DISCUSSION PAPER SERIES

IZA DP No. 3145

The Formal Sector Wage Premium and Firm Size

Eliane Badaoui
Eric Strobl
Frank Walsh

November 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Formal Sector Wage Premium and Firm Size

Eliane Badaoui
THEMA, Université de Cergy-Pontoise

Eric Strobl
Ecole Polytechnique Paris
and IZA

Frank Walsh
University College Dublin

Discussion Paper No. 3145
November 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3145
November 2007

ABSTRACT

The Formal Sector Wage Premium and Firm Size

We show theoretically that when larger firms pay higher wages and are more likely to be caught defaulting on labour taxes, then large high-wage firms will be in the formal sector and small low-wage firms will be in the informal sector. The formal sector wage premium is thus just a firm size wage differential. Using data from the South African labour force survey we illustrate that firm size is indeed the key variable determining whether a formal sector premium exists.

JEL Classification: J42

Keywords: informal sector, wage premium, firm size

Corresponding author:

Frank Walsh
School of Economics
University College Dublin
G216 John Henry Newman Building
Belfield, Dublin 4
Ireland
E-mail: frank.walsh@ucd.ie

1 Introduction

One of the main differences between labour markets in developing compared to developed economies is the existence of large informal sectors. For example, in Africa the informal sector is estimated to absorb about 60% of the urban labour force.[1] Importantly in this regard, it is generally assumed, and empirically substantiated by much of the literature, that workers in the informal sector are paid less than their formal sector counterparts.[2] However, theoretically it is not clear why this should be the case. While a tax wedge would explain differences in gross wages, if workers can move between sectors then net wages should surely be equalised. Earlier papers in the literature such as Lewis (1954) or Harris and Todaro (1970) assumed a dual labour market structure where workers earned rents in the primary sector and secondary sector workers queued for good jobs. There are of course many models that could be used to justify why workers in particular sectors would earn wage premiums - as, for example, efficiency wage and union models[3] - but applying these to explain a wage premium for formal sector employees would mean arbitrarily assuming that formal sector workers earn rents because of some exogenously imposed feature that for some reason is more relevant to the formal rather than the informal sector.

In this paper we start off of by demonstrating that in essentially any labour market model where in equilibrium larger firms pay higher wages, if larger firms are more likely to be caught defaulting on labour taxes, then large-high wage firms will be in the formal and small-low wage firms will be in the informal sector. The formal sector premium is thus just a firm size premium. In order to solve for the wage distribution explicitly in the case where the firm size wage premium emerges endogenously, we then incorporate taxes on labour income and an enforcement technology into the equilibrium search model of Burdett and Mortensen (1998). More specifically, firms post wages and workers may work in the formal sector or may opt for a tax free outside option, which could be viewed as informal sector employment, as discussed by Albrecht *et al.* (2005). We find that in this set-up formal sector employees do indeed earn rents relative to their informal counterparts in the model. However, this is not because they are formal sector employees, but because in our model large firms will pay higher wages and have the incentive to stay in the formal sector. In this regard, it arguably makes intuitive sense that small firms would be the most difficult for the government to find and the most likely to stay in the informal sector. Indeed, a number of theoretical models [Fortin *et al.* (1997) and Rauch (1991), for example] impose this assumption. Moreover, many empirical studies seem to confirm that informal sector workers are concentrated in small firms.[4] As a matter of fact, small enterprise size is part of the ILO definition of the informal sector and has been used in a number of papers as a proxy for such.

[1] See http://www.ilo.org/public/english/employment/skills/informal/who.htm

[2] See, for example, Mazumdar (1981), Heckman and Hotz (1986), Pradhan and Van Soest (1995), Tansel (1999), and Gong and Van Soest (2002).

[3] Jones (1983) uses the shirking efficiency wage model to characterise the formal sector in a model with minimum wages.

[4] See, for instance, Tybout (2000).

A search model where it is difficult for workers and firms to find each other seems like a natural way to model the labour market with an informal sector in developing countries, where it is often argued that there are no clear channels for the exchange of labour market information and search costs are high.[5] As a matter of fact, there have been other papers in the literature that use a search-matching framework to model the informal labour market. For instance, Albrecht *et al.* (2005) extend the Mortensen and Pissarides (1994) matching model to incorporate a self-employed informal sector where there is heterogeneity in workers' productivity in that more productive workers may opt to wait for a formal sector job, while others may select into the informal sector. Also, Boeri and Garibaldi (2005) develop a matching model with supervision where workers in the informal sector cannot avail of unemployment benefits, and show that matches found not paying tax are dissolved. Their model suggests that policies aimed at reducing the size of the shadow economy may increase unemployment. Alternatively, Fugazza and Jacques (2001) incorporate psychic costs as part of the costs of being in the informal economy in a matching model where workers direct their search at informal sector firms. However, it is important to emphasize that while these papers use the matching framework, they just focus on exogenously given worker heterogeneity. In the equilibrium search framework we adopt here the firm size premium and the informal sector emerge endogenously without arbitrarily imposing any differences between the two sectors other than that larger firms are more likely to be caught defaulting on their tax.[6]

A key prediction of the equilibrium search framework is that large firms pay more even when there is no heterogeneity amongst either workers or firms ex ante. It is only in the case where there are no search frictions that the labour market is competitive and the formal/large firm size premium disappears. There is already some evidence that suggests that firm size may be a driving factor behind the often observed formal sector wage premium. For example, Pratap and Quintin (2005) find, using Argentinean data and semi-parametric techniques to deal with the selectivity issue inherent in estimating the possibility of a formal sector wage premium, that there is no difference in gross wages between informal workers and their formal sector counterparts and that the employer's size is crucial in making the wage premium 'disappear'.[7] Using a similar econometric technique and rich South African data that allows a relatively precise measure of informal employment we confirm that firm size can explain away the formal sector wage premium, but only if one assumes, as appears reasonable and as we do in our model, that informal sector workers do not pay taxes.

[5] See, for example, Hussmanns (1994) or Byrne and Strobl (2004).

[6] In our paper, we interpret informality to mean tax avoidance rather than just any illegal activity. Schneider and Enste (2000) provide a survey of the general literature on shadow economies and its various definitions.

[7] Amaral and Quintin (2006) outline a theoretical framework where the only difference between informal and formal sector firms is that informal sector firms are seen as more likely to default on loans, have difficulty accessing credit and because of this, rely on self financing. Because of the complementarity between skill and capital, high skill capital intensive firms enter the formal sector and hire high skill workers. Thus in contrast to our model labour markets are competitive and wage differentials can be explained by differences in ability.

Our equilibrium search framework also allows us to do comparative static analysis on the policy parameters and predict the long run change in the equilibrium wage distribution accounting for firm entry and exit. Some of the comparative static results may be considered surprising. In particular, we find that in the long run, when we account for the impact of firm exit on the shape of the distribution, an increase in the tax rate may reduce the share of the informal sector for plausible parameter values. Also, an increase in the enforcement/punishment parameter tends to reduce the share of the informal sector as one would expect. Given the amount of structure we impose to solve the equilibrium search model explicitly, we view these comparative static results as examples that illustrate interesting possibilities in a reasonable framework.

The remainder of the paper is organised as follows. In the next section we present our model. In Section II we describe our data. Empirical evidence in support of the results derived from our model are shown in Section III. Concluding remarks are given in the final section.

2 The Model

2.1 Exogenous Firm Size Wage Premium

We start with a general model where there is a positive and continuous relationship between a firm's employment n and the wage w, $n(w)$, in a stationary equilibrium, but initially do not specify why this positive relationship exists.[8] More specifically, firms have production function $q(n)$ and p is the price of output. There is a tax rate t on wages. There will be a Poisson arrival rate of tax inspectors, which is increasing in the number of employees at the firm: $\Theta[n(w)]$. If firms are caught not paying their taxes they must pay a fine $\Omega[wtn(w)]$. This function is increasing in the per period tax bill $wtn(w)$. The flow values of defaulting (d) and complying (c) firms in a stationary equilibrium at any wage w are:

$$\begin{aligned} rV^d &= pq[n(w)] - wn(w) - \Theta[n(w)]\Omega[wtn(w)] - \delta V^d \\ rV^c &= pq[n(w)] - w(1+t)n(w) - \delta V^c \end{aligned} \tag{1}$$

The flow value of the firm where r is the discount rate is the dividend stream (flow of profits) plus any capital gain/loss in the value of the firm. It is instructive to look at the difference in the value of complying and defaulting at any wage w:

$$(V^c - V^d) = \frac{\Theta[n(w)]\Omega[wtn(w)] - wtn(w)}{r+\delta} \tag{2}$$

We will denote the tax liability as $B = wtn(w)$ for shorthand. From (2) we can establish our first proposition.

[8]Workers are identical in the model since the firm size and formal sector premiums that interest us remain after we control for worker characteristics.

Proposition One: *We assume that there is a stationary equilibrium where there is a continuous positive relationship between employment and the wage rate $n(w)$. If the elasticity of punishment with respect to the tax bill is greater than or equal to unity: $\frac{\partial \Omega(B)}{\partial B}\frac{B}{\Omega} \geq 1$ and there are some compliant and some non-compliant firms in equilibrium, then there will be a cut-off point in firm size below which all firms will default on their taxes, and above which firms will be compliant. In other words, we will have a wage distribution with small-low wage firms in the informal sector and large-high wage firms in the formal sector.* The proof is given in Appendix One.

The assumption that the elasticity of the punishment with respect to the tax bill is greater than or equal to unity seems plausible. This just means that when a firm is caught defaulting on their tax the punishment they pay increases at least proportionately with the amount of tax they owe. It is worth noting at this stage that in equation (1) we assume that the punishment for non-payment of taxes is dependent on the current tax bill. Because we look at a stationary equilibrium it would be difficult to make the punishment depend on the total tax liability incurred by a firm since it began defaulting, which would be more plausible. While it would be difficult to model this formally, we might expect that if we did it would make it even more likely that the elasticity in Proposition One would exceed unity.

Proposition One shows that in a general setup where we have a distribution of firms, small firms pay low wages, and large firms pay high wages, then small low-wage firms will be in the informal and large-high wage firms in the formal sector. It is worth noting that when taxes are on labour income, we also expect the punishment function to be determined by labour income. In other words, a firm's decision on whether to be in the formal or informal sector depends only on the cost side of the firm's objective function. This makes Proposition One fairly general. Firms of different sizes may be in different markets, have different market structures, or have different production functions etc. But, the decision on whether to default depends only on comparing the tax bill with the expected punishment for a firm of a given size, and this depends only on the firm size and wage, which are jointly determined.

One should note that there exists an extensive literature showing a positive relationship between firm size in both developing and developed countries.[9] While there are different models that seek to explain this premium, Proposition One implies that if we are in a country with weak enforcement and a sizeable informal sector that, because the informal sector firms are located in small firms, the formal sector wage premium is determined by the firm size premium. In Section III we show empirical support for this proposition.

[9] See Oi and Idson (1999), for instance, for a review of the literature on the firm size premium. Examples of studies of developing countries include Schaffner (1998), Velenchik (1997), and Marcelle and Strobl (2003).

2.2 Endogenous Firm Size Wage Premium

2.2.1 Burdett and Mortensen (1998) Model

While the above proposition is very general, it may be useful to solve the model explicitly. One can think of different models that rationalise why firms would pay higher wages. In Rebitzer and Taylor (1995) the efficiency wage premium is increasing in firm size, while in recent decades dynamic monopsony models where firms pay higher wages to lower turnover or attract more workers have been prevalent in the literature [see Manning (2003)]. In these latter type of models firms that pay higher wages retain and attract more workers and so larger firms pay higher wages.[10] One obvious candidate in this regard is the equilibrium search framework outlined by Burdett and Mortensen (1998). As a matter of fact, Mortensen (2003) argues that this model is a convincing candidate to explain the firm size and industry wage differentials that are empirically widely documented. In this model the firm size premium emerges endogenously and we can solve for the equilibrium wage distribution explicitly.

We first derive the labour supply curve in a model where there are search frictions and workers receive on the job offers.[11] There is a mass of M identical employers and a mass L of identical workers in the economy. The non-employment outside option is b.[12] Workers receive offers according to a Poisson arrival rate, λ, at each instant. There is random matching so that any offer is equally likely to come from any firm irrespective of the firm's size.[13] The distribution of wage offers which we will solve for is $F(w)$. We assume that the arrival rate of job offers is the same for employed and unemployed workers. Burdett and Mortensen (1998) assume $r = 0$ in their derivation of the labour supply curve and we follow this assumption. The separation rate at any firm, $d(w)$, is just the sum of the job destruction rate δ plus the arrival rate of offers to each worker times the probability the offer comes from a higher wage firm: $\lambda[1 - F(w)]$

$$d(w) = \delta + \lambda[1 - F(w)] \tag{3}$$

In a stationary equilibrium inflows and outflows to unemployment are equal, implying the following relationship between the unemployment rate u and the arrival rates:

$$\frac{u}{1-u} = \frac{\delta}{\lambda} \tag{4}$$

If $N(w)$ is aggregate employment at wage w or less, stationarity also ensures that the inflows to this stock (the number of offers less than w accepted by unemployed workers)

[10] In more traditional monopsony models, such as the company town model, large firms have more monopsony power and pay lower wages [see Boal and Ransom (1997) for a survey]

[11] See Mortensen (2003) and Burdett and Mortensen (1998) for a detailed derivation of the labour supply curve.

[12] Traditionally this outside option b is viewed as unemployment benefits. In the context of developing countries it is perhaps more appropriately seen as self-employment or support for the non-employed by their family which is a relatively common feature of the developing world.

[13] See Manning (2003) pp284-286 for a discussion on the matching technology.

and outflows (the separation rate times the stock) are equal:

$$\lambda F(w)uL - \{\delta + \lambda[1 - F(w)]\}N(w) = 0 \tag{5}$$

$F(w)$ is the wage offer distribution, while we define $G(w)$ as the wage distribution of employed workers. Since the employment rate times the wage distribution equals the stock of workers working for a wage less than w, we can use (5) to define the relationship between the wage and wage offer distributions:

$$G(w) = \frac{N(w)}{1-u} = \frac{\delta F(w)}{\delta + \lambda[1 - F(w)]\}} \tag{6}$$

The number of offers received by workers from any firm is the offer arrival rate times $\frac{L}{M}$. If we multiply this by the fraction unemployed plus the fraction of employed workers earning less than w, we get the number of offers accepted to a firm offering a wage w at any point in time:

$$h(w) = \lambda \frac{L}{M}[u + (1-u)G(w)] \tag{7}$$

Next, recognising that since the separation rate times employment must equal new hires [given in (7)] for a firm to be in a stationary equilibrium, we use (6) in (7) and divide by (3) to get the labour supply curve:[14]

$$n(w,F) = \frac{h(w)}{d(w)} = \frac{\delta\lambda}{M\{\delta + \lambda[1 - F(w)]\}^2} \tag{8}$$

Given that employed and unemployed workers have the same arrival rate of job offers, the reservation wage is just the benefit level b. The employment levels of firms paying the reservation wage and the highest wage $\overline{w}$ are:

$$n(b,F) = \frac{\delta\lambda}{M(\delta+\lambda)^2} \qquad \text{and} \qquad n(\overline{w},F) = \frac{\lambda}{M\delta} \tag{9}$$

The derivation of the labour supply curve shows that when there are search frictions and an equilibrium where some firms wish to be larger than others, we will have wage dispersion and large firms will pay higher wages even when workers are identical. Burdett and Mortensen (1998) made an important contribution to this literature by solving for the unique stationary wage equilibrium in a model with constant productivity. Our next step is to use the Burdett and Mortensen model to illustrate Proposition One. One should note that while the model imposes a lot of structure, it allows us to explicitly solve for the wage distribution of formal/informal firms.

[14] We note here that the labour supply curve in Burdett and Mortensen allows for different arrival rates for unemployed (λ_0) and employed (λ_1) workers. The labour supply curve in this case, not normalising the mass of workers L to unity is:

$$n(w,F) = \frac{\delta\lambda_1}{M\big\{\delta + \lambda_1[1 - F(w)]\big\}^2}\Big[\frac{\lambda_0}{\lambda_1}\Big(\frac{\delta+\lambda_1}{\delta+\lambda_0}\Big)L\Big]$$

That is it is just the labour supply curve in (2) (with λ replaced by λ_1) times a constant.

2.2.2 Burdett and Mortensen (1998) Model with Formal and Informal Firms

We normalise the mass of workers to unity for simplicity. Burdett and Mortensen (1998) assume firms are identical ex-ante and workers have a constant productivity p at any firm. Here we modify the Burdett and Mortensen model by introducing a tax rate t on wage income that is paid by firms. Labour supply is given by (8) once we solve for the wage distribution. We assume the Poisson arrival rate of tax inspectors is a constant z times employment to the power of a constant β so that large firms are more visible and more likely to be caught defaulting: $zn(w)^{\beta}$. We specify the penalty for defaulting as x times the firm's per period tax bill: $wtn(w)$. To save on notation we define $s = xz$ as the parameter that when multiplied by employment to the power of $\sigma = \beta + 1$ determines the expected punishment for defaulters at any point in time. We can rewrite equation (1), i.e., the value of complying and defaulting firms as:

$$\begin{aligned} V^d &= \frac{(p-w)n(w) - swtn(w)^{\sigma}}{\delta} \\ V^c &= \frac{[p - w(1+t)]n(w)}{\delta} \end{aligned} \tag{10}$$

We note that the two policy instruments the government has are the tax rate t and the degree of punishment/enforcement s. While Proposition One makes no assumption about firm entry, to solve the model explicitly we assume free entry. This ensures that in equilibrium the value of all firms along the wage distribution, compliant and non-compliant, will be equalised. Using this condition in (10) gives us the level of employment below which firms will default:

$$V^d > V^c \quad \text{if} \quad \frac{1}{s} > n^{\sigma-1} \qquad \text{and} \qquad V^d < V^c \quad \text{if} \quad \frac{1}{s} < n^{\sigma-1} \tag{11}$$

We can use the expression for labour supply (8) in (11) to calculate the cut-off value of the wage offer distribution below which firms will be defaulting:[15]

$$F^* = \frac{\delta + \lambda}{\lambda} - \sqrt{\frac{s^{\frac{1}{\sigma-1}}\delta}{M\lambda}} \tag{12}$$

Free entry ensures that $V^d = V^c = k$. Imposing this free entry condition using (6) and (7) for the value of firms and (1) for labour supply we can calculate the relationship between the wage and offer distribution for defaulting and compliant firms:

$$w^d = \left\{p - \frac{kM}{\lambda}[\delta + \lambda(1-F)]^2\right\}\left\{\frac{M^{\sigma-1}[\delta+\lambda(1-F)]^{2(\sigma-1)}}{M^{\sigma-1}[\delta+\lambda(1-F)]^{2(\sigma-1)} + st(\lambda\delta)^{\sigma-1}}\right\} \tag{13}$$

$$w^c = \frac{p}{1+t} - \frac{kM}{\lambda}\frac{[\delta+\lambda(1-F)]^2}{1+t} \tag{14}$$

[15]It is worth noting from (10) that even with a general production function $y = y(n)$,where y is output, equation (11) will hold.

The wage in the lowest wage firm is b and since all other firms pay higher wages the value of the wage offer distribution will be zero at a wage b. Using $w = b$ and $F = 0$ in (6) and setting the value of the lowest wage firm equal to entry costs k we can solve for the relationship between entry costs and the mass of firms in terms of the exogenous parameters:

$$k = \frac{(p-b)n(b) - bstn^{\sigma}(b)}{\delta} = \frac{(p-b)\lambda}{M(\delta+\lambda)^2} - \frac{stb\delta^{\sigma-1}\lambda^{\sigma}}{M^{\sigma}(\delta+\lambda)^{2\sigma}} \tag{15}$$

In Figure 1, we graphically depict the inverse wage offer distribution of our model for two different tax rates, 10% and 30%, using (10) for values of F between zero and F^* and (11) for values of F between F^* and unity under assumed values for the exogenous parameters. The graph illustrates a wage offer distribution which is consistent with the stylised facts. Small low wage firms are in the informal sector and large high wage firms in the formal sector. While we will do some comparative static analysis later where both arrival rates of job offers λ and entry costs k are dependent on the mass of firms in equilibrium, Figure 1 plots the response to a tax change under the simpler assumption that these parameters are fixed when the mass of firms changes in accordance with (12). The wage distribution becomes more compressed in response to the higher tax rate as we would expect. Firms paying high wages must adjust there wage downwards in response to the tax, while the lowest wage firms are already paying the reservation wage and cannot lower the wage any further.

One should note that the basic Burdett and Mortensen model with homogeneous productivity across firms predicts a wage distribution with a lot of weight on the upper tail of the distribution, whereas empirically it has been observed that the wage distribution generally has a long right hand tail. Mortensen (2003) discusses this issue and outlines a number of generalisations to the basic Burdett and Mortensen model where productivity varies across firms. These generalisations generate wage distributions that are more in keeping with empirically observed wage distributions. This can be where there is exogenous variation in firms' productivity and firms can choose the number of contacts with workers, or, alternatively, where firms may be allowed to invest in costly match specific or general capital, which generates differences in productivity. We will take the case where firms invest in match specific capital and apply our model of the informal sector to this set-up.

Within this framework we look at the model analysed earlier where the risk of detection for defaulters rises with firm size so that small low wage firms are in the informal sector. We will set up the profit function in general terms before distinguishing between the defaulting and compliant sectors. We assume that $j \in [d, c]$ so that $w_j = w$ when $j = d$ and $w_j = w(1+t)$ when $j = c$.

We note from equation (8) that the labour supply curve is:

$$n^j(w, F^j) = \frac{\lambda h^j(w)}{d^j(w)} \tag{16}$$

In this section the ability to distinguish between the separation and offer acceptance rates will be important. Firms invest in match specific human capital T, which also costs the firm T. These sunk costs will be incurred every time an offer is accepted and a new worker is hired. Human capital enhances the productivity of a match according to the concave function $p(T)$, but the productivity gain of the investment is lost as soon as the worker leaves this firm. The cost of the investment T is multiplied by the number of matches but is unaffected by the separation rate. The profit function in this case is:

$$\begin{aligned}\pi^j(w_j, F^j) &= \lambda h^j(w)\Big[\frac{p(T) - w_j}{d^j(w)} - T\Big] \\ &= \frac{\lambda\delta}{M\{\delta + \lambda[1 - F^j(w)]\}}\Big[\frac{p(T) - w_j}{\delta + \lambda[1 - F^j(w)]} - T\Big]\end{aligned} \tag{17}$$

We assume that $p(T) = pT^\alpha$ and from the first order condition for the optimal choice of training T:

$$T = (p\alpha)^{\frac{1}{1-\alpha}}\Big\{\delta + \lambda[1 - F^j(w)]\Big\}^{\frac{1}{1-\alpha}} \tag{18}$$

Substituting (18) into the profit function one obtains:

$$\begin{aligned}\pi^j(w_j, F^j) = \frac{\lambda\delta}{M}\Big[&(\frac{1-\alpha}{\alpha})(p\alpha)^{\frac{1}{1-\alpha}}\Big\{\delta + \lambda[1 - F^j(w)]\Big\}^{\frac{2-\alpha}{\alpha-1}} \\ &- w_j\Big\{\delta + \lambda[1 - F^j(w)]\Big\}^{-2}\Big]\end{aligned} \tag{19}$$

Equation (10) can be amended to:

$$\begin{aligned}V^d &= \frac{\pi_c(w) - swtn(w)^\sigma}{\delta} \\ V^c &= \frac{\pi_d(w)}{\delta}\end{aligned} \tag{20}$$

Using (19) in (20) the value of defaulting and complying firms respectively can be written as:

$$\begin{aligned}V^d = &\frac{\lambda}{M}\Big[(\frac{1-\alpha}{\alpha})(p\alpha)^{\frac{1}{1-\alpha}}\Big\{\delta + \lambda[1 - F^d(w)]\Big\}^{\frac{2-\alpha}{\alpha-1}} - w\Big\{\delta + \lambda[1 - F^d(w)]\Big\}^{-2}\Big] \\ &- swt\frac{\lambda^\sigma\delta^{\sigma-1}\Big\{\delta + \lambda[1 - F^d(w)]\Big\}^{-2\sigma}}{M^\sigma}\end{aligned} \tag{21}$$

$$\begin{aligned}V^c = \frac{\lambda}{M}\Big[&(\frac{1-\alpha}{\alpha})(p\alpha)^{\frac{1}{1-\alpha}}\Big\{\delta + \lambda[1 - F^c(w)]\Big\}^{\frac{2-\alpha}{\alpha-1}} \\ &- w(1+t)\Big\{\delta + \lambda[1 - F^c(w)]\Big\}^{-2}\Big]\end{aligned} \tag{22}$$

One can see by comparing (21) and (22) that in equilibrium at a given wage and value of the distribution, equation (11) still gives the condition that determines whether a firm can profit from moving to the defaulting from the compliant sector or *vice versa*. Firms below the critical level of employment will default and firms above the critical level will comply. Given that (11) still holds equation (12) continues to give the fraction of wage offers in the defaulting sector. The equilibrium value of firms is given by looking at (21) for the lowest wage firm where $F = 0$ and $w = b$.

$$k = \frac{\lambda}{M}\Big[(\frac{1-\alpha}{\alpha})(p\alpha)^{\frac{1}{1-\alpha}}(\delta+\lambda)^{\frac{2-\alpha}{\alpha-1}} - b(\delta+\lambda)^{-2}\Big] - sbt\frac{\lambda^{\sigma}\delta^{\sigma-1}(\delta+\lambda)^{-2\sigma}}{M^{\sigma}} \tag{23}$$

Next one can equate $V^c = V^d = k$ to solve for the equilibrium relationship between the wage and the wage distribution for both firm types:

$$\begin{aligned} w = &\frac{M^{\sigma-1}(\frac{1-\alpha}{\alpha})(p\alpha)^{\frac{1}{1-\alpha}}\Big[\delta+\lambda(1-F^d)\Big]^{\frac{\alpha+2(\sigma-1)(\alpha-1)}{\alpha-1}}}{M^{\sigma-1}\Big[\delta+\lambda(1-F^d)\Big]^{2(\sigma-1)}+st(\delta\lambda)^{\sigma-1}} \\ &-\frac{k\lambda^{-1}M^{\sigma}\Big[\delta+\lambda(1-F^d)\Big]^{2\sigma}}{M^{\sigma-1}\Big[\delta+\lambda(1-F^d)\Big]^{2(\sigma-1)}+st(\delta\lambda)^{\sigma-1}} \end{aligned} \tag{24}$$

$$w(1+t) = (\frac{1-\alpha}{\alpha})(p\alpha)^{\frac{1}{1-\alpha}}\Big[\delta+\lambda(1-F^c)\Big]^{\frac{\alpha}{\alpha-1}} - \frac{kM}{\lambda}\Big[\delta+\lambda(1-F^c)\Big]^{2} \tag{25}$$

One can also solve for the highest wage by setting $F = 1$ in (21).

We plot the distribution for the same assumed parameter values of Figure 1 in Figure 2 at two different tax rates. Once again the graph illustrates a wage distribution that is consistent with the stylised facts. Low wage small informal firms, and large, high wage formal firms. In this case the inverse wage offer distribution is convex, indicating a small amount of weight in the upper tails, which is more in keeping with the empirically observed wage distributions. The higher tax rate compresses the wage distribution as in Figure 1.

2.2.3 Comparative Statics

Next we investigate the effect of changes in the policy variables on the percentage of employed workers who will be in the informal sector. It is apparent from (12) that an increase in the tax rate or the punishment/enforcement parameters may have a direct affect, but may also affect the size of the formal sector by causing firm exit. The mass of firms may also plausibly affect fixed entry costs and the arrival rate of job offers to workers which, in turn, will also determine the size of the informal sector.

Before making our next proposition we allow for the fact that the arrival rate of job offers λ and fixed entry costs k may depend on the mass of firms. Specifically we assume that $\frac{\partial k}{\partial M} \geq 0$ and $\frac{\partial \lambda}{\partial M} \geq 0$. We define $= \varepsilon_{\lambda M}$ as the elasticity of the arrival rate with respect to the mass of firms and define the following condition:

- Condition one: $p > b(1+\sigma t)$

Proposition Two: *When productivity is exogenously given, then if* $\varepsilon_{\lambda M} < 1$ *and Condition One holds, this ensures that:* $\frac{dG*}{ds} < 0$ *and* $\frac{dG*}{dt} < 0$. The proof is given in Appendices Two and Three.

Proposition Two shows conditions under which we can unambiguously say that an increase in punishment/enforcement rate reduces the fraction of employed workers in the informal sector. More surprisingly, it shows that the same conditions are sufficient for an increase in the tax rate to reduce the fraction of employed workers in the informal sector.

As we noted earlier, we would not argue that the comparative static results here, in particular the result that a higher tax rate reduces the share of the informal sector, is a general result. It is nevertheless informative. If we look at equations (10) we see that the reason that the tax rate cancels out in equation (12) is because the tax bill enters the costs of complying firms and the punishment of defaulting firms linearly. If the tax rate in (10) had an exponent greater than unity, for example, we would expect t to enter (12) and an increase in the tax rate to directly increase the size of the informal sector offsetting the impact of firm exit in increasing the size of this sector. We could think of the comparative static results as illustrating that for plausible parameter values higher tax and enforcement rates typically cause firm exit which in the long run changes the shape of the distribution in a way that increases the share of the informal sector. If there is no direct effect where the higher tax rate reduces the share of the informal sector, the impact of firm exit can dominate and the share of the informal sector will increase.[16]

3 Data

An important result of our theoretical model is that, even when there is no heterogeneity amongst either workers or firms ex ante, large firms will operate in the formal sector and will pay higher wages than smaller firms, which are predicted to conduct business in the informal sector. To investigate whether there is empirical support for these predictions we use the example of South Africa, where our data source is the South African Labour Force Survey (SALFSS)[17]. The SALFSS is a twice-yearly rotating panel household survey conducted since September 2000, specifically designed to measure the dynamics of employment and unemployment in the country. For our analysis we use the waves September 2001, March 2002, September 2002, March 2003, and September 2003.[18]

In terms of classifying informal sector activity, the SALFSS explicitly asks individuals that are employed whether their main activity is in the informal sector. More precisely, each employed individual is asked whether 'the organisation/business/enterprise/branch

[16]In the case of an increase in punishment/enforcement parameter s both the direct and indirect affects go in the same direction.

[17]One should note that this is not an official abbreviation; we simply use it for ease of notation.

[18]We restrict our analysis to these waves because they allow us to link households over time.

where he/she works is in the formal sector or in the informal sector (including domestic work)'.[19] Additionally, there are a number of other questions regarding fringe benefits of a job that allow us to further verify the individual's informal sector status. These include questions regarding whether the firm is registered, provides medical aid, deducts unemployment insurance contributions, and is registered for VAT. If an individual answers in the affirmative to any of these questions, we change his/her sector status to being of the formal sector even if they classify themselves as working in the informal sector.[20]

An important feature of our model described above is that of firm size. In the SALFSS employed individuals provide explicit information on the size of their employer as it falls within six categories: 1 employee, 2-4 employees, 5-9 employees, 10-19 employees, 20-49 employees, 50 or more employees. We create a set of zero-one dummy variables that captures these differences in employer size.

Since we are specifically interested in the pay differential associated with working in the informal sector, a third important piece of information required from our data is that concerning remuneration. For those persons in paid employment, the SALFSS explicitly asks the remuneration in their main activity. More precisely, the SALFSS provides a person's weekly, monthly, or annual income and hours worked in the previous week in their main job, and we use this information to calculate hourly wage rates. One should note that for about 23% of individuals who were in paid employment the salary was reported in income categories.[21] For these we used the mid-point between category thresholds, except for the first and last category where we simply used the threshold itself as the salary value. The derived nominal hourly wage rate data was then converted into real wages (September 2001 values) by using the South African consumer price deflator. We also checked the observations for those individuals that claimed to be in paid employment, but for whom information on remuneration was missing. This turned out to be a little over 7% of the total sample, of which 10 per cent indicated that they worked in the informal sector (as defined above).[22]

An important assumption of our model is that individuals working in the informal sector are not subject to taxation. Ideally we would like to take account of this, however, it is difficult from simple labour force data, where there is no information on non-labour income and where we cannot easily link immediate family members within a household, to accurately estimate the amount on labour income that is likely to be deducted in terms of taxes for most labour market groups. In order to be able to calculate reason-

[19]According to the questionnaire, 'Formal sector employment is where the employer (institution, business or private individual) is registered to perform the activity. Informal sector employment is where the employer is not registered'.

[20]However, in the end there were only 2.1% of observations where we needed to change their status. The correlation between the two classifications was about 0.96

[21]Of these about 6 per cent reported to be in the informal sector according to our definition above.

[22]Thus, compared to our final sample, there does not appear to have been a disproportionate allocation of missing salary observations for those in the informal sector.

ably accurate net (after taxes) income from employment for those working in the formal sector, we thus limit our sample to single men for which we can relatively easily infer their income tax liabilities for a given annual income.[23] More precisely, we calculated gross monthly labour income and then used the tax tables relevant for that period as published by the South African Revenue Service to calculate net monthly income for those working in the formal sector and assumed that informal sector workers do not pay taxes on their earnings from employment.[24] From this we derived net hourly wage rates.

Apart from an explicit definition of the formality of an individual's employer and a precise measure of their remuneration, the SALFSS can also be regarded as relatively rich in other information potentially relevant to an individual's labour market status. We thus compiled information on those factors that are likely to be important for determining a person's pay, as well as whether he/she works in the informal sector. The ones used in the current analysis are grouped for convenience sake into those related to human capital (age, gender, race, marital status, education level, occupation) and job characteristics like job training, region, tenure, and industry (eleven dummies). We provide a comprehensive list of these and their definitions in Table 1.

An important aspect of the data is its rotating panel nature. In this regard, it is easy to link households across waves when they are re-surveyed since they are given a unique household identifier. In contrast, although individuals are likely also to be resurveyed across waves if they remain within the same household, there is no straightforward way to link these across waves. Thus, by pooling all data across waves, we would be using multiple observations across at least some individuals in our analysis without being able to control for this. We thus, instead, in order to ensure that this is not the case, only used information taken from one wave per household, arbitrarily chosen as the latest date at which the household was surveyed. Finally, we reduced our sample to non-self employed males, between the ages of 15 and 70, working in sectors other than the public sector. One should note that focusing only on males allows us to abstract from the often more complex labour force participation decision that is generally associated with females. Also, while comparing self-employed informal sector workers to their formal sector counterparts may be of interest in its own right, one could argue that the decision of whether to register one's own enterprise is likely to be less constrained or at least determined by different criteria than attempting to get a formal sector job. Apart from this self employed workers earnings would be expected to have greater measurement error and incorporate returns to risk etc., that would not be included in wages of employees. Analyzing this group would thus require a separate analysis which is beyond the scope of the current paper.

Overall our selection criteria left us with a sample of 7,249 single males of which 1,529

[23]One should note that by focusing only on single males allows us also to abstract from the often more complex labour force participation decision that is generally associated with females or married males.

[24]Further details are available from the authors.

work in the informal sector. We provide some simple summary statistics of these in Table 2. As can be seen, formal sector workers earn substantially more than their informal sector counterparts in terms of gross log wages, namely about 76%. When one allows for the income tax deductions from the earned income for those working in the formal sector, this discrepancy is reduced (to about 54%) but nevertheless remains. We also provide the distribution of the formal and informal sector workers by the given employer sizes in the same table. Accordingly, only about 31% in the formal sector work for firms with less than 10 employees. In contrast, in the informal sector the equivalent figure is about 90%. We also calculated the ratio of the formal relative to the informal log wage rate within firm size categories in Table 3. Here it can be seen that in terms of gross wages the relative log wage rate differences are largest in the very small and the very large employer size categories, while formal sector workers earn between 20 and 35% more in the employer size categories that lie between these two. However, once one allows for tax payments for formal sector workers, the discrepancy is reduced in the largest and the smallest categories, while it virtually disappears for the intermediate ones, especially for those working with employers of size 2-10 workers.

4 Econometric Analysis

Our simple summary statistics suggested that it is important to take account of tax payments by those working in the formal sector when calculating the formal sector wage premium as is assumed in our model. Moreover, comparing wages across the formal and informal sector within categories suggested that at least some of the difference in total mean wages may be due to the different distributions of employer size across the two sectors. This would be supportive of our theoretical result that the formal wage premium may just be due to differences in firm sizes in these two sectors. In order to obtain support for these assertions more formally we now proceed to test them econometrically.

In terms of measuring the wage premium associated with the informal sector one may be tempted to simply run OLS on a standard Mincerian wage equation where one regresses logged wages on an indicator of formal sector employment while controlling for other relevant and available (as from the data) determinants of earnings. However, as recently shown by Pratap and Quintin (2005), not properly taking account of the selection bias in estimating such a parametric regression could bias the results. More specifically, the authors implement a semi-parametric propensity score matching estimator that allows one to explicitly deal with the problem of common support common in standard OLS, where one may be comparing very dissimilar workers. As a matter of fact under OLS Pratap and Quintin (2005) find evidence of a gross wage informal sector premium using Argentinian data, but no such earnings differential is detectable under the semi-parametric propensity score matching estimator.We thus similar follow Pratap and Quintin (2005) and resort to this semi-parametric approach in investigating the formal sector wage premium.

Using a similar notation to Pratap and Quintin (2005), we define the average formal sector premium as what is in the matching literature known as the Average Treatment Effect on the Treated (ATT), where treatment refers to employment in the formal sector F:

$$ATT = E(wage^F|X, sector = F) - E(wage^I|X, sector = F) \tag{26}$$

where X are vector of observed individual and job related characteristics and workers i may be employed in the formal sector, $i \in F$, or in the informal sector, $i \in I$. If one assumes that the conditional independence assumption holds:

$$wage^F, wage^I \perp sector|X \tag{27}$$

i.e., that selection only occurs in terms of the observed characteristics, then (29) can be estimated by:[25]

$$ATT = E(wage^F|X, sector = F) - E(wage^I|X, sector = I) \tag{28}$$

Rosenbaum and Rubin (1983, 1994) have shown that if the conditional independence assumption holds then conditioning on propensity scores, defined as $P(sector = F|X_i)$, is the same as conditioning on the covariates themselves. One can then use these propensity scores to create a sample of 'matched' similar individuals, where matching is done via a chosen matching algorithm. In our case we use the caliper method, using a caliper δ of size 0.001, although it must be noted that we obtained similar results also using nearest neighbor and kernel matching methods.[26] More specifically, each formal sector worker is matched with a set of informal sector workers whose propensity scores lie within 0.001 of the formal worker in question. Assuming reasonable matches the ATT is then just:

$$ATT = \frac{1}{N^M} \sum_{i \in F^M} \left\{ w_i^F - \sum_{i \in I^M} (n_{ij} w_i^I) \right\} \tag{29}$$

where F^M and I^M are the sets of matched formal and informal sector employees, respectively that could be matched, N^M is the total number of these, and for all $(i, j) \in F \times I$:

$$n_{ij} = \begin{cases} 0 & \text{if } |p_i - p_j| > \partial \\ \frac{\frac{1}{|p_i - p_j|}}{\sum_{\{i,j:|p_i-p_j| \leq \partial\}} \frac{1}{|p_i-p_j|}} & \text{otherwise} \end{cases} \tag{30}$$

In order to generate the propensity score to match formal sector workers we estimate a probit model of formal sector employment conditional on all characteristics as listed in Table 2, alternatively with and without the firm size dummies. Importantly for (33) to be an unbiased estimator of the formal sector wage premium it must be emphasized,

[25] See Rosenbaum and Rubin (1983).

[26] Details are available from the authors upon request.

however, that the conditional independence assumption must hold and, thus, that one can argue that the set of covariates X that we use to generate the propensity scores captures all factors that determine both selection into formal sector employment and earnings. While it is not possible for us to test this, given our rich set of characteristics we feel reasonably confident that we are indeed likely to be satisfying the conditional independence assumption.

Matching on our set of covariates according to the algorithm above reduced our sample in the case with the firm size dummies to $5,563$ and for the one without to $5,587$ single men. To assess our success in matching, we, as suggested by Rosenbaum and Rubin (1985), calculated and compared the standardized bias (SB) of the propensity scores for our overall and matched sample using:

$$SB = 100 * \frac{|\overline{p}_F - \overline{p}_I|}{\sqrt{0.5 * (V(p_F) + V(p_I))}} \tag{31}$$

where $p_{F,I}$ is the average propensity score and $V(p_{F,I})$ its variance for the two sectors. Using this we found that the percentage bias reduction was considerable from matching, around 50% when either including or excluding the firm size dummies. We also, as suggested by Sianesi (2004), compared the pseudo R-squared of our matching equation with the pseudo R-squared from re-estimating this on our matched sample. This was found that to be reduced from 0.41 to 0.14 when we did not include firm size dummies, and from 0.53 to 0.25 when these were included. Thus the matching procedure was able to create a sample for which in terms of our explanatory variables much the decision on participation in the formal sector remains random. In order to see if the matching can be substantially improved with a more restrictive calliper, we also experimented with $\delta = 0.0001$. While this further reduced the sample by about 16%, there was no noticeable reduction in the bias or in lower pseudo r-squared values.

Using our matched sample we then proceeded to calculate the ATT as in (31) first for the gross hourly wage rate without using firm size dummies in the matching procedure, the results of which are given in the first row of Table 4. Accordingly, the earnings premium associated with working in the informal sector is 50.2% and statistically significant. Using net rather than gross wages, as shown in the second row, reduces this premium substantially to 35%, but it still remains statistically significant. Matching with the set of our covariates including the firm size dummies in the subsequent row, the ATT on gross wages reduces by 7.7 percentage points, but again lies within standard significance levels. It is only once we assume that informal sector workers do not pay taxes on their wage earnings and use firm size dummies in our matching procedure that the wage premium becomes statistically insignificant. Thus our results suggest, in congruence with our theoretical framework, that in terms of net (of tax) wages, differences in the distribution across employer sizes for informal and formal sector workers and the effect of this firm size wage effect can account for any observed formal sector wage premium.

As a further robustness check we also redid our matching within firm size categories and then calculated out the net wage premium associated with working in the formal sector in the final six rows of Table 4. One should note that this meant matching on small samples, particularly for the very small and the very large categories where there were not many formal and informal sector workers, respectively. Our results show that even within firm size categories there is no significant (net) wage premium. Thus, once one reduces our sample to more homogenous sub-samples in terms of the size of employer there is also no earnings premium for working in the formal sector.

5 Concluding Remarks

The presence of a firm size wage premium in developing countries is well documented in the literature, as is the prevalence of a large informal sector where workers tend to be concentrated in small-low wage firms. In this paper we have drawn these two strands of the literature together to explain the existence of wage premiums for workers in the formal sector. We have shown in a fairly general framework that if there is a firm size wage premium and if large firms are more likely to be caught defaulting on labour taxes, theory predicts what one observes: informal sector firms will be small and thus have low wages, while large firms will pay higher wages and be in the formal sector. In this model the formal sector wage premium is then just a firm size premium. Using South African data we find empirical support for this result.

We also use the equilibrium search model, which in the literature has already been proposed as a natural framework in which a firm size wage premiums may arise endogenously, in our formal/informal sector context and show again that the formal sector wage premium is just a firm size wage differential. One should note that we also find in this example that, because of the impact of firm exit on the shape of the wage distribution, a higher tax rate can reduce the fraction of informal workers in the long run. Less surprisingly, an increase in enforcement or punishment of defaulters is found to reduce the size of the informal sector for a wide range of parameter values.

References

[1] Albrecht, James, and Lucas Navarro, and Susan Vroman. 2005. "Modeling the Effects of Labor Market Reform in a Developing Economy." Mimeo.

[2] Amaral, Pedro, and Erwan Quintin. 2006. "A Competitive Model of the Informal Sector." *Journal of Monetary Economics* 53 (October): 1541-1553.

[3] Boeri, Tito, and Pietro Garibaldi. 2006. "Shadow Sorting." CEPR Discussion Paper no. 5487.

[4] Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer size, and Unemployment." *International Economic Review* 39 (May): 257-273.

[5] Byrne, David, and Eric Strobl. 2004. "Defining Unemployment in Developing Countries: Evidence from Trinidad and Tobago." *Journal of Development Economics* 73 (February): 465-476.

[6] Fugazza, Marco and Jean-Francois Jacques. 2004. "Labor Market Institutions, Taxation and the Underground Economy." *Journal of Public Economics* 88 (January): 395-418.

[7] Gong, Xiaodong and Arthur Van Soest. 2002. "Wage Differentials and Mobility in the Urban Labour Market: A Panel Data Analysis for Mexico." *Labour Economics* 9: 513-529.

[8] Heckman, James J., and V. Joseph Hotz. 1986. "An Investigation of the Labor Market Earnings of Panamanian Males: Evaluating the Sources of Inequality." *Journal of Human Resources* 21 (Autumn): 507-542.

[9] Hussmanns, R. 1994. "International Standards on the Measurement of Economic Activity, Employment, Unemployment and Underemployment." *Bulletin of Labour Statistics*, ILO, Geneva, 1994-4, p.p.

[10] Johnson, Simon, and Daniel Kaufmann, and Pablo Zoido-Lobaton. 1998. "Regulatory Discretion and the Unofficial Economy." *American Economic Review* 88 (May): 387-392.

[11] Manning, Alan. 2003. *Monopsony in Motion, Imperfect Competition in Labor Markets.* Princeton University Press.

[12] Manning, Alan. 2001. "Labour Supply, Search and Taxes." *Journal of Public Economics* 80 (June): 409-434.

[13] Mazumdar, Dipak. 1982. "The Urban Labor Market and Income Distribution: A Study of Malaysia." *Journal of Economic Literature* 20 (June): 626-628.

[14] Mortensen, Dale T. 2003. *Wage Dispersion: Why are Similar Workers paid differently.* MIT Press.

[15] Pradhan, Menno, and Arthur Van Soest. 1995. "Formal and Informal Sector Employment in Urban Areas in Bolivia." *Labour Economic* 2 (September): 275-297.

[16] Pratap, Sangeeta, and Erwan Quintin. 2006. "Are Labor Markets Segmented in Argentina? A Semiparametric Approach." *European Economic Review* 50 (October): 1817-1841.

[17] Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (April): 41-55.

[18] Rosenbaum, Paul, and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *American Statistican* 39 (February): 33-38.

[19] Sianesi, Barbara. 2004. "An Evaluation of the Active Labour Market Programmes in Sweden." *The Review of Economics and Statistics* 86 (September): 133-155.

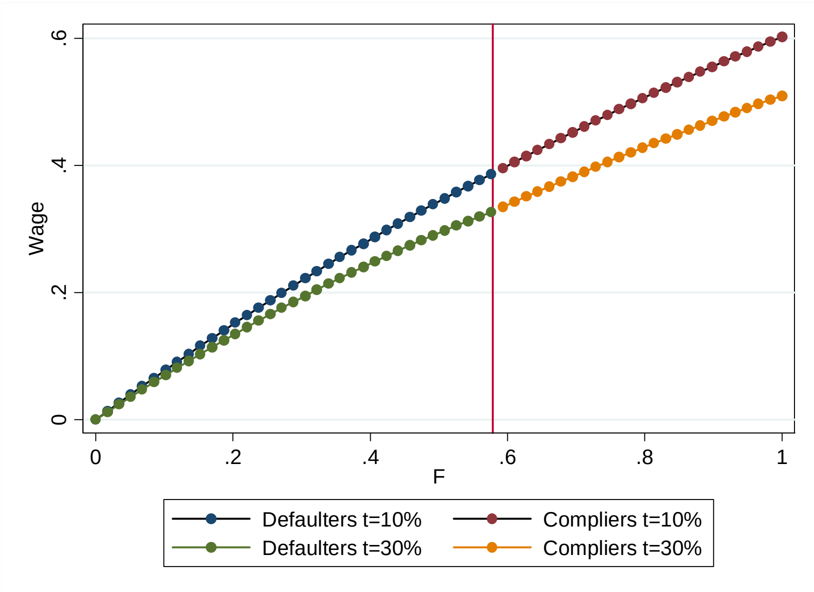
[20] Sianesi, Barbara. 2001. "Implementing Propensity Score Matching Estimators with STATA." Program available at: www.stata.com.

[21] Schneider, Friedrich, and Dominik H. Enste. 2000. "Shadow Economies: Size, Causes, and Consequences." *Journal of Economic Literature* 38 (March): 77-114.

[22] Tansel, Aysit. 1999. "Formal versus Informal Sector Choice of Wage Earners and Their Wages in Turkey." Economic Research Forum Working Paper No. 9927, February 1999.

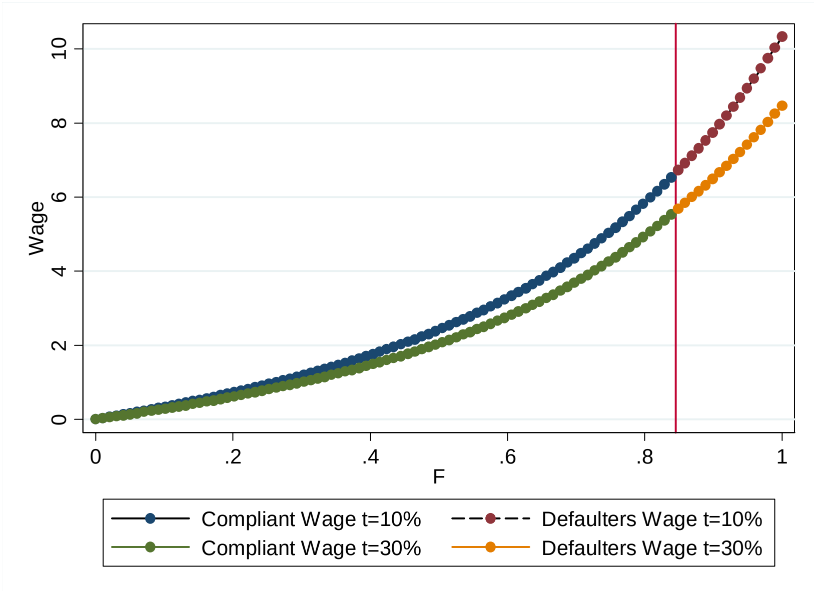
[23] Tybout, James. 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature* 38 (March): 11-44.

Figure 1: Defaulters and compliers inverse wage offer distributions for tax rates of 10% and 30% and exogenous productivity.



Notes: For both graphs we assume $s=0.2$, $b=0$, $p=1$, and $k=1$, and follow Mortensen (2003) and assume $\lambda=0.287$ and $\delta=0.207$. One should note in particular that the assumption $b=0$ simplifies the derivation of M and causes the equilibrium mass of firms and cut-off value of F to be constant when t changes in both graphs.

Figure 2: Defaulters and compliers inverse wage offer distributions for tax rates of 10% and 30% and endogenous productivity.



Notes: We make the additional assumption that σ=2 for this graph

Table 1: List of Variables of Interest

Variable name	Definition of the variable
Hourly Wage	Real hourly logged wage calculated using a person's income, hours worked in their main job and the South African consumer price deflator.
Black White Coloured	Three dummies related to a person's race (the population group that the worker belongs to).
Married	Variable defining the marital status of a person as married.
Afrikaans English	Two dummies defining the most often spoken language of the worker at home.
No primary (can not read and write), No primary (can read and write), Primary, Secondary, NTC, University	Six dummies associated to a person's education level (the highest level of education completed).
Age	A worker's age (restricted to the interval 15-70).
Job training	The possibility for the worker to be trained in skills that can be used for work.
Occupation	Ten dummies for the occupation variables.
Urban area	Dummy for whether living in an urban area.
Tenure	The period (in years) during which the person was working with the same employer he/she mentioned.
Tools	Dummy for whether the person owns the tools and/or the equipment that he/she uses at work.
Supervision	Dummy variable for whether the work is supervised.
Part-time job	Classifying the job as a full-time job or part-time job (part-time work dummy).
1 worker, 2-4 workers, 5-9 workers, 10-19 workers, 20-49 workers and $\geq$ 50 workers	Six dummies related to the firm size.
Industry	Eleven dummies for the industry variables (the eleventh industry dummy 'Exterior organizations and foreign government' is omitted).

Table 2: General Summary Statistics

		Formal	Informal
log(Gross Wage)	Mean	1.39	0.79
log(Net Wage)	Mean	1.22	0.79
1 employee	% of total	0.03	0.45
2-4 employees	% of total	0.12	0.35
5-9 employees	% of total	0.16	0.10
10-19 employees	% of total	0.20	0.06
20-49 employees	% of total	0.21	0.03
50+ employees	% of total	0.28	0.01

Table 3: Ratio of the Formal Relative to the Informal log Wage Rate by Employer Size

Firm Size	log(Gross Wage) Ratio	log(Net Wage) Ratio
1 employee	1.66	1.38
2-4 employees	1.28	1.07
5-9 employees	1.21	1.05
10-19 employees	1.34	1.17
20-49 employees	1.34	1.17
50+ employees	1.92	1.71

Table 4: Estimate of ATT of the Formal Sector Wage Premium

Sample	Wage	Firm Size DVs Included	ATT	Standard Error	Matched Sample
Total	Gross	No	0.502**	0.057	5587
Total	Net	No	0.350**	0.055	5587
Total	Gross	Yes	0.423*	0.186	5563
Total	Net	Yes	0.241	0.157	5563
1 employee	Net	-	0.010	0.228	126
2-4 employees	Net	-	0.010	0.103	451
5-9 employees	Net	-	-0.006	0.157	367
10-19 employees	Net	-	-0.079	0.156	449
20-49 employees	Net	-	0.020	0.244	366
50+ employees	Net	-	0.299	0.515	51

(1) ** and * stand for one and five per cent significance levels, respectively;

(2) Standard errors generated via bootstrapping using 500 replications;

(3) Matching done separately for individual firm size categories.

Appendix 1: Proof of Proposition One

To show this we take the derivative of the difference in the value of firms in equation (2) to get:

$$\frac{\partial(V^c - V^d)}{\partial w} = \frac{\Omega\Theta\frac{\partial n}{\partial w} + (nt + wt\frac{\partial n}{\partial w})(\Theta\frac{\partial\Omega}{\partial B} - 1)}{r + \delta} \tag{32}$$

As long as $\frac{\partial B}{\partial\Omega(B)} < \Theta$, then (32) will be positive. From equation (2) in the text we see that a firm will comply if $(V^c - V^d) > 0$, which implies:

$$\Theta > \frac{B}{\Omega} \tag{33}$$

We also note that if the elasticity of punishment with respect to the tax bill is greater than unity, i.e., $\frac{\partial\Omega}{\partial B}\frac{B}{\Omega} \geq 1$, then using (33) and this elasticity we see that if $(V^c - V^d) > 0$ the following inequality holds:

$$\frac{\partial B}{\partial\Omega(B)} \leq \frac{\Omega}{B} < \Theta \tag{34}$$

Next we assume that in equilibrium there are some defaulting and some compliant firms. Now we find the smallest firm where $(V^c - V^d) > 0$ and call it firm*. Since inequality (34) holds for firm* then the derivative in (32) is positive for firm*. This implies that the firm that is just bigger than firm* will also choose the formal sector, and so on, for all larger firms up to the largest firm. All firms below firm* must be defaulters since firm* is the smallest compliant firm by assumption.

Appendix 2: The Impact of a Change in t or s on the Mass of Firms when Productivity is Exogenous

While the Section 2.2.2 derives the wage offer distribution, one generally observes the wage distribution in the data, i.e., the fraction of workers paid different wages or the fraction of workers in the informal sector etc. We note from (6) though that the wage distribution is a monotonic transformation of the wage offer distribution. For $z \in (s, t)$ one sees that:

$$\frac{dG^*}{dz} = \frac{\delta \frac{dF^*}{dz}[\delta + \lambda(1 - F^*)] + \delta F^*[-\frac{d\lambda}{dM}\frac{dM}{dz}(1 - F^*) + \frac{dF^*}{dz}\lambda]}{[\delta + \lambda(1 - F^*)]^2} \tag{35}$$

From (35) we see that when λ is fixed the sign of the derivative of the wage distribution is the same as the sign of derivative of the wage offer distribution. Moreover, if λ is increasing in the mass of firms $\frac{dM}{dz} < 0$ is a sufficient condition for

$$sgn\frac{dG^*}{dz} = sgn\frac{dF^\star}{dz} \tag{36}$$

This means comparative static results given for the wage offer distribution below will also apply to the wage distribution when $\frac{dM}{dz} < 0$. In the remainder of this appendix we will determine conditions where $\frac{dM}{dz} < 0$, which in turn ensures (36) holds even if λ depends on the mass of firms.

We define two conditions which we will use in the comparative static analysis:

- Condition one: $p > b(1 + \sigma t)$
- Condition two: $\frac{\lambda}{\delta} > \frac{\varepsilon_{\lambda M} - 1}{1 + \varepsilon_{\lambda M}}$

Totally differentiating (15) with respect to t and M and setting the derivative equal to zero, we get the following expression:

$$\begin{aligned} &\Big\{\frac{\partial k}{\partial M} \\ &+\Big[\frac{(p-b)n(b)}{\delta} - \sigma\frac{bstn^\sigma(b)}{\delta}\Big]\frac{1}{M} \\ &+\Big[\frac{(p-b)n(b)}{\delta} - \sigma\frac{bstn^\sigma(b)}{\delta}\Big]\Big(\frac{\lambda-\delta}{\delta+\lambda}\Big)\frac{\partial\lambda}{\partial M}\frac{1}{\lambda}\Big\}dM \\ &+\frac{bsn^\sigma(b)}{\delta}dt = 0 \end{aligned} \tag{37}$$

The first line constitutes the change in fixed entry costs from a change in the mass of firms, the second term is the direct impact of a change in the mass of firms, the third line is the derivative from a change in offer arrival rates resulting from a change in firm entry, and the fourth line provides the derivative with respect to a change in the tax rate t. One should note that if one totally differentiates (15) with respect to the punishment/enforcement rate s and M one would get the same expression as (37) except that the final term would be: $\frac{btn^{\sigma}(b)}{\delta}ds$. Also, we assume that $\frac{\partial k}{\partial M} > 0$ and $\frac{\partial \lambda}{\partial M}$ and define $\frac{\partial \lambda}{\partial M}\frac{M}{\lambda} = \varepsilon_{\lambda M}$ as the elasticity of the arrival rate with respect to firm. We can multiply (37) by M and rewrite it as:

$$\Big\{\frac{\partial k}{\partial M}M + \Big[\frac{(p-b)n(b)}{\delta} - \sigma\frac{bstn^{\sigma}(b)}{\delta}\Big]$$
$$\Big[\frac{\lambda(1+\varepsilon_{\lambda M}) + \delta(1-\varepsilon_{\lambda M})}{\delta+\lambda}\Big]\Big\}dM + \frac{bsn^{\sigma}(b)}{\delta}dt = 0 \tag{38}$$

First we will take the left hand side term in square brackets from the second line of (38):

$$\frac{(p-b)n(b)}{\delta} - \sigma\frac{bstn^{\sigma}(b)}{\delta} \tag{39}$$

We note from (12) that in any equilibrium where there are some defaulting firms $s < \frac{1}{n^{\sigma-1}(b)}$. Substituting the right hand side in for s in (39) we see Condition One is a sufficient condition for this expression to be positive. Next we see that the second term in squared brackets in (38) is positive if Condition Two holds. We can conclude that $\frac{dM}{dt} < 0$ if Conditions One and Two hold.

While Condition Two may not hold, for reasonable parameter values the indication is that it will hold unless $\varepsilon_{\lambda M}$ is very large. For example if $\lambda > \delta$ Condition Two certainly holds, or, taking the values $\lambda = 0.207$ and $\delta = 0.287$ used by Mortensen (2003) in his simulations, Condition Two will hold as long as $\varepsilon_{\lambda M} < 6.17$. One should also note that this is a sufficient condition, so there is a range of parameter values where Conditions One or Two fail but continues to hold. We also remark that in the simpler case where $\frac{dM}{dt} < 0$ is not dependent on the mass of firms Condition Two always holds so that Condition One is sufficient for $\frac{dM}{dt} < 0$.

Appendix 3: Comparative Static Results

The sign of $\frac{dG*}{ds}$ and $\frac{dG*}{dt}$ when productivity is exogenous.

The derivatives of (12) are:

$$\frac{dF^*}{dt} = \frac{\partial F^*}{\partial M}\frac{dM}{dt} = \frac{1}{M}\left\{\frac{1}{2}\sqrt{\frac{s^{\frac{1}{\sigma-1}}\delta}{\lambda M}}\Big[1+\frac{d\lambda}{dM}\frac{M}{\lambda}\Big] - \frac{\delta}{\lambda}\frac{M}{\lambda}\frac{d\lambda}{dM}\right\}\frac{dM}{dt}$$
$$= \frac{1}{M}\left\{\frac{1}{2}\sqrt{\frac{s^{\frac{1}{\sigma-1}}\delta}{\lambda M}}[1+\varepsilon_{\lambda M}] - \frac{\delta}{\lambda}\varepsilon_{\lambda M}\right\}\frac{dM}{dt} \tag{40}$$

One should also note from (12):

$$1 - F^* = \frac{-\delta}{\lambda} + \sqrt{\frac{s^{\frac{1}{\sigma-1}}\delta}{\lambda M}} \tag{41}$$

Using this (40) can be written:

$$\frac{dF^*}{dt} = \frac{\partial F^*}{\partial M}\frac{dM}{dt} = \frac{1}{M}\left\{\sqrt{\frac{s^{\frac{1}{\sigma-1}}\delta}{\lambda M}}\Big(\frac{1-\varepsilon_{\lambda M}}{2}\Big) + (1-F^*)\varepsilon_{\lambda M}\right\}\frac{dM}{dt} \tag{42}$$

We note that if λ is inelastic with respect to firm entry, i.e. $\varepsilon_{\lambda M} < 1$, this ensures that $sgn\frac{dF^*}{dt} = sgn\frac{dM}{dt}$. Since $\varepsilon_{\lambda M} < 1$ ensures that Condition Two is satisfied, this means that $\varepsilon_{\lambda M} < 1$ and condition one are sufficient for $\frac{dM}{dt} < 0$ and by implication $\frac{dF^*}{dt} < 0$ and $\frac{dG^*}{dt} < 0$ when productivity is exogenous.

$$\frac{dF^*}{ds} = \frac{\partial F^*}{\partial s} + \frac{\partial F^*}{\partial M}\frac{dM}{ds}$$
$$= \frac{1}{M}\left\{(1-F^*)\varepsilon_{\lambda M} + \sqrt{\frac{s^{\frac{1}{\sigma-1}}\delta}{\lambda M}}\Big(\frac{1-\varepsilon_{\lambda M}}{2}\Big)\right\}\frac{dM}{ds} - \frac{1}{2s(\sigma-1)}\sqrt{\frac{s^{\frac{1}{\sigma-1}}\delta}{\lambda M}} \tag{43}$$

If $\varepsilon_{\lambda M} < 1$ and Condition One holds, this ensure that both terms in (42) are negative and $\frac{dF^*}{ds} < 0$ and $\frac{dG^*}{ds} < 0$ when productivity is exogenous.