Is There An Informal Sector Wage Penalty?

Evidence from South Africa

Eliane Badaoui

THEMA - Université de Cergy-Pontoise

Eric Strobl*

Ecole Polytechnique Paris

Frank Walsh

University College Dublin

***** Corresponding author: Dept. of Economics, Ecole Polytechnique, 91128 Palaiseau Cedex, France. Email: eric.strobl@shs.polytechnique.fr

Abstract

We estimate the wage penalty associated with working in the informal sector. To this end we use a rich South African data set on males that allows one to accurately distinguish workers employed in the informal sector from those employed in the formal sector and link individuals over time. Implementing various econometric approaches we find that there is a gross wage penalty of a little over 12 per cent for working in the informal sector. However, once we reduce our sample to a group for which we can reasonably calculate earnings net of taxes the wage penalty disappears.

JEL Classification: J31; O17

Keywords: informal sector, wage penalty, South Africa

Section I: Introduction

The informal sector has traditionally been viewed as a temporary alternative to unemployment and poverty (Fields, 1975) which tends to disappear as an economy matures and becomes more developed. It is not surprising then that many economists initially associated informal economic activity with developing countries (De Soto, 1989; Marshall, 1987) where decent work deficits are most pronounced and social security nets are relatively underdeveloped. However, in contrast to such a view of the informal sector as a transitional marginal phenomenon, recent evidence seems to indicate that it may be more of a long term feature of developing economies (Bekkers and Stoffers, 1995; Charmes, 2000), particularly in Africa and Latin America where there seem to be expansionary tendencies.[1] If informal sector employment is indeed a more permanent, not necessarily self eradicating, feature of developing countries, then clearly understanding its workings is essential to comprehending labor markets and, more generally, poverty in developing countries.[2] Unfortunately, data constraints have generally not allowed researchers to clearly identify informal sector employment. Moreover, these constraints have limited the number of comprehensive empirical studies of the informal sector to a handful, so that even stylized facts about informal labor markets remain disputed.

One seemingly stylized fact is that informal sector workers, even if equally productive, are subject to lower remuneration than their formal sector counterparts, as suggested by many of the earlier empirical studies.[3] There a number of explanations that has been offered in this regard, mostly based on a segmented view of the labor market. For instance, the presence of barriers to entry into the formal sector could pose a possible cause, so that working in the informal sector is associated with a negative wage premium even for equally productive workers; see Fields (1975) and Mazumdar (1975). However, several more recent studies postulate that it is more efficient for

---

[1] According to Maloney (2004), the informal sector comprises 30-70 per cent of the labour market of most Latin American countries. Friedman *et al.* (2000) estimate that 14-62 per cent of output comes from the informal sector across a range of transition economies.

[2] Poverty is a key characteristic of the informal sector (Pradhan *et al.* 1999).

[3] See, for example, Mazumdar (1982), Heckman and Hotz (1986), Roberts (1989), Pradhan and Van Soest (1995), Tansel (1999), and Gong and Van Soest (2002).

entrepreneurs to remain outside the often underdeveloped and inefficient regulatory umbrella of the formal sector; see Tybout (2000). Similarly, Maloney (1998) introduces a dualistic perspective according to which workers may find informal sector employment a desirable alternative, both due to inefficiencies in the formal sector and low levels of labor productivity. A wage penalty for informal sector employment may also be due to sorting, where those with low levels of human capital are also those more likely to work in the informal sector (Tokman, 1982). Such sorting may arise in part because firms' access to financing is relatively more limited in the informal sector and because employers with low degrees of capitalization tend to recruit less able workers; see, for example, Amaral and Quintin (2006).

As with the theory, many of the empirical studies in more recent years seem to hint at the possibility that wage differences between formal and informal workers may not be a stylized fact as previously thought. For example, Marcouiller, Ruiz and Woodruff (1997) applied wage regressions to calculate unexplained wage gaps between the two sectors. The results show that significant wage premia are associated with work in the formal sector in El Salvador and Peru, whereas, in contrast, a premium is associated with informal work in Mexico. Tansel (2000) carried out an analysis for men and women workers separately using the 1994 Turkish Household Expenditure Survey, defining uncovered wage earners and self-employed as part of the informal sector, while covered wage earners are interpreted to be part of the formal sector. The results indicate substantial earnings differences between the formal and informal sectors for men, but not so for women. Also for Mexico, Gong and Van Soest (2002) find that wage differentials between the formal and informal sector are typically small for the lower educated, and only arise with increasing levels of education. In addition evidence from Tannuri Pianto and Pianto (2002) suggests that differences in returns to attributes explain around 30 per cent of their earnings gap at low quantiles, while the gap is completely explained by differences in their individual characteristics at high quantiles of the distribution. Pratap and Quintin (2006), on the other hand, find that after controlling for selection no wage premium remains and that job satisfaction is no lower in the informal sector for Argentinean data.

In the current paper we study the possible existence of a wage penalty for males working in the informal sector in the South African labor market. Arguably, one of the major stumbling blocks to being able to more precisely identify and understand features specific to the informal sector has been the difficulty in properly classifying activities as 'informal'. For one, while the ILO adopted a number of recommendations on this issue, there are still some major deviations from the international definition in use. In particular, data restrictions have not generally allowed a more uniform use of the term in empirical analysis. For example, some authors have defined 'informality' by a priori assuming that certain characteristics that belong to formal activities, such as high income and wages, job protection and social security system (see, for example, Banerjee, 1985). Informal workers have also been considered to be those who are not in regular employment (Kingdon and Knight, 2001; Barrientos and Barrientos, 2002) or those who have activities that are not reported to the government and to the tax authorities (Masatlioglu and Rigolini, 2005). Moreover, some empirical papers used small firm size as a criterion to define the informal sector either alone or in combination with criteria such as the type of location of the workplace or the health insurance provided for workers. In this regard, our data at hand, the South African Labour Force Survey, explicitly asks individuals if they are working in the informal sector. It also allows some cross-checking to verify if they are truly working outside the formal regulatory umbrella, thus giving us a more accurate measure of informal sector activity than most previous studies. Moreover, the data is relatively rich in other information pertaining to levels of human capital and job characteristics.[4] Finally, the rotating panel nature of the survey allows us to follow individuals' labor market experience over time. This information and appropriate econometric techniques arguably enables us to give a more precise estimation of the wage penalty associated with working in the informal sector than previous studies.

[4] For example, some studies related to the informal sector in South Africa show that it is dominated by black South Africans and that education levels for workers in the informal activities are low (Devey, Skinner and Valodia, 2002).

Arguably South Africa makes a good case study of the wage penalty associated with informal sector employment in that while there is extremely high and rising unemployment, the informal sector remains small compared to other developing countries.[5] As a matter of fact, Kingdon and Knight (2003) have argued that currently unemployed workers in South Africa are involuntarily unemployed in the sense that they would accept formal sector jobs at the going wages, but may voluntarily choose not to enter the informal sector because of low incomes associated with it.[6]

The remainder of the paper is organized as follows. In the following section we describe our data set and provide some summary statistics. In Section 3 we isolate the wage penalty associated with being employed in the informal sector using various econometric approaches. Concluding remarks are provided in the final section.

Section II: Data Description and Summary Statistics

A. Data Description

Our data source is the South African Labour Force Survey (LFS)[7]. The LFS 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.[8]

In terms of classifying informal sector activity, the LFS 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 where he/she works is

---

[5] The informal sector in South Africa has been estimated to represent between 16 and 40 per cent of GDP (Abedian & DeSmidt 1990; DeSmidt 1988; Thomas 1989).

[6] There have, however, been some notable changes in the importance of the informal sector over time. For example, over the period 1997 until 2001 the number of formal employees marginally increased by about 6.8 per cent, whereas informal employment grew by about 48.5 per cent. More recently, there appears to be decline in the informal sector.

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

[8] We limited our analysis to these waves since starting in March 2004 a completely new sample of households was introduced, hence 'breaking' the rotation with the previous waves.

in the formal sector or in the informal sector (including domestic work)'.[9] Additionally, there are a number of other questions regarding fringe benefits and other aspects of the 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.[10] Incorporating these reclassifications with the direct information on formality gives us our benchmark definition, which we refer to as Definition A.

Since we are specifically interested in the pay differential associated with working in the informal sector, a second important piece of information required from our data is that concerning remuneration. For those persons in paid employment, the LFS explicitly asks the remuneration in their main activity. More precisely, the LFS 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 per cent of individuals who were in paid employment the salary was reported in income categories.[11] 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 per cent of the total sample, of which 10 per cent indicated that they worked

[9] 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'.
[10] However, in the end there were only 2.1 per cent of observations where we needed to change their status. The correlation between the two classifications was about 0.96
[11] Of these about 6 per cent reported to be in the informal sector according to our definition above.

in the informal sector (as defined above).[12] A final important point with regard to our earnings data is that we do not have explicit information on the value of other benefits of the job, such employer contributions to pension, health insurance, job stability, etc., only on the reported gross income. We can thus only estimate any existent informal sector wage penalty in terms of the monetary remuneration to the worker.

Part of our estimation strategy is to be able to control for unobserved individual time invariant effects. In this regard one should note that the LFS is a rotating panel in that in every round about 20 per cent households are replaced. However, while households can be identified over time via a unique household identifier, there is no such identifier for individuals within the household. In order to link individuals across rounds over time we thus used the method proposed by Madrian and Lefgren (1999) utilizing information on individuals' sex, race, and age. Each wave this allowed us to link on average 72.1 per cent of observations of individuals who claimed to have been in the same household at the previous survey.[13, 14] We restrict our analysis mainly this linked data.

Apart from an explicit definition of the formality of individual's employer and a precise measure of remuneration, the LFS can also be regarded as relatively rich in other information potentially relevant to an individual's labor market status. In this regard, we 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 (such as age, gender, race, marital status, education level, occupation, whether they ever received any job training) and job characteristics (like firm size, industry, supervision, urban area, part-time status, tools) etc.. We provide a comprehensive list of these and their definitions in Table A1 in the Appendix.

---

12 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.

13 Using this methodology Byrne and Strobl (2004) were able to link 80 per cent of individuals for the developing country of Trinidad and Tobago.

14 On average 4.9 per cent state to have not been in the same household during the previous survey.

Finally, we reduced our sample to non-self employed males aged between 15 and 65 working in industries other than the public sector. One should note that focusing only on males allows us to abstract from the often more complex labor 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.[15]

B. Summary Statistics

All in all, we, after dropping observations with missing values for any of our variables, were left with a total number of 11,571 observations on 5,829 number of linked males. Of these observations 11.1 per cent were for individuals working in the informal sector as defined above. It is also noteworthy that overall the share of informal sector in total employment fell over our sample period, starting at 11.5 per cent in September 2001, and ending in 9.5 per cent in September 2003.

Table 1 presents some basic summary statistics concerning the dispersion of characteristics of our sample as broken down into those working in the formal and those employed in the informal sectors. Accordingly, one finds that gross logged wages in the formal sector are on average 126 per cent times larger than those found in the informal sector. For convenience sake, we divide these into those that are related to human capital characteristics and generally time invariant (except for age), and those that are likely to be easily transferable across employers of all

[15] Additionally, as will be seen, an important part of our econometric analysis focuses on job movers, which we identify by their length of tenure, the information of which is not available for the self-employed. Dropping the self-employed resulted in dropping 1230 individuals, of which 34 per cent stated that their business was not formally registered.

types, and thus that are more job specific. As can be seen, there is little difference in the average age of employees in the two sectors. In contrast, there are some notable differences in the dispersion of race across the formality of sectors. For example, a much larger proportion of the workforce in the informal sector is Black, while there are fewer White employees. One possible reason may be the racial historic differences in South African society and the apartheid racial discrimination which touched every aspect of social life, sanctioned "white-only" jobs, and inhibited the development of entrepreneurial skills and social networks (Devey, Skinner and Valodia, 2003), and thus may have acted as a barrier to entry for these to formal sector employment.

One also discovers that workers in the formal sector are more likely to be married than their informal sector counterparts. Additionally, there are some differences in the linguistic abilities of workers across the two broad sectors. Specifically, on average both Afrikaans and English are more likely to be the primary language of workers in the formal sector than for those in the informal economy. Measures of education serve not only as a fairly direct proxy of general human capital, but are also likely to be correlated with other unobserved abilities. In this regard, one finds that workers in the informal sector are less likely to be able to read and write and are more likely to be equipped with an education level below that of those employed in the formal sector. One may also note that on average they receive less on the job training. In terms of the occupational structure one finds that informal sector workers are most likely to be in Elementary Occupations[16] (33 per cent), followed by Skilled Agricultural And Fishery (23 per cent) and Craft and Related Trade (22 per cent). Those in the formal sector are most likely to be found in the Elementary Occupations (27 per cent), Plant and Machinery Operators and Assemblers (24 per cent), and Craft and Related Trade (20 per cent).

[16] Elementary occupations include street vendors and related workers; shoe cleaning and other street services; domestic and related helpers, cleaners and launderers; building caretakers, window and related clearners; messengers, porters, doorkeepers and related workers; garbage collectors and related labourers.

The characteristics of a job may also be driving the perceived 'negative' average wage premium of being employed in the informal sector. For example, the dummies indicating employer size show that firms with less than five regular employees are substantially more likely to be part of the informal sector. In contrast, much of the formal sector labor force is employed by firms with a workforce greater than five. One should note in this regard that it has been shown that there is a substantial wage premium associated with working for 'large' employers; see Oi and Idson (1999). Whether the work is directly supervised may also explain some of the discrepancy of payment across firms and sectors, since according to the efficiency wage theory (Shapiro and Stiglitz, 1984) employers may choose to pay workers a premium above the market rate in order to discourage workers from 'shirking' on the job. However, our data suggest that, on average, formal sector workers are more likely to be supervised than their informal sector counterparts.

One also discovers that, for the informal sector, the probability that jobs are part-time is much higher. Moreover, workers tend to accumulate on average much less tenure in the informal sector, hence suggesting lower stability of employment. Finally, one finds, that slightly over 41 per cent are employed in the Private Households Sector, while other high informal worker employers are located in Construction and Agriculture, and Hunting, Forestry, and Fishing. One should note that of those employed by private households 51 per cent are in Skilled Agricultural and Fishery, 30 per cent in Elementary, and 10 per cent in Domestic occupations. Thus it appears that most of the Skilled Agricultural and Fishery male workers that state that they are employed by private households are likely to be in small scale farming/fishing run by private households. In contrast, the formal sector is much more evenly distributed than in the informal counterpart, with Agriculture, Hunting, Forestry, and Fishing, Mining and Quarrying, Manufacturing, and Wholesale and Retail trade constituting the most frequent employers.

Section III: Econometric Analysis

The problem of measuring any potential informal sector wage penalty boils down to trying to answer the counterfactual question: what wage would a person employed in the informal sector have if he/she were instead employed in a similar job in the formal sector? In this regard, our problem is similar to those commonly undertaken in laboratory experiments where the effect of a 'treatment' is assessed by comparing treatment and control groups, i.e., where we try to measure the informal sector penalty by comparing the remuneration of those in the informal sector to those employed in the formal sector. However, unlike laboratory experiments, and as it is the case for most economic questions of this nature, the data at hand is non-experimental and it is well recognized that the estimate of a causal effect obtained by comparing treatment groups with non-experimental control groups may be biased because of problems such as self-selection or some systematic judgment by researchers in selecting units to be part of the treatment group.[17] For example, our summary statistics suggested that while informal sector employees are likely to earn less than their formal sector counterparts, they are also different in other human capital and job aspects that may at least in part be driving this differential. Of course we can think of many reasons why the unobserved as well as observed characteristics of informal and formal sector workers would differ. For example, the unusually high unemployment rate in South Africa has been argued to be due to the possibility that for some potential workers the wage in the informal sector is too low[18], given their possible high reservation wage. While we have no information on reservation wages, we do however roughly control for changes in the unemployment rate with time and regional dummies in our analysis.

One should note that dealing with the possible sample selection bias is the main challenge of any evaluation study with non-experimental data. In the context of measuring the informal sector wage penalty, the approaches have differed widely. Most of the earlier studies have simply either implicitly or explicitly assumed that the information available on workers was sufficient to

[17] See Dehejia and Wahba (2002).
[18] See Kingdon and Knight (2003)

control for sample selection bias and run OLS on a mincerian wage equation with, amongst other controls, an indicator of the formality of the job.[19] Others have implemented more sophisticated two stage models, where participation is jointly specified with the wage regressions; see, for example, Gong and Van Soest (2002), and Tannuri-Pianto and Pianto (2002). More recently researchers have resorted to matching estimators to deal with the sample selection problem inherent in the analysis; see, Pratap and Quintin (2006). Also, Guenther and Launov (2006) use a mixture model which is a generalization of a Heckman regression that allows for two types of informal sectors, a competitive one and a segmented one.

In order to estimate the impact of informal sector jobs, we specify the following standard Mincerian type wage equation:

$$w_{it} = \alpha + \beta_1 I_{it} + \beta_2 X_{it} + \beta_3 Z_{it} + \varepsilon_{it} \tag{1}$$

where $\varepsilon_{it} = \mu_i + \eta_{it}$, i=1,…,N represents individual units and t=1,…, T time periods. w_{it} denotes the logarithm of real hourly wages, X_{it} and Z_{it} are set of human capital and job characteristics, respectively. I_{it} is a dummy variable taking value 1 if the firm of worker i is in informal sector and 0 otherwise. ε_{it} is the error term.

A. Ordinary Least Squares

We start by estimating (1) with OLS in levels. Initially we include only the informal dummy for the whole sample of linked individuals without any other explanatory variables except seasonal and year dummies. Our results of this exercise, given in the first row of Table 2, show that the wage penalty for workers employed in the informal sector is found to be around 112.8 per cent. However, as noted earlier, wage differences can be attributed to many different levels

[19] See, for example, Marcouiller *et al.* (1997).

of human capital and various characteristics of the jobs that may also be correlated with selection into the informal sector. We thus proceeded by first including human capital variables and, as can be seen, the negative premium falls substantially to 53.6 per cent. Moreover, introducing all our available job characteristics further decreases the coefficient on the informal sector dummy to -0.372. Thus, differences in the distribution of observable human capital and job characteristics can explain nearly three quarters of the average differences in wages between informal and formal sector.

One advantage, at least in the short run, of working in the informal sector is arguably that one is not subject to income taxation like employees of registered firms would normally be. Thus, from an immediate net return perspective, one may be overestimating the premium associated with working for a formal sector employer if one solely examines gross earnings. While one would ideally like to take account of this, it is difficult from simple labor force data, where there is no information on non-labor income and where we cannot easily link immediate family members within a household, to accurately estimate the amount of labor income that is likely to be deducted in terms of taxes for most labor market groups.

In order to nevertheless gain insight into how taxation may affect the negative premium for informal sector work we thus focus on single persons for which we can relatively easily calculate what income tax liabilities are for a given annual income from employment.[20] In order to ensure that single individuals are not that different in terms of the gross earnings informal sector penalty, we first determined the OLS estimate of the wage penalty for single persons only and found this, as shown in Table 2, to be almost identical to that of the total sample, with a coefficient of -0.363. We subsequently calculated the log of the net (after taxes) real hourly wage rate for formal sector workers by using the income tax tables provided by the South African

[20] When reporting their earnings individuals are explicitly asked to state the amount including overtime, allowances and bonus, but before any tax or deductions and hence we can reasonably interpret this amount to be gross earnings. One possible concern may nevertheless be that household heads who are answering these questions for other members may only know their net earnings. Unfortunately we have no way to investigate this possibility.

Revenue Service.[21] One should note, from Table 1, that after taxes, formal sector wages are only 101.1 per cent larger than their informal counterparts. Using this new series for formal sector workers, while retaining the log of the real gross hourly wage rate for informal sector employees, we estimated the OLS effect of informal sector status on net earnings and found the coefficient to be 48 per cent lower (-0.188), which indicates that looking solely at gross wages may substantially overestimate any difference in earnings.

B. Difference-in-Differences Estimator

While our data set is arguably relatively rich in individual and worker level characteristics, there may still be a considerable probability that there are other unobserved factors that determine both selection into the informal sector and wages. One obvious example would be unobserved productivity that is not correlated with the educational level. The failure to account for such could then lead to biased estimates of the informal wage sector penalty. Given the panel nature of our data, one natural way to purge such unobservables, if they are time invariant, is to take first differences of our wage equation, thus essentially isolating what is known as the difference-in-differences (DID) effect of informal sector status.[22]

The results of taking first differences of the wage equation for our whole sample are provided in the middle panel of Table 2.[23] As can be seen, this considerably lowers the estimated wage penalty to 12.3 per cent, thus suggesting that time invariant unobservables are an important factor behind the observed informal wage penalty even after controlling for a rich set of characteristics. In order to investigate whether the distinction between gross and net wages may still be important after controlling for time invariant unobservable factors, we once again

---

[21] See http://www.sars.gov.za/;

[22] Another possibility would be to use an instrumental variables approach. However, it is notably difficult to find instruments that determine selection into different sectors of the labour market that convincingly do not determine wages.

[23] One should note that in general our human capital variables are purged from the DID equation because they are mostly time invariant over our short time period. We thus only include the first differences in the job characteristics which are more likely to change across jobs.

reduced our sample to single workers and compared the informal sector wage penalty using alternatively our 'gross' and 'net' definitions. As can be seen in the second and third rows of the middle panel of Table 2, the wage penalty is a little over seven percentage points higher for the single sub-sample, but then completely disappears when one uses net log wages as the dependent variable for formal sector workers. This again suggests that the wage premium for working in the formal sector observed with gross wages may simply be the result of not taking account of the income taxation that informal sector workers are likely able to avoid.

One worry in using DID to isolate a wage penalty in the informal sector is that variation in the informal sector dummy may be solely due to job movers that move between sectors, so that a large part of the control group consists of job stayers.[24] For example, from those working in the formal sector only about 6.5 per cent in any period move to a new job, while the corresponding figure for the informal sector is 24.5 per cent. Job movers may, however, be very different from job stayers and the literature has proposed a number of reasons in this regard. For example, human capital models highlight the importance of employer-specific human capital, part of which is not transferable and hence movers are likely to experience earnings losses.[25] In contrast, job matching models would predict positive gains since workers leave their employers in search of better matches.[26] Importantly, under either of these explanations, differences in the propensity in job mobility across formal/informal sectors could very well then be driving the results observed in our DID estimation. As a first step to investigate this we ran a simple probit regression to investigate the determinants of the probability to move jobs for informal and formal sector workers separately, the results of which are given in the first two columns of Table. Accordingly, one finds that for the formal sector a greater wage, tool use, supervision, and greater firm size (except for those working in the largest category) reduce the probability of moving to a new job. Similarly workers in the formal sector with higher age and

[24] As a matter of fact over 60 per cent of our observations refer to job stayers. Of the job movers slightly over 10 per cent move between the informal and formal sector.
[25] Hashimoto (1981) for a theoretical model or Farber (2005) for evidence of wage penalties to displaced workers.
[26] See Jovanovic (1979).

tenure are less likely to change employers, although at a decreasing rate. In contrast, none of the reported factors have a significant influence on the probability to move for informal sectors.

A simple manner to investigate whether the inclusion of stayers as control group is biasing results in the DID estimation, is to just reduce our sample to job movers, so that the control group is now only those that moved jobs but remained within their sector. In order to define such job movers we utilize the information on their tenure and classify any person who reports less than six months of tenure as a person who has started a new job.[27] Raw averages show that from those with formal employment who switch jobs about 9.6 per cent move to the informal sector, while 47.5 per cent of job movers in the informal sector on average move in the opposite direction. To investigate what can predict the probability of whether a person moves to a job in another rather than the same sector, we also ran simple probits of the probability to move to a different (in terms of formality) sector for formal and informal sector workers separately, as shown in columns 3 and 4 in Table. As can be being Black, speak Africaans or English, greater tenure (at a decreasing rate), supervision at the current job, and greater firm size (up to a threshold) decrease the probability that the job change will be from the formal to the informal job rather than another form type of employment, while part-time employment has the opposite effect. In contrast, for informal workers only workers at enterprises with between 5 and 9 employees are more likely to switch to a formal job rather than just take up another informal type of employment, although one may be cautious in these results given the small sample size.

Replication of our DID results of the informal sector wage premium for job movers are shown in the last three rows of the middle panel of Table 2. One should note in this regard that any estimated coefficient for this movers sub-sample, while arguably serving as a solution to the problem of differences in job mobility propensity across sectors, can only be interpreted as conditional on an individual being a job mover, and thus not necessarily representative of the

[27] It should be pointed out that this also drops all observations of those persons who remained in the same job but whose employer changed formality status. One would suspect, however, that many of these are likely to be due to measurement and reporting error.

entire sample. As can be seen, the coefficient on the informal sector wage penalty is similar to that for the entire sample, indicating that the inclusion of stayers does not to lead to a noticeable bias in the estimated penalty involved in moving to the informal sector. Examining single persons, however, one still finds evidence that any observed penalty may simply be due to the failure to take account of tax avoidance in the informal sector, as indicated by the continuing insignificance of the coefficient on the informal sector dummy when using our measure of 'net' wages.

C. Combined Propensity Score Matching and Difference-in-Differences Estimator

One feature of using OLS or the DID estimators is that because of their linearity assumption underlying the effect of controls one can use all observations for which there are non-missing values on the chosen controls. This may, however lead to the case where one is comparing very different treatment and control group individuals, i.e., where there is little of what is known as 'common support' among the two groups. An extreme example in this regard may be where age is an important determinant of selection into the informal market and wages, but only young people work in the informal and the elderly are employed in formal sector jobs.

Thus, one would ideally like to ensure that one has comparable groups where for each group of similar individuals there are some working in the formal while others are employed in the informal sector. One difficulty in ensuring that one does have such 'comparable' groups is that usually there is a whole set of characteristics that may determine selection into treatment and pay and hence one faces the problem of matching individuals on multiple dimensions. A possible solution to this problem is the utilization of a summary statistic to match 'similar' individuals. In this regard, Rosenbaum and Rubin (1983) suggest the use of a propensity score generated from modeling the probability of the treatment effect of interest, known as propensity score matching (PSM). Accordingly, in our context one first identifies the probability of working in the informal compared to working in the formal sector (also known as the 'propensity score') conditional on a set of observables that determine selection into the informal sector using a probit model and then uses the resultant propensity score to match 'similar' individuals.

An important aspect in estimating the propensity score to be used to 'match' individuals is the choice of covariates to be utilized as determinants of working in the informal sector. In this regard, one should use variables that could potentially influence participation in the informal market and the wage rate.[28] We thus include all our explanatory variables used in the OLS in levels equation as determinants of being employed in the informal sector.[29] With the propensity scores at hand, the next step is to choose the matching algorithm. In this regard we choose the caliper matching method, using a caliper of size 0.1 without replacement, where an individual from the control group is matched with an individual from the treatment group that lie within this chosen caliper and who is closest in terms of the propensity score.[30] This resulted in matching in total 388 observations. The choice of caliper can be pertinent in determining the size and quality of the matched sample in terms, but it may difficult to judge a priori what level of lack of similarity is tolerable; see Caliendo and Kopeining (2005). In assessing the match quality of our chosen caliper, we plotted kernel density estimates of the estimated propensity scores of the unmatched and matched samples in Figure 1. As can be seen from comparing these, the distribution of propensity scores is much more similar after matching.[31]

An important assumption behind the validity of the propensity score matching estimator (PSM) is that after conditioning on the chosen set of observable characteristics, mean outcomes are conditionally mean independent on treatment. However, Smith and Todd (2005), using the case study where one can compare actual random matching with that created by the estimator, showed that this is unlikely to hold. Rather they demonstrate that combining a PSM estimator

---

[28] One may be tempted to simply include as many variables as possible, even if there is no a priori reason to believe that they affect both selection and outcome. However, Bryson *et al.* (2002) note that this may just exacerbate the overlap problem and lead to inconsistent estimates by increasing the variance.
[29] Detailed results of this probit estimation are available from the authors upon request.
[30] The matching is performed in STATA Version 9 using the software provided by Sianesi (2001).
[31] We experimented with smaller calipers but this reduced the size of our already small matched sample considerably without providing any gains in terms of matching quality, which we assessed using the test suggested by Rosenbaum and Rubin (1985); details are available from the authors.

with DID estimator can produce much superior estimates in a non-experimental context and we thus similarly proceed along these lines. The results of this combined PSM-DID estimator with our data are shown in the final panel of Table 2.[32] One finds in this regard that the coefficient from the combined PSM and DID estimate of the informal sector wage penalty is only marginally higher than for the simple DID estimator for the entire sample and for job movers. While this difference is somewhat larger for single men and single male movers, once one takes account of the likely tax avoidance by informal sector workers, the penalty becomes insignificant as before. Given the similarity of the results with the DID estimator on its own, the substantially reduced sample size when using the matched sample, and the fact that, as Smith and Todd (2005) point out, results from a 'matched' sample cannot necessarily be interpreted as representative of the whole sample, we proceed to using just DID estimates of the penalty for the remainder of our analysis.

C. Other Robustness Checks

There are a number of issues concerning our use of the data that require further robustness checks. First of all, the failure to link some males across time and jobs may be due to non-random factors that differ across sectors, such as migration or flows into unemployment or out of the labor force. Again this could bias our results. To investigate this, we re-estimated our wage equation in OLS in levels, including job and human capital characteristics for all those observations that we were unable to link across waves. Reassuringly, however, the result of this exercise given in the first row of Table 3 shows a similar coefficient on the informal sector dummy to that found for the linked sample using OLS in levels.

Our definition of informal sector employment, based on the registration of the firm, is arguably a stringent one and may thus not be capturing all informal sector workers. For example, registration may not be a sufficiently discerning proxy of informality of employers in when there

[32] One should note that Pratap and Quintin (2006) only used the PSM on its own to estimate the wage penalty associated with working in the informal sector.

is increased outsourcing and subcontracting of employment.[33] We hence experimented with a broader definition based on employment conditions. In particular, we classified persons as working in the formal sector if they received paid leave, pensions, and/or Unemployment Insurance Fund Contributions, and informal otherwise. This meant reclassifying 13 per cent of our sample. We then proceeded to implement our DID estimator with this new definition of 'informality' (Definition B), the results of which are given in the second panel of Table 3. Accordingly, however, this leaves the results qualitatively similar and changes only little quantitatively.

As noted in Section II, over 40 per cent of informal sector employees are working in the Private Households sector, while this industry only constitutes 1 per cent of the formal sector. To check whether our estimates of the informal sector wage penalty are being driven by workers in Private Households, we re-calculated our DID estimates for non-Private Household industries only, the results of which are shown in the third panel of Table 3. Accordingly, the estimated coefficients are marginally lower for all samples. Once again, however, the findings suggest that taking account of tax payments in the formal sector makes any informal sector penalty vanish.

One may also be concerned that most of those that reported their salaries in categories rather than actual values were persons employed by formal employers, allowing for the possibility of substantially more measurement error in this sector. To investigate this we report estimates of the sub-sample of individuals who reported actual earnings in the final panel of Table 3. Again, the findings are qualitatively the same and quantitatively similar to our entire linked sample.

A final worry may be with regard to other benefits that workers are likely to receive in formal sector jobs, such as medical, unemployment insurance fund, paid leave, and pension contributions. More specifically, it may be that reported gross wage for those workers that receive such benefits may be stated net of such employer contributions, thus creating measurement error that is likely to be systematically correlated with informal

[33] Thanks are due to a referee for making this point.

sector status. One would expect this to create a downward (in absolute value) bias in the estimated coefficient on our informal sector dummy. To investigate this we re-estimated our benchmark OLS and DID regressions including dummies for whether the worker reported receiving medical, unemployment insurance fund, and pension contributions and paid leave from the employer, the results of which are shown in Table 5. As can be seen, the coefficient estimates using gross wages are lower than compared to not including the dummies (as reported in Table 2). However, for movers in general and those that are single the estimated coefficient is indeed higher (in absolute value). Thus the evidence with regard to a priori expectations of a possible downward bias is rather mixed. At any rate, and most importantly, once one takes account of taxes the disappearance of the informal sector wage penalty remains.

IV. Conclusion

In this paper we re-examined whether individuals working in the informal sector suffer from a wage penalty as is commonly believed. To this end, we used rich South African data on males that includes information that allow one to explicitly classify workers as being employed in the informal sector and link workers over time. Our analysis unearthed a number of potentially interesting findings. First of all, nearly 70 per cent of the wage penalty that is observed by comparing simple average gross earnings across the informal and formal sectors is due to observable differences in human capital and job characteristics. Once these characteristics are controlled for, however, gross logged wages are still 37 per cent lower for those employed in the informal sector. Controlling for unobservable time invariant factors further reduces the informal sector wage penalty to just over 12 per cent. Importantly, however, when we focus our analysis on single men for which we can easily calculate net, after tax, returns to employment, and assume that informal sector employees can avoid tax payments on their employment earnings, any pay discrepancy disappears. Thus, there seems to be little support that, once other factors are controlled for, the formality of employers is a feature of pay differences in the labor market in South Africa.[34]

One should note, nevertheless, that simply deducting taxes ignores the long-run advantages from being part of the social welfare system in terms of receiving benefits, such as pension payments or unemployment compensation some time in the future. Moreover, we have not been able to examine differences in other workplace related characteristics, such as stability, safety, and fringe benefits that may differ between formal and informal jobs. Clearly, not taking account of these factors would tend to underestimate the 'true' value of working in the formal sector.

[34] Having said this, one should note that some of these very control variables may be highly correlated with informal sector status. For example, Badaoui *et al.* (2006) imbed informal firms in an equilibrium search mode and show theoretically and empirically that firm size is the crucial variable in that large firms pay more but also will choose to be in the formal sector.

References

Abedian, I. and DeSmidt, M. 1990. "The Informal Economy in South Africa." *The South African Journal of Economics* 58 (December): 258-268.

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

Badaoui, Eliane, Eric Strobl and Frank Walsh. 2006. "An Equilibrium Search Model of the Informal Sector." Mimeo.

Banerjee, Nirmala. 1985. *Women Workers in the Unorganized Sector*. Sangam Books: Hyderabad.

Barrientos, Armando, and Stephanie Ware Barrientos. 2002. "Extending Social Protection to Informal Workers in the Horticulture Global Value Chain." Social Protection Discussion Paper no. 0216, The World Bank, Washington, DC, June.

Bekkers, Hans, and Wim Stoffers. 1995. "Measuring Informal Sector Employment in Pakistan: Testing a New Methodology." *International Labour Review* (Geneva) 134 (1): 17-36.

Bryson, Alex., Richard Dorsett, and Susan Purdon. 2002. "The Use of Propensity Score Matching in the Evaluation of Labour Market Policies." Working Paper no. 4, Department for Work and Pensions.

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

Caliendo, Marco, and Sabine Kopeinig. 2005. "Some Practical Guidance for the Implementation of Propensity Score Matching." IZA Discussion Paper no. 1588.

Charmes, Jacques. 2000. "The Contribution of Informal Sector to GDP in Developing Countries: Assessment, Estimates, Methods, Orientations for the Future." OECD-EUROSTAT-State Statistical Committee of the Russian Federation, Non-Observed Economy Workshop, Sochi (Russia), October 16-20.

Dehejia, Rajeev, and Sadek Wahba. 2002. "Propensity Score Matching Methods for Non-experimental Causal Studies." *Review of Economics and Statistics* 84 (February): 151-161.

DeSmidt, M. 1988. "The Informal Sector, in Theory and Practice: An Examination of the Informal Activities in Block C, Site C, Khayelitsha, Cape Town." Economic Honours Long Paper, School of Economics, University of Cape Town, Cape Town.

De Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World*. New York: Harper & Row.

Devey, R.ichard, Caroline Skinner, and Imraan Valodia.. 2002. "The Informal Economy in South Africa: Who, Where, What and How Much?" Conference on Labour Markets in South Africa, Johannesburg, 22-24 October 2002.

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

Farber, Henry S. 2005. "What Do We Know About Job Loss in the United States? Evidence from the Displaced Workers Survey, 1984-2004." Working Paper no. 498.

Fields, Garry S. 1975. "Rural-Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDC's." *Journal of Development Economics* 2 (June): 165-87.

Friedman, Eric, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton. 2000. "Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 countries." *Journal of Public Economics* 76 (June): 459-493.

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 (September): 513-529.

Guenther, Isabel, and Andrey Launov. 2006. "Competitive and Segmented Informal Labor Markets." IZA Discussion Paper no. 2349.

Hashimoto, Masanori. 1981. "Firm Specific Human Capital as a Shared Investment", *American Economic Review* 71 (June): 475-482.

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.

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover," *Journal of Political Economy* 87(October): 972-990.

Kingdon, Geeta G., and John Knight. 2001. "Why High Open Unemployment and Small Informal Sector in South Africa?" Working Paper Series 2000/2, Centre for the Study of African Economies, University of Oxford.

Kingdon, Geeta G., and John Knight. 2003. "Well-Being Poverty versus Income Poverty and Capabilities Poverty." Working Paper Series, Centre for the Study of African Economies, University of Oxford, July, mimeo.

Krueger, Alan, and Lawrence Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56 (March): 259-293.

Madrian, Brigitte C., and Lars John Lefgren. 1999. "A Note on Longitudinally Matching Current Population Survey (CPS) Respondents." NBER Technical Working Paper no. 247, November 1999.

Marcouiller, Douglas, Veronica Ruiz de Castilla, and Christopher Woodruff. 1997. "Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru." *Economic Development and Cultural Change,* 45 (January): 367-392.

Marshall, Adriana. 1987. "Non-Standard Employment Practices in Latin America." Discussion Paper no. 6, Labour Market Programme. Geneva: International Labour Office.

Masatlioglu, Yusufchan, and Jamele Rigolini. 2005. "Labor Dynamics and the Informal Economy." Conference Paper, United Nations University, 13 May 2005.

Maloney, William. 1998. "Are LDC Labor Markets Dualistic?" The World Bank, February 1998.

Maloney, William. 2004. "Informality Revisited." *World Development* 32 (July): 1159–1178.

Mazumdar, Dipak. 1975. "The Theory of Share-Cropping with Labour Market Dualism." *Economica* 42 (August): 261-271.

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

Oi, Walter, and Todd Idson. 1999. "Firm Size and Wages," in O. Ashenfelter and D. Card, Eds., *Handbook of Labor Economics*, 3rd Ed. Amsterdam: North-Holland.

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

Pradhan, Basant K., P. K. Roy, and M. R. Saluja. 1999. "Informal Sector in India: A Study of Household Saving Behaviour." Contribution of the Informal Sector to the Economy, Report no. 1, National Council of Applied Economic Research, New Delhi August 1999.

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

Roberts, Bryan. 1989. "The Other Working Class: Uncommitted Labor in Britain, Spain, and Mexico." pp. 352-372 in *Cross-National Research in Sociology*, edited by M. L. Kohn. Newbury Park, CA: Sage Publications.

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.

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.

Sianesi, Barbara. 2001. "Implementing Propensity Score Matching Estimators with STATA." program available at: www. stata.com

Smith, Jeffrey, and Petra Todd. 2005. "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators." *Journal of Econometrics* 125 (March-April): 305-353.

Tannuri-Pianto, Maria, and Donald Pianto. 2002. "Informal Employment in Brazil - A Choice at the Top and Segmentation at the Bottom: A Quantile Regression Approach." Paper presented at the XXIV Brazilian Econometrics Meeting, Rio de Janeiro, Brazil.

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.

Tansel, Aysit. 2000. "Wage Earners, Self Employed and Gender in the Informal Sector in Turkey." Policy Report Research on Gender and Development, Working Paper Series no. 24, November 2000.

Thomas, W.H. 1989. "South Africa's Growing Black Market: Challenges for the 1990s." Speech at the Spotlight on the Black Market Conference sponsored by the Small Business Development Centre in Johannesburg, South Africa, June 6, 1989.

Tokman, Victor. 1982. "Unequal Development and the Absorption of Labour: Latin America 1950-1980." *CEPAL Review* 17:121-133.

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

Table 1: Summary Statistics

	Formal Sector		Informal Sector	
	Mean	**SD**	**Mean**	**SD**
WAGE				
Gross Log of Hourly Wage	1.85	1.07	0.87	0.81
Net Log of Hourly Wage (single males only)	1.60	0.94	0.81	0.80
HUMAN CAPITAL CHARACTERISTICS				
Age	37.92	10.61	38.18	12.29
Asian/Indian	0.05	0.11	0.01	0.08
Black	0.60	0.49	0.87	0.33
White	0.13	0.34	0.00	0.03
Colored	0.22	0.42	0.12	0.32
Married	0.71	0.45	0.56	0.50
Afrikaans	0.29	0.45	0.12	0.33
English	0.12	0.33	0.02	0.12
Read	0.91	0.28	0.82	0.38
Write	0.91	0.28	0.82	0.39
Job training	0.16	0.37	0.07	0.25
Education level:				
No primary (can not read and write)	0.08	0.28	0.18	0.38
No primary (can read and write)	0.11	0.32	0.19	0.39
Primary	0.48	0.50	0.54	0.50
Secondary	0.27	0.44	0.08	0.27
NTC	0.01	0.11	0.00	0.00
University	0.03	0.18	0.00	0.00
Occupation:				
Legislators, senior officials and managers	0.05	0.23	0.00	0.03
Professionals	0.02	0.15	0.00	0.02
Technicians and associated professionals	0.05	0.24	0.00	0.06
Clerks	0.06	0.26	0.00	0.05
Service workers, shop and sales workers	0.10	0.28	0.03	0.15
Skilled agricultural and fishery workers	0.02	0.12	0.23	0.26
Craft and related trade workers	0.20	0.40	0.21	0.23
Plant and machine operators and assemblers	0.24	0.43	0.10	0.18
Elementary occupations (except Domestic Workers)	0.27	0.43	0.34	0.30
Domestic workers	0.01	0.04	0.07	0.45
JOB CHARACTERISTICS				
Urban area	0.63	0.48	0.52	0.50
Tenure	8.16	8.21	4.37	6.42
Supervision	0.99	0.09	0.95	0.23
Part-time job	0.91	0.29	0.75	0.43
Firm Size:				
Worker 1	0.02	0.14	0.44	0.50
Workers 2 – 4	0.09	0.28	0.38	0.49
Workers 5 – 9	0.13	0.33	0.11	0.31
Workers 10 – 19	0.16	0.37	0.04	0.20
Workers 20 – 49	0.19	0.40	0.02	0.16
Workers ≥50	0.41	0.49	0.01	0.10
Industry:				
Agriculture, Hunting, forestry and fishing	0.22	0.41	0.14	0.19
Mining and quarrying	0.16	0.34	0.00	0.03
Manufacturing	0.21	0.44	0.04	0.12
Electricity, Gas and water supply	0.01	0.07	0.00	0.00
Construction	0.06	0.22	0.19	0.20
Wholesale and retail trade	0.16	0.37	0.09	0.19
Transport, storage and communication	0.05	0.21	0.09	0.07
Financial intermediation, insurance, real estate and business services	0.09	0.28	0.02	0.11
Community, social and personal services	0.03	0.17	0.02	0.10
Private household	0.01	0.09	0.41	0.38

Table 2: Estimates of the Informal Sector Wage Penalty

Method	Sample	Wage	Controls	IS Def.	Coeff.	Std. Err.	R-Sq.	Obs.
OLS	*All*	*Gross*	*None*	*A*	-1.131***	0.037	0.07	11607
OLS	*All*	*Gross*	*HC*	*A*	-0.536***	0.030	0.57	11607
OLS	*All*	*Gross*	*HC, JC*	*A*	-0.372***	0.031	0.68	11607
OLS	*Single*	*Gross*	*HC, JC*	*A*	-0.363***	0.048	0.61	3483
OLS	*Single*	*Net*	*HC, JC*	*A*	-0.188***	0.047	0.59	3481
DID	*All*	*Gross*	*JC*	*A*	-0.123***	0.033	0.05	7890
DID	*Single*	*Gross*	*JC*	*A*	-0.206***	0.054	0.07	2310
DID	*Single*	*Net*	*JC*	*A*	-0.025	0.053	0.06	2306
DID	*Movers*	*Gross*	*JC*	*A*	-0.131***	0.045	0.06	3315
DID	*Movers Single*	*Gross*	*JC*	*A*	-0.171**	0.071	0.10	1326
DID	*Movers Single*	*Net*	*JC*	*A*	0.012	0.070	0.08	1324
DID - PSM	*All*	*Gross*	*JC*	*A*	-0.144**	0.070	0.09	388
DID - PSM	*Single*	*Gross*	*JC*	*A*	-0.279**	0.110	0.24	154
DID - PSM	*Single*	*Net*	*JC*	*A*	-0.123	0.109	0.19	154
DID - PSM	*Movers*	*Gross*	*JC*	*A*	-0.141*	0.084	0.11	304
DID - PSM	*Movers Single*	*Gross*	*JC*	*A*	-0.237*	0.126	0.29	132
DID - PSM	*Movers Single*	*Net*	*JC*	*A*	-0.083	0.126	0.25	132

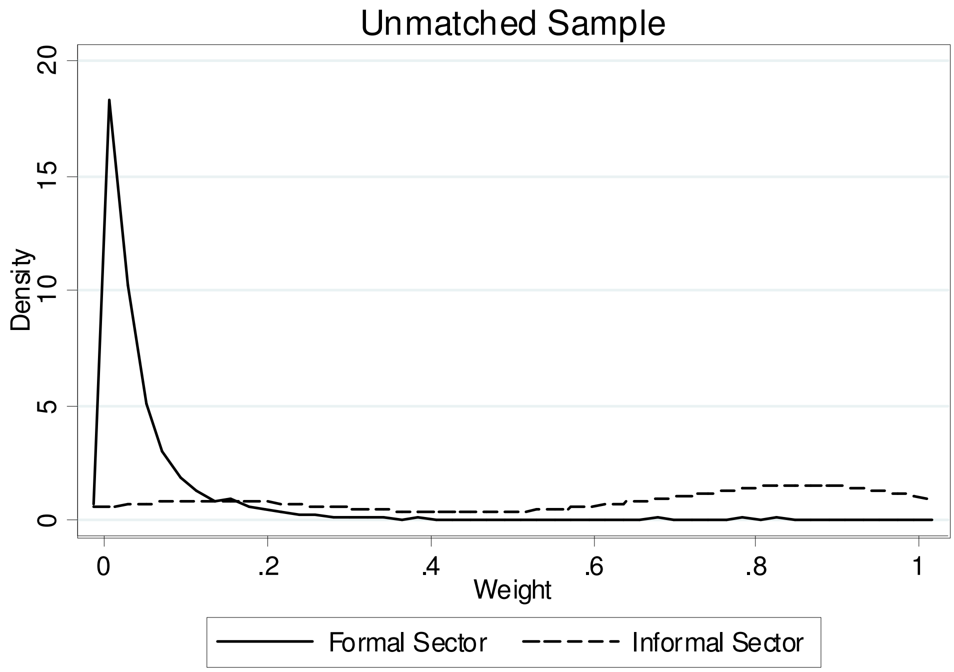
Notes: (a) ***, **, and * are one, five, and ten per cent significance levels, respectively (b) IS: Informal Sector; HC: Human Capital Controls; JC: Job Characteristics Controls; OLS: Ordinary Least Squares; DID: Difference-in-Differences; A: Standard Informal Sector definition (see text). (c) HC controls drop out in the DID estimations.

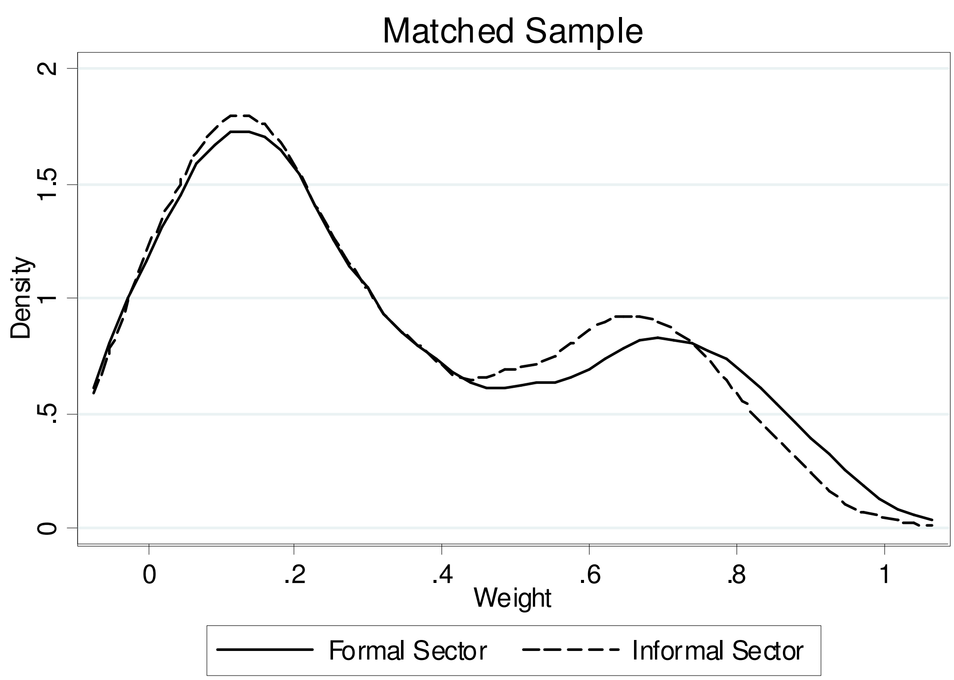
Table 3: Estimates of the Informal Sector Wage Penalty – Robustness Checks

Method	Sample	Wage	Controls	IS Def.	Coeff.	Sd.Er.	R-Sq.	Obs.
OLS	*Excluded*	*Gross*	*HC, JC*	*A*	-0.354***	0.020	0.63	18396
DID	*All*	*Gross*	*JC*	*B*	-0.143***	0.017	0.06	8236
DID	*Single*	*Gross*	*JC*	*B*	-0.185***	0.030	0.08	2430
DID	*Single*	*Net*	*JC*	*B*	-0.022	0.029	0.07	2428
DID	*Movers*	*Gross*	*JC*	*B*	-0.162***	0.025	0.07	3513
DID	*Movers Single*	*Gross*	*JC*	*B*	-0.191***	0.037	0.10	1415
DID	*Movers Single*	*Net*	*JC*	*B*	-0.033	0.037	0.08	1414
DID	*NPH*	*Gross*	*JC*	*A*	-0.132***	0.042	0.06	4225
DID	*NPH - Single*	*Gross*	*JC*	*A*	-0.207***	0.068	0.09	1425
DID	*NPH - Single*	*Net*	*JC*	*A*	-0.047	0.065	0.08	1421
DID	*NPH - Movers*	*Gross*	*JC*	*A*	-0.134**	0.051	0.06	2025
DID	*NPH - Movers Single*	*Gross*	*JC*	*A*	-0.157*	0.088	0.09	831
DID	*NPH - Movers Single*	*Net*	*JC*	*A*	0.001	0.086	0.08	829
DID	*NIB*	*Gross*	*JC*	*A*	-0.118***	0.034	0.06	6214
DID	*NIB - Single*	*Gross*	*JC*	*A*	-0.214***	0.057	0.09	1861
DID	*NIB - Single*	*Net*	*JC*	*A*	-0.033	0.056	0.08	1857
DID	*NIB - Movers*	*Gross*	*JC*	*A*	-0.123***	0.048	0.07	2669
DID	*NIB - Movers Single*	*Gross*	*JC*	*A*	-0.195**	0.078	0.10	1076
DID	*NIB - Movers Single*	*Net*	*JC*	*A*	-0.011	0.078	0.09	1074

Notes: (a) ***, **, and * are one, five, and ten per cent significance levels, respectively; (b) IS: Informal Sector; JC: Job Characteristics Controls; DID: Difference-in-Differences; PSM: Propensity Score Matching; NPH: Non-Private Households; NIB: Non Income Bracket; A: Standard Informal Sector definition (see text); B: Alternative Informal Sector Definition (see text); (c) HC controls drop out in the DID estimations.

Figure 1: Kernel Density Estimates of the Estimated Propensity Scores





Appendix - Table A1: List of the variables of interest

Variable name	Definition of the variable
Hourly Gross wage	Real hourly logged gross wage calculated using a person's income, hours worked in their main job and the South African consumer price deflator.
Hourly Net wage	Real hourly logged net wage calculated using a person's income, hours worked in their main job, the South African consumer price deflator, and the income tax bracket (if in the formal sector).
Black White Colored	Three dummies related to a person's race (the population group that the worker belongs to), where Asians/Indians are the excluded base category.
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	Whether the worker had ever been 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	Dummy for Part-time Work.
Firm Dummies	Six Dummies for Firm Size.
Union Dummy	Dummy for whether employee belongs to a union.
Industry	Eleven dummies for the industry variables (the eleventh industry dummy 'Exterior organizations and foreign government' is omitted).