Session Number: Session 4b (Parallel)
Time: Thursday, September 24, 14:00-15:30

Paper Prepared for the Special IARIW-SAIM Conference on
"Measuring the Informal Economy in Developing Countries"

Kathmandu, Nepal, September 23-26, 2009

The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data

Olivier Bargain, University College Dublin, Geary Institute and IZA
Prudence Kwenda, University College, Dublin

This paper is posted on the following website: http://www.iariw.org

IZA Discussion Paper No. 4286
July 2009

ABSTRACT

The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data*

This paper provides new evidence on the wage gap between informal and formal salary workers in South Africa, Brazil and Mexico. We use rich datasets that allow us to define informality in a relatively comparable fashion across countries. We compute precise wage differentials by accounting for taxes paid in the formal sector. For each country, we analyze how the sector wage gap varies within groups, between groups and over time. To account for unobserved heterogeneity, we use large (unbalanced) panels to estimate fixed effects models at the mean and at different quantiles of the wage distribution. We find that unobserved heterogeneity explains a large part of the (conditional) wage gap. The remaining informal sector wage penalty is large in the lower part of the distribution but almost disappears at the top. The penalty primarily concerns young workers and is found to be procyclical. We carefully investigate the robustness of these results and discuss their policy implications as well as regularities across countries.

JEL Classification: J21, J23, J24, J31, C14, O17

Keywords: wage gap, informal sector, quantile regression, fixed effects model, selection

Corresponding author:

Olivier Bargain
University College Dublin
Newman Building
School of Economics
Dublin 4
Ireland
E-mail: olivier.bargain@ucd.ie

* We are grateful to participants to seminars/conferences at CEPS-INSTEAD, UCD, IZA and Brunel University for useful advices. Many thanks to Blaise Melly for his comments and his kind help regarding the estimator and to Eliane Badaoui, Bill Maloney, Martin Rama, Arthur van Soest and Elizabeth Villagomez for their kind advices regarding the data. The usual disclaimers apply.

1 Introduction

The existence of large informal sectors in developing countries has often been cited as a central factor underlying wage inequality, persistent poverty and labor market inefficiencies. According to the traditional view (Fields, 1975, Dickens and Lang, 1985), workers enter informality to escape unemployment and because they are rationed out of the formal sector as a result of an overly regulated labor market. They earn less than *identical* workers in the formal sector – wages in the latter are set above market-clearing prices because of minimum wages, higher unionization or efficiency-wage explanations. Some authors have recently questioned the traditional 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.[1] If labor markets are competitive, wage equalization should eventually occur – or remaining wage gaps could be justified by compensating differentials in one or the other sector.[2] Recent labor market modeling has combined the two polar views into a dual representation of the informal sector where a competitive or 'voluntary entry' sector coexists with a rationed 'lower-tier' segment (Funkhouser, 1997, Maloney, 2004, Fields, 2005).

In this context, accurate measures of wage differentials across sectors represent an important aspect of the analysis of labor markets in developing countries. Admittedly, they do not constitute a comprehensive measure of welfare nor do they allow testing directly the hypothesis of segmentation on the labor market (Heckman and Sedlacek, 1985, Magnac, 1991, Gindling, 1991, Maloney, 1999); yet they provide a first important step on the way. If inter-sectoral pay inequalities are too large to reflect pure compensating differentials across sectors, they may suggest the existence of rigidities and the need for policy action to restore efficiency and to improve the financial conditions of the poorest workers.

Importantly, robust measures of earnings differentials are typically hampered by two caveats. *Firstly*, unobserved individual characteristics of workers are rarely accounted for. Since unobserved skills may be correlated with both sector choice and earnings, recent studies have implemented two-stage models where selection is jointly specified with the wage regressions, possibly in a dynamic framework (e.g., Gong and Van Soest, 2002, on Mexican data). Also, recent studies suggest controlling for time invariant unobservables using panel data (Badaoui et al., 2008) or propose to apply matching estimators (Pratap and Quintin, 2006) or models of essential heterogeneity (Arias and Khamis, 2007). Interestingly, several of these authors find that unobserved characteristics explain a great deal of the average informal sector wage penalty. *Secondly*, it may be empirically difficult to draw a line between high- and low-tier informal sectors. While such a dual representation is convenient for modeling purposes, the informal sector is likely to present a high degree of heterogeneity.[3] As a result, the informal wage gap may change gradually along the wage distribution or with workers' attributes. Recent empirical contributions thus depart from simple

[1] Evidence of voluntary selection into the informal sector has been particularly compelling for Latin America and for self-employed workers. See Maloney (1999, 2004) for Mexico, Yamada (1996) and Saavedra and Chong (1999) for Peru and Brazil, among others.

[2] An apparent informal sector wage penalty could in fact be compensated by non-wage job characteristics in this sector (e.g., independence and job flexibility, training area for young workers, etc.). Inversely, informal workers could enjoy higher wage rates to compensate for the non-receipt of social security benefits (medical coverage, pension) in the formal sector. Yet this is not necessarily the case as the perceived value of these benefits may be low, either because these services are traditionally provided through family support or because workers may be aware of inefficiencies in formal sector protection. They may also balance these benefits against the possibility to avoid taxes in the informal sector (Maloney, 1999).

[3] Notwithstanding, several interesting studies show that the dual representation of the informal sector proves to be a better alternative than polar models (see in particular Cunningham and Maloney, 2001, and Guenther and Launov, 2006).

estimation of the wage gap at the mean, which may suffer from heteroscedasticity and fail to capture important information. Quantile regression (QR) techniques unveil more complex patterns and allow for rich distributional analyses (e.g., Tannuri-Pianto and Pianto, 2002, and Tannuri-Pianto et al., 2004, for Brazil and Bolivia respectively).

This paper provides new evidence on the informal-formal wage gap among salary workers in South Africa, Brazil and Mexico. Our first contribution is an accurate account of raw wage differentials; in particular, we refine usual measures by adjusting wages for the taxes paid in the formal sector.[4] Secondly, we use rich datasets to define informality in the most comparable way across countries, adopting the legalistic view based on the receipt of social security by formal sector employees – and consistent with the payment of taxes and social security contributions by these workers. Our third and main contribution derives from a methodological perspective. We exploit the availability of large (rotating) panels to estimate the informal wage penalty along the distribution while accounting for workers' unobserved characteristics. Previous attempts have relied on the application of instrumental variable techniques to quantile regression (IV-QR), as suggested for instance by Buchinsky (1998). Yet, finding convincing instruments for the sector selection is not an easy task. We suggest here an alternative approach relying on time variation in individual wages and labor market sector histories. The idea is simply to use 'fixed effects' panel regressions at different points of the wage distribution. To our knowledge, this paper is the first application of the fixed effects quantile regression estimators of Koenker (2004) and Canay (2008) to the measure of the informal wage gap.[5]

New results complete the existing literature for the three countries under investigation. Despite potential differences in the functioning of the labor markets and the nature of informality across countries, interesting similarities are revealed. In all three countries, we observe a raw wage penalty for informal salary workers throughout the wage distribution, partly explained by 'better' observed and unobserved characteristics in the formal sector. Yet a significant penalty remains at the bottom while it tends to disappear at the top in all three countries. Finally, we investigate between-group heterogeneity and time variations of the wage gap. We show in particular that informal wage penalties primarily concern younger workers; also, fluctuations of the wage gap are smoothed out when accounting for unobserved heterogeneity and reveal the procyclical nature of the informal sector penalty. We discuss the policy implications of these results.

The paper is organized as follows. Section 2 briefly presents the labor markets in the three countries under study and the related literature on informality. Section 3 describes the data, the identification of informality and the construction of net wages. The econometric approach is detailed in section 4. Section 5 discusses the empirical results and section 6 reports robustness checks and extensions. Section 7 concludes.

[4] The role of taxes on employment and wages in formal and informal sectors has recently received attention in theoretical work (e.g., Albrecht et al., 2006) and empirical studies (e.g., Badaoui et al., 2008, for South Africa).

[5] Koenker's estimator is used by Lamarche (2006) to evaluate private school vouchers and by Bargain and Melly (2007) to gauge the public sector wage gap in France.

2 Labor Markets and Informality in Brazil, South Africa and Mexico

The question of informality has received a lot of attention in the literature. A large amount of evidence is summarized in Leontaridi (1998), Perry et al. (2007), Jütting et al. (2007), Ruffer and Knight (2007), among others. In this section, we simply provide a brief background description for each country. We show that the evidence on wage differentials and inter-sector mobility is mixed.

In *Brazil*, the informal sector comprises more than ten million firms; 70% of them are located in local commerce and small services (cf., Informal Urban Economy Survey 2003). The stringent labor legislation is usually blamed for the large informal sector, especially following the 1988 constitutional changes (Barros and Corseuil, 2001). Several macroeconomic crises, with alternating periods of recession and high inflation, may have also contributed to the expansion of the informal sector – the latter accounts for 87% of the jobs created between 1992 and 2002.[6] For the recent period, the Monthly Employment Survey indicates that informal employment remains high, with a share of total employment fluctuating between 30% and 35% over 2002-2005.

Carneiro and Henley (2001) and Menezes-Filho et al. (2004) show that for some workers, the informal sector may be a desirable form of employment in Brazil; they also find that the large informal wage gap can be explained by selection bias and consequently favor the competitive markets hypothesis. This view seems to be supported by studies on sectoral mobility. Barros et al. (1990) find high mobility rates between sectors in the Sao Paulo region while Ruffer and Knight (2007) argue that there cannot be wage segmentation if there is such free mobility between sectors. In contrast, other studies report evidence of significant wage differentials – that may favor the segmentation hypothesis – in the lower part of the wage distribution (Tannuri-Pianto and Pianto, 2002).

South Africa is somewhat different from Latin American or other African countries. It is indeed characterized by a relatively small informal sector which coexists with high unemployment. According to Kingdon and Knight (2007), the overall proportions of informal employment and unemployment are estimated to 24% and 29% respectively in 2003; focusing on salary workers, Badaoui et al. (2008) evaluate the informal sector at 11% of the total labor force. The relatively small size of the informal sector is partly on account of the potential hidden costs in the high-tier informal segment (due in particular to land/credit constraints, inhibition of entrepreneurial skills and high crime rate against self-employed owners, cf. Fields, 2006). Another reason is that reservation wages may be higher in South Africa than in lower income countries – the unemployed who receive some support from within or beyond the household may prefer to remain outside the low-tier informal sector where real income is very low (Kingdon and Knight, 2001).

Several authors point toward sharp segmentation between the formal and informal segments of the labor market (Hofmeyr, 2002, Kingdon and Knight, 2007), highlighting the role of trade unions, collective bargaining and labor standards (work hours, minimum wages) in 'registered' employment. Informal sector wages, being more subject to market forces, are about 60% lower according to Kingdon and Knight (2007). Yet informality seems to be a rather dynamic segment of the South African labor market according to some other studies. For the region of KwaZulu-Natal, Valodia et al. (2006) and Cichello et al. (2005),

[6] Trade liberalization in the early 1990s must have put some pressure on the tradable good sector, resulting in large movements of labor out of the (formal) manufacturing sector and into the informal part of the service sector, with relatively contained unemployment (Hoek, 2007).

find that for many workers, the informal sector has generated more employment and shown faster wage progression in the 1990s.

The study of informality in *Mexico* has received a lot of attention in the literature. Marcouiller et al. (1997) show that this sector represents 31% of total employment when defined according to firm size but more than 43% when the social security definition is used. Maloney (1999) reports that movement from formal salaried work to self-employment is associated with a wage increase. Studying mobility patterns across business cycles for Argentina, Brazil and Mexico, Bosch and Maloney (2007) suggest that a substantial part of the informal sector, particularly the self-employed, likely corresponds to voluntary entry while informal salaried work may correspond more closely to the standard queuing view, especially for younger workers. Gong et al. (2004) find that entry and exit rates for the formal sector are lower than for the informal sector; the probability of formal sector employment increases with the education level, possibly in response to higher returns to education attached to formal jobs. Gong and van Soest (2002) confirm this view, suggesting that the dual structure is supported for highly educated workers but not for low-educated ones. They also find that the lagged sector state does not affect current wages, once wage differentials and unobserved heterogeneity are accounted for.

3 Measuring the Raw Wage Gap

3.1 Data

For *Brazil*, we make use of the Monthly Employment Survey (*Pesquisa Mensal de Emprego*, PME) conducted by the Brazilian Statistical Agency (*Instituto Brasileiro de Geografiae Estatistica*, IBGE). This household survey covers 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, hence workers are observed at most twice over a two-year period. We use the first and fifth interviews, creating a panel with observations that are a year apart. We focus on years 2002 to 2007. For *South Africa*, we use the labor Force Survey (LFS), a rotating panel conducted by Statistics South Africa (Stats SA) and covering all provincial areas, both urban and rural. 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 rely on the National Occupation and Employment Survey (ENOE) conducted by the *Instituto Nacional de Estadistica, Geographica e Informatica* (INEGI). This is a quarterly survey with a rotation scheme of 20%, i.e., workers are observed at most five times over a five-quarter period. The ENOE is a recent version of the *Encuesta Nacional de Empleo Urbano* (ENEU) which now includes information on rural areas. 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. Also, households are identified over time but individuals are not. Therefore, we construct the individual panel by linking individuals within households over time on the basis of gender, race and age. The attrition resulting from this procedure is relatively high in Brazil (58%) and smaller in South Africa and Mexico (28% and 32% respectively). While 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, we find that sample attrition does not relate to labor market status (see also Antman and McKenzie, 2007, for Mexico and Badaoui et al., 2008, for South Africa).

3.2 Defining Informality

An important aspect concerning the data is the possibility to identify informality in a fairly consistent way. There is generally no consensus on how to define the phenomenon of labor market informality in developing economies but most studies opt for either the *productive* view (based on job types or firm size) or the so-called *legalistic* or *social protectionist* view. In the latter, informality corresponds to the lack/avoidance of formal registration, taxation, regulation of maximum working hours or worker protection standards. These aspects are important for welfare considerations as informal sector workers may experience bad work conditions (e.g., no social protection) at the same time as lower wages. We opt here for the legalistic view using definitions which are as comparable as possible across countries.[7]

For *Brazil*, the PME does not have explicit information on benefits but workers are asked whether they hold a formal/registered labor contract (i.e., have a signed labor card or *carteira assinada*). This contract entitles them to receive state-mandated benefits such as medical coverage and a pension. Workers whose job is not regulated by a formal labor contract are then classified as belonging to the informal sector. Similar choices are made by Amuedo-Dorantes (2004) and Tannuri-Pianto and Pianto (2002) using the 1999 Brazilian household survey. The latter study and Henley et al. (2007) show that this simple definition seems to capture some of the other features commonly used when defining informality (including firm size and job types). For *South Africa*, the LFS contains several questions regarding fringe benefits and other aspects of the job that can be used to identify the sector, in particular questions regarding whether the firm provides medical aid and deducts unemployment insurance contributions. This measure of informality significantly overlaps with the self-reported status also provided in the data. The informal sector in *Mexico* is frequently defined along the productive view, both in recent studies (e.g., Maloney, 1999, Gong and van Soest, 2002) and by Mexican authorities. In contrast, and to improve comparability with other countries, we opt for a characterization more in line with the legalistic view and based upon whether employees contribute to (and benefit from) social security (see also Martin, 1999, Bosch and Maloney, 2007, 2008, Gong et al., 2004).

3.3 Sample Selection

We restrict samples to urban male workers aged 15-65 and not engaged in any form of education or training. We focus on men because a large proportion of women in the three countries under study 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., 2004). 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 restrict the sample to workers that are observed at least twice in the data and whose observations are consecutive over the periods of the

[7] The challenge arising from the difficulty to define informality in a uniform fashion given the non-uniformity of the data sources and the more fundamental differences across labor markets is discussed in other comparative studies like Marcouiller et al., (1997), Duryea et al. (2007), Bosch and Maloney (2007), Jütting et al. (2007) and Perry et al. (2006).

survey.[8] In South Africa (resp. Brazil), whites and asians (resp. asians) are excluded from the sample as they represent less than 1% of the informal sector. Results do not change significantly when including these groups.

An important step in the selection is the focus on salary workers only, a choice not specific to the present study (see for instance Badaoui et al., 2008). We argue that self-employed workers form a vastly heterogeneous group, from street vendor to professional independent workers, and deserve a particular study. Also, self-employment income is typically subject to substantial measurement error and incorporates other elements (e.g., returns to risk) that would not be included in wages.

This selection leaves a sample size of *13,710* men for Brazil, *9,099* men for South Africa, and *100,868* men for Mexico. Summary statistics are reported in Table 2 in the Appendix and discussed below.

3.4 Adjusted Wages

Real hourly wages are calculated from the gross monthly wages and reported work hours in the primary job. For the sake of comparability between countries and over time, earnings are converted into 2002 international dollars using relevant CPI deflators and PPP adjustment factors drawn from the World Development Indicators. The premium associated with formal sector employment is overestimated if taxes paid by registered workers are ignored. Thus we use available information to adjust gross wages for taxes in this sector, which is consistent with the chosen definition of formality.[9]

Adjusting for taxes is sometimes seen as a difficult exercise because of data limitation. Yet we argue that the datasets at hand and the nature of the tax systems in the countries under study allow for a reasonable approximation of the taxes paid on labor income. Tax rules are summarized in Table 1 in the Appendix.[10] The tax system is progressive in all three countries but the top marginal tax rates are not very high by international standards in Brazil (27.5%) and Mexico (28%). A flat rebate (depending on age) is applied in South Africa while in Mexico, a refundable (and progressive) tax credit benefits those earning less than twice the minimum wage. In these countries, income taxation is purely individualized while in Brazil, taxpayers can also file jointly and benefit from a deduction for each dependent relative (spouse, if inactive, and children aged under 22, or 25 if in education). We have used available information on family links for the main adults in the household and assumed that other adults were single. For the latter, we thus potentially overestimate tax liabilities; yet most of them are young workers with low wages, and hence likely exempt from tax payment (as detailed below, only the top 20% of the gross wage distribution is liable for income tax). Another usual limitation to tax calculation is the absence of information concerning capital income, which is therefore excluded from the tax base in our simulations. This should concern only a limited number of people at the very top of the distribution.

[8] In the final selection, all Brazilian workers are observed only twice; for South Africa and Mexico, respectively 40% and 80% of the workers are observed at least three times.

[9] As often, it is difficult to evaluate the value of medical coverage for formal sector employees and almost impossible to account for the present value of future benefits such as pensions. Consistently, we do not adjust wages for social security contributions, arguing that social security coverage can be seen as a pure insurance mechanism. We have nonetheless calculated these contributions in order to compute more accurately taxable income, often based on gross income minus part of social security contributions.

[10] Detailed descriptions of the tax-benefit systems in force in South Africa, Brazil and Mexico are available from the South African Revenue Service (http://www.sars.gov.za), Immervoll et al (2007) and Absalón and Urzúa (2008) respectively. The precise description of the imputation process adopted in the present study is available upon request.

3.5 Data Description

Tables 2 and 3 in the Appendix describe the selected sample as homogenously as possible across countries. Table 2 shows that wages are on average larger in the formal sector in all three countries, with a larger average gap in South Africa. Using previous definitions, we find that informality as a fraction of total salary work is large in Mexico (43% of our selected sample) and more modest in South Africa (around 11%) and Brazil (15%). For Brazil in particular, this is lower than the share reported in Section 2 (30%) because of the selection – indeed, women and self-employed, excluded from our final sample, are disproportionately represented in the informal sector. In terms of location in the wage distribution, we find that informal workers are more concentrated in lower quantiles in South Africa and Brazil; in Mexico, workers of informal and formal sectors are more evenly distributed.

We also estimate the propensity to be in the informal sector using a simple probit model. Results reported in table 3 point toward a U-shaped relationship between age and informality, that is, the young and the old workers are more likely to be in the informal sector. In South Africa and Mexico, the probability of being formal increases with education. For Brazil, only secondary schooling or higher (i.e., more than 11 years of schooling) guarantees a significantly smaller probability of being informal. A weaker link between informality and lower educational attainment in this country is also reported in Henley et al. (2007).

Figure A1 in the Appendix reports the results of our tax calculations in the form of average tax rates faced by workers at different points of the gross wage distribution. It clearly shows that the redistributive effect of taxes in Brazil and South Africa is limited to the top of the distribution. The progressive effect is more substantial in Mexico thanks to the refundable tax credit which subsidizes the first 70% of the formal sector distribution. Positive taxation kicks in for all three countries at about the same level, i.e., around 1.2-1.3 times the median wage. As a result, taxation is responsible for slightly reducing the informal wage penalty for the top quarter of the distribution in all three countries, while it actually increases it for the first 70% in Mexico.

4 Econometric Approach

We first estimate standard Mincer wage equations at the mean and at various quantiles using pooled years data for each country. Explanatory variables comprise standard human capital information (age, age squared, education) and other individual/household characteristics as reported in table 2 (race, number of children, marital status, region). Ideally we would like to compare formal and informal sector workers on a like-for-like basis *within* a certain industry. Because this would reduce sample size too much, we conduct estimations on the whole selection of workers and simply add broad industry dummies to control for the possible structural differences between formal and informal sectors.

Next, we rely on panel data to identify (time-invariant) unobserved heterogeneity. We first estimate a fixed effects model on (unbalanced) panel data for each country and compare the result to standard OLS. Denote I the informal sector dummy, x_{it} a set of controls, 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 model is simply written:

$$y_{it} = \alpha_i + \gamma_t + x_{it}\beta + I_{it}\delta + \varepsilon_{it}$$

where $E\left[\varepsilon_{it}\left|\alpha_i, x_{it}, I_{it}\right.\right] = 0$ for all individuals i and periods t. The fixed effects (FE) estimator is consistent even if unobserved characteristics are correlated with both selection and wages, as long as those characteristics are constant over time.[11] The estimated coefficient $\widehat{\delta}$ is interpreted as a measure of the informal sector wage penalty. Using the 'stayers' as the reference group, this wage penalty is derived from the groups of people moving in or out of the informal sector. The intuition for the identification of the wage gap is best illustrated with a simple two-period example. Assume that, with a strictly positive probability, some individuals move from the informal sector to the formal sector and others move in the opposite direction between period 1 and period 2. Asymptotically, we can observe:

$$\begin{aligned}
E\left[y_{i2} - y_{i1}|I_{i1} = k, I_{i2} = k\right] &= \Delta \text{ for } k = 0,1 \\
E\left[y_{i2} - y_{i1}|I_{i1} = 0, I_{i2} = 1\right] &= \Delta + \delta \\
E\left[y_{i2} - y_{i1}|I_{i1} = 1, I_{i2} = 0\right] &= \Delta - \delta \\
\text{with } \Delta &= \gamma_2 - \gamma_1 + (x_{i2} - x_{i1})\beta.
\end{aligned}$$

Identification on the population of 'movers' (second and third lines above) is standard. Nonetheless, we provide important checks regarding the frequency of the moves and the nature of the movers (section 5); we also verify that recorded moves correspond to genuine job changes rather than to measurement error (section 6). We assume for now that the wage penalty δ is constant over time but relax this assumption later on (section 6).

Next, we consider the extension of the standard QR model to longitudinal data. For any worker i, we can write the τ^{th} quantile of the y distribution conditionally on observables as:

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

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). 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. Importantly, however, the effects of the covariates are permitted to depend on the quantile of interest, in particular the informal sector premium/penalty $\delta(\tau)$. Following Koenker (2004), we can estimate this model for several quantiles simultaneously by solving:

$$\min_{\alpha,\beta,\gamma,\delta} \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{t=1}^{T} w_j \rho_{\tau_j}(y_i - \alpha_i - \gamma_t(\tau_j) - x_{it}\beta(\tau_j) - I_{it}\delta(\tau_j)) \tag{1}$$

[11] We believe that this is an acceptable approximation for the short panels used in the present study, even if it does not cover the whole range of possible unobserved characteristics (see the discussion in the concluding section). Arguably, all methods have their limitations. Simple OLS and QR estimations assume that the choice of sector is random. IV approaches aim to relax this assumption but face the usual problem of finding instruments which explain sector choice and are uncorrelated with wage determination. The FE estimator used here allows correlation between sector choice and wages at a given period but does not account for the possibility that workers change sector in response to a change in unobservables (ex: following anticipated shocks on individual or sector-specific productivity). That is, sector changes observed in the panel must be random to some extent. Lemieux (2008) suggests considering only involuntary job changers, i.e., people who changed jobs because of plant closing, family responsibilities, dismissal, etc. We refrain from doing so primarily because the reasons for job changes are usually unknown. Note however that better wage prospects only partly explain why people move. We find that transitions (from formal to informal sector or in the other direction) are associated with wage increases for some workers only, mainly grouped in the upper part of the distribution; transitions can be accompanied by wage losses for others, especially at the bottom and for those leaving the formal sector. These additional results are available upon request.

where $\rho_\tau(u) = u(\tau - 1(u \leq 0))$ denotes the quantile loss function, with $1(\cdot)$ the indicator function. The w weights control for the relative influence of the J quantiles on the estimation of the fixed effects (in the application we simply use Tukey's trimean weights: $w_j = 0.5 - |j - 0.5|$). As the dimensions of this problem are extremely large, it is not possible to time-demean the data as often done for FE estimations. Fortunately, the design matrix has a very sparse structure – the majority of its elements are equal to 0 – and can be handled by the algorithm of Koenker and Ng (2005). An alternative and simpler approach to estimate fixed effects quantile regression (FE-QR) has been recently suggested by Canay (2008). It 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 $\widehat{\alpha}_i$ in order to regress corrected wages $\widehat{y}_i = y_i - \widehat{\alpha}_i$ on the other covariates by traditional QR. We found that both methods lead to very similar results (a detailed comparison is available upon request).

5 Empirical Results

5.1 Characterizing the Movers

As explained above, movers play a key role in the identification of the wage gap and deserve some attention. We first check that the number of transitions across sectors is large enough for a valid use of the FE estimator. We find that 8% of panel observations in Brazil, 12% in South Africa and 24% in Mexico correspond to sector changes, which are reassuring numbers regarding the possibility to identify fixed effects.[12] In section 6, we verify that these transitions are associated with job changes and are not driven by measurement error.

We also check that movers are not too specific. Firstly, we verify that transitions across sectors are large enough at quantile levels and are not restricted to certain groups of workers. Figure A2 in the Appendix depicts the number of movers in and out of the informal sector between two periods, averaged over the different waves of the panel. For the sake of comparability, it is expressed as a proportion of the size of base-period informal sector quintiles. It turns out that a substantial number of workers move in both directions and do so at all earnings levels. Transitions are slightly more frequent in the upper quintiles in South Africa and Mexico and occur more often from informal to formal sector (especially in Mexico and in lower quintiles in Brazil).[13] Overall, however, they do not seem to be overly concentrated at certain levels of the wage distribution. Secondly, we characterize the movers by running additional probits (dependent variable equals to one if the worker moves). It turns out that movers are not extremely different from the overall selected population in terms of their observed characteristics (pseudo-R2 are around 0.02 for Brazil, 0.06 for South Africa and 0.01 for Mexico).[14] We also find that those moving from formal to informal sectors are not significantly different from workers going in the other direction.

[12] The lower rate for Brazil partly translates the fact that workers can move at most once, as explained in the data section. When ignoring workers who move more than once in the two other countries, flows become more comparable.

[13] We refrain from drawing any conclusions – in particular about why people move – 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 (1997) and Maloney (1999).

[14] Only a few characteristics are significant. Movers seem to be younger and more often single. Moves occur more frequently within certain sectors (e.g., construction and trade in Brazil). There is no clear evidence for the role of education.

5.2 Estimation Results

Our main results are represented graphically in the rest of this section. For each country, we report the estimated coefficient $\widehat{\delta}$ (the informal sector wage penalty) from OLS, QR, FE and FE-QR. In table 4 in the Appendix, we also report the wage penalty at the mean, the median and two extreme quantiles as well as the (bootstrapped) standard errors.[15]

Brazil

Figure 1 confirms the existence of an informal sector wage penalty for Brazil. It shows that pooled QR estimates are not contained in the confidence interval surrounding the OLS coefficient and reveal important differences along the wage distribution. In brief, the penalty faced by informal sector workers is larger at lower quantiles and disappears at the top. Estimates of the FE-QR give similar results but display a smaller penalty along the whole distribution.[16] Precisely, the (conditional) wage penalty for informal workers ranges between 8% at the bottom (compared to 17.5% with pooled QR) to around zero at the top. Interestingly, this pattern is qualitatively close to that in Tannuri-Pianto and Pianto (2002) who use an IV-QR approach – we discuss this point further in the concluding section. Yet we find that the penalty is relatively small for most workers and even in lower quantiles. Hence our results are in line with previous studies which cast doubt upon the presence of segmentation in this country (see section 2).

South Africa

For South Africa, figure 2 reports a wage penalty of around 69% at the median according to QR on the pooled sample, slightly larger than the average penalty. It is in line with recent results of Kingdon and Knight (2007). When accounting for unobserved heterogeneity using FE, the (conditional) wage penalty decreases at all levels, down to around 18% on average.[17] The wage gap is not uniform and a similar pattern to Brazil emerges: the conditional wage gap is very moderate at the top but larger at the bottom. The similarity between South Africa and the two other countries in this respect is interesting and somewhat in contrast with the idea that this country is very specific because of a relatively smaller informal sector and the presence of unemployment. Actually, the penalty faced by informal sector workers is especially large at lower quantiles in this country. For these workers, our results are therefore consistent with some of the evidence presented in section 2 concluding at potential labor market segmentation – but this may not apply to workers at the top of the distribution.

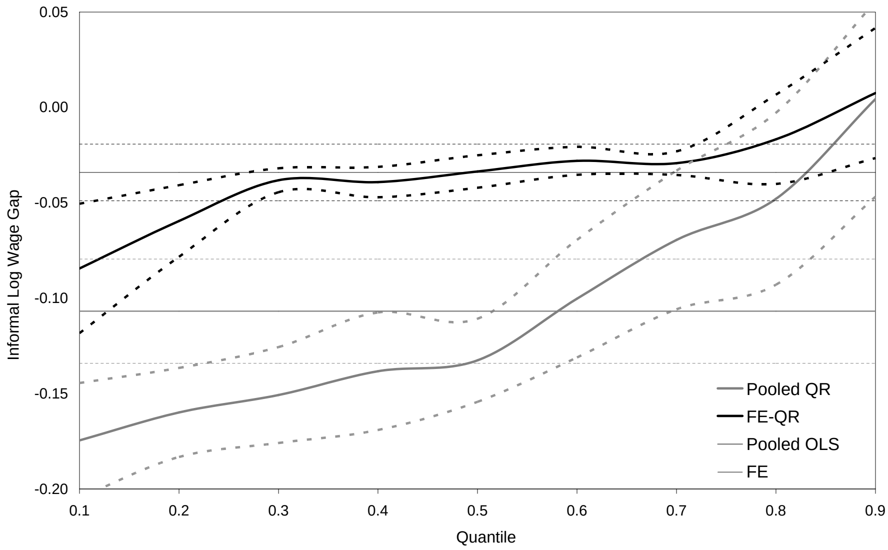
Mexico

Results for Mexico are presented in figure 3. Pooled QR give very similar estimates to what we obtain for Brazil, with a wage penalty ranging between 30% at the bottom and 5% at the top. Like in Brazil and

[15] Because of space limitation, we have not reported the full estimation tables, which are available upon request. Their findings can be summarized as follows. Returns to education typically increase with the education level. Returns to experience (here proxied by age) generally increase as we move to higher quantiles; the same is true for education with a few exceptions (i.e., at lower education levels in Mexico and for university education in South Africa). Many interpretations are possible: higher ability workers may benefit from higher school quality or obtain higher returns to a given experience/education level. More country-specific results also appear, for instance regional differences (e.g., workers in Sao Paulo benefit from higher pay) and differences by race in Brazil and South Africa. Results of FE regressions are less easy to comment upon since only time-variant regressors are included.

[16] Note that standard errors are smaller when using standard FE estimations since many less variables are used.

[17] This result is also found in Badaoui et al. (2008). Yet these authors find a smaller average wage gap as they consider the earlier period of 2001-2003. This is consistent with our results on the time-varying wage gap in the next section.



Note: bootstrapped 95% confidence intervals represented by dashed lines.

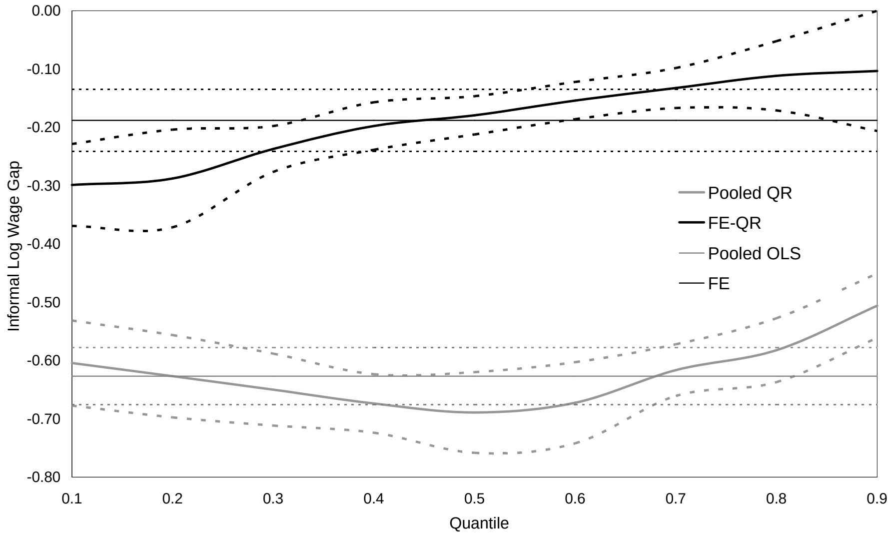
Figure 1: Fixed Effects Estimations for Brazil

South Africa, accounting for FE considerably decreases the extent of the penalty. The remaining wage gap tends to disappear at the top while it is significant and still large (around 15%) at the bottom. Our results thus indicate that the nature of the Mexican labor market is not genuinely different compared to the two other countries under consideration. This conclusion is not to be opposed to other results describing the informal sector as a desirable segment of the Mexican labor market and reporting informal sector wage premia (e.g., Maloney, 1999, 2004). Indeed, these studies focus on another group, namely the self-employed workers, for whom labor market conditions are substantially different.

6 Robustness Checks and Extensions

6.1 Job Movers

Admittedly, inter-sector moves could reflect mere measurement error, i.e., flaws in reporting the correct sector status at certain periods. To check this point, we first verify whether sectoral transitions are accompanied by actual job changes, as indicated by changes in occupation type, industry type, firm size or tenure. Table 5 in the Appendix shows that of all inter-sector moves (which potentially include several moves per worker over the relevant period), 75% in Brazil, 87% in South Africa and 80% in Mexico are accompanied by a change in *at least* one of these characteristics. Notably, a third of sectoral moves in South Africa and Mexico are concomitant with changes in firm size only, which does not fully guarantee that actual job change has occurred. However, even if the worker does not actually move to a different firm, a dramatic change in firm size/organization might be treated as such. Indeed, genuine changes in the formal/informal nature of jobs may still occur. As it expands, a firm becomes more at risk of being caught defaulting on stipulated regulation and is therefore more likely to register its workers. At the same time, it may also change its wage setting policy.



Note: bootstrapped 95% confidence intervals represented by dashed lines.

Figure 2: Fixed Effects Estimations for South Africa

Furthermore, we aim to ensure that our results are robust to measurement error and we reestimate FE-QR solely on *job movers*. According to figure 4, results appear not to be fundamentally different in this case. We further restrict estimations to job moves which are not solely defined by a firm size change; we find that results do not change dramatically in Brazil and Mexico. The informal wage gap tends however to increase in South Africa, especially in the upper part.

6.2 A Closer Look at the Top

One may genuinely wonder why informal sector workers at the top of the distribution do not face a large wage penalty. To characterize these workers, we run a probit on the population of informal workers where the binary variable takes a value of one if the worker is in the top quintile. In all countries, the top paid are significantly older, are more often located in economically active areas (e.g., the Sao Paulo region in Brazil), generally have higher education levels (with the exception of South Africa where they more often hold a vocational degree) and more often hold managerial or administrative positions. These simple descriptive facts convey that at the top of the distribution, informal sector workers share similar characteristics with their formal sector counterparts but are nonetheless categorized in the informal sector according to the legalistic definition at use.

Badaoui et al. (2007) theoretically show that the informal sector wage penalty is essentially due to a firm size effect: larger firms pay higher wages and, at the same time, have higher incentives to be registered since they are more likely to be caught defaulting. Probit estimations show that compared to their formal sector counterparts, informal workers at the top of the distribution tend be in smaller firms in South Africa and Mexico.[18] In Brazil, however, these workers tend to be more often in large

[18] In South Africa, half of top informal workers are located in firms of less than 4 employees compared to 3% of their formal sector counterparts. In Mexico, 18% of them work in firms of more than 16 employees compared to 72% of top formal sector employees.

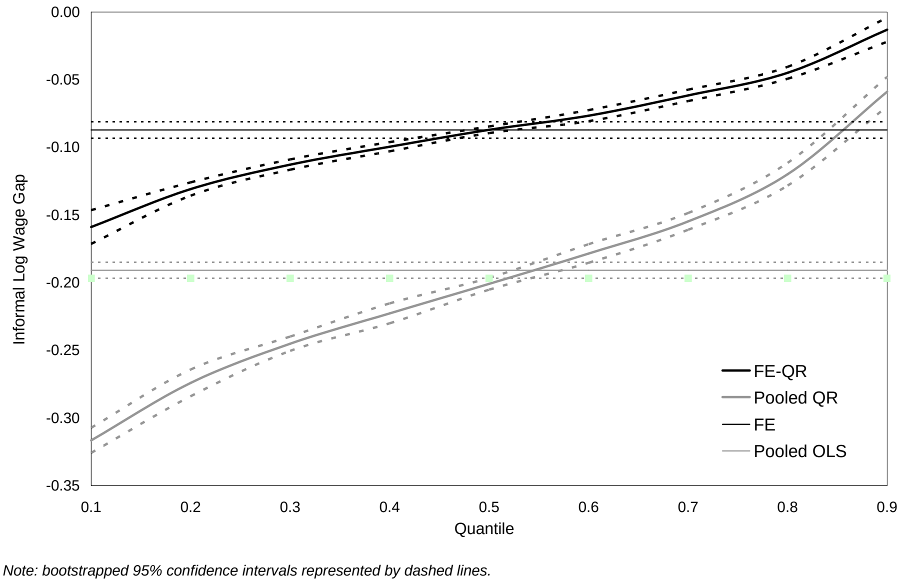


Figure 3: Fixed Effects Estimations for Mexico

firms and less frequently in very small firms, making them even more comparable to top formal sector employees.[19] Consequently, if the firm size explanation is valid, we should expect *some* penalty at the top of the distribution in South Africa and Mexico but not in Brazil. Corrected measures of the informal wage gap reported in figure 4 point in that direction. Thus, our results seem to support empirically the prediction made in Badaoui et al. (2007), not only on average but also for top wage workers.[20]

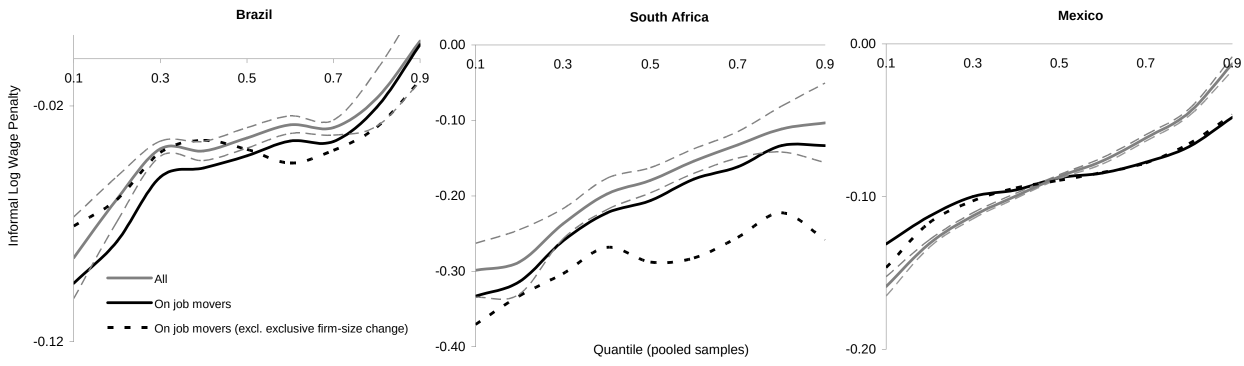
6.3 Between-group and Time Variations

The FE-QR model simply uses a dummy variable for the informal sector and may be seen as misspecified. While it is well known that, in case of misspecification, least square regression provides a minimum mean squared error linear approximation to the true functions, Angrist et al. (2006) provide a similar result for quantile regression. Therefore our findings have meaningful interpretation even if the true informal wage penalty depends on the covariates.

Nonetheless, it is desirable to relax the assumption that returns to education and experience are identical in the two sectors. We examine the heterogeneity of the informal wage penalty by simply interacting it with workers' age and education levels. Figure 5 essentially shows that younger workers

[19] According to Kenyon and Kapaz (2005), tax evasion in Brazil is not limited to small and medium-size enterprises, as is commonly believed. Even large and very large firms report only moderate compliance. Note also that in our data, around 86% (resp. 96%) of informal (resp. formal) sector workers in the top quintile are located in firms with 11 or more employees, compared to 53% for informal workers in lower quintiles.

[20] We also find that informal workers in the lower part of the distribution tend to work in small firms compared to their formal sector counterparts, which is consistent with the significant penalty that we have reported for these workers. This also indicates some overlap between the legalistic definition and the firm-size definition. Yet, findings for the top workers in Brazil give support to the shift in the literature (and in ILO practice) from the productive view to the legalistic view. Indeed, the latter recognizes that informal employment can also be found in large firms (Perry et al., 2006). Note that we refrain from using firm size as an explanatory variable for obvious reasons regarding endogeneity issues; also, restricting the estimations to workers located in small firms would pose a problem of sample size and representativity since few formal sector employees do work in small firms.



Note: bootstrapped 95% confidence intervals for the FE-QR on the whole selected sample represented in grey dashed lines.

Figure 4: Robustness Check: Estimations on Job Movers

face larger penalties, especially in Brazil and South Africa. This is in line with previous results by Bosch and Maloney (2007) who suggest that informal salaried work may correspond more closely to the standard queuing view for younger workers. Education levels seem to affect the wage gap only at the two extremes of the distribution. At the top, the informal wage penalty is smaller in all countries – and even turns into a premium in Brazil – for those with higher education. This could be related to the fact that some of the top informal workers share similar characteristics with their formal sector counterparts, as previously discussed. In the lower part of the distribution, a larger penalty is observed for high education groups in Brazil and Mexico. This possibly reflects that education has a higher return in the formal sector, either because it acts as a signaling device or because this sector is capital-intensive and highly rewarded as a complement to capital inputs (Gong and Van Soest, 2002).

We have assumed so far that the penalty is constant over time. Yet it may be necessary to relax this assumption as our data spans several years (up to 7 years for South Africa). We can estimate a time-varying wage gap δ_t by simply interacting the informal sector dummy with year dummies. In Figure 6, we report the penalty estimated by FE-QR for quantiles .2, .5 and .8. The trend appears to be very stable in Mexico, with a median penalty around 9% over the period 2004-2007. For Brazil, the informal wage penalty slightly oscillates around 5% at the median. Yet the period of economic slowdown of the early 2000s is characterized by a smaller penalty (and even a premium in the upper half of the distribution), reflecting higher wage depression in the formal sector. In South Africa, the median gap stays close to around 16% in the first years of the 2000s but doubles in more recent years characterized by better economic conditions (declining unemployment from 2004 onwards and faster economic growth with a peak in 2006).

Thus, it seems that wages are more procyclical in the formal sector. This interpretation is consistent with the fact that workers in this sector benefit from the surge in the prices of export commodities, as opposed to informal sector workers employed more frequently in domestic activities like construction and small services.[21] It is also in apparent contrast with the view that formal sector wages are less responsive to market forces because of labor market regulations like minimum wages. However, evidence for Brazil in figure 6 is consistent with the fact that downward rigidity in the formal sector occurs in lower quantiles

[21] However, this description is static. A more in-depth analysis would require accounting for the effect of the business cycle on inter-sector flows (i.e., the size of the informal sector) – see Bosch and Maloney (2007, 2008).

(the penalty is relatively constant over time for the 10th centile) while formal sector wages (and hence the wage gap) are procyclical in the second half of the distribution (captured by quantiles .5 and .8 here). Evidence for South Africa is consistent with the fact that formal sector workers in upper quantiles have been the first to benefit from the economic upturn starting 2004 while workers at lower wage levels catch up later. Another consequence for both countries is the compression of the formal sector wage distribution – resulting in smaller differences in the penalty between top and low quantiles – when economic conditions improve.

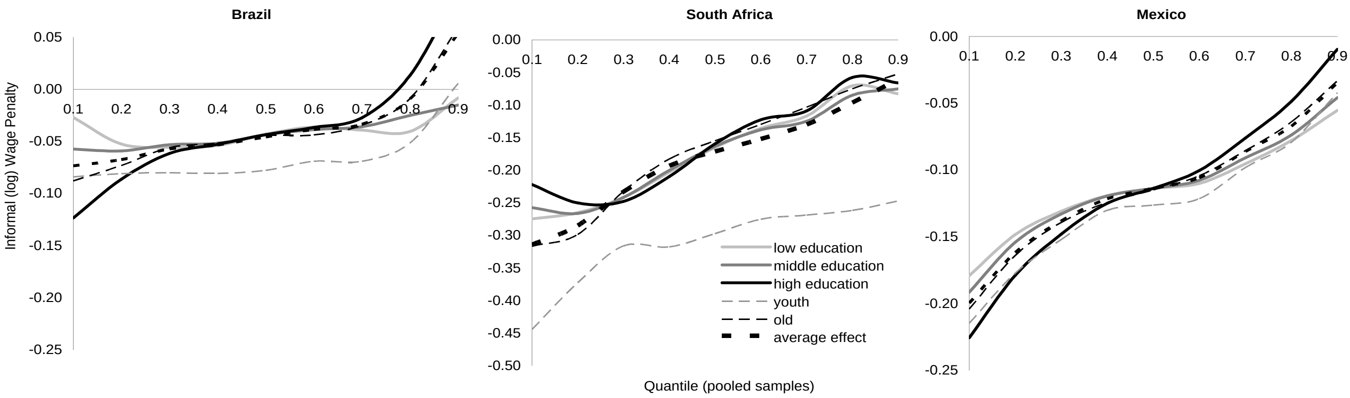


Figure 5: Informal Wage Penalty: Interactions

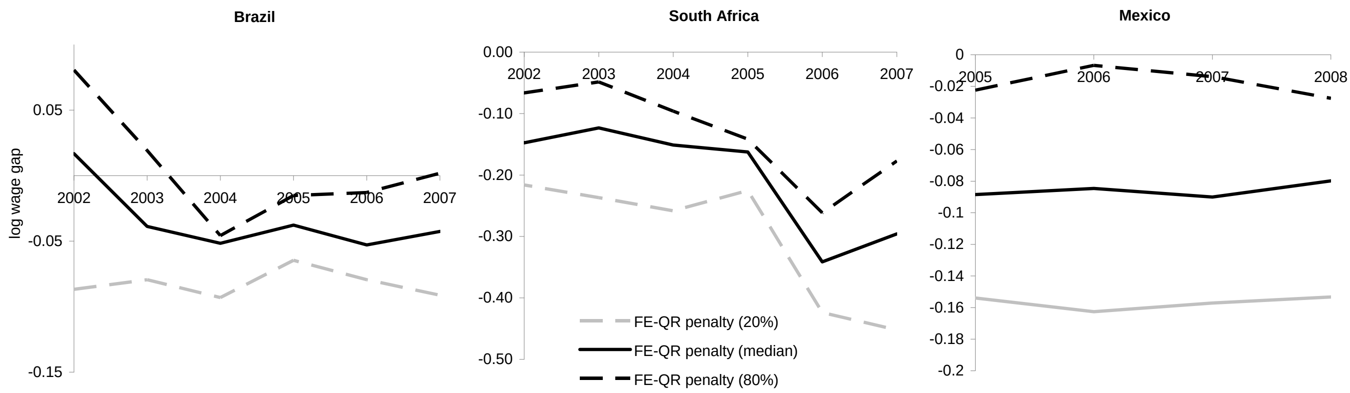


Figure 6: Informal Wage Penalty: Time Variations

7 Concluding Discussion

This study complements the existing literature on informality by measuring the informal sector wage gap in Brazil, South Africa and Mexico. Reported earnings have been adjusted to account for taxes paid in the formal sector. Fixed effects quantile estimations are used to perform a distributional analysis while accounting for workers' unobserved heterogeneity. A few interesting conclusions and policy implications are derived from this exercise. Firstly, our results conform to the stylized fact that workers in the formal sector have 'better' observed characteristics at all wage levels. They also point toward better unobserved characteristics, which seem to play an important role in explaining the informal sector wage penalty. Secondly, we illustrate the importance of distributional analyses. Standard measures of the informal sector penalty at the mean fail to capture the important within-group heterogeneity found in our results. Interestingly, the distributional pattern obtained by FE-QR is qualitatively similar across all countries. Precisely, most of the wage gap at the top of the distribution disappears. In the lower part, and for younger workers at every quantile, large (unexplained) informal wage penalties remain and *could* be consistent with some segmentation for these workers (see Tannuri-Pianto and Pianto, 2002). The fact that the wage penalty is not constant along the distribution shows that policies aimed to levy labor market regulations should not be applied in a blanket fashion. The key to better functioning and more equitable labor markets may also pertain to additional efforts towards building workers' capabilities.

Some of the limitations of the present study are well-known. In particular, wage gap measures are only part of a more complete welfare analysis. As Badaoui et al. (2008), we have attempted to account for taxes to improve the rending of financial situations in the formal sector. Yet some efforts should be made to account for other cash or non-pecuniary advantages attached to a particular sector. This is a considerable challenge, given data limitation and the difficulty to measure welfare (for instance to impute future benefits like pensions). Still, this is an important one. Bourguignon et al. (2007) show for instance that what makes the Brazilian distribution of income so unequal, in addition to the structure of returns to human capital, is the poor access to non-labor incomes like pensions.

Another issue concerns the potential limitations in the way unobserved heterogeneity and selection issues are accounted for. Firstly, while IV estimations face the usual problem of finding convincing instruments, the FE approach is potentially subject to measurement error, an issue carefully investigated in our analysis. Yet it is encouraging to find that for Brazil, the pattern obtained by IV-QR (cf. Tannuri-Pianto and Pianto, 2002) is relatively similar to ours when using FE-QR. More systematic comparisons of the two methods for the same country and the same period should be carried out. Also, it is possible to combine the two approaches and to explicitly estimate sector selection in a first stage. While this is technically feasible in the FE-QR context (cf. Harding and Lamarche, 2008), this would however require instruments which vary over time and hence make the search for appropriate instruments even more challenging. Finally, the range of unobservables accounted for in our approach is naturally limited to time-invariant characteristics. While we account for important dimensions that are unlikely to change over the few years of our panels (talent, risk aversion, etc.), some other unobserved worker/job characteristics may change over a short period and may concern one sector more than the other (e.g., received training). These limitations should motivate future research.

References

[1] Absalón, C. and C. M. Urzúa (2008): "LATINMOD: preliminary report on Mexico", Latinmod working paper

[2] Albrecht, J.W., A van Vuuren and S. Vroman (2004): "Decomposing the gender wage gap in the Netherlands with sample selection adjustments", IZA Discussion paper 1400

[3] Albrecht, J.W., L. Navarro and S. Vroman (2006): "The effects of labor market policies in an economy with an informal sector", IZA Discussion paper 2141

[4] Amuedo-Dorantes, C. (2004) “Determinants and Poverty Implications of Informal Sector Work in Chile”, in *Economic Development and Cultural Change*, 52(2), 349-368.

[5] Angrist, J., V. Chernozhukov and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure", *Econometrica*, 74(2), 539-563

[6] Antman, F. and D. J. McKenzie (2007) “Earnings Mobility and Measurement Error: A Pseudo-Panel Approach”, *Economic Development and Cultural Change*, 56(1):125-161.

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

[8] Badaoui, E. Strobl, E. and Walsh, F (2007): "The Formal Sector Wage Premium and Firm Size", IZA DP No. 3145

[9] Badaoui, E. Strobl, E. and Walsh, F (2008): "Is There an Informal Employment Wage Penalty? Evidence from South Africa", *Economic Development and Cultural Change*, 56, 683–710

[10] Bargain, O. and B. Melly (2007): "Public Sector Pay Gap in France: New Evidence using Panel data", IZA working paper

[11] Barros, R.P. and Corseuil, C.H. (2001), “The impact of regulations on Brazilian labor market performance”, Research Network Working Paper No. R-427, Inter-American Development Bank.

[12] Barros, R., G. Sedlacek, and S. Varanda (1990). Segmentação e mobilidade no mercado de trabalho: a carteira de trabalho em são paulo. *Pesquisa e Planejamento Econômico* 20 (1).

[13] Bosch, M. and W. Maloney (2007): "Comparative Analysis of labor Market Dynamics Using Markov Processes: An Application to Informality", World Bank.

[14] Bosch, M. and W. Maloney (2008): "Cyclical Movements in Unemployment and Informality in Developing Countries", World Bank

[15] Bourguignon, F., F. Ferreira and P. Leite (2007): "Beyond Oaxaca–Blinder: Accounting for differences in household income distributions", *Journal of Economic Inequality.*

[16] Buchinsky, M. (1998): “The dynamics of changes in the female wage distribution in the USA: A quantile regression approach,” *Journal of Applied Econometrics*, 13, 1–30.

[17] Canay, I. A. (2008): "A Note on Quantile Regression for Panel Data Models", Department of Economics, Northwestern University

[18] Carneiro,F,G, A. and A. Henley (2001) : "modeling Formal Vs Informal employment and earnings: microeconometric evidence for Brazil", *Annals of the XXIX National Meeting of Economics of ANPEC.*

[19] Cichello, P.L., G.S. Fields, M. Leibbrandt (2005): "Earnings and Employment Dynamics for Africans in Post-apartheid South Africa: A Panel Study of KwaZulu-Natal", *Journal of African Economies*, 1-48.

[20] Cunningham, W. and W.F. Maloney (2001): "Heterogeneity among Mexico’s microenterprises: an application of factor and cluster analysis", *Economic Development and Cultural Change*, 50(1) 131-56.

[21] Dickens, W. and K. Lang (1985): "A Test of the Dual labor Market Theory", *American Economic Review*, Vol. 75, No. 4, pp.792-805

[22] Duryea, S., G. Marquez, C. Pages and S. Scarpetta (2007): "For Better or for Worse? Job and Earnings Mobility in Nine Middle and Low Income Countries", mimeo IDB.

[23] Fields,G. S. (1975): "Rural-Urban Migration, Urban Unemployment and Underemployment, and job search activity in LDCs" *Journal of Development Economics,* Vol. 2, pp. 165-187.

[24] Fields,G. S. (2005): "A Guide to Multisector labor Markets Models", Paper prepared for the World Bank labor Market Conference.

[25] Fields, G. S. (2006). 'modeling labor market policy in developing countries: A selective review of the literature and needs for the future', mimeo., December, Cornell University.

[26] Funkhouser, E (1997): "Mobility and labor Market Segmentation: The Urban labor Market in El Salvador", *Economic Development and Cultural Change*, Vol. 46, No.1, pp.123-153

[27] Gindling, T. H. (1991): "Labor Market Segmentation and the Determinants of Wages in the Public, Private-Formal and Informal Sectors in San-Jose, Costa Rica", *Economic Development and Cultural Change*,Vol. 39, No. 3, pp. 585

[28] Gong, X., A. van Soest, and E. Villagomez (2004): "Mobility in the urban labor market: a panel data analysis for Mexico", *Economic Development and Cultural Change*

[29] Gong, X. and A. van Soest (2002): "Wage differentials and mobility in the urban labor market: a panel data analysis for Mexico", *Labor Economics*, Vol. 9, Issue 4, pp. 513-529

[30] Gunther, I and A. Launov (2007): "Competitive and Segmented Informal labor Markets", IZA Discussion Papers 2349, Institute for the Study of labor (IZA).

[31] Harding, M. and C. Lamarche (2008): "A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables", Stanford University, mimeo.

[32] Heckman, J., Sedlacek, G. (1985): "Heterogeneity, aggregation and market wage functions: an empirical model of self-selection in the labor market", *Journal of Political Economy* 93, 1077– 1125.

[33] Henley, A., G.R. Arabsheibani, and F.G. Carneiro (2007): "On defining and measuring the informal sector", Policy research working paper 3866, World Bank.

[34] Hoek, J. (2007): "Labor Flows in Formal and Informal labor Markets in Brazil", mimeo

[35] Hofmeyr, J.F. (2002): "The importance of segmentation in the South African labor market", University of Natal.

[36] Immervoll, H., H. Levy, J.R. Nogueira, C. O'Donoghue and R. Bezerra de Siqueira (2007): "The Impact of Brazil's Tax-Benefit System on Inequality and Poverty", IZA DP No. 2114

[37] Jütting, J., T. Xenogiani and J. Parlevliet (2007): "Work and Well-Being: Informal employment revisited", OECD Development Centre

[38] Kenyon T. and E. Kapaz (2005): "The Informality Trap: Tax Evasion, Finance, and Productivity in Brazil", Public policy for the private sector, Note 301, World Bank.

[39] Kingdon, G., and Knight, J. (2001) "Why high open unemployment and small informal sector in South Africa?" Centre for the Study of African Economies, Department of Economics, University of Oxford, October 2001.

[40] Kingdon, G. and J. Knight (2007). 'Unemployment in South Africa, 1995-2003: Causes, problems and policies', *Journal of African Economies*, forthcoming.

[41] Koenker, R. (2004): "Quantile Regression for Longitudinal Data", *Journal of Multivariate Analysis*, 91, 74-89.

[42] Koenker, R. and P.T. Ng (2005): "A Frisch-Newton Algorithm for Sparse Quantile Regression", *Acta Mathematicae Applicatae Sinica*, 21, 225-236.

[43] Lamarche, C. (2006): "Private School Vouchers and Student Achievements: A Fixed Quantile Regression Evaluation of the Milwaukee Parental Choice Program", forthcoming in *Labour Economics*.

[44] Lemieux, T.(1998): "Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection", *Journal of Labor Economics*, 16 (2), 261-291

[45] Leontaridi, M. (1998) "Segmented labor Markets: Theory and Evidence." *Journal of Economic Surveys* 12(1): pp. 103–109.

[46] Magnac,T. (1991): "Segmented or Competitive labor Markets?", *Econometrica*,Vol.59. No.1, pp.165-187.

[47] Maloney,W.F. (1999): "Does Informality Imply Segmentation in Urban labor Markets? Evidence from Sectoral Transitions in Mexico", *World Bank Economic Review*, Vol.13, No.3, pp.275-302

[48] Maloney,W.F. (2004): "Informality revisited", *World Development*, 32 (7), 1159-78

[49] Marcouiller, D. and V. Ruiz de Castilla and C. Woodruff (1997): "Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru", *Economic Development and Cultural Change*, University of Chicago Press, Vol. 45, No.2, pp 367-92.

[50] Martin, P. (1999): "Trade and Migration: The Mexican–US Experience", University of California, Davis mimeo

[51] Menezes-Filho, N., N. Mendes, and E. Almeida (2004): "O diferencial de salarios formal-informal no brasil: Segmentação ou vies de seleção?", *Revista Brasileira de Economia* 58 (2).

[52] Pratap, S. and Quintin, E. (2006) "Are labor Markets Segmented in Developing Countries? A Semiparametric Approach." *European Economic Review*

[53] Perry, G., W. Maloney, O. Arias, P. Fajnzylber, A. Mason, J. Saavedra-Chanduvi (2006, eds.), *Informality: Exit and Exclusion*, World Bank latin America and Caribbean Studies.

[54] Ruffer, T. and J. Knight (2007): "Informal sector labor markets in developing countries", University of Oxford

[55] Saavedra, J. and A. Chong (1999): "Structural reforms, institutions and earnings: evidence from the formal and informal sectors in urban Peru", *Journal of Development Studies*, 35(4), 95-116

[56] Tannuri-Pianto, M.E. and D.M. Pianto (2002): "Informal Employment in Brazil – A Choice at the Top and Segmentation at the Bottom: A Quantile Regression Approach", Department of Economics Working Paper 236, University of Brasilia.

[57] Tannuri-Pianto, M.E., D.M. Pianto and O. Arias (2004): "Informal Employment in Bolivia: A Lost Proposition", World Bank, mimeo.

[58] Valodia, I., L. Lebani and C. Skinner (2006): "Low-waged and informal employment in South Africa", mimeo

[59] Yamada, G. (1996) : "Urban Informal Employment and Self-employment in Developing Countries: Theory and Evidence." *Economic Development and Cultural Change*, 44, 289-314.

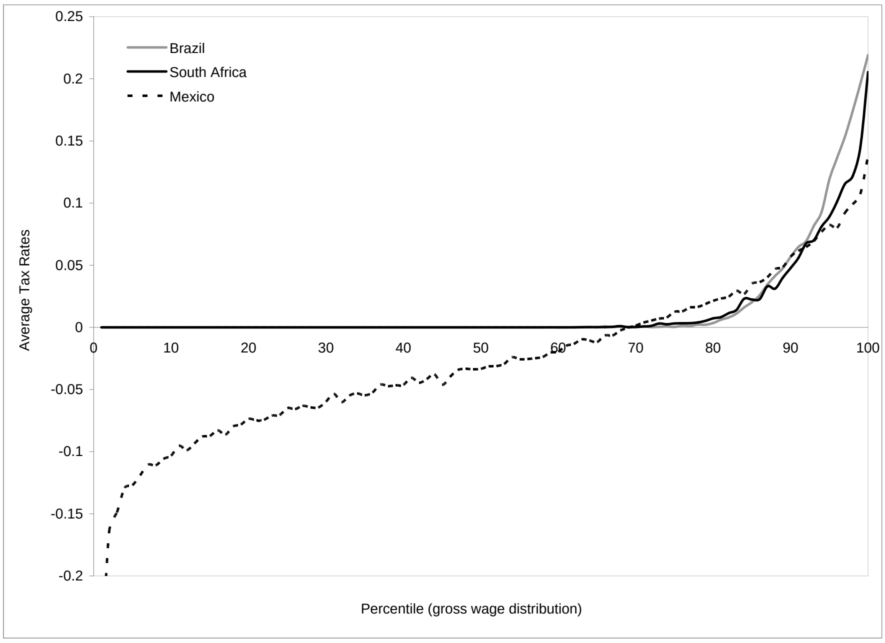


Figure A1: Distribution of Average Tax Rates

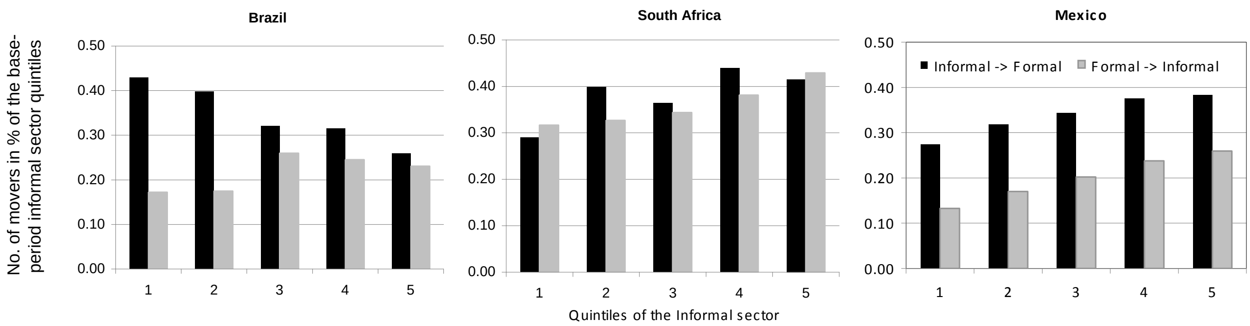


Figure A2: Distribution of Movers in/out of the Informal Sector

Table 1: Tax Schedules

	brackets (annual income)		marginal rate	Others
	in 2002 PPP$	in % of median income		
Brazil	0 … 10,485	0.0 … 1.3	0%	
	10,486 … 20,971	1.3 … 2.6	15%	
	20,971 +	2.6 +	27.5%	
South Africa	0 … 9,091	0.0 … 0.6	0%	
	9,091 … 13,468	0.6 … 0.9	18%	A tax rebate of PPP$ 1,636 also applies for all. The upper threshold of the first positive-rate bracket (0.6 times the median) is effectively higher due to the rebate @
	13,468 … 26,936	0.9 … 1.7	25%	
	26,936 … 37,037	1.7 … 2.4	30%	
	37,037 … 57,239	2.4 … 3.7	35%	
	57,239 … 80,808	3.7 … 5.2	38%	
	80,808 +	5.2 +	40%	
Mexico	0 … 656	0.0 … 0.1	1.9%	
	656 … 5,570	0.1 … 1.0	6.4%	
	5,570 … 9,789	1.0 … 1.8	10.9%	People with earnings in the first 3 brackets also receive a refundable tax credit from PPP$ 538 (for zero earnings) down to PPP$ 288 @ #
	9,789 … 11,379	1.8 … 2.1	16%	
	11,379 … 13,624	2.1 … 2.5	17.9%	
	13,624 … 27,478	2.5 … 5.0	19.9%	
	27,478 … 43,309	5.0 … 7.9	22%	
	43,309 +	7.9 +	28%	

Notes: this table summarizes the tax schedules in force in Brazil, South Africa and Mexico in years 2002, 2002 and 2007 respectively. Our calculations also account for structural changes and nominal adjustments of tax bands occurring at other years. We also account for different treatments of different groups. E.g., for persons aged 65+ in South Africa, there is no second bracket, the theshold to the third is 14,356 and the rebate is increased by 1,010.

@ The upper threshold of the first positive-rate bracket (0.6 and 0 times the median income in South Africa and Mexico respectively) is effectively higher (around 1.2) because of the rebate/tax credit.

 # Hence someone at the end of the 2nd bracket (5,570) has a negative net tax liability of -178; someone close to the end of the 3rd bracket has to pay a net tax of about 500

Table 2: Selected Samples: Descriptive Statistics

Variable	Brazil Formal	Brazil Informal		South Africa Formal	South Africa Informal		Mexico Formal	Mexico Informal
Gross hourly wage	4.77	3.53		2.54	0.99		2.77	2.30
	(6.94)	(4.92)		(3.67)	(1.60)		(2.33)	(1.96)
Net hourly wage	4.43	3.53		2.39	0.99		2.75	2.30
	(5.52)	(4.92)		(3.07)	(1.60)		(1.91)	(1.96)
Demographics								
Age	36.5	35.9		38.5	38.9		34.6	32.1
# children	3.2	3.3		1.7	2.1		1.8	1.6
household size	3.8	3.9		5.9	6.3		4.6	4.9
% married	0.64	0.54		0.63	0.47		0.62	0.44
Black	0.07	0.07	Black	0.74	0.86			
Brown	0.32	0.33	Coloured	0.26	0.14			
White	0.61	0.60						
Education								
No Schooling	0.01	0.02	No schooling	0.09	0.15	No Schooling	0.02	0.04
1-3 years	0.04	0.06	Primary	0.31	0.40	1-3 years	0.04	0.08
4-7 years	0.24	0.31	Secondary	0.53	0.42	4-7 years	0.24	0.34
8-10 years	0.18	0.19	Vocational	0.07	0.03	8-10 years	0.45	0.40
11+ years	0.53	0.42	University	0.001	0.00	11+ years	0.25	0.13
Province								
Recife	0.06	0.04	Western Cape	0.21	0.11	> 100,000 Inhab.	0.72	0.56
Salvador	0.07	0.06	Eastern Cape	0.09	0.16	15,000-99,999	0.11	0.17
Belo Horizonte	0.16	0.11	Northern Cape	0.08	0.05	2,500-14,999	0.08	0.14
Rio de Janeiro	0.27	0.35	Free State	0.11	0.08	< 2,500	0.08	0.13
Sao Paulo	0.25	0.29	Kwazulu-Natal	0.11	0.14			
Porto Alegre	0.18	0.15	North West	0.11	0.13			
			Gauteng	0.12	0.11			
			Mpumalanga	0.11	0.10			
			Limpopo	0.05	0.12			
Economic sector								
Manufacturing	0.32	0.19		0.36	0.08		0.39	0.20
Construction	0.07	0.15		0.12	0.25		0.11	0.37
Trade & Retail	0.24	0.30		0.28	0.15		0.26	0.16
Services	0.24	0.24		0.17	0.37		0.15	0.09
Transport and Comm.	0.13	0.11		0.07	0.15		0.10	0.18
# panel observations	27,420			20,053			260,878	
# workers	13,710			9,099			100,868	
Share of informal sector	15%			11%			43%	

Statistics concern the selected sample of male aged 15-65, neither in education nor in the public sector. Data covers the period 2002-2007 for Brazil, 2001-2007 for South Africa and 2005-2008 for Mexico. Log hourly wages in 2002 PPP international $. Standard deviations in brackets.

Table 3: Probit: Informal Sector

	Variable	Brazil			South Africa			Mexico	
Demographics	*Ref:*	*white, single*			*black, single*			*Single*	
	Age	-0.217	(0.017)		-0.068	(0.015)		-0.130	(0.005)
	Age squared	0.003	(0.000)		0.001	(0.000)		0.001	(0.000)
	# children	-0.028	(0.038)		0.091	(0.021)		0.003	(0.005)
	household size	0.067	(0.034)		-0.022	(0.010)		0.039	(0.004)
	Married	-0.414	(0.061)		-0.281	(0.061)		-0.729	(0.026)
	Black	-0.157	(0.115)	Coloured	-0.246	(0.087)			
	Brown	-0.086	(0.070)						
Education	*Ref:*	*no schooling*			*no schooling*			*no schooling*	
	1-3 Years	-0.079	(0.256)	Primary	-0.331	(0.080)	1-3 Years	-0.291	(0.061)
	4-7 Years	0.068	(0.227)	Secondary	-0.976	(0.087)	4-7 Years	-0.686	(0.053)
	8-10 Years	-0.136	(0.232)	Vocational	-1.501	(0.143)	8-10 Years	-1.281	(0.054)
	11+ Years	-0.575	(0.228)	University	-1.597	(0.849)	11+ Years	-1.759	(0.057)
Province	*Ref:*	*Recife*			*Western Cape*			*>100,000 Inhab.*	
	Salvador	0.085	(0.155)	Eastern Cape	0.938	(0.105)	15,000-99,999	0.728	(0.026)
	Belo Horizonte	-0.193	(0.136)	Northern Cape	0.203	(0.119)	2,500-14,999	0.962	(0.029)
	Rio de Janeiro	0.473	(0.128)	Free State	0.161	(0.121)	< 2,500	0.685	(0.029)
	Sao Paulo	0.530	(0.131)	Kwazulu-Natal	0.515	(0.113)			
	Porto Alegre	0.001	(0.142)	North West	0.608	(0.115)			
				Gauteng	0.412	(0.114)			
				Mpumalanga	0.223	(0.120)			
				Limpopo	0.994	(0.130)			
Economic secto	*Ref:*	*Construction*			*Construction*			*Construction*	
	Manufacturing	-1.092	(0.092)		-1.521	(0.088)		-1.818	(0.022)
	Trade & Retail	-0.451	(0.090)		-1.041	(0.082)		-1.549	(0.023)
	Services	-0.577	(0.093)		0.075	(0.077)		-1.382	(0.026)
	Transport and Comm	-0.835	(0.106)		0.143	(0.100)		-0.140	(0.027)
Period	*Ref:*	*year 2002*			*year 2001*			*year 2005*	
	2003	-0.040	(0.083)	2002	-0.242	(0.069)	2006	-0.021	(0.015)
	2004	-0.045	(0.092)	2003	-0.180	(0.080)	2007	-0.099	(0.018)
	2005	-0.095	(0.093)	2004	-0.072	(0.090)	2008	-0.716	(0.020)
	2006	-0.146	(0.094)	2005	0.082	(0.091)			
	2007	-0.330	(0.103)	2006	0.115	(0.092)			
				2007	0.171	(0.093)			
Constant		2.533	(0.418)		0.685	(0.316)		4.394	(0.100)

Dependent variable = 1 if informal sector. Standard errors are in brackets.

Table 4: Informal Wage Gap: Estimation Results

Estimation methods		Mean		Q=0.2		Q=0.5		Q=0.8	
		coef.	std.err.	coef.	std.err	coef.	std.err	coef.	std.err
OLS and pooled QR									
	Brazil	-0.107	0.014	-0.160	0.012	-0.133	0.011	-0.048	0.023
	South Africa	-0.627	0.025	-0.627	0.036	-0.689	0.035	-0.582	0.028
	Mexico	-0.191	0.003	-0.274	0.005	-0.201	0.002	-0.120	0.004
FE and FE-QR									
	Brazil	-0.034	0.008	-0.060	0.010	-0.034	0.004	-0.017	0.012
	South Africa	-0.188	0.027	-0.288	0.043	-0.179	0.017	-0.112	0.030
	Mexico	-0.087	0.003	-0.131	0.002	-0.087	0.001	-0.045	0.002

Informal wage penalty = estimated coefficient of the informal sector dummy. All estimations based on the variables reported in the descriptive statistics, except time-invariant characteristics (race, education and region) in the fixed effects estimations.

Table 5: Sector Moves versus Job Changes

	Brazil			South Africa			Mexico		
	N	%		N	%		N	%	
Inter-sector moves*	2,312	0.08		2,405	0.12		63,646	0.24	
Job changes according to changes in:**		inclusive	exclusive		inclusive	exclusive		inclusive	exclusive
Occupation	1,040	0.45	0.09	449	0.19	0.02	20,641	0.32	0.03
Industry	854	0.37	0.06	897	0.37	0.03	20,802	0.33	0.03
Firm size ***	700	0.30	0.10	1,834	0.76	0.34	45,555	0.72	0.34
Tenure	763	0.33	0.08	458	0.19	0.03	n/a		
Unexplained	580	**0.25**		319	**0.13**		12,730	**0.20**	

** in number of moves across sectors, either way and as % of all panel observations (note: there are potentially several moves per worker)*

*** job moves in number and in % of sector moves (inclusive = job change according to at least this characteristic; exclusive = job change according to this characteristic only). Ex: 45% of Brazilian sector moves are concomitant with job changes including occupational change; 9% of Brazilian sector moves are accompanied by occupational change only.*

**** Firm size: change in reported firm size category (Brazil: 1-5, 6-10, 11+; South Africa: 1-4, 5-9, 10-19, 20-49, 50+; Mexico: 1-5, 6-10, 11-15, 16-50, 51+)*