DISCUSSION PAPER SERIES

IZA DP No. 6727

# Is There an Informal Employment Wage Premium? Evidence from Tajikistan

G. Reza Arabsheibani
Anita Staneva

July 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Is There an Informal Employment Wage Premium? Evidence from Tajikistan

**G. Reza Arabsheibani**
*Swansea University*
*and IZA*

**Anita Staneva**
*Swansea University*

Discussion Paper No. 6727
July 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 6727
July 2012

# ABSTRACT

## Is There an Informal Employment Wage Premium? Evidence from Tajikistan

This paper defines informal sector employment and decomposes the difference in earnings distributions between formal and informal sector employees in Tajikistan for 2007. Using the quantile regression decomposition technique proposed by Machado and Mata (2005), we find a significant informal employment wage premium across the whole earnings distribution. This contrast with earlier studies and casts doubt on the recent literature showing that the informal sector is poorly rewarded. It seems to be the case that the informal employment in Tajikistan is the main source of income.

JEL Classification: C14, J21, J30

Keywords: formal/informal employment, quantile regression decomposition

Corresponding author:

G. Reza Arabsheibani,
School of Business, Economics and Law
Richard Price Building
Swansea University
Singleton Park
Swansea, SA2 8PP
United Kingdom
E-mail: G.Arabsheibani@swan.ac.uk

# 1. INTRODUCTION

The size of the informal sector, cited as the central factor underlying wage inequality, poverty and labour market inefficiency, has always attracted attention in all transition and developing countries and Tajikistan is not an exception. The National Human Development Report (2009) for Tajikistan shows that during 2008-2009 over 47% of employment in the country was in the informal sector and the number of people employed in this sector was estimated to be over 1 million.

It is generally assumed, and empirically supported by much of the literature, that workers in the informal sector are paid less than their formal sector counterparts[1]. Empirically and theoretically it is not clear why this should be the case. There are a number of explanations offered, most of which are based on a segmented view of the labour market (Badaoui*et al*., 2007). The presence of barriers to entry into the formal sector could be a possible cause (Fields, 1975; Mazumdar, 1975). A wage penalty for informal workers might be due to sorting, where those with lower human capital are more likely to work in the informal sector (Tokman, 1982). Marcouiller*et al.* (1997) show a significant wage premium in the informal sector in Mexico, while in El Salvador and Peru the formal workers have a higher wage premium. However, evidence in favour of the existence of a wage premium for formal worker depends on the category of informal job (e.g. self-employed are in general better paid than salaried workers). Using the Russian Longitudinal Monitoring Survey data, Braithwaite (1994) and Kolev (1998) find higher wage premium in secondary jobs, compared to the main job, which are the main locus of the informal economic activities. Other studies find evidence of a significant earnings differential in the lower part of the earnings distribution (Tannuri-Pianto and Pianto, 2002). Carneiro and Henley (2001) indicate that differences in earnings are strongly explained by the non-observable characteristics of workers who

decide to join each sector. Wu and Li (2006) show that about 90% of the observed informal-formal monthly income differential can be explained by differences in observed characteristics in urban China.

One of the major difficulties in being able to identify and understand the informal sector has been the lack of consensus on how to define and measure informal sector activities. Whilst there is a large literature on the informal economy, very few studies have applied it so far to the transition economies. In this regard the present paper attempts to capture the diversity of informal sector in Tajikistan by estimating earnings differential along two main dimensions. First, we try to distinguish and define informal employment in Tajikistan – a country that has received little attention in the literature.Second, in order to provide more detailed insight into the formal-informal pay differential, we look across the entire conditional earnings distribution.

The remainder of the paper is structured as followed. In section 2, we provide a discussion of the existing literature on defining and measuring informal sector employment and propose different measures of informality for Tajikistan. Section 3 discusses the data, and section 4 presents the econometric methodology. Finally, sections 5 and 6 discuss the main results and conclusions.

## 2. DEFINING INFORMALITY

Informal employment, sometimes known as undeclared, hidden or grey employment, can be broadly described as employment engaged in producing legal goods and services where one or more of the legal requirements associated with employment are not complied with (OECD, 2008). The informal economy is traditionally viewed as the disadvantaged sector employing unskilled, less educated, urban migrants or ethnic minorities (Mazumdar, 1983). It is supposed to play a negative role in the economy by

decreasing official output, reducing government tax revenue and constraining the growth of the private sector (Johnson *et al*., 1997; Lacko, 2000). The informal employees lack social security coverage and some or all of the protections provided by labour contracts. It is particularly important, given recent developments in the debate on measuring informality, to carefully distinguish between informal and formal sector employment. Although the term has been very widely used, its meaning is not clear. There is no precise definition and consensus over what constitutes informal sector employment and how to measure it. In many cases researchers' choice of definition is determined by the availability of data. Thus, the informal sector has been referred to as street vendors, domestic workers or unregistered small-scale activities in developing countries, and drug trafficking and prostitution in western countries. Some authors argue that all self-employed workers should be included in the informal sector, while others include only those who are not paying social security contributions and are outside any employment protection. Legalistic definition of informality refers to the avoidance of formal registration, taxation and the lack of social security protection. Merrick (1976) defines informal sector workers by their lack of social security status. International Labour Office's (ILO) definition is based on employees in small establishment size of fewer than 5-10 employees (depending on the country). Pradhan and van Soest (1995, 1997) use a definition of fewer than 6 employees for Bolivia and Mexico; Funkhouser (1997) uses fewer than 5 employees for an analysis of five Central American countries. Under the guidelines of the statistical definition of informal employment established by the International Conference of Labour Statisticians in 2003, informal employment includes casual jobs or jobs with limited short duration; jobs with hours of work or wages below a specified threshold; employment of persons in households; jobs where the employee's place of work is outside the premises of the

employer's enterprise; or jobs for which labour regulations are not applied (Daza, 2005).

Henley *et al*. (2009) indicate three definitions of informality – first, according to the employment contract status, second, according to the social security protection and finally according to the nature of the employment and the characteristics of the employer. More specifically, the way of defining informality in developing countries does matters and the conditional impact of particular factors on the likelihood of informality varies considerably from one definition to another. Where information on contract status is not available, the alternative definition of informal employment is by the lack of social protection. For instance, Portes*et al*. (1989), Marcouiller*et al*. (1997), and Maloney (1999) define workers as informal employees by the criterion of no social protection and security. Bernabe (2000) applies the typology of informal employment in Georgia by using proxies for "household enterprises" and "non-regular employment". A study by Gasparini and Tornarolli (2007) defines as informal employees those engaged in low-productive jobs, often family-based activities. There are a group of researchers who view the informal labour market in line with the 'survival-oriented' informal activities. For instance, Desai and Idson (1998) and Rose and McAllister (1996), show that Russian households rely mainly on informal economic activities in order to cope with the dramatic deterioration in their life circumstances. Johnson *et al.* (1997) identify six types of survival strategies used in Russia, which they refer to as informal activities: having a second job; using land to grow food; working as a private taxi driver; renting out one's apartment; business trips abroad for resale and renting out one's garage.

The informal employment literature is moving away from the traditional view of informality as evidence of labour market segmentation. Rather than seeing informal employment as a survival mechanism for low-productivity workers who are queuing

until they find a better paid, formal job opportunity, recent empirical studies argues that some informal workers "choose" informal employment. They do so because informal employment offers them the best financial return on their skills or experience (OECD, 2008). Additionally, informal sector offers flexibility in hours, place of work, and allows small businesses with entrepreneur ability to made a successful career.

The diversity in definitions of the informal sector is a result of the fact that different units of observation and different criteria of informality have been used. Bernabe (2002) summarises four main units of observation (enterprises, activities, income, and people) and the main criteria to identify informality (registration and regulation). In developing countries, informality has largely been associated with urban household enterprises. In western industrialised countries, the term has been used to describe all income or production that avoids taxation. Finally, there has been little debate on how to define informal sector in transition countries.

*Defining informal employment in Tajikistan*

In the process of transition to market economy, the structure and character of informal employment in Tajikistan has changed and informal employment has reached a considerable scale. The informal economy in Tajikistan contributes approximately 35% to Gross National Product (OECD, 2007). Some commentators argue that the country seems to have entered into a process that has already taken hold in other countries – Latin America, sub-Saharan Africa and many Asian countries,where there is a tendency towards the 'informalisation' of the formal sector (Wallenborn, 2009). According to the Labour Force Survey (LFS) in 2004 the share of employment in the informal sector of the country was 53.3%.The NHDR (2007) Survey on the informal economy in Tajikistan shows that the main reason why firms prefer to use informal methods, are

overcomplicated and burdensome formal procedures and low quality of civil servants. The overcomplicated,costly, and time-consuming formal procedures seem to encourage firms to try avoiding them by resorting to informal methods in an effort to minimize costs of operation.

Does employment in informal sector offer advantages? Maloney (1999) suggests that the informal sector in developing countries may be a desirable choice. Individuals working in the informal sector benefit from flexibility in terms of working hours, and in some cases choice of work location. This aspect may be especially valued by women with children. At the same time, because wages of the "unofficially employed" are not subject to taxation, salaries in the informal sector may be significantly higher. This suggests that workers may face an improved wage offer in the informal sector compared to the formal sector (Henley*et al*., 2009).

On the question what is the main reason to accept a job without social security benefits, 26% of respondents from the 2007 Tajikistan survey indicate that social security benefits are not important and 20% answered that they can receive more money by working in the informal sector. About 57% of the respondents working without a signed contract reported that they are satisfied with their current job conditions. However, not all people in the informal sector are there by choice. Many individuals may be displaced involuntarily into this sector because this is their only chance of paid employment. Around 44% of the respondents in the 2007 survey stated that they actually have no other choice and therefore would accept a job without social security benefits.

In line with Henley *et al.* (2009), we adopt three different definitions of informality for Tajikistan[2]. Table 1 presents proportions of workers, who report wages and who are classified as informal under each of the following:

Definition *A – contract status:*there is a widespread belief that employment relations should be legitimized in a written employment contract. We classify a person as an informal worker if he/she works without a signed labour contract. The data indicates that 50.4% of all economically active in the sample are employed without any written contract. Overall, 39.3% of employees, 93.7% of self-employees, and 94.1% of domestic workers work without any written contract (see Table 1).

Definition*B – social security affiliation status:*a person is defined as an informal employee if he/she has no affiliation to the social security scheme. Based on this definition 52.4% of all workers who report wages have no social security affiliation. Again almost all of the self-employed and domestic workers operate without social security affiliation.

Definition *C – establishment size*:the most commonly used definition of the informal sector is based on the size of the enterprise. A person is defined as informal if he/she is employed in an establishment of less than five employees. However, including only enterprises with less than five employees results in the inclusion of professionals and managerial (2.3%) who could have relatively high incomes and who are considered to be in the formal sector. Therefore we exclude professional group from definition *C*. This measure provides the smallest estimates at around 31% of all economically active work as informal. Around 19% of economically active employees and 70% of self-employed are covered by this definition (see Table 1).

[Table 1 here]

Figure 1 presents a Venn diagram of the three definitions across the full sample of economically active individuals who report wages. About 25% of the workforce can be classified as informal by all three measures and around 40% of workers are classified as informal on the basis of having no signed labour contract and no social security

coverage. The percentage of workers classed as informal under measure *A* and measure *B*, but not under measure *C* is 14.4%. Around 27% of the economically active are classified as informal workers on the basis of having no social security membership and working in the small establishment size.

[Figure 1 here]

Table 2 reveals the different proportions of males and females for these three definitions of informality. Informal employment amongst males is highest when measured by no social security affiliation and lowest when measured by establishment size. Female informality is highest when measured by contract status. Around 44% of females work without any written contract. Overall, female informality is much lower. Defining informality by all three approaches indicates that 31.9% of males and 14.9% of females in Tajikistan work as informal.

[Table 2 here]

It is clear from the above discussions that different definitions of informality are capturing different groups of workers. Each indicator on its own has conceptual and statistical shortcomings as a proxy for informal employment but taken together they may provide a robust approximation. In this way we consider as informal all individuals who operate without social security affiliation, who are employed in small enterprises and who work without any written contract. At the same time, we place in the formal sector those workers who are not classified as informal under any of the three measures.

## 3. THE DATA

The data used in this paper is from the 2007 Tajikistan Living Standard Measurement Survey (LSMS), which provides comprehensive information on education, health, employment, housing, migration and income. Appendix Table A.1 reports the means and standard deviations of the chosen variables broken down into

those working in the formal and those employed in the informal sector, defined as above.Samples are tested to see whether the data sets support separation. The reported *t*-statistics in the descriptive table indicate that almost all variables are significantly different from each other. After dropping observations with missing values for the variables used in the estimation we were left with a total number of 8,123 individuals. Of these observations 25% were for individuals who fall into all three informal definitions. Hourly net earnings are defined as reported monthly earnings divided by 4.33 and then divided by reported weekly hours of work.The control variables used in the analysis include education (total number of years in education), age (linear and quadratic terms), variables for individuals years of tenure with the current firm, controls for marital status, Tajik ethnicity, private sector employment, occupations, rural settlement, and a set of regional variables to pick up regional effects. The omitted categories are workers with more than 10 years within the firm and administrative occupations.

The descriptive statistics highlight some interesting patterns. The logarithm of hourly net earnings in the informal sector is higher than those found in the formal sector. The kernel density, which plots both densities, shows that the informal wage curve is situated to the right of the formal sector one. Empirical evidence suggests that higher minimum wages are associated with lower formal sector employment, at least in countries where the minimum wage is binding in the formal sector (e.g. Carneiro, 2004). Examining the earnings distribution of formal and informal employees provides an indication of whether the minimum wage is binding for formal employees, a key determinant of whether minimum wage has an impact on informality. The vertical line in Figure 2 represents the minimum wage. Very few formal employees and only a small proportion of informal employees earn less than the minimum wage. Based on this

evidence, it seems unlikely that the minimum wage is a particular important cause of informality in Tajikistan[3].

[Figure 2 here]

The quantile-quantile (Q-Q) plot[4] shows that all observations are below the diagonal line, implying that wages for informal workers are higher than their formal workers counterparts. Income inequality is also stronger for higher earners (see Figure 3). This preliminary evidence illustrates significant discrepancy between formal and informal workers in Tajikistan.

[Figure 3 here]

In addition, descriptive statistics show that the informal sector is strongly male dominated. Around 77% of informal workers are males. The average age of people in informal sector is lower than the age of individuals in the formal sector.There are marked differences between sectors in terms of education. The total number of years in education is higher for formal workers. Around 65% of those working in the private sector and only 2% of public sector workers are informally employed. Informal employment is more rural than urban with around 56% of pure defined informal workers in rural areas. Although most studies write specifically about the informal sector in the urban areas, it is surely not possible to deny the existence of similar enterprises in rural areas as in our case. Finally, comparing wages across the formal and informal sectors might suggest that at least some of the differences in wages may be due to the different distributions of occupations across the two sectors. Around 31% of pure informal workers are employed in trade and sales, and around 15% are in elementary occupational jobs (unskilled workers, street vendors, cleaners). Some regions have higher shares of informal employment. In particular the informal employment is highest in Sogd region.

## 4. ECONOMETRIC METHODOLOGY

A common feature of much of the literature is that the analysis is conducted at the mean of the earnings distribution, with no attention paid to how predicted earnings differentials may vary across the distribution. Therefore, to decompose the differentialin the formal and informal log wages into a component due to differences in labour market characteristics between the sectors and a component due to difference in the rewards formal and informal workers receive for those labour market characteristics, we utilise Machado and Mata's (2005) technique.Based on Koenker-Basset (1978), Machado and Mata (2005) propose a method to extend the traditional Oaxaca-Blinder decomposition method. Their main methodological procedure is to simulate a conditional marginal wage distribution estimated through quantile regressions. We are interested in the wage gap measuring the effect of different returns to formal and informal workers when informal sector characteristics are used in counterfactual calculations. A positive wage gap implies that the returns to informal workers' characteristics are lower than those of formal workers, and a negative gap implies the reverse.

The modeling strategy begins by assuming a sample of observations on log earnings, $y_i$ , $i$ = 1,…,$n$, where $y_i$ is dependent on $X_i$, a $K$ x 1 vector comprising education, experience and other control characteristics typically employed as earnings function covariates. The quantile regression model can be expressed as:

$$y_i = X_i'\beta_\theta + u_{\theta i}, \;\; \text{Quant}_\theta(y_i \mid X_i) = X_i'\beta_\theta, \;\; \theta \in (0,1) \qquad (1)$$

where $\text{Quant}_\theta(y_i \mid X_i)$ denotes the quantile $\theta$ of log earnings conditional on the vector of regressors. The regression quantile $\theta$ can be defined as the solution to the problem (Koenker and Bassett; 1978):

$$\min_\beta \frac{1}{n}\left[\sum_{i:y_i \geq X_i'\beta}\theta\left|y_i - X_i'\beta_\theta\right| + \sum_{i:y_i < X_i'\beta}(1-\theta)\left|y_i - X_i'\beta_\theta\right|\right] = \min_\beta \frac{1}{n}\sum_{i=1}^{n}\rho_\theta(u_{\theta i})$$

(2)

where $\rho_\theta(.)$ is a check function, defined as:

$$\rho_\theta(u_{\theta i}) = \begin{cases} \theta u_{\theta i} & \text{if } u_{\theta i} \geq 0 \\ (\theta - 1)u_{\theta i} & \text{if } u_{\theta i} < 0 \end{cases} \tag{3}$$

Estimates can be obtained by minimising the sum of weighted absolute deviations using linear programming methods (Buchinsky 1998), with the estimated variance-covariance matrix obtained using a bootstrap re-sampling. The regression coefficients provide estimates of the marginal change in the $\theta$th conditional quantile due to a marginal change in a particular regressor, on the assumption that a particular individual remains in the same quantile following this marginal change.

In order to undertake a decomposition analysis equation (1) is estimated for quantiles across the distribution for both formal and informal sector workers and the counterfactual density distribution is generated following the Machado and Mata (2005) method. Specifically,the marginal earnings distributions can be briefly described as follows:

Step 1: Generate a random sample of size*m* from a uniform distribution$U[0,1]u_1...,u_m$. This will give a series of numbers telling us which percentiles are to be estimated.

Step 2: Estimate for the formal and informal employees separately quantile regression coefficients: $\hat{\beta}_{u_i}^{form}(\theta), \hat{\beta}_{u_i}^{in}(\theta), i = 0.01,....0.99,$ where $\hat{\beta}_{u_i}^{form}(\theta)$ are $u_i^{th}$ quantile regressions estimates taken from the (log) hourly earnings equation for formal sector workers and $\hat{\beta}_{u_i}^{in}(\theta)$ are $u_i^{th}$ quantile regression estimates taken from the hourly earnings equation for informal sector workers.

Step 3: Sampling with replacement: a random sample of size *m* is taken of formal and informal sector worker's characteristics that were used to estimate quantile regression coefficients. The vectors of characteristics for formal $\{\tilde{X}_i^{form}\}_{i=1}^m$ and informal $\{\tilde{X}_i^{in}\}_{i=1}^m$ sector workers are then used to predict (log) hourly earnings in the formal $\{\tilde{w}_{u_i}^{form} = \tilde{X}_i^{form}\hat{\beta}_{u_i}^{form}\}_{i=1}^m$ and informal $\{\tilde{w}_{u_i}^{in} = \tilde{X}_i^{in}\hat{\beta}_{u_i}^{in}\}_{i=1}^m$ sectors. These predicted wages are equivalent to a random sample of size *m* drawn from the marginal wage distributions of formal $(w^{form})$ and informal $(w^{in})$ sector workers. Counterfactual density is found as $\{\tilde{w}_i^{cf} = \tilde{X}_i^{in}\hat{\beta}_{u_i}^{form}\}_{i=1}^m$, which is the density that would arise if informal sector workers retained their own labour market characteristics but were paid like formal workers[5].

Step 4:Differences in $\theta^{th}$ percentiles of the estimated marginal wage distribution are then used to decompose the formal-informal sector wage gap into an effect due to characteristics in the formal and informal sector being rewarded differently (coefficient effect) and an effect due to differences in the distribution of worker characteristics in the two sectors (characteristic effect).

The difference in the logarithm of hourly earnings between formal and informal sector workers at the $\theta^{th}$ percentile is given by:

$$
\begin{aligned}
&Q_\theta(\ln w^{form}) - Q_\theta(\ln w^{in}) = \\
&= [Q_\theta(\tilde{X}^{form}\hat{\beta}^{form}) - Q_\theta(\tilde{w}^{cf})] + [Q_\theta(\tilde{w}^{cf}) - Q_\theta(\tilde{X}^{in}\hat{\beta}^{in})] + resid \\
&= \underbrace{[Q_\theta(\tilde{X}_i^{form}\hat{\beta}^{form}) - Q_\theta(\tilde{X}_i^{in}\hat{\beta}^{form}(\theta_i))]}_{characteristics\ component} + \underbrace{[Q_\theta(\tilde{X}_i^{in}\hat{\beta}^{form}(\theta_i)) - Q_\theta(\tilde{X}_i^{in}\hat{\beta}^{in})]}_{coefficient\ effect} + resid
\end{aligned}
\tag{3}
$$

where $Q_\theta$ is the $\theta^{th}$ percentile of the earnings distribution. The first term of the right hand side of expression (3), the characteristics component, shows the contribution of the differences in covariates between formal and informal employees to the earnings gap at the quantile $\theta$ and the second term is the contribution due to differences in coefficients

(coefficient effect). The difference between $\theta^{th}$ quantile of the marginal wage densities between formal and informal distributions weighted by the characteristics of workers randomly chosen in the economy does contain an additional component, which we treat as a residual. The residual term is typically of second order of importance and tends to became smaller with a larger number of simulations. Standard errors for the reported components of the decomposition were obtained using a standard bootstrapping method[6].

## 5. EMPIRICAL RESULTS

The results from the decomposition analysis are reported in Table 3. In the secondcolumn, we present the raw wage gap estimates, calculated as the difference in log hourly wages between formal and informal sector employees at certain points of the wage distribution. In the next columns, we give the contribution of the coefficients and the covariates to the difference between the $\theta^{th}$ quantile of the formal sector wage distribution and $\theta^{th}$ quantile of the informal sector wage distribution, together with the percentage of the gap that is attributable to the coefficient and covariate effects, and residual terms due to the differences between the empirical and simulated densities. The interesting part is the differences in rewards. If the two sectors reward the same characteristic differently, it might be indication of different wage setting mechanisms at the two labour markets. The bootstrapped standard errors for these contributions are given in parentheses. Estimates at the $10^{th}$, $20^{th}$, $30^{th}$, $40^{th}$, $50^{th}$, $60^{th}$, $70^{th}$, $80^{th}$, and $90^{th}$ percentiles are reported. The results over the whole distribution are best viewed graphically. In Figure 4 we plot the estimated coefficient effect with the 95% confidence interval.

The estimates show that the raw wage differential is negative but diminishes and is considerably narrower at the top of the distribution compared to the bottom of the

distribution. The raw wage gap is sizeable, especially at the low end of the distribution. Both coefficients and covariates contribute to the actual wage gap and their effect is significantly different from zero. The largest fraction of the formal-informal wage gap is attributable to the differences in characteristics. Overall, the model works fairly well, as the residuals account for relatively small portion of the total wage gap.

The results indicate that the 'unexplained' component of the inter-sector wage differential works in favour of the informal sector. The coefficient effect is negative across the whole earning distribution. The penalty faced by formal sector workers is especially large at the lower end of the distribution where the informal sector employees earn substantially higher wage premium. The wage penalty for formal workers ranges between 48% of the relevant gap at the bottom of the earnings distribution to around 20% at the top of the distribution. Overall, at the top of the distribution the informal workers face a lower wage premium. This might be due to the fact that at the top of the earnings distribution formal workers tend to be in larger firms which pay higher wages and we might expect higher incentives to be registered. The large informal earnings found here are in line with Marcouiller*et al.* (1997) who find a wage premium associated with work in informal sector in Mexico which they explain with different benefit systems in the two sectors.

[Table 3 here]

[Figure 4 here]

Note also that our sample is strongly male dominated[7]. However, a detailed examination of formal versus informal wage differential among males indicates that identical male workers still earned more in the informal sector than in the formal one.

Interestingly, the covariate effect is negative across the whole earnings distribution, indicating that informal workersstrongly dominate with their endowment

component. Moreover, at the top end of the distribution, the large proportion of the raw gap–about 76% – appears to be the result of characteristics effect.However as higher education seems to be associated with the formal employment, we might indicate that factors other than human capital endowments explain the wage disparities in the country – generally attributed to labour market imperfections.

## 6. CONCLUSIONS

This paper contributes to the previous literature by defining informal sector employment and decomposing the differences in earnings distribution between formal and informal sector employment for Tajikistan, a country where no empirical evidence on informal sector employment and earnings currently exists. The Machado and Mata (2005) method is applied, which is design to simulate the counterfactual distribution that would arise if informal sector workers retained their own labour market characteristics but were paid like formal workers.The decomposition analysis amounts to examining the extent to which the observed earnings differential is attributable to differences in the observable characteristics and differences in returns to these characteristics.

We find a significant high level of informal employment in Tajikistan. Around 65% of the 2007 Tajik sample is classified as informal in at least one of the proposed definitions, with males having higher informal employment rates than females. Informal sector workers tend to be less qualified and more likely to be employed in the trade and sales services. Around 65-67% of private sector workers are informally employed.

The wage differential decomposition results indicate a strong wage penalty for formal sector workers throughout the whole earning distribution.The penalty is especially large at the lower end of the distribution, where 48% of the observed wage gap is attributed to differences in returns. Following Marcouiller*et al*. (1997) we can

attribute the informal sector premium in Tajikistan on the ground of compensating differentials theory, which would lead one to expect informal sector wages to be higher than formal sector wages. Benefits obviously differ between workers covered by social security and those who are not. We found that approximately 60-70% of the observed differential can be ascribedto differences in distribution of characteristics between formal and informal sector workers.At the top end of the conditional earnings distribution, the characteristics effect plays a larger role in explaining the formal sector wage gap. Nonetheless, most of the formal sector wage gap across the distribution continues to be accounted for by differences in how the two sectors are rewarded.

Our findings contradict the previous literature and cast doubt on the accepted notion that the informal sector is always poorly rewarded compared to the formal sector. The implication is that some informal jobs are better than some formal jobs with respect to earnings. It might be the case that Tajikistan requires a different story and policy implications.


## Acknowledgements

We would like to thank the comments we received from Andrew Henley and Alan Marin throughout the development of this paper.


**Notes**

1. See for example Mazumdar (1981), Heckman and Hotz (1986), Pradhan and Van Soest (1995), Tansel (1999).
2. The current paper focuses on the informal employment using household leveldata and does not look at the informal economic activities at macro-level.
3. For similar findings and more discussions see the OECD (2008).
4. The graph relates quantiles of log hourly formal wage on the vertical axis to quantiles of log hourly informal wage on the horizontal axis. A point on the symmetry line indicates that quantile of one distribution has the same value as corresponding quantile of the other distribution.

5.The decomposition can also be made with the counterfactual $\{\tilde{y}_i^{cf} = \tilde{X}_i^{form}\hat{\beta}^{Inf}\}_{i=1}^{m}$ which is a counterfactual earnings density that would have prevailed if informal workers were given formal workers' labour market characteristics, but still receive the returns of informal workers to those characteristics.
6. Bootstrap estimates are based on 800 replications.
7. 77% of the informal employees are males and this might affect the results.

**References**


Arabsheibani, G. R. and Henley, A. (2009)."How are informal sector workers disadvantaged? Quantile regression evidence from Brazil".Unpublished manuscript.

Badaoui, E., Strobl, E., and Walsh, F. (2007)."Is there an informal employment wage penalty? Evidence from South Africa".*IZA Discussion Paper* N. 3151.

Bernabe, S. (2002)."Informal employment in countries in transition: A conceptual framework". *CASEPaper* N.56.

Braithwaite, J. (1994). "From second economy to informal sector: The Russian labour market in transition".*ESP Discussion Paper* N. 58.

Buchinsky, M. (1998). "Recent advances in quantile regression models: A practical guide for empirical research". *Journal of Human Resources*, 33(1): 88-126.

Buchinsky, M. (2001), "Quantile regression with sample selection: Estimating women's return to education in the US".*Empirical Economics*, 26:87-113.

Carneiro, F., and Henley, A. (2001)."Modelling formal vs. informal employment and earnings: micro econometrics evidence for Brazil". Annals of the XXIX National Meeting of Economics of ANPEC.

Daza, J. (2005). "Informal economy, undeclared work and labour administration". *DIALOGUE Paper* N 9, Interantional Labour Office-Geneva.

Desai, P., and Idson, T. (1998)."Wage arrears, poverty and family survival strategies in Russia".Mimeo, Columbia University.

Duval-Hernandez, R. (2006)."Informality, segmentation and earnings in urban Mexico".Unpubished paper, Center for US-Mexican Studies, University of California, San Diego.

Fields, G. (1975). "Rural- urban migration, urban unemployment and underemployment, and job search activity in LDC's".*Journal of Development Economics*, 2: 165-187.

Funkhouser, E. (1997). "Mobility and labour market segmentation: The urban labour market in El Salvador". *Economic Development and Cultural Change*, 46(1): 123-153.

Gasparini, L. and L. Tornarolli (2007). "Labour Informality in Latin America and the Caribbean: Patterns and trends from household survey microdata". WorkingPaper No.0046, Facultad de Ciencias Económicas Universidad Nacional de La Plata (CEDLAS), Argentina.

Heckman, J., and Hotz, V. (1986)."An investigation of labour market earnings of Panamanian males".*Journal of Human Resources*21: 507-542.

Henley, A., Arabsheibani, G., and Carneiro, F. (2009)."On defining and measuring the informal sector: Evidence from Brazil". *World Development*, 37(5): 992-1003.

International Conference of Labour Statisticians ILO (2003). In report of the 17th Conference, Geneva: ILO.

Johnson, S., Kaufmann, D., and Shleifer, A. (1997). "The unofficial economy in transition", *Brookings Papers of Economic Activity*: 159-221.

Koenker, R., and Bassett, J. (1978)."Regression quantiles".*Econometrica*, *46*(1): 33-50.

Kolev, A. (1998). "Labour supply in the informal economy in Russia during transition". *CEPR Discussion Paper N* 2024. London: CEPR.

Kuddusov, D. (2009). "The survey of the labour market in Tajikistan".Working version of the document. Dushanbe: 86.

Lacko, M. (2000)."Hidden economy-an unknown quantity?"*Economics of Transition*, 8(1): 117-149.

Machado, J., and Mata, J. (2005)."Counterfactual decomposition of changes in wage distribution using quantile regression".*Journal of Applied Econometrics*, 20(4): 445-465.

Maloney, W. (1999)."Does informality imply segmentation in urban labour markets? Evidence from sectoral transition in Mexico".*World Bank Economic Review*, 13(2): 275-302.

Marcouiller, D., Ruiz de Castilla, V., and Woodruff, C. (1997)."Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru". *Economic Development and Cultural Change*, 45(2): 367-392.

Mazumdar, D. (1975). "The theory of urban employment in less developed countries".*World Development*, 4: 655-679.

Mazumdar, D. (1981). "The urban labour market income distribution: a study of Malaysia".Oxford University press, Oxford.

Mazumdar, D. (1983). "Segmented labour markets in LDCs".*American Economic Review*: Papers and Proceedings,73(2): 254-259.

Merrick, T. (1976)."Employment and earnings in the informal sector in Brazil: The case of Belo Horizonte". *Journal of Developing Areas*, 10(3): 337-353.

National Human Development Report (2009). "Employment in the context of human development".

National Human Development Report.(2007). "Informal economy in Tajikistan".

OECD (2007)."Development centre studies: Informal institutions: How social norms help or hinder development".

OECD (2008).OECD employment outlook."Declaring work or staying underground: Informal employment in seven OECD countries".

Pisani, M.J. and Pagán, J.A. (2004). "Sectoral selection and informality: A Nicaraguan case study".*Review of Development Economics*, 8(4): 541-556.

Portes, A., Castells, M., and Benton, L. (1989)."The informal economy: Studies in advanced and less developed countries". Baltimore: John Hopkins University Press.

Pradhan, M., and Van Soest, A. (1995)."Formal and informal sector employment in urban areas of Bolivia".*Labour Economics*, 2: 275-297.

Pradhan, M., and Van Soest, A. (1997)."Household labour supply in urban areas of Bolivia".*Review of Economics and Statistics*, 79(2): 300-310.

Rose, R., and McAllister, I. (1996)."Is money the measure of welfare in Russia?"*Review of Income and Wealth*, 42: 75-90.

Tannuri-Pianto, M., and Pianto, D. (2002)."Informal employment in Brazil - a choice at the top and segmentation at the bottom: a quantile regression approach". *Working Paper* N 236, Department of Economics, University of Brasilia.

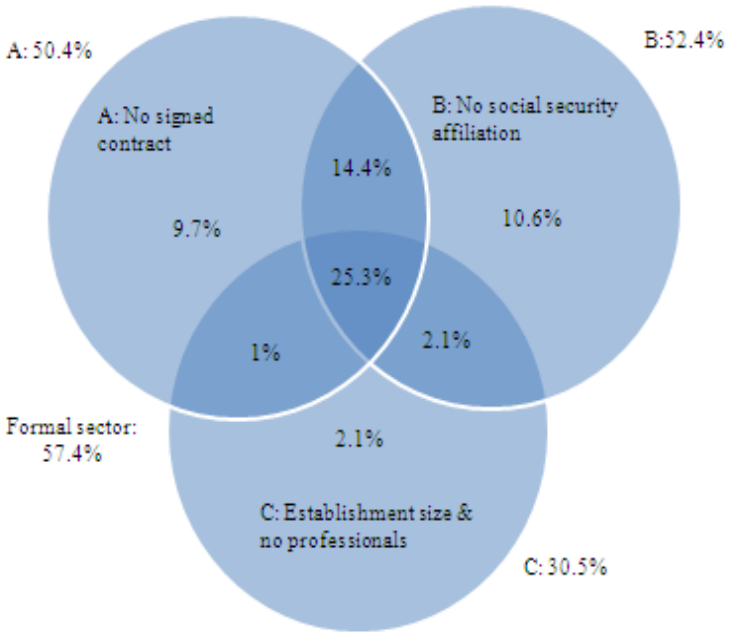
Tansel, A. (1999). "Formal versus informal sector choice of wage earners and their wages in Turkey".*Economic Research Forum Working Paper* N. 9927.

Tokman, V. (1982)."Unequal development and the absorption of labour: Latin America 1950-1980". *CEPAL Review*17:121-133.

Wallenborn, M. (2009)."Skills development for poverty reduction (SDRP): The case of Tajikistan". *International Journal of Educational Development,* 29(6): 550-557.

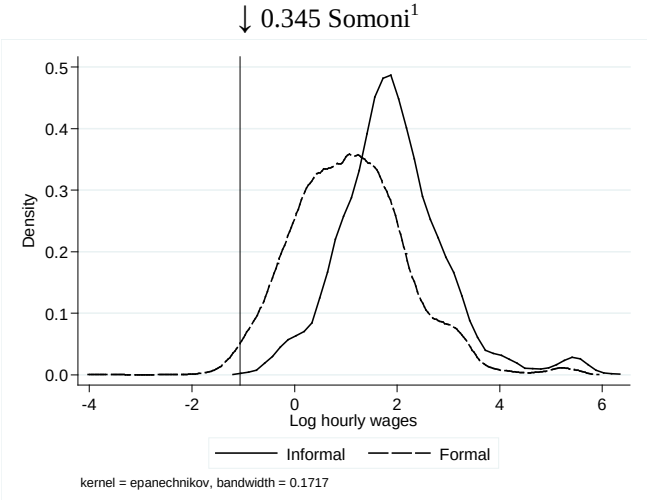
Wu, Y., and Li, T. (2006)."Employment situation and trends in China". In Cai, F. and Gu, B. (Ed.), *Green Book of Population and Labour*. Social Sciences Academic Press: 20-43.

**Figure 1**Coincidence of alternative definitions of informality, 2007 Tajikistan



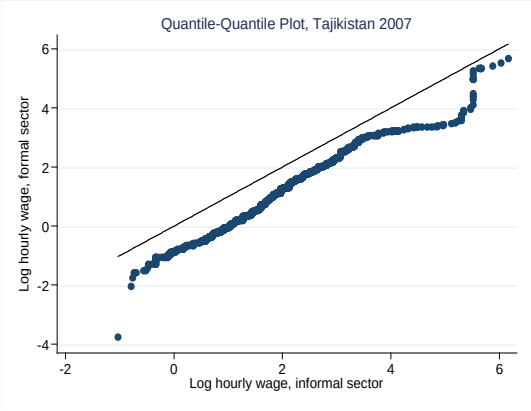

*Source:* Tajikistan 2007 LSMS

**Figure 2** Kernel density estimates of log hourly wage in formal and informal sector



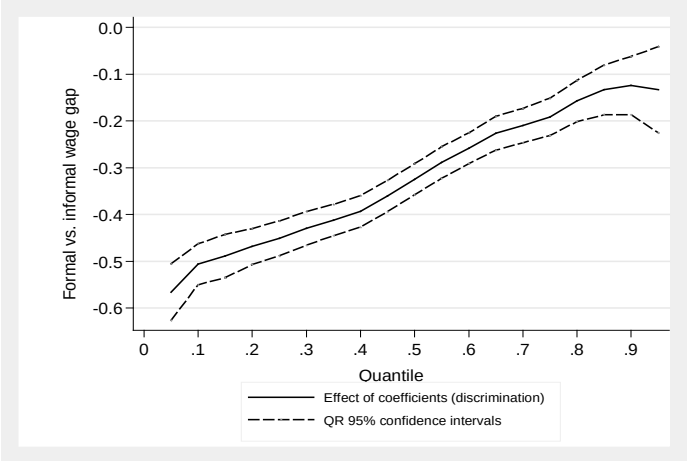

*Source:* Tajikistan 2007 LSMS.
[1]2007 Minimum salary per hour (0.345 Tajik Somoni= Log value of -1.0635).
The somoni (Tajik: сомонӣ) is the currency of Tajikistan.

**Figure 3**Comparing formal and informal wage distributions in Tajikistan, Q-Q plot




*Source*: Tajikistan 2007 LSMS.

**Figure 4**Formal-informal sector wage differential, 2007 Tajikistan




*Source*: Tajikistan 2007 LSMS.
*Notes*: Machado and Mata decomposition estimates.

**Table 1** Proportion of informal employees by main employment status, 2007 Tajikistan

| ***Economically active*** | **All** | **Employees** | **Self-employed** | **Domestic workers** |
|---|---|---|---|---|
| **A.** No signed labour contracts | 50.4% | 39.3% | 93.7% | 94.1% |
| **B.** No social security affiliation | 52.4% | 41.5% | 94.4% | 96.2% |
| **C**. Establishment size and no professionals | 30.5% | 19.3% | 70.2% | 85.3% |
| Pure informal workers | 25.3% | 13.9% | 65.9% | 78.6% |
| Informal at least in one of the three approaches | 65.4% | 56.9% | 98.1% | 99.7% |
| Employees in the sample | 80.1% | - | - | - |
| Self-employees in the sample | 15.3% | - | - | - |
| Domestic workers in the sample | 4.6% | - | - | - |

*Source*: Authors' calculations from Tajikistan 2007 LSMS.
*Notes*: The proportions are based on a sample of workers that report wages.

**Table 2** Proportion of informal employees by gender, 2007 Tajikistan

| ***Economically active*** | **Males** | **Females** | ***t*-test** |
|---|---|---|---|
| **A**. No signed labour contracts | 54.6% | 43.8% | 9.51 |
| **B**. No social security affiliation | 59.4% | 41.2% | 16.25 |
| **C**. Establishment size and no professionals | 38.2% | 18.3% | 20.42 |
| Pure informal workers | 31.9% | 14.9% | 18.48 |
| Informal at least in one of three approaches | 70.8% | 56.7% | 12.91 |

*Source*: Authors' calculations from Tajikistan 2007 LSMS.
*Notes*: The proportions are based on a sample of workers that report wages; *t*-test for difference between women and men.

**Table 3**Decomposition of changes in parameters of the distribution, 2007 Tajikistan

| Percentiles | Raw gap | Coefficients | Covariates | Residuals |
|---|---|---|---|---|
| $10^{th}$ | -1.050 | -0.506 [48%] | -0.541 [52%] | -0.003 [0%] |
| | | (0.022) | (0.022) | |
| $20^{th}$ | -1.059 | -0.469 [44%] | -0.548 [52%] | -0.042 [4%] |
| | | (0.019) | (0.020) | |
| $30^{th}$ | -1.022 | -0.430 [42%] | -0.549 [54%] | -0.043 [4%] |
| | | (0.018) | (0.019) | |
| $40^{th}$ | -0.916 | -0.393 [43%] | -0.502 [55%] | -0.021 [2%] |
| | | (0.017) | (0.020) | |
| $50^{th}$ | -0.783 | -0.325 [42%] | -0.467 [60%] | 0.009 [-1%] |
| | | (0.016) | (0.020) | |
| $60^{th}$ | -0.758 | -0.258 [34%] | -0.457 [60%] | -0.043 [6%] |
| | | (0.016) | (0.019) | |
| $70^{th}$ | -0.665 | -0.210 [32%] | -0.459 [69%] | 0.004 [-1%] |
| | | (0.018) | (0.019) | |
| $80^{th}$ | -0.631 | -0.157 [25%] | -0.485 [77%] | 0.011 [-2%] |
| | | (0.022) | (0.023) | |
| $90^{th}$ | -0.652 | -0.133 [20%] | -0.499 [76%] | -0.020 [3%] |
| | | (0.027) | (0.026) | |

*Source*: Tajikistan 2007 LSMS.
*Notes*: (i)Bootstrapped standard errors in parenthesis; (ii) The differential is calculated by every $5^{th}$ percentile. The main percentile levels are presented. Full decomposition results are available on request.

# Appendix

**Table A.1** Summary statistics, 2007 Tajikistan

| | | *Formal* | | *Informal* | | |
|---|---|---|---|---|---|---|
| *Variables* | | Mean | Std. Dev. | Mean | Std. Dev. | t-stat |
| lhwage | Log of hourly wage | 1.091 | 1.196 | 1.913 | 1.036 | -27.83 |
| school | Years in education | 10.467 | 3.059 | 9.343 | 2.221 | 15.35 |
| age | Age | 36.163 | 12.995 | 34.802 | 12.009 | 4.18 |
| agesq | Age squared | 1476.620 | 1001.952 | 1355.303 | 885.154 | 4.88 |
| tenure1 | 1 if less than 7 months | 0.115 | 0.319 | 0.163 | 0.370 | -5.67 |
| tenure2 | 1 if 1 if 7-12 months | 0.085 | 0.279 | 0.121 | 0.327 | -4.91 |
| tenure3 | 1 if 1-2 years | 0.173 | 0.379 | 0.232 | 0.422 | -5.86 |
| tenure4 | 1 if 3-5 years | 0.223 | 0.416 | 0.247 | 0.431 | -2.20 |
| tenure5 | 1 if 6-10 years | 0.119 | 0.323 | 0.114 | 0.318 | 0.61 |
| married | 1 if married | 0.619 | 0.486 | 0.676 | 0.468 | -4.59 |
| ethnicity | 1 if Tajik ethnicity | 0.738 | 0.440 | 0.776 | 0.417 | -3.50 |
| female | 1 if Female | 0.440 | 0.496 | 0.228 | 0.420 | 17.42 |
| private | 1 if in Private sector | 0.436 | 0.496 | 0.653 | 0.476 | -17.32 |
| occupat1 | 1 if Administrative | 0.018 | 0.133 | 0.001 | 0.031 | 5.78 |
| occupat2 | 1 if Skilled &Trade | 0.061 | 0.239 | 0.311 | 0.463 | -31.54 |
| occupat3 | 1 if Service | 0.310 | 0.462 | 0.110 | 0.313 | 18.25 |
| occupat4 | 1 if Sales | 0.116 | 0.320 | 0.306 | 0.461 | -20.70 |
| occupat5 | 1 if Machines &operators | 0.042 | 0.200 | 0.121 | 0.327 | -13.10 |
| occupat6 | 1 if Elementary | 0.098 | 0.298 | 0.150 | 0.357 | -6.45 |
| rural | 1 if Rural | 0.643 | 0.497 | 0.558 | 0.497 | 6.88 |
| region1 | 1 if Dushanbe region | 0.188 | 0.391 | 0.233 | 0.423 | -4.46 |
| region2 | 1 if Sogd region | 0.165 | 0.371 | 0.272 | 0.445 | -10.71 |
| region3 | 1 if Khatlon region | 0.389 | 0.488 | 0.204 | 0.403 | 15.51 |
| region4 | 1 if Rrp region | 0.158 | 0.365 | 0.239 | 0.427 | -8.33 |
| region5 | 1 if Gbao region | 0.100 | 0.301 | 0.052 | 0.222 | 6.71 |
| *N* | | *6065* | | *2058* | | |

*Notes:* Tajikistan 2007 LSMS.