FE) and dashed lines (FE-QR).

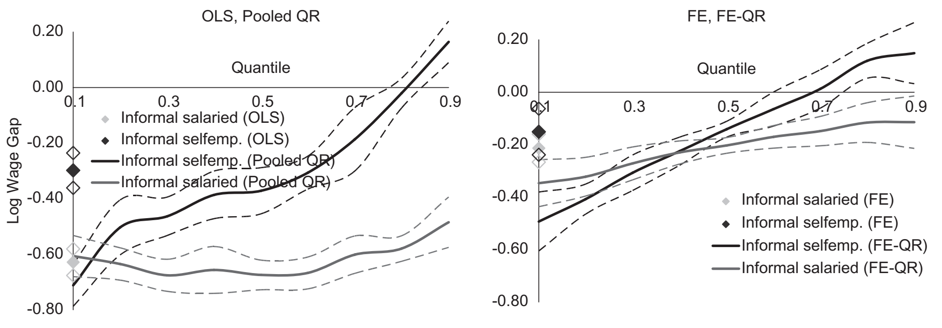

Figure 4. Estimated Earnings Gaps at the Mean and at Conditional Quantiles (South Africa)

Note: Bootstrapped 95% confidence intervals represented by dashed lines and empty diamonds.

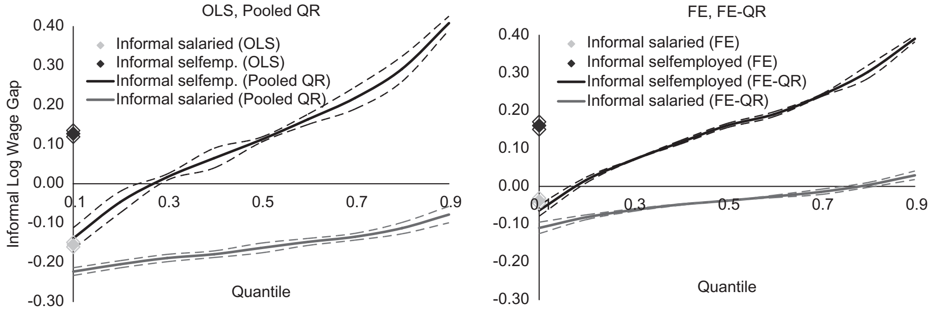

Figure 5. Estimated Earnings Gaps at the Mean and at Conditional Quantiles (Mexico)

Note: Bootstrapped 95% confidence intervals represented by dashed lines and empty diamonds.

TABLE 3
Conditional Earnings Gaps: Summary

Country	Sector	Method	Mean Coef.	Mean Std.Err.	Q = 0.2 Coef.	Q = 0.2 Std.Err.	Q = 0.5 Coef.	Q = 0.5 Std.Err.	Q = 0.8 Coef.	Q = 0.8 Std.Err.
Brazil	Informal salaried	(1) Pooled OLS	–0.093	(0.011)	–	–	–	–	–	–
		(2) Pooled QR	–	–	–0.130	(0.008)	–0.080	(0.014)	–0.045	(0.016)
		(3) Panel, FE	–0.052	(0.012)	–	–	–	–	–	–
		(4) Panel, FE-QR	–	–	–0.078	(0.004)	–0.046	(0.002)	–0.024	(0.004)
	Informal self-empl.	(1) Pooled OLS	0.119	(0.010)	–	–	–	–	–	–
		(2) Pooled QR	–	–	–0.013	(0.011)	0.123	(0.009)	0.184	(0.013)
		(3) Panel, FE	–0.002	(0.015)	–	–	–	–	–	–
		(4) Panel, FE-QR	–	–	–0.060	(0.002)	0.000	(0.001)	0.057	(0.004)
South Africa	Informal salaried	(1) Pooled OLS	–0.629	(0.024)	–	–	–	–	–	–
		(2) Pooled QR	–	–	–0.635	(0.030)	–0.674	(0.027)	–0.579	(0.024)
		(3) Panel, FE	–0.213	(0.027)	–	–	–	–	–	–
		(4) Panel, FE-QR	–	–	–0.322	(0.019)	–0.203	(0.008)	–0.116	(0.019)
	Informal self-empl.	(1) Pooled OLS	–0.298	(0.032)	–	–	–	–	–	–
		(2) Pooled QR	–	–	–0.503	(0.051)	–0.372	(0.042)	–0.020	(0.027)
		(3) Panel, FE	–0.151	(0.045)	–	–	–	–	–	–
		(4) Panel, FE-QR	–	–	–0.407	(0.015)	–0.141	(0.008)	0.120	(0.017)
Mexico	Informal salaried	(1) Pooled OLS	–0.155	(0.003)	–	–	–	–	–	–
		(2) Pooled QR	–	–	–0.205	(0.003)	–0.163	(0.002)	–0.113	(0.003)
		(3) Panel, FE	–0.038	(0.003)	–	–	–	–	–	–
		(4) Panel, FE-QR	–	–	–0.082	(0.002)	–0.038	(0.000)	0.004	(0.002)
	Informal self-empl.	(1) Pooled OLS	0.127	(0.004)	–	–	–	–	–	–
		(2) Pooled QR	–	–	–0.047	(0.003)	0.112	(0.002)	0.292	(0.003)
		(3) Panel, FE	0.161	(0.005)	–	–	–	–	–	–
		(4) Panel, FE-QR	–	–	0.017	(0.002)	0.161	(0.000)	0.303	(0.002)

Notes: Figures represent the conditional earnings gaps between informal salary and formal salary workers (δ) or between informal self-employed and formal salary workers (ρ). Estimations are based on the variables reported in the descriptive statistics, except time-invariant characteristics (race, education, and region) in the fixed effects estimations. "Pooled" OLS and QR mean estimations on the pooled years (in contrast to FE and FE-QR which are estimated using panel information). Numbering (1)–(4) corresponds to the equations (1)–(4) in the text. Standard errors are in parentheses.

A more complete characterization of the role of returns versus endowments to explain the unconditional earnings gap across sectors would require an Oaxaca-type decomposition based on unconditional quantile counterfactuals, as suggested for instance in Firpo *et al*. (2009). Yet the extensions of this method to fixed effects estimations, and *a fortiori* to FE-QR, is not straightforward.

Nonetheless, we suggest an informal comparison between the conditional earnings differentials estimated without fixed effects (left panels of Figures 3, 4, and 5) and the unconditional earnings differentials represented in Figure 2. This shows striking similarities, for both informal salary penalties and self-employment premia. Arguably these two sets of results represent very different things since (i) Figure 2 accounts for both the differences in returns and the differences in workers' characteristics across sectors, and (ii) quantiles are not defined in exactly the same way (because workers of a given sector are not distributed evenly across deciles of the pooled sample). There are however noticeable similarities in the *shape* of conditional and unconditional earnings differentials, which suggests that differences in returns between sectors play an important role in explaining the dispersion of the raw earnings differentials. Admittedly, the *levels* are not exactly comparable between the two sets of graphs. For instance, the raw penalties for informal wage earners are larger than the conditional penalties in all three countries, suggesting that formal salary workers also have "better" observable skills.[9] Unconditional premia for the self-employed (first graph of Figure 2) are larger than conditional premia in Brazil (Figure 3), suggesting that the self-employed have better characteristics (from Table 1, we see in particular that they have higher education levels).

In the rest of this section, we specifically comment on the unexplained earnings gaps across sectors (the conditional penalties or premia). In perfectly competitive markets with randomly distributed workers, we could expect that (conditional) earnings equalization may eventually occur across sectors. Possible reasons for persisting conditional earnings gaps pertain to segmentation (some workers randomly end up in the unregulated sector and do not benefit from the impact of unionization, minimum wage, etc.), and the presence of unmeasured/unobserved job attributes (risk, independence, in-kind rewards, etc.) and unmeasured/unobserved workers' characteristics (skills, preferences). We suggest several interpretations along those lines.

Unobserved Heterogeneity

An important contribution of the present paper is that we account for unobserved heterogeneity. We first attempt to narrow down what these fixed effects could represent, in particular among the possible unmeasured/unobservable characteristics mentioned above. Individual effects captured in our FE/FE-QR estimations are time-invariant and cannot be sector-specific (identification would require

[9]In particular in South Africa, the mean raw penalty is as high as 80 percent while the conditional penalty is around 63 percent (when controlling for observed characteristics but not for fixed effects).

many more years of observation).[10] Hence they do not correspond to some unobservables like the particular job place the worker is in, for instance, but are interpretable in terms of workers' unobserved skills. There is another element that suggests it is the case, namely the fact that unobserved characteristics seem to complete observables in explaining raw differentials. For instance, penalties faced by informal salary workers are larger when fixed effects are ignored and larger still when observed characteristics are not controlled for (unconditional earnings gap)—hence both observed *and* unobserved skills are poorer among informal salary workers compared to formal sector wage earners. The reverse is true for the Brazilian self-employed (we find no or very moderate self-employment premia in the upper half of the conditional distribution after controlling for unobservables).

The important result is that accounting for unobservables changes the picture quite considerably in many of the situations described in the paper, even after controlling for a rich set of observable characteristics. This is very much in line with the findings of Badaoui *et al.* (2008) for informal wage earners in South Africa. Yet our results also extend this conclusion to informal wage earners in Brazil and Mexico: after purging estimations from fixed effects, conditional penalties faced by informal salary workers appear much smaller than previously thought and extremely moderate in Latin America—and to the Brazilian self-employed especially.

Nevertheless, different types of explanation must be found regarding the large remaining penalties in the South African informal sector and the self-employment premia in Mexico. As suggested above, obvious candidates are labor market segmentation and job attributes (which are not captured in fixed effects and could justify compensating differentials). Hereafter, we simply suggest tentative interpretations drawing from the cross-country comparisons in our results and from differences in the legal and historical contexts of these countries.

Cross-Country Comparisons and Legal Contexts

Informal sector penalties are substantial in South Africa compared to the two other countries. This is in line with the role of factors internal to the labor market as highlighted in several studies (e.g. Kingdon and Knight, 2007). Hofmeyr (2002) notes that "since the early 1980s a wide variety of workers' rights has been entrenched, in particular since 1994. A probable effect of these measures is to protect existing formal-sector workers, in particular unionised workers, at the expense of those who seek such jobs, creating classes of relatively privileged insiders and increasingly marginalised outsiders. This would be expected to result in segmentation of the market with wage differentials between the segments which cannot be explained by differences in skills and working conditions." Note however that this explanation alone does not justify the very large earnings

[10]Recall that another limitation of what we could capture in terms of individual effects is the fact that they are pure location shifters and cannot cause variation across the conditional spectrum of relative earnings.

penalties found in the literature (about 60 percent at the mean in Kingdon and Knight, 2007, which is close to our median/mean estimates of the conditional penalty without fixed effects). We have indeed found that a great deal of it is explained by the poor unobserved characteristics of informal workers.[11]

Results also show interesting cross-country differences concerning the relative position of the self-employed. Several authors often point to a "Mexican exception" (Marcouiller *et al.*, 1997; Gonzalez and Maloney, 1999; Maloney, 1999). Based on our results, we would rather think of a continuum of experiences concerning the nature of self-employment. Mexico and South Africa appear as two polar cases within this small sample of countries while Brazil lies somewhere in-between.[12]

At one extreme, self-employment penalties concern a large part of the conditional earnings distribution in South Africa. As argued for the informal salary gap, the presence of unionization in formal employment must play a role. Also, the relatively small size of South African self-employment and the lower productivity in this sector—even after controlling for workers' characteristics—have been partly explained by the difficulties in doing business, due especially to land/credit constraints, inhibition of entrepreneurial skills resulting from the apartheid era, restrictive bye-laws (e.g. licensing, environmental, and health regulations), and high crime rate against business owners (Devey *et al.*, 2003). Formal salaried work remains on average the most lucrative mode of employment (Hofmeyr, 2002; Kingdon and Knight, 2004, 2007; Ruffer and Knight, 2007).[13]

At the other extreme, we find large conditional premia for most of the self-employed in Mexico. They do not necessarily translate into welfare premia, even when conditioning on workers' characteristics, and are no proof that segmentation and rationing do not exist in Mexico. While the previous literature indicates the competitive nature of labor markets in this country, the massive differences between South African and Mexican experiences are simply suggestive. Nonetheless, the different situation in Mexico is partly explained by the relatively low remunerations in the formal sector. According to Maloney (2004, p. 1159): "The usual sources of wage rigidity that would segment the market seem absent: minimum wages have not been binding for the last decade, unions to date have primarily been concerned about preserving employment rather than raising remuneration, and wages have shown extraordinary downward flexibility during crises." Arguably this does not completely justify self-employment premia, and other interpretations must rely on the existence of compensating differentials in

[11]Note that earnings differentials are partly explained by sorting, where workers with low levels of human capital are also more likely to work in the informal sector. Indeed firms' access to financing is relatively more limited in the informal sector and employers with low degrees of capitalization tend to recruit less able workers (Amaral and Quintin, 2006). This factor is to some extent controlled for by observable characteristics and the fixed effects.

[12]Marcouiller *et al.* (1997) also find more intermediate situations for El Salvador and Peru compared to Mexico.

[13]Results nonetheless show that self-employment penalties vanish completely for a third of the self-employed and some authors actually point to dynamic segments of the informal labor market in South Africa (see, e.g. Cichello *et al.*, 2005).

self-employment, reflecting formal sector benefits and demand for risk premia due to job uncertainty in informal self-employment.[14]

The Role of Education and Experience

Our estimations have simply used dummy variables for informal salary work and self-employment, which may seem too restrictive.[15] Thus we examine the heterogeneity of the conditional premia/penalties by simply interacting the informal sector dummies *I* and *S* with workers' age and education groups. Results are reported in Panel A of Figure 6 where we focus on the self-employment penalty/premium. At each conditional quantile, we represent earnings gaps conditional on being old or young, with high or low education. This should help to understand how each characteristic affects the size of conditional premia/penalties. Yet, these are conditional distributions and should be interpreted with caution (e.g. the 90th wage centile of young workers cannot be directly compared to the 90th centile of the old worker distribution). In Mexico, the between-group variation (between young/old or low/high educated) is not important compared to within-group heterogeneity (i.e. the variation of the earnings gap along the conditional distribution). Variation in age and education can affect the conditional earnings premium/penalty more significantly in the other countries. In particular, being a younger worker in South Africa increases the conditional penalty by around 20 percentage points. We also find that higher education leads to higher earnings premia (or lower penalties) while the opposite is true for low education levels. This result is in line with one possible interpretation suggested by Rees and Shah (1986). These authors show that education serves as a "filter" such that the more educated tend to be better informed and more efficient at assessing opportunities and risks and hence are able to run their businesses well relative to low educated entrepreneurs. In our results, this is true for all countries but especially for Brazil, where being low-educated and self-employed would systematically lead to penalties. Overall, however, education and age may well affect the conditional distribution of penalties/premia but do not seem to explain their very nature which is to be looked for in interpretations as suggested above.[16]

5. Robustness Checks

We provide a series of robustness checks of our main results. First, the identification of FE on movers is standard but one must verify that the number

[14]If there is voluntary entry into self-employment as suggested by several studies (e.g. Bosch and Maloney, 2008, 2010), those opting for it do not seem to value formal social security (maybe because informal support networks can substitute for unemployment insurance or retirement funds at lower cost).

[15]Models may also be misspecified. However, least square regression provides a minimum mean squared error linear approximation to the true functions in case of misspecification, and Angrist *et al.* (2006) provide a similar result for quantile regression. Therefore our findings have meaningful interpretation even if the true informal earnings penalty/premium depends on the covariates.

[16]Here again, a more complete characterization of the role of demographic heterogeneity could be obtained by using a decomposition approach like Firpo *et al.* (2009). As said, the extension of this method to fixed effects estimations is unfortunately not straightforward and should be the topic of future research.

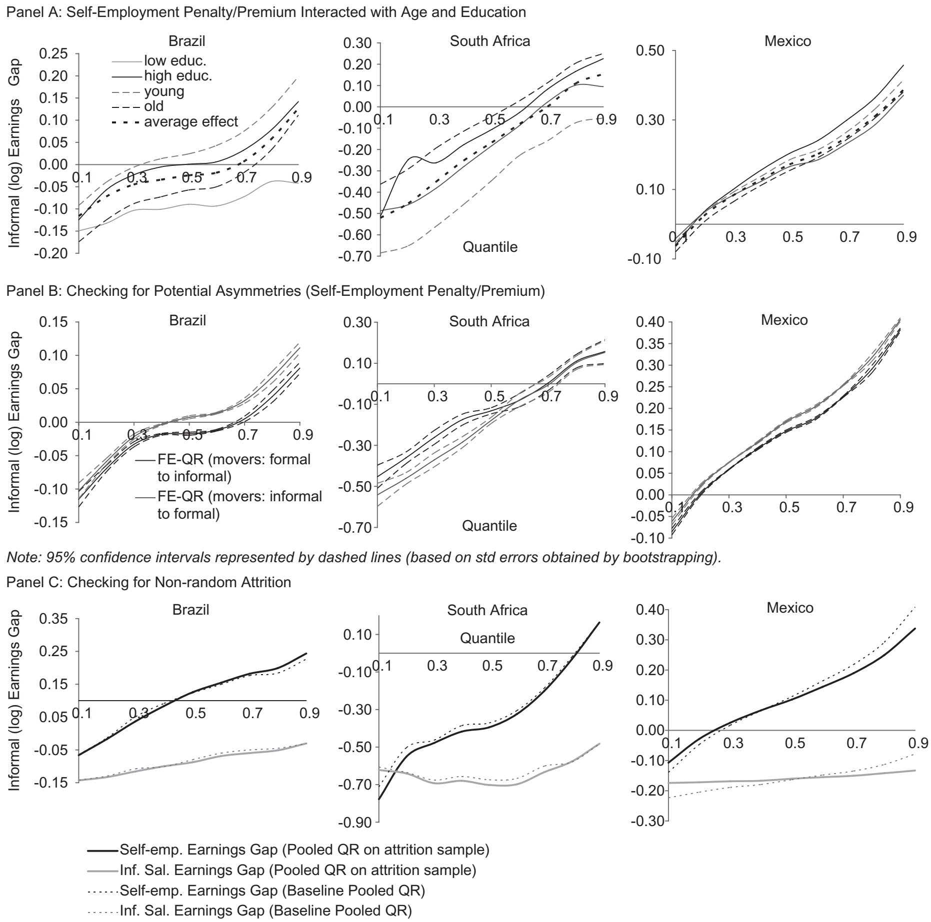

Figure 6. Extension and Robustness Checks

of moves across sectors is large enough for a valid use of this estimator. We calculate the proportion of all panel transitions (i.e. all pairs of observations for the same workers) that correspond to a move. We find that 5, 8, and 10 percent of all transitions are moves between formal and informal salary work (either way) for Brazil, South Africa, and Mexico, respectively, which correspond to 1117, 1088, and 25,028 movers, respectively. Moves between formal salary work and self-employment represent 4, 4, and 3 percent of all transitions, corresponding to 807, 500, and 6650 movers, respectively. These are reassuring numbers regarding the possibility of identifying the parameters of interest.[17]

We also check that movers are not too specific. In particular, one may expect that cross-sector moves are limited to specific groups of workers, for instance those

[17]Note that the identification of the earnings gaps is completed by the moves between informal salaried work and informal self-employment, which correspond to 5, 3, and 9 percent, respectively, of all transitions.

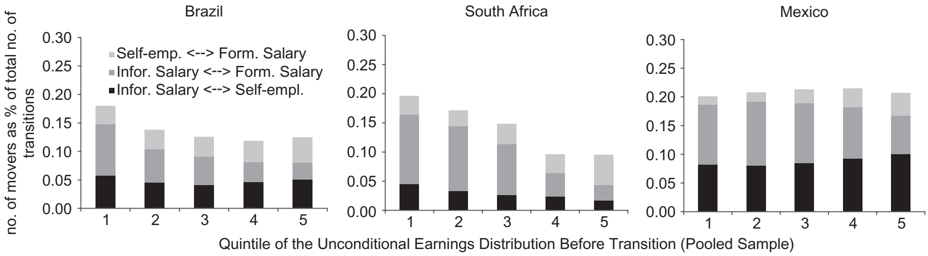

Figure 7. Transitions across Employment States

Notes: The graphs represent the number of movers in % of the total number of transitions in the panels (including both stayers and movers-type of transitions). Quintiles are calculated on the basis of the period ahere workers are first observed in the pooled sample. Total number of transitions = 22.186 in Brazil: 14.026 in South Africa: 263.082 in Mexico.

with the least earnings who are in search of better job prospects. In this case, our estimates could be biased. Figure 7 depicts the proportion of movers per quintile of the "initial" unconditional earnings distribution, i.e. using the periods where workers are first observed in the pooled years sample. This picture shows that moves are relatively spread throughout the unconditional distribution in Brazil and South Africa, although a larger number of moves between formal and informal salary work occurs at the bottom. The high frequency of moves in Mexico (around 20 percent of all panel transitions) occurs at all earnings levels—even if moves between formal sector and self-employment are more frequent in the upper part. This overall picture is reassuring in the sense that moves are not overly concentrated in some parts of the earnings distribution.[18]

Another aspect of the identification strategy that merits discussion is the assumption that the earnings penalty/premium is the same for those that move from informal self-employment into formal salary work as it is for those that move in the opposite direction. If all the unobservable heterogeneity is time-invariant, as assumed in the FE estimator, then this is not an issue. However, with the traditional view of self-employment as a safety net for those losing preferred formal sector jobs, one would expect that becoming self-employed is more often the result of time-specific negative shocks (e.g. productivity shocks), and that moves in this direction capture larger penalties (or smaller premia) than moves in the opposite direction. We replicate our results when including only one type of move at a time. Panel B of Figure 7 shows that results are not fundamentally asymmetrical.

Finally, we check that panel attrition does not lead to some bias by selecting out a specific type of worker. Indeed our baseline estimation excluded all workers observed only once in the data. However, it might be expected that workers in the informal sector are more likely to exit from the panel because of higher migration or higher misreporting. To check for possible non-random attrition, we simply estimate QR on pooled years for those observed only once in the data and compare

[18]We refrain from drawing any conclusions based on these "raw" transitions. A more in-depth interpretation of inter-sector flows would require some adjustments for turnover and job creation as performed in Bosch and Maloney (2010) and Maloney (1999), which is naturally beyond the scope of the present paper.

the estimated earnings gaps to baseline results. Panel C of Figure 7 shows that results are very similar in both cases, conveying that sample attrition does not relate to labor market states.

6. Concluding Discussion

This study provides a comprehensive analysis of male earnings structure in Brazil, South Africa, and Mexico based on the estimation of conditional earnings gaps between formal and the informal sectors, distinguishing between informal wage earners and the informal self-employed. We use rich panel datasets to define informality in a comparable way across countries and run fixed effects estimations at different points of the conditional earnings distribution. We find several interesting results. First, once estimates are purged from time-invariant unobserved skills, the penalties faced by informal salaried workers are not as large as usually thought, even in South Africa where they remain significant. Nonetheless, all informal salary workers in all countries are systematically underpaid vis-à-vis formal salary workers, *ceteris paribus*. Second, self-employment generates significant conditional premia (penalties) in Mexico (South Africa), except at the bottom (top) of the conditional distribution. Such a difference between these two countries is not explained by unmeasured characteristics of workers (at least not time-invariant ones), but possibly may be due to other factors such as different legal contexts (segmentation and binding minimum wages in South Africa, premia in Mexican self-employment to compensate job insecurity or formal social security). Third, in all countries, we find that self-employment contributes to a more dispersed earnings distribution possibly due to the heterogeneous nature of activities in this sector that range from small traders to self-employed professionals.

We conclude with a series of comments on the present approach. First, we have used panel information to purge our estimations from time-invariant unobservable heterogeneity. Extending the approach to time-varying unobservables would require selection to be modeled explicitly. Yet it seems extremely challenging to find proper instruments, i.e. instruments that can convincingly explain selection into a given sector (but not earnings) and that also vary over time. Second, as in many studies, we have compared self-employment income to formal sector wages on the basis of hourly earnings. It may be desirable, however, to distinguish between wages and profits for the self-employed workers.[19] Finally, gross earnings gaps could be corrected for income taxes and social contributions paid in the formal sector (as in Badaoui *et al.*, 2008, and Bargain and Kwenda, 2009)—yet informal self-employment should be better identified for that purpose.

[19]Following Headen (1990) and under the assumption of equal returns to labor for both salary and self-employed workers in the informal sector, we can interpret the earnings gap between self-employed and informal employees as a measure of the returns to capital. Using our estimates (relative to the formal sector), we find an insignificant difference at the bottom of the distribution, which is consistent with little capital-intensive businesses in the lower-tier sector; it increases gradually with earnings levels. At the top of the distribution, this difference amounts to around 10 percent of the informal sector wages in Brazil, 20 percent in South Africa, and 40 percent in Mexico. Yet the assumption of equal returns to labor remains to be discussed. One may also argue that similar corrections should be made for salary workers, i.e. adjusting wages of salary workers for private investment in human capital.

Accounting for non-pecuniary advantages attached to a particular sector, and above all for medical benefits and pensions paid in the formal sector, represents a considerable challenge, given data limitations (cf. Bourguignon *et al.*, 2008). Yet this is a fundamental and necessary improvement for welfare analyses to become possible.

References

Albrecht, J. W., A. van Vuuren, and S. Vroman, "Decomposing the Gender Wage Gap in The Netherlands with Sample Selection Adjustments," *Labour Economics*, 16, 383–96, 2009.

Almeida, R. and P. Carneiro, "Enforcement of Regulation, Informal Labor and Firm Performance," *Journal of Comparative Economics*, 37, 28–46, 2009.

Amaral, P. and E. Quintin, "A Competitive Model of the Informal Sector," *Journal of Monetary Economics*, 53, 1541–53, 2006.

Amuedo-Dorantes, C., "Determinants and Poverty Implications of Informal Sector Work in Chile," *Economic Development and Cultural Change*, 52, 349–68, 2004.

Angrist, J., V. Chernozhukov, and I. Fernández-Val, "Quantile Regression under Misspecification, with an Application to the US Wage Structure," *Econometrica*, 74, 539–63, 2006.

Arias, O. and M. Khamis, "Comparative Advantage, Segmentation and Informal Earnings: A Marginal Treatment Effects Approach," IZA DP No. 3916, 2008.

Aronson, R. L., *Self-Employment: A Labor Market Perspective*, ILR Press, Ithaca, NY, 1991.

Badaoui, E., E. Strobl, and F. Walsh, "Is there an Informal Employment Wage Penalty? Evidence from South Africa," *Economic Development and Cultural Change*, 56, 683–710, 2008.

Bargain, O. and P. Kwenda, "The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data," IZA Discussion Paper No. 4286, 2009.

Barros, R. P. and C. H. Corseuil, "The Impact of Regulations on Brazilian Labor Market Performance," in J. Heckman and C. Pages (eds), *Law and Employment: Lessons from Latin America and the Caribbean*, The University of Chicago Press, Chicago, 273–350, 2004.

Blunch, N-H., S. Canagarajah, and D. Raju, "The Informal Sector Revisited: A Synthesis across Space and Time," Social Protection Discussion Paper Series No. 0119, Social Protection Unit, Human Development Network, The World Bank, 2001.

Bosch, M. and W. F. Maloney, "Cyclical Movements in Unemployment and Informality in Developing Countries," Policy Research Working Paper Series 4648, World Bank, 2008.

———, "Comparative Analysis of Labor Market Dynamics Using Markov Processes: An Application to Informality," *Labour Economics*, 17, 621–31, 2010.

Bosch, M., E. Goni, and W. Maloney, "The Determinants of Rising Informality in Brazil: Evidence from Gross Worker Flows," IZA Discussion Paper 2970, 2007.

Bourguignon, F., F. Ferreira, and P. Leite, "Beyond Oaxaca—Blinder: Accounting for Differences in Household Income Distributions," *Journal of Economic Inequality*, 6, 117–48, 2008.

Canay, I. A., "A Note on Quantile Regression for Panel Data Models," Department of Economics, Northwestern University, 2010.

Carneiro, F. G. and A. Henley, "Modeling Formal vs Informal Employment and Earnings: Microeconometric Evidence for Brazil," *Annals of the XXIX National Meeting of Economics of ANPEC*, 2001.

Carrington, W. J., M. Kristin, and P. Brooks, "The Role of Employer–Employee Interactions in Labor Market Cycles: Evidence from the Self-Employed," *Journal of Labor Economics*, 14, 571–602, 1996.

Charmes, J., "The Contribution of Informal Sector to GDP in Developing Countries: Assessment, Estimates, Methods, Orientations for the Future," 4th Meeting of the Delhi Group on Informal Sector Statistics, Geneva, 28–30, 2000.

Cichello, P. L., G. S. Fields, and M. Leibbrandt, "Earnings and Employment Dynamics for Africans in Post-Apartheid South Africa: A Panel Study of KwaZulli-Natal," *Journal of African Economies*, 14, 1–48, 2005.

Cunningham, W. and W. F. Maloney, "Heterogeneity among Mexico's Microenterprises: An Application of Factor and Cluster Analysis," *Economic Development and Cultural Change*, 50, 131–56, 2001.

Devey, R., C. Skinner, and I. Valodia, "Informal Economy Employment Data in South Africa: A Critical Analysis," Report prepared for the Employment Data Research Group, Human Sciences Research Council. 2003.

Dickens, W. and K. Lang, "A Test of the Dual Labor Market Theory," *American Economic Review*, 75, 792–805, 1985.

Fields, G. S., "Rural-Urban Migration, Urban Unemployment and Underemployment, and Job Search Activity in LDCs," *Journal of Development Economics*, 2, 165–87, 1975.

———, "A Guide to Multisector Labor Markets Models," Paper prepared for the World Bank Labor Market Conference, 2005.

Firpo, S., N. Fortin, and T. Lemieux, "Unconditional Quantile Regressions," *Econometrica*, 77, 953–73, 2009.

Funkhouser, E., "Mobility and Labor Market Segmentation: The Urban Labor Market in El Salvador," *Economic Development and Cultural Change*, 46, 123–53, 1997.

Gong, X. and A. van Soest, "Wage Differetials and Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico," *Labour Economics*, 9, 513–29, 2002.

Gonzalez, P. A. and W. F. Maloney, "Logit Analysis in a Rotating Panel Context and an Applications to Self-Employment Decisions," Policy Research Working Paper Series 2069, The World Bank, 1999.

Hamilton, B. H. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment," *Journal of Political Economy*, 108, 604–31, 2000.

Headen, A. E., "Wage, Returns to Ownership, and Fee Responses to Physician Supply," *Review of Economics and Statistics*, 72, 30–37, 1990.

Henley, A., G. R. Arabsheibani, and F. G. Carneiro, "On Defining and Measuring the Informal Sector," Policy Research Working Paper 3866, The World Bank, 2006.

Hofmeyr, J. F., "The Importance of Segmentation in the South African Labor Market," University of Natal, 2002.

Jütting, J., T. Xenogiani, and J. Parlevliet, "Work and Well-Being: Informal Employment Revisited," OECD Development Centre, 2007.

Kingdon, G. and J. Knight, "Unemployment in South Africa: The Nature of the Beast," *World Development*, 32, 391–408, 2004.

———, "Unemployment in South Africa, 1995–2003: Causes, Problems and Policies," *Journal of African Economies*, 16, 813–48, 2007.

Koenker, R., "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89, 2004.

Leibenstein, H., "Entrepreneurship and Development," *American Economic Review*, 58, 72–83, 1968.

Leontaridi, M., "Segmented Labor Markets: Theory and Evidence," *Journal of Economic Surveys*, 12, 103–09, 1998.

Maloney, W. F., "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico," *World Bank Economic Review*, 13, 275–302, 1999.

———, "Informality Revisited," *World Development*, 32, 1159–78, 2004.

Marcouiller, D., V. Ruiz de Castilla, and C. Woodruff, "Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru," *Economic Development and Cultural Change*, 45, 367–92, 1997.

Moro, S., F. Chein, and A. F. Machado, "Self-Employment in Brazil and its Determinants: A Spatial Analysis," Textos para Discussao Cedeplar-UFMG td204, Cedeplar, Universidade Federal de Minas Gerais, 2003.

Perry, G., W. Maloney, O. Arias, P. Fajnzylber, A. Mason, and J. Saavedra-Chanduvi (eds), *Informality: Exit and Exclusion*, World Bank Latin America and Caribbean Studies, 2007.

Rees, H. and A. Shah, "An Empirical Analysis of Self-Employment in the UK," *Journal of Applied Econometrics*, 1, 95–108, 1986.

Ruffer, T. and J. Knight, "Informal Sector Labor Markets in Developing Countries," University of Oxford, 2007.

Saavedra, J. and A. Chong, "Structural Reforms, Institutions and Earnings: Evidence from the Formal and Informal Sectors in Urban Peru," *Journal of Development Studies*, 35, 95–116, 1999.

Sullivan, D. H. and T. M. Smeeding, "All the World's Entrepreneurs: The Role of Self-Employment in Nineteen Nations," Luxembourg Income Study Working Paper Series No. 163, 1997.

Tannuri-Pianto, M. E. and D. M. Pianto, "Informal Employment in Brazil—A Choice at the Top and Segmentation at the Bottom: A Quantile Regression Approach," Serie de Textos para Discussao n 236, Universidade de Brasilia, 2008.

Yamada, G., "Urban Informal Employment and Self-Employment in Developing Countries: Theory and Evidence," *Economic Development and Cultural Change*, 44, 289–314, 1996.