



<b>Title</b>	Is there an informal employment wage penalty? Evidence from South Africa
<b>Authors(s)</b>	Walsh, Frank, Badaoui, Eliane, Strobl, Eric
<b>Publication date</b>	2008-04
<b>Publication information</b>	Walsh, Frank, Eliane Badaoui, and Eric Strobl. "Is There an Informal Employment Wage Penalty? Evidence from South Africa." University of Chicago Press, April 2008. <a href="https://doi.org/10.1086/533550">https://doi.org/10.1086/533550</a> .
<b>Publisher</b>	University of Chicago Press
<b>Item record/more information</b>	<a href="http://hdl.handle.net/10197/182">http://hdl.handle.net/10197/182</a>
<b>Publisher's statement</b>	All rights reserved
<b>Publisher's version (DOI)</b>	10.1086/533550

Downloaded 2026-05-02 00:25:58

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd\_oa)



© Some rights reserved. For more information

# Is There an Informal Employment Wage Penalty? Evidence from South Africa

**ELIANE EL BADAOU**

THEMA—Université de Cergy-Pontoise

**ERIC STROBL**

Ecole Polytechnique Paris

**FRANK WALSH**

University College Dublin

## **I. Introduction**

The informal sector has traditionally been viewed as a temporary alternative to unemployment and poverty (Fields 1975) that 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 (Marshall 1987; De Soto 1989), 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.<sup>1</sup> If informal-sector employment is indeed a more permanent and 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.<sup>2</sup> 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.

<sup>1</sup> According to Maloney (2004), the informal sector comprises 30%–70% of the labor market of most Latin American countries. Friedman et al. (2000) estimate that 14%–62% of output comes from the informal sector across a range of transition economies.

<sup>2</sup> Poverty is a key characteristic of the informal sector (Pradhan et al. 1999).

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 earlier empirical studies.<sup>3</sup> A number of explanations have been offered in this regard, most of which are 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; Mazumdar 1975). However, several more recent studies postulate that it is more efficient for 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, due to both 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, e.g., 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 as much of 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 premiums 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 considered part of the formal sector. The results indicate substantial earnings differences between the formal and informal sectors for men but not for women. Also, for Mexico, Gong and van Soest (2002) find that wage differentials between the formal and informal sectors are typically small for the lesser 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%

<sup>3</sup> See, e.g., Mazumdar (1982), Heckman and Hotz (1986), Roberts (1989), Pradhan and van Soest (1995), Tansel (1999), Gong and van Soest (2002).

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. However, Pratap and Quintin (2006) find that, after controlling for selection, no wage premium remains and that job satisfaction is not lower in the informal sector for Argentinean data.

In this article we reexamine the possible existence of a wage penalty for working in the informal sector, using the case study of South African non-self-employed males. In this regard, our contribution to the current literature is derived from both data quality and a methodological perspective. 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. While the International Labour Organisation has adopted a number of recommendations on this issue, data restrictions have not generally allowed a more uniform use of the term in empirical analyses. “Informality” has been, for example, proxied by assuming that only certain characteristics such as high wages, job protection, or access to a social security system belong to formal-sector activities (see, e.g., Banerjee 1985); that those not in regular employment are informal workers (Kingdon and Knight 2001; Barrientos and Barrientos 2002); that those who have activities that are not reported to the government and to the tax authorities are involved in the informal sector (Masatlioglu and Rigolini 2005); that all self-employment in developing countries is informal (Yamada 1996); that wage earners who are not registered in social security institutions are informal salaried workers (Calderon-Madrid 2000); or that small firms operate outside the “formal” umbrella of economic activity (Marcouiller et al. 1997; Maloney 1999; Gong and van Soest 2002). The source of our data, the South African Labour Force Survey, in contrast, explicitly asks individuals if they are working in the informal sector and allows indirect verification of this.

From a methodological standpoint, one of the major challenges in estimating the existence of a penalty has been to deal with the sample selection bias that is likely to be inherent in the empirical analysis. More precisely, the formality of jobs is unlikely to be randomly assigned across wage earners. In this regard one should note that the data we use are relatively rich in information pertaining to levels of human capital and job characteristics that are likely to determine the level of earnings.<sup>4</sup> In addition, we take advantage of the rotating-panel nature of the survey and control for time-invariant unobservables that

<sup>4</sup> 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).

may be biasing results under standard estimation techniques. Finally, we explicitly take into account that, in contrast to their formal-sector counterparts, informal employees are unlikely to be paying taxes, and we adjust reported earnings accordingly to give a more accurate measure of the wage penalty.

Of course, our empirical analysis must be considered within the context of the labor market in question in order to potentially generalize any of our results. In this regard, one should note that South Africa can be considered somewhat peculiar in that it has a relatively small informal sector (around 11%) but high unemployment (ranging between 26.6% and 38.8%, depending on the definition of unemployment used).<sup>5</sup> As suggested by Kingdon and Knight (2001), one possible reason for this could be justified under Harris-Todaro type models in which, for some workers, the probability of securing formal employment is higher if conducted from open unemployment because of search costs and/or the disdain associated with working in informal-sector jobs. However, the authors find that the unemployed are less happy than their sector counterparts and interpret this as evidence that this is unlikely to serve as a plausible explanation. Another possibility may be that there are barriers to entry not only to the formal sector but also to the informal sector, and Kingdon and Knight (2001) make an argument in favor of this explanation. It is important to note, however, that if entry barriers are indeed the factor behind the small informal sector and this significantly affects the nature and composition of employment in the informal sector in South Africa compared to other countries, then the informal-sector wage penalty should be smaller than it would be if no barriers existed.<sup>6</sup> But, as we shall show, comparisons of raw figures as well as estimates under ordinary least-squares (OLS) standard techniques suggest that the informal wage penalty in South Africa is not unusually low compared to other studies for other countries. It is only after we control for time-invariant unobservables and calculate earnings net of tax payments that the wage penalty “disappears.”

The remainder of the article is organized as follows. In the following section

<sup>5</sup> More recently, the unemployment rate according to the official definition declined from 26.7% in September 2005 to 25.5% in September 2006. This decline in unemployment was accompanied by a decline in the number of discouraged work seekers, suggesting an expansion of employment. There have been, however, some notable changes in the importance of the informal sector over time. For example, over the period 1997–2001, the number of formal wage employees marginally increased by about 6.8%, whereas informal wage employment grew by about 48.5% (Devey et al. 2003). More recently, there appears to be a decline in the informal sector, with the number of informal workers decreasing by about 29% over the period from March 2001 to March 2006 (and the percentage of informal employment in total employment declining from 23.1% to 17.6%).

<sup>6</sup> One could argue that it may be more appropriate to use the term “informal employment wage penalty.”

we describe our data set and provide some summary statistics. In Section III 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.

## II. Data Description and Summary Statistics

### A. Data Description

Our data source is the South African Labour Force Survey (LFS).<sup>7</sup> The LFS is a semiannual rotating-panel household survey that has been 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.<sup>8</sup>

In terms of classifying informal-sector activity, the LFS explicitly asks individuals who 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 in the formal sector or in the informal sector (including domestic work).”<sup>9</sup> In addition, 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, or is registered for value-added tax. If individuals answer in the affirmative to any of these questions, we change their sector status to being of the formal sector, even if they classify themselves as working in the informal sector.<sup>10</sup> 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

<sup>7</sup> The South African Labour Force Survey is conducted by Statistics South Africa, Pretoria (<http://www.statssa.gov.za>).

<sup>8</sup> 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. Prior to September 2001, linking households is not possible.

<sup>9</sup> 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.”

<sup>10</sup> However, in the end there were only 2.1% of observations for which we needed to change the status. The correlation between the two classifications was about 0.96.

income, and hours worked in the previous week in his or her 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.<sup>11</sup> For these we used the midpoint between category thresholds, except for the first and last categories, for which we simply used the threshold itself as the salary value.<sup>12</sup> The derived nominal hourly wage rate data were then converted into real wages (September 2001 values) by using the South African consumer price deflator. We also checked the observations for those individuals who 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% indicated that they worked in the informal sector (as defined above).<sup>13</sup> 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 as employer contributions to pension, health insurance, or job stability; we only have information on the reported gross income.<sup>14</sup> 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% of 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), using information on individuals' sex, race, and age. Each wave allowed us to link, on average, 72.1% of observations of individuals who claimed to have been in the same household during the previous survey.<sup>15</sup> We restrict our analysis mainly to this linked data.

<sup>11</sup> From those with income given in the categories, about 6% reported being in the informal sector. Also, 3% and 1.3% of those in categories were in the lower and upper income brackets, where the corresponding proportion of informal-sector workers in these were 37% and 5%, respectively.

<sup>12</sup> An interesting alternative to this would have been to use the multiple imputation techniques suggested by Ardington et al. (2006).

<sup>13</sup> 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.

<sup>14</sup> It is usually assumed that formal-sector workers are more likely to be provided with nonwage benefits such as social insurance systems, a higher level of job security, paid sick leave, and pensions. However, the informal sector may have desirable nonwage features (Maloney 2004). Several studies have pointed out that nonwage remunerations may also be associated with informal work (Marcouiller et al. 1997; Günther and Launov 2006).

<sup>15</sup> Using this methodology, Byrne and Strobl (2004) were able to link 80% of individuals for the

Apart from an explicit definition of the formality of an 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 her or she works in the informal sector. Those used in the current analysis are grouped for the sake of convenience into those related to human capital (such as age, gender, race, marital status, education level, occupation, and whether they ever received any job training) and job characteristics (such as firm size, industry, supervision, urban area, part-time status, and tools). We provide a comprehensive list of these and their definitions in table A1 in the appendix.

Finally, we reduced our sample to non-self-employed males ages 15–65 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 the 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 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 a greater measurement error and incorporate returns to risk, and so forth, that would not be included in wages of employees. Analyzing this group would thus require a separate analysis, which is beyond the scope of this article.<sup>16</sup>

### **B. Summary Statistics**

All in all, after dropping observations with missing values for any of our variables, we were left with a total number of 11,571 observations on 6,064 linked males. Of these observations, 11.03% were for individuals working in the informal sector as defined above.<sup>17</sup> It is also noteworthy that, overall, the share of informal sector in total employment fell over our sample period,

---

developing country of Trinidad and Tobago. On average, 4.9% state to have not been in the same household during the previous survey.

<sup>16</sup> In addition, as will be shown, an important part of our econometric analysis focuses on job movers, which we identify by their length of tenure, information that is not available for the self-employed. Dropping the self-employed resulted in dropping 1,230 individuals, of which 34% stated that their business was not formally registered.

<sup>17</sup> Given that our sample is not completely representative of all informal-sector employees (e.g., self-employed), one could argue that "informal employment" rather than "informal sector" may be a more accurate term to use. Nevertheless, we continue to use the term "informal sector" in order to be consistent with the general terminology in the literature.

**TABLE 1**  
**SUMMARY STATISTICS**

	Formal Sector		Informal Sector	
	Mean	SD	Mean	SD
Gross log of hourly wage	1.94	1.04	.87	.81
Net log of hourly wage	1.75	1.01	.87	.81
Human capital characteristics:				
Age	37.26	10.70	37.36	12.31
Asian/Indian	.04	.19	.01	.09
Black	.66	.47	.89	.31
White	.12	.32	.00	.06
Colored	.18	.38	.10	.30
Married	.65	.48	.47	.50
Afrikaans	.24	.43	.10	.30
English	.10	.31	.02	.14
Read	.90	.30	.81	.39
Write	.90	.30	.81	.39
Job training	.15	.36	.07	.25
Education level:				
No primary (cannot read and write)	.10	.30	.19	.39
No primary (can read and write)	.12	.33	.20	.39
Primary	.48	.50	.53	.50
Secondary	.25	.44	.08	.27
National Technical Certificate	.01	.10	.00	.03
University	.03	.17	.00	.01
Occupation:				
Legislators, senior officials, and managers	.04	.21	.00	.04
Professionals	.02	.14	.00	.03
Technicians and associated professionals	.05	.22	.00	.06
Clerks	.06	.23	.00	.07
Service workers, shop and sales workers	.10	.30	.03	.17
Skilled agricultural and fishery workers	.02	.13	.23	.42
Craft and related trade workers	.20	.40	.22	.41
Plant and machine operators and assemblers	.24	.43	.10	.30
Elementary occupations (except domestic workers)	.27	.44	.33	.47
Domestic workers	.01	.04	.07	.26
Job characteristics:				
Urban area	.60	.49	.52	.50
Tenure	7.49	8.05	4.37	5.81
Tools	.99	.08	.93	.25
Supervision	.91	.28	.72	.45
Part-time job	.05	.21	.27	.44
Firm size:				
Worker 1	.02	.15	.43	.49
Workers 2–4	.09	.29	.36	.48
Workers 5–9	.12	.33	.11	.32
Workers 10–19	.16	.37	.06	.24
Workers 20–49	.19	.39	.03	.16
Workers ≥50	.41	.49	.01	.10
Industry:				
Agriculture; hunting, forestry, and fishing	.22	.41	.14	.35
Mining and quarrying	.16	.37	.00	.04
Manufacturing	.21	.41	.04	.19
Electricity, gas, and water supply	.01	.07	.00	.01

TABLE 1 (Continued)

	Formal Sector		Informal Sector	
	Mean	SD	Mean	SD
Construction	.06	.24	.19	.39
Wholesale and retail trade	.16	.37	.09	.28
Transport, storage, and communication	.05	.21	.09	.29
Financial intermediation, insurance, real estate and business services	.09	.29	.02	.13
Community, social, and personal services	.03	.17	.02	.13
Private household	.01	.10	.41	.49

starting at 11.45% in September 2001 and ending at 9.45% in September 2003.

Table 1 presents some basic summary statistics concerning the dispersion of characteristics of our sample, broken down into those working in the formal as well as informal sectors. Accordingly, one finds that gross logged wages in the formal sector are on average 123% larger (or 2.23 times larger) than those found in the informal sector. One should note that the large earnings found here are in line with several other studies of other countries. For example, in Brazil, Tannuri-Pianto and Pianto (2002) find that the average hourly earnings for formal workers is almost as high as that for informal workers. In El Salvador, the log wages in the formal sector are 81% larger in the formal sector for men, in Peru the wage gap between the formal and the informal sectors is 35%, and in Mexico the wage differential for men is about 2.6% (Marcouiller et al. 1997). In their study, Pratap and Quintin (2006) find that the average real hourly earnings are significantly higher in the formal sector than in the informal sector (37.5%) for different definitions of informal employment. Using data from Côte d'Ivoire, Günther and Launov (2006) report that the means of monthly formal earnings are 154% larger than informal-sector earnings.

For the sake of convenience, we divide our other variables 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 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, compared to white employees. One possible reason may be the racial and historical differences in South African society and the apartheid racial discrimination that touched every aspect of social life,<sup>18</sup> sanctioned

<sup>18</sup> See Klasen (2002) for a discussion on this.

“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 to formal-sector employment.

One also discovers that workers in the formal sector are more likely to be married than their informal-sector counterparts. In addition, 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 languages of workers in the formal sector than of 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 (33%),<sup>19</sup> followed by skilled agricultural and fishery (23%) and craft and related trade (22%). Those in the formal sector are most likely to be found in the elementary occupations (27%), plant and machinery operators and assemblers (24%), and craft and related trade (20%).

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 in 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 are 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 av-

<sup>19</sup> 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 cleaners; messengers, porters, doorkeepers, and related workers; and garbage collectors and related laborers.

erage, much less tenure in the informal sector, hence suggesting lower stability of employment. Finally, one finds that 41% 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% are in skilled agricultural and fishery, 31% in elementary, and 17% in domestic occupations. Thus, it appears that most of the skilled agricultural and fishery male workers who 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 the informal counterpart, with agriculture; hunting, forestry, and fishing; mining and quarrying; manufacturing; and wholesale and retail trade constituting the most frequent employers.

### III. Econometric Analysis

The problem of measuring any potential informal-sector wage penalty boils down to trying to answer the following counterfactual question: what wage would a person employed in the informal sector have if he or she were instead employed in a similar job in the formal sector? In this regard, our problem is similar to that commonly undertaken in laboratory experiments, in which the effect of a “treatment” is assessed by comparing treatment and control groups, that is, when 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 is the case for most economic questions of this nature, the data at hand are non-experimental, and it is well recognized that the estimate of a causal effect obtained by comparing treatment groups with nonexperimental 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 (see Dehejia and Wahba 2002). 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 instance, it has been argued that the unusually high unemployment rate in South Africa could be due to the fact that for some potential workers, the wage in the informal sector is too low, given their possible high-reservation wage (see Kingdon and Knight 2003). While we have no information on reservation

wages, we do roughly control for changes in the labor market with time and regional dummies in our analysis.

One should note that dealing with possible sample selection bias is the main challenge of any evaluation study with nonexperimental 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 control for sample selection bias and have run OLS on a Mincerian wage equation with, among other controls, an indicator of the formality of the job (see, e.g., Marcouiller et al. 1997). Others have implemented more sophisticated two-stage models, for which participation is jointly specified with the wage regressions (see, e.g., Gong and van Soest 2002; 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, Günther and Launov (2006) use a mixture model that 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:

$$w_{it} = \alpha + \beta_1 I_{it} + \beta_2 X_{it} + \beta_3 Z_{it} + \varepsilon_{it}, \quad (1)$$

where  $\varepsilon_{it} = \mu_i + \eta_{it}$ ,  $i = 1, \dots, N$  represents individual units and  $t = 1, \dots, T$  time periods. The variable  $w_{it}$  denotes the logarithm of real hourly wages;  $X_{it}$  and  $Z_{it}$  are the sets of human capital and job characteristics, respectively. The dummy variable  $I_{it}$  takes a value of one if the firm of worker  $i$  is in informal sector and zero otherwise;  $\mu_i$  is an unobservable individual fixed effect, and  $\varepsilon_{it}$  is the error term.

#### A. Ordinary Least Squares

We start by estimating equation (1) with OLS in levels. Initially, we include only the informal dummy for the whole sample of linked individuals, without any explanatory variables except seasonal and year dummies. Our results of this exercise, given in the first row of the first section of table 2, show that the wage penalty for workers employed in the informal sector is found to be around 112.8%. However, as noted earlier, wage differences can be attributed to many different levels of human capital and various characteristics of the jobs that may also be correlated with selection to 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%. Moreover, introducing all our available job characteristics further decreases the coefficient on the informal-

TABLE 2  
ESTIMATES OF THE INFORMAL-SECTOR WAGE PENALTY

Method	Sample	Wage	Controls	IS Def.	Coeff.	SE	R <sup>2</sup>	Obs.
OLS	All	Gross	None	A	-1.128***	.037	.07	11,571
OLS	All	Gross	HC	A	-.536***	.030	.57	11,571
OLS	All	Gross	HC, JC	A	-.370***	.031	.68	11,571
OLS	Single	Gross	HC, JC	A	-.363***	.048	.61	3,475
OLS	Single	Net	HC, JC	A	-.188***	.047	.59	3,473
DID	All	Gross	JC	A	-.179***	.031	.03	8,253
DID	Single	Gross	JC	A	-.238***	.051	.04	2,405
DID	Single	Net	JC	A	-.066	.050	.03	2,401
DID	Movers	Gross	JC	A	-.218***	.085	.08	578
DID	Movers (single)	Gross	JC	A	-.189**	.113	.12	262
DID	Movers (single)	Net	JC	A	-.026	.112	.10	262
DID-PSM	All	Gross	JC	A	-.151*	.066	.07	412
DID-PSM	Single	Gross	JC	A	-.251*	.099	.18	160
DID-PSM	Single	Net	JC	A	-.092	.101	.14	160
DID-PSM	Movers	Gross	JC	A	-.262*	.134	.26	90
DID-PSM	Movers (single)	Gross	JC	A	-.374*	.184	.43	39
DID-PSM	Movers (single)	Net	JC	A	-.214	.184	.40	39

**Note.** IS = informal sector; HC = human capital controls; JC = job characteristics controls; OLS = ordinary least squares; DID = difference-in-differences; PSM = propensity score matching; A = standard informal-sector definition (see text). HC controls drop out in the DID estimations.

\* 10% significance level.

\*\* 5% significance level.

\*\*\* 1% significance level.

sector dummy to  $-0.370$ , although one may want to keep in mind that some of these job characteristics may almost exclusively exist for informal-sector jobs and hence including them may arguably constitute “overcorrecting” of the true “penalty.” Nevertheless, this exercise allows one to conclude that the distribution of observable human capital and job characteristics can explain nearly three quarters of the average differences in wages between the informal and formal sectors.

One should also note that our estimate of the informal-sector penalty is not too out of line with previous studies that have used OLS and controlled for characteristics of the individual and the job. For example, Schumann et al. (1994) find that both worker and job characteristics are important determinants of the wage differential. Pratap and Quintin (2006) run OLS to compute the wage differential between the formal and informal sectors in Argentina and find that, after controlling for individual and establishment characteristics, the formal premium is 25%. Other studies controlled also for the observed characteristics of all workers in their estimation of the wage differentials between the formal and informal sectors. Tansel (2000) found that the wage differential for men in Turkey between the covered and uncovered wage earners was 40.68% (and it is partly due to the higher levels of human

capital endowments of covered wage earners). Marcouiller et al. (1997) estimated the wage gaps between the formal and the informal sectors. The formal wage premium is weighted by the average characteristics of the formal workers and informal workers separately. When they use formal worker characteristics, the premium in El Salvador is 26%, in Peru 14%, and in Mexico -11%. When they use informal worker characteristics, the premium is 36%, 12%, and -13%, respectively.

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 overestimation, it is difficult from simple labor-force data—since there is no information on nonlabor income and since 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 gain insight into how taxation may affect the negative premium for informal-sector work, we focus on single persons for which we can relatively easily calculate what income tax liabilities are for a given annual income from employment.<sup>20</sup> 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 the first section of 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 Revenue Service.<sup>21</sup> One should note from table 1 that, after taxes, formal-sector wages are only 101% 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% lower (-0.188), which indicates that looking solely at gross wages may substantially overestimate any difference in earnings.

<sup>20</sup> When reporting their earnings, individuals are explicitly asked to state the amount, including overtime, allowances, and bonus, but before any tax or deductions; 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.

<sup>21</sup> See <http://www.sars.gov.za>.

### 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.<sup>22</sup>

The results of taking first differences of the wage equation for our whole sample are provided in the second section of table 2.<sup>23</sup> As can be seen, this considerably lowers the estimated wage penalty to 17.9%, 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 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 second section of table 2, the penalty in terms of gross wages is a little below six percentage points higher for the single subsample 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 into account 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 a large part of the control group consists of job stayers.<sup>24</sup> For example, from those working in the formal sector, only about 6.5% in any period move to a new job, while the corresponding figure

<sup>22</sup> 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 labor market that convincingly do not determine wages.

<sup>23</sup> 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.

<sup>24</sup> As a matter of fact, over 60% of our observations refer to job stayers. Of the job movers, slightly over 10% move between the informal and formal sectors.

for the informal sector is 24.5%. Job movers may, however, be very different from job stayers, and the literature has proposed a number of reasons in this regard. For instance, 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.<sup>25</sup> In contrast, job-matching models would predict positive gains, since workers leave their employers in search of better matches (see Jovanovic 1979). It is important to note that, for either of these explanations, differences in the propensity in job mobility across formal/informal sectors could very well be driving the results observed in our DID estimation.<sup>26</sup> As a first step to investigate this, we ran a simple probit regression of the determinants of the probability to move jobs, including initial wage (of the original job) and all our human capital and job characteristics variables as controls. In addition, we included interaction terms of all of these with an informal-sector dummy, where the interaction terms can be interpreted as the differential effect of characteristics in terms of being a job mover who leaves a job in the informal rather than the formal sector. The results, not reported here,<sup>27</sup> showed that a greater level of wages, job training, union membership, greater age, and higher tenure all act to decrease the probability of moving, with the effect of the latter two aspects being at a decreasing rate. Thus, there appears to be support for both the job-matching and human capital explanations of job changing. In terms of the interaction terms, however, only those on the black and Afrikaans-speaking dummies were significant, with their negative sign suggesting that informal-sector workers with these characteristics are less likely to move jobs than are equivalent workers employed in the formal economy.

While the probit results just discussed may alleviate some of the concern that differing propensities to move across sectors could be driving our results, another simple manner to investigate whether this is true is to just reduce our sample to job movers so that the control group is now only those who moved jobs but remained within their sector. In order to define such job movers, we use the information on their tenure and classify any person who

<sup>25</sup> See Hashimoto (1981) for a theoretical model or Farber (2005) for evidence of wage penalties to displaced workers.

<sup>26</sup> Sample selection into unemployment could also be biasing our results. In this regard, one should note that, on average, 3.3% and 21.5% of individuals became unemployed over any wave for the formal and informal sectors, respectively. Ideally, if one had had information on layoffs rather than quits, in terms of both layoffs and job movers, one could have attempted to model this selection process.

<sup>27</sup> Detailed results are available from the authors.

reports less than 6 months of tenure as a person who has started a new job.<sup>28</sup> Raw averages show that from those with formal employment who switch jobs, about 9.6% move to the informal sector, while 47.5% of job movers in the informal sector, on average, move in the opposite direction.

To first investigate how individuals differed in the sense of moving from the formal to the informal sector rather than vice versa, we ran a simple probit of the determinants of changing sectors, conditional on being a job mover, for which the explanatory variables included the wage of the initial job, the human capital proxies, the job characteristics, as well as interaction terms of all of these with informal-sector status. The interaction terms can thus be interpreted in terms of what factors are more likely attributable to persons moving from the informal to the formal sector rather than in the opposite direction. The estimates, not reported here, indicated that higher wages and primary or secondary education reduced the probability of switching sectors. In terms of the interaction terms, the positive and significant coefficient on the two educational dummy variables implied, however, that these educational effects were lower for those moving from the informal to the formal sector. In addition, part-time job movers were less likely to be moving from the informal to the formal sector, rather than vice versa. Thus, apart from being part-time employed, only educational levels are able to predict whether workers will switch sectors, and this latter effect is lower for those employed in the informal economy. Hence, in terms of most of our observables, there appears to be little difference in the propensity to switch sectors among job movers.

We next replicated our DID results of the informal-sector wage premium for job movers only, as shown in the last three rows of the second section of table 2. One should note in this regard that any estimated coefficient for this movers subsample, 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 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 lead to a noticeable bias in the estimated penalty involved in moving to the informal sector. By examining single persons, however, one still finds evidence that any observed penalty may simply be due to the failure to take into account tax avoidance in the informal sector,

<sup>28</sup> 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.

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 nonmissing values on the chosen controls. This may, however, lead to the case in which one is comparing very different treatment- and control-group individuals, that is, where there is little of what is known as “common support” among the two groups. An extreme example in this regard may be when age is an important determinant of selection into the informal market and wages but only young people work in the informal sector jobs and the elderly are employed in formal-sector jobs. Moreover, as noted earlier, many of the job characteristics that we include may not be considered because of informal employment.

Thus, one would ideally like to ensure that one has comparable groups in which, for each group of similar individuals, there are some working in the formal sector 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 use 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 sector compared to working in the formal sector (also known as the propensity score), conditional on a set of observables that determines 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 used 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.<sup>29</sup> We thus include all of our explanatory variables used in the OLS-levels

<sup>29</sup> 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, Dorsett, and Purdon (2002) note that this may just exacerbate the overlap problem and lead to inconsistent estimates by increasing the variance.

equation as determinants of being employed in the informal sector.<sup>30</sup> 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 0.05 without replacement, by which an individual from the control group is matched with an individual from the treatment group who lies within this chosen caliper and who is closest in terms of the propensity score.<sup>31</sup> This resulted in matching, in total, 412 observations. The choice of caliper can be pertinent in determining the size and quality of the matched sample in terms, but it may be difficult to judge a priori what level of lack of similarity is tolerable (see Caliendo and Kopeinig 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.<sup>32</sup>

An important assumption behind the validity of the PSM estimator 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 a case study in which 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 with DID estimator can produce superior estimates in a nonexperimental context, and we thus similarly proceed along these lines. The results of this combined PSM-DID estimator with our data are shown in the lower section of table 2.<sup>33</sup> 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

<sup>30</sup> Detailed results of this probit estimation are available from the authors upon request.

<sup>31</sup> The matching is performed in STATA version 9 using software provided by Sianesi (2001).

<sup>32</sup> 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.

<sup>33</sup> 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.

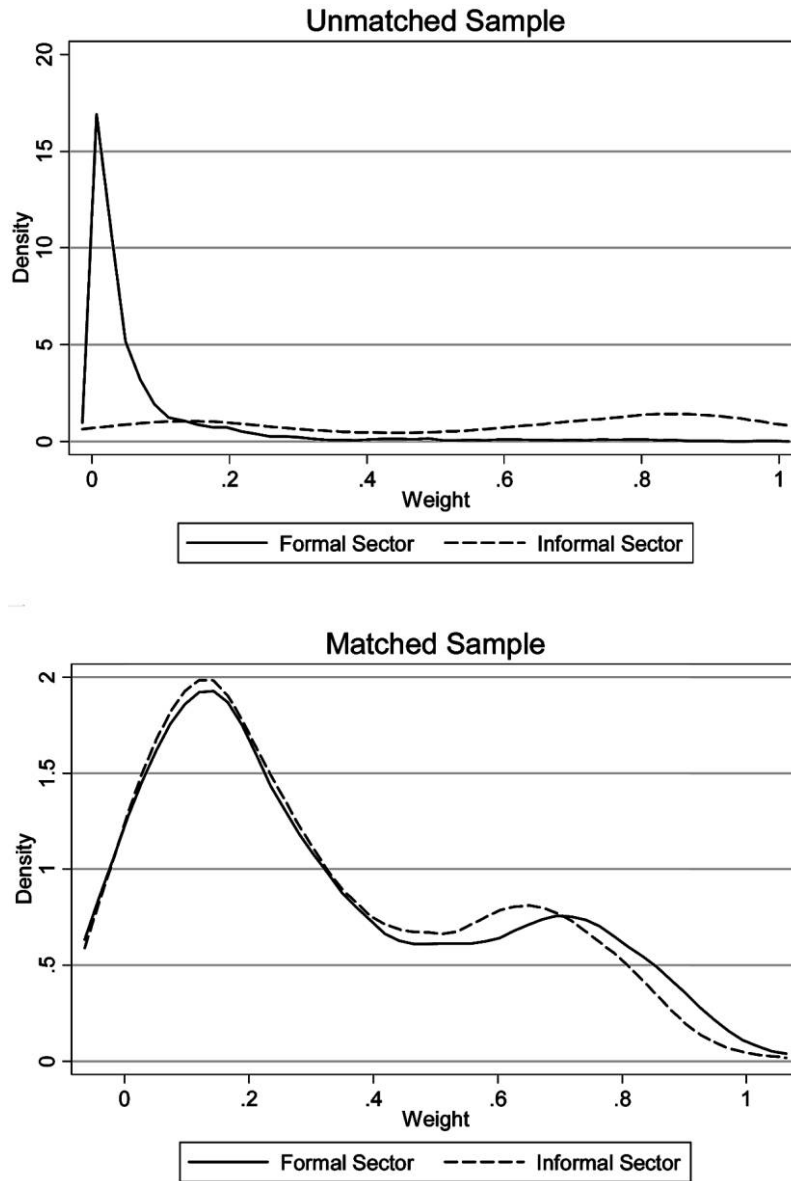


Figure 1. Kernel-density estimates of the estimated propensity scores

as representative of the whole sample, we proceed to using just DID estimates of the penalty for the remainder of our analysis.

#### **D. 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 nonrandom 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 reestimated 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 the first section of table 3 shows a similar coefficient on the informal-sector dummy to that found for the linked sample.

Our definition of informal-sector employment, based on the direct question on informality and all the listed benefits, 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 when there is increased outsourcing and subcontracting of employment.<sup>34</sup> 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% of our sample compared to definition A.<sup>35</sup> We then proceeded to implement our DID estimator with this new definition of informality (definition B), the results of which are given in the remaining rows of the first section of table 3. However, this leaves the results qualitatively similar, and the changes are quantitatively small.

As noted in Section II, over 40% of informal-sector employees are working in the private households sector, while this industry only constitutes 1% of the formal sector. To check whether our estimates of the informal-sector wage penalty are being driven by workers in private households, we recalculated our DID estimates for non-private household industries only, the results of which are shown in the second section 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.

<sup>34</sup> Thanks are due to a referee for making this point.

<sup>35</sup> A large portion of reclassifications were those that stated they were in formal employment but did not receive any of these benefits and thus were reclassified as informal.

**TABLE 3**  
ESTIMATES OF THE INFORMAL-SECTOR WAGE PENALTY—ROBUSTNESS CHECKS

Method	Sample	Wage	Controls	IS Def.	Coeff.	SE	R <sup>2</sup>	Obs.
OLS	Excluded	Gross	HC, JC	A	-.355***	.020	.63	18,279
DID	All	Gross	JC	B	-.159***	.017	.04	8,614
DID	Single	Gross	JC	B	-.201***	.029	.05	2,523
DID	Single	Net	JC	B	-.036	.028	.04	2,521
DID	Movers	Gross	JC	B	-.181***	.057	.08	612
DID	Movers (single)	Gross	JC	B	-.230***	.077	.12	283
DID	Movers (single)	Net	JC	B	-.055	.075	.09	283
DID	NPH	Gross	JC	A	-.171***	.038	.03	8,043
DID	NPH-single	Gross	JC	A	-.212***	.063	.04	2,281
DID	NPH-single	Net	JC	A	-.049	.061	.04	2,277
DID	NPH-movers	Gross	JC	A	-.250***	.094	.08	529
DID	NPH-movers (single)	Gross	JC	A	-.207**	.125	.14	238
DID	NPH-movers (single)	Net	JC	A	-.024	.037	.15	238
DID	NIB	Gross	JC	A	-.174***	.031	.04	6,463
DID	NIB-single	Gross	JC	A	-.244***	.053	.05	1,927
DID	NIB-single	Net	JC	A	-.075	.052	.04	1,923
DID	NIB-movers	Gross	JC	A	-.244***	.092	.11	470
DID	NIB-movers (single)	Gross	JC	A	-.266*	.127	.16	216
DID	NIB-movers (single)	Net	JC	A	-.102	.126	.12	216
DID	All	Gross	...	A	-.106***	.028	.01	8,736
DID	Single	Gross	...	A	-.137***	.044	.01	2,524
DID	Single	Net	...	A	-.064	.043	.01	2,520
DID	Movers	Gross	...	A	-.220***	.079	.02	618
DID	Movers (single)	Gross	...	A	-.193**	.109	.01	281
DID	Movers (single)	Net	...	A	-.091	.107	.01	281
DID	All	Gross	JC, BF	A	-.160***	.030	.04	8,422
DID	Single	Gross	JC, BF	A	-.208***	.049	.05	2,428
DID	Single	Net	JC, BF	A	-.039	.048	.05	2,424
DID	Movers	Gross	JC, BF	A	-.275***	.077	.08	580
DID	Movers (single)	Gross	JC, BF	A	-.278***	.100	.12	267
DID	Movers (single)	Net	JC, BF	A	-.113	.099	.10	267

**Note.** IS = informal sector; HC = human capital controls; JC = job characteristics controls; DID = difference-in-differences; NPH = nonprivate households; NIB = nonincome bracket; A = standard informal-sector definition (see text); B = alternative informal-sector definition (see text); BF = benefit dummies included. HC controls drop out in the DID estimations.

\* 10% significance level.

\*\* 5% significance level.

\*\*\* 1% significance level.

One may also be concerned that most of those who 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 subsample of individuals who reported actual earnings in the third section of table 3. Again, the findings are qualitatively the same and quantitatively similar to our entire linked sample.

As noted earlier, controlling for all job characteristics may arguably constitute “overcorrecting,” since many of these features are the very aspects that one normally associates with working in the informal sector.<sup>36</sup> To see whether such overcorrecting may be driving our result of no remaining penalty, we reran our DID estimation for single-job movers but excluded all (first differenced) job-related characteristics in the fourth section of table 3. As can be seen, the wage penalty for gross wages is indeed higher than when these are excluded. It is important to note, however, that once we account for tax payments in the formal sector, any discrepancy in wages disappears.

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 wages for those workers who receive such benefits may be stated net of such employer contributions, thus creating measurement error that is likely to be systematically correlated with informal-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 reestimated our benchmark OLS and DID regressions, including dummies for whether the worker reported receiving medical, unemployment insurance fund, or pension contributions and paid leave from the employer, the results of which are shown in the lower section of table 3. As can be seen, the coefficient estimates are sometimes lower and sometimes higher than compared to not including the dummies (as reported in table 2). 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 taxes into account, the disappearance of the informal-sector wage penalty remains.

#### **IV. Conclusion**

In this article we reexamined 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 include information that allows one to explicitly classify workers as being employed in the informal sector and links workers over time. Our analysis unearthed a number of potentially interesting findings. First of all, nearly 75% 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

<sup>36</sup> One should note that the PSM estimator arguably addresses this lack of “overlap” problem to some extent.

wages are still 37% lower for those employed in the informal sector. Controlling for unobservable time-invariant factors further reduces the informal-sector wage penalty to just over 18%. It is important to note, however, that when we focus our analysis on single men for whom 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. One should note, nevertheless, that simply deducting taxes ignores the long-run advantages of being part of the social welfare system in terms of receiving benefits such as pension payments or unemployment compensation some time in the future. In addition, 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. Moreover, one should not forget that our data refer to a subsample of total informal-sector employment, namely, non-self-employed males.

Nevertheless, at least in the short run and in terms of remuneration, our results suggest that the formality of employers may not be a feature of unexplained pay differences in the labor market in South Africa.<sup>37</sup> One possibility is that there are actually no barriers of entry between formal and informal jobs, as has been traditionally modeled, but that workers simply sort themselves according to their productivity into these sectors and that this requires the proper estimation methodology and data to be accounted for. Another reason could be that there are not only barriers to entry to the formal sector, but also, as proposed by Kingdon and Knight (2001), barriers to entry to the informal sector, so that the least-productive workers remain unemployed. Arguably, the relatively high unemployment rate in South Africa provides some indirect support for this explanation. Finally, Günther and Launov (2006) provide evidence that some persons might be voluntarily employed in the informal sector while others work there involuntarily, so the calculation of an average informal-sector penalty may be distortive. Clearly, all of these aspects should be the subject of future research.

<sup>37</sup> Having said this, one should note that some of these very controlled variables may be highly correlated with informal-sector status. For example, Badaoui, Strobl, and Walsh (2006) embed informal firms in an equilibrium search model and show theoretically and empirically that firm size is the crucial variable, in that large firms pay more but also choose to be in the formal sector.

## 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 his or her 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 his or her 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 (cannot read and write), no primary (can read and write), primary, secondary, National Technical Certificate, university	Six dummies associated with 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 has 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 or she mentioned.
Tools	Dummy for whether the person owns the tools and/or the equipment that he or 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).

## References

- Amaral, Pedro, and Erwan Quintin. 2006. "A Competitive Model of the Informal Sector." *Journal of Monetary Economics* 53 (October): 1541–53.
- Ardington, Cally, David Lam, Murray Leibbrandt, and Matthew Welch. 2006. "The Sensitivity to Key Data Imputations of Recent Estimates of Income Poverty and Inequality in South Africa." *Economic Modelling* 23, no. 5:822–35.
- Badaoui, Eliane, Eric Strobl, and Frank Walsh. 2006. "An Equilibrium Search Model

- of the Informal Sector." Working Paper no. 06/29, School of Economics, University College Dublin.
- Banerjee, Nirmla. 1985. *Women Workers in the Unorganized Sector*. Hyderabad: Sangam.
- 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 (June), World Bank, Washington, DC.
- Bekkers, Hans, and Wim Stoffers. 1995. "Measuring Informal Sector Employment in Pakistan: Testing a New Methodology." *International Labour Review* (Geneva) 134, no. 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, London.
- Byrne, David, and Eric Strobl. 2004. "Defining Unemployment in Developing Countries: Evidence from Trinidad and Tobago." *Journal of Development Economics* 73 (February): 465–76.
- Calderon-Madrid, Angel. 2000. "Job Stability and Labor Mobility in Urban Mexico: A Study Based on Duration Models and Transition Analysis." Inter-American Development Bank Research Network Working Paper no. R-419, Washington, DC.
- Caliendo, Marco, and Sabine Kopeinig. 2005. "Some Practical Guidance for the Implementation of Propensity Score Matching." IZA Discussion Paper no. 1588, Bonn.
- 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 Nonexperimental Causal Studies." *Review of Economics and Statistics* 84 (February): 151–61.
- De Soto, Hernando. 1989. *The Other Path: The Invisible Revolution in the Third World*. New York: Harper & Row.
- Devey, Richard, Caroline Skinner, and Imraan Valodia. 2002. "The Informal Economy in South Africa: Who, Where, What and How Much?" Paper presented at the Conference on Labour Markets in South Africa, Johannesburg, October 22–24.
- . 2003. "Informal Economy Employment Data in South Africa: A Critical Analysis." Report prepared for the Employment Data Research Group, Human Sciences Research Council, Pretoria.
- 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, Federal Reserve Bank of Chicago.
- 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): 459–93.
- 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): 513–29.
- Günther, Isabel, and Andrey Launov. 2006. "Competitive and Segmented Informal Labor Markets." IZA Discussion Paper no. 2349, Bonn.
- Hashimoto, Masanori. 1981. "Firm Specific Human Capital as a Shared Investment." *American Economic Review* 71 (June): 475–82.
- 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): 507–42.
- Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87 (October): 972–90.
- 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.
- . 2003. "Well-Being Poverty versus Income Poverty and Capabilities Poverty." Working Paper Series (July), Centre for the Study of African Economies, University of Oxford.
- Klasen, Stephan. 2002. "Low Schooling for Girls, Slower Growth for All?" *World Bank Economic Review* 16:345–73.
- 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), National Bureau of Economic Research, Cambridge, MA.
- Maloney, William. 1998. "Are LDC Labor Markets Dualistic?" Working paper (February), World Bank, Washington, DC.
- . 1999. "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico." *World Bank Economic Review* 13: 275–302.
- . 2004. "Informality Revisited." *World Development* 32 (July): 1159–78.
- 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): 367–92.
- Marshall, Adriana. 1987. "Non-standard Employment Practices in Latin America." Discussion Paper no. 6, Labour Market Programme, International Labour Office, Geneva.
- Masatlioglu, Yusufchan, and Jamele Rigolini. 2005. "Labor Dynamics and the Informal Economy." Conference Paper (May 13), United Nations University, Tokyo.
- Mazumdar, Dipak. 1975. "The Theory of Share-Cropping with Labour Market Dualism." *Economica* 42 (August): 261–71.
- . 1982. "The Urban Labor Market and Income Distribution: A Study of Malaysia." *Journal of Economic Literature* 20 (June): 626–28.
- Oi, Walter, and Todd Idson. 1999. "Firm Size and Wages." In *Handbook of Labor Economics*, 3rd ed., ed. O. Ashenfelter and D. Card. Amsterdam: North-Holland.
- 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 (August), National Council of Applied Economic Research, New Delhi.
- Pradhan, Menno, and Arthur van Soest. 1995. "Formal and Informal Sector Employment in Urban Areas of Bolivia." *Labour Economics* 2 (September): 275–97.
- Pratap, Sangeeta, and Erwan Quintin. 2006. "Are Labor Markets Segmented in Argentina? A Semiparametric Approach." *European Economic Review* 50 (October): 1817–41.
- Roberts, Bryan. 1989. "The Other Working Class: Uncommitted Labor in Britain, Spain, and Mexico." In *Cross-National Research in Sociology*, ed. M. L. Kohn, 352–72. Newbury Park, CA: Sage.
- 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.
- . 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *American Statistician* 39 (February): 33–38.
- Schumann, Paul L., Dennis A. Ahlburg, and Christine Brown Mahoney. 1994. "The Effects of Human Capital and Job Characteristics on Pay." *Journal of Human Resources* 29, no. 2:481–503.
- Shapiro, Carl, and Joseph Stiglitz. 1984. "Unemployment as a Worker Discipline Device." *American Economic Review* 74 (June): 433–44.
- Sianesi, Barbara. 2001. "Implementing Propensity Score Matching Estimators with STATA." Report prepared for UK STATA Users Group (May), VII Meeting, London. <http://www.stata.com/meeting/7uk/sianesi.pdf>.
- Smith, Jeffrey, and Petra Todd. 2005. "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125 (March–April): 305–53.
- 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.
- Tansel, Aysit. 1999. "Formal versus Informal Sector Choice of Wage Earners and Their Wages in Turkey." Economic Research Forum Working Paper no. 9927 (February), Cairo.
- . 2000. "Wage Earners, Self Employed and Gender in the Informal Sector in Turkey." Working Paper Series no. 24 (November), Policy Research Report on Gender and Development, Washington, DC.
- Tokman, Victor. 1982. "Unequal Development and the Absorption of Labour: Latin America 1950–1980." *CEPAL Review* 17:121–33.
- Tybout, James. 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature* 38 (March): 11–44.
- Yamada, Gustavo. 1996. "Urban Informal Employment and Self-Employment in Developing Countries: Theory and Evidence." *Economic Development and Cultural Change* 44 (January): 289–314.