
MENDELU Working Papers
in Business and Economics

48/2014

**Formal and Informal Sector Wage Differences in
Transition Economies: Evidence from Tajikistan**

Peter Huber, Ulugbek Rahimov

MENDELU Working Papers in Business and Economics

Research Centre

Faculty of Business and Economics

Mendel University in Brno

Zemědělská 1, 613 00 Brno

Czech Republic

<http://vyzc.pef.mendelu.cz/en>

+420 545 132 605

Citation

Huber, P., Rahimov, U. (2014): Formal and Informal Sector Wage Differences in Transition Economies: Evidence from Tajikistan. *MENDELU Working Papers in Business and Economics* 48/2014. Mendel University in Brno. Cited from: <http://ideas.repec.org/s/men/wpaper.html>

Abstract

Peter Huber, Ulugbek Rahimov: **Formal and Informal Sector Wage Differences in Transition Economies: Evidence from Tajikistan**

Analyzing the self-selection of workers into formal and informal sector employment in Tajikistan, a poor transition economy, with higher informal sector than formal sector wages and an informal sector employment share exceeding 50 percent, we find that the selection of formal and informal sector workers is based on comparative advantages rather than labor market segmentation. Furthermore, labor supply to the two sectors reacts rather elastically to relative wages. Policies increasing relative wages in the formal sector could thus be effective in reducing the high informal sector employment share in this country.

Key words

formal/informal sector wages, self selection, segmentation

JEL: J31, J42, J21

Contacts

Peter Huber, Austrian Institute for Economic Research, Arsenal Objekt 20, A-1030 Vienna, Austria and Mendel University Brno - Faculty of Economics and Business Administration. e-mail: Peter.Huber@wifo.ac.at.

Ulugbek Rahimov, Westminster International University in Tashkent, Faculty of Economics, e-mail: urahimov@wiut.uz

Introduction

The literature on labor markets in developing countries often characterizes these as consisting of two sectors: a formal sector which offers relatively high earnings and secure employment and an informal sector with low earnings and insecure employment (see Maloney 2004, Loayza and Rigolini 2006, Amaral and Quintin 2006, Bennett and Estrin 2007 and Bennett 2008 for theoretical work and World Bank 2007 and Battini et al. 2010 for survey). Furthermore two competing hypotheses seek to explain why individual may be employed in the informal employment. The segmentation hypothesis assumes that individuals escape involuntary unemployment and opt for the informal sector. By contrast, the comparative advantage hypothesis claims that informal employment is the result of voluntary income or utility maximization. One important question with respect to the workings of informal labor markets is therefore whether individuals choose to participate in the informal sector voluntarily or if they are forced into informal employment because of exclusion from the formal sector. Given the policy relevance of this question quite a few empirical contributions test for the comparative advantage hypothesis. For instance Magnac (1991) tests for this by estimating a generalized Roy model. In complementary papers, Gindling (1991) uses regressions with sample selection and Pratap and Quintin (2006) use a semi-parametric approach to test for the labor market segmentation hypothesis. The results of these papers support the competitive labor market hypothesis and suggest that individuals self-select into different sectors according to comparative advantages.

A further implication of the segmentation hypothesis is that individuals are poor as a result of informal sector employment (e.g., Harris and Todaro, 1970). In contrast to this, the comparative advantage hypothesis suggests that individuals can only obtain informal employment because they are poorly endowed with characteristics necessary for high returns employment in the formal sector (e.g., Maloney, 2004). Another important question with respect to informal sector employment in developing countries is therefore, whether it is the most able or the least able that work in the informal sector and how changes in relative remuneration in the formal and informal sector affect labor supply to these two sectors. With respect to this question recent studies find that formal-informal sector wage differentials are lower among high income groups than among low income groups (e.g. Bargain and Kwenda, 2009, Nguyen et al. 2013, Tansel and Kan 2012). This is usually interpreted as evidence that informal labor markets consist of an upper tier and a lower tier part, with the upper tier part representing the competitive part of the informal sector and the lower tier consisting of persons rationed out of the formal sector (see: Cunningham and Maloney 2001, Fields 2005, Günther and Launov 2012).

Only few of these contributions have, however, focused on transition countries. This is somewhat of a shortcoming given the rather different role of the informal sector in post-Soviet economies, where the informal sector is much more closely linked to the market economy than in many developing countries (Benarbe, 2002). Two exceptions are Arabsheibani and Staneva (2012) on Tajikistan and Nguyen et al. (2013) on Vietnam. These contributions primarily focus on measuring and decomposing formal – informal sector wage differentials with Arabsheibani and Staneva (2012) - in contrast to the typical findings for most developing countries - finding higher informal sector wages in all parts of the wage distribution in Tajikistan.

In this paper we explore this atypical finding in more detail. We use Tajikistan as an example of a poor transition economy in which the informal sector – aside from offering higher wages also accounts for more than half of the employment according to World Bank (2007) estimates, to analyze whether the self-selection of workers into the formal and informal sector accords with the competitive advantage or segmentation hypothesis. Furthermore, we ask to what degree differences in returns to education and other productive characteristics of workers contribute to explaining wage differentials between these sectors. Finally, we assess how changes in the remuneration of the formal and informal sector affect labor supply in these sectors. To achieve these goals we extend on previous analysis by estimating a generalized three state Roy-model of the self-selection of workers into formal and informal sector employment as well as into non-employment by a Heckman-type two-step estimator based on multinomial choice suggested by Bourguignon et al. (2007). This aside from providing structural estimates of wage determination and self-selection of workers in these two sectors also allows us to determine the contribution of observed and unobserved variables to total formal-informal sector wage difference.

Our results suggest that even in this untypical country stylized facts accord with the self-selection on comparative advantages rather than labor market segmentation hypothesis. The self-selection of workers on unobservable characteristics increases wages in both sectors. Furthermore Oaxaca-Blinder decompositions of the formal-informal sector wage differential suggest that self-selection on such unobservable characteristics is the main reason for higher wages in the informal than the formal sector. In addition we find a number of differences between formal and informal sector wage setting in Tajikistan: Returns to education as well as to experience are higher in the formal than in the informal sector and wage penalties for women and non-Tajik workers are higher in the informal than the formal sector. Furthermore, young, less educated, male Tajik workers self-select into the informal sector. Finally we find that labor supply in the two sectors reacts rather elastically to relative wage changes. Increases in formal sector wages by 1 percent reduce the probability of working in the

informal sector by 0.03 to 0.26 percentage points, but increase the probability to work in the formal sector by 0.41 to 0.69 percentage points. Increases in informal sector wages by 1 percent, by contrast, increase the probability of informal sector employment by around 0.25 percentage points but reduce the probability of working in the formal sector by around 0.31 percentage points. We argue that these results suggest that policies that increase relative wages in the formal sector would in all likelihood also be effective in reducing the high informal sector employment share in Tajikistan.

1 The Model

As a theoretical framework for our analysis we consider a model in which workers (indexed by j) can choose to be in one of three states. These are non-employment, informal and formal sector employment (indexed by $k \in \{n, i, f\}$). Each of these states is associated with a payoff y_{jk} consisting of wages (w_{jk}) plus non-pecuniary benefits (Π_{jk}) from being in this state. We assume that wages in sector k have mean ω_{jk} and random component u_{jk} , which is known to the individual but unobserved by the researcher, with variance σ_k and covariance $\sigma_{kk'}$. Non-pecuniary benefits, by contrast, have mean π_{jk} and random component ζ_{jk} (also known to the individual but unobserved by the researcher) with variance s_k and covariance $s_{kk'}$. Furthermore, we assume that ω_{jk} can be written as $\omega_{jk} = X_j \beta_k$ with X_j a set of observable productive characteristics of individuals determining wages and β_k a vector of state specific coefficients measuring the returns to these characteristics in sector k , while no wages are observed if the individual is not employed and that $\pi_{jk} = Z_j \gamma_k$, with Z_j a set characteristics determining the non-pecuniary benefits received in a particular sector and γ_k a set of parameters.

In this setup workers will select into the state in which they can expect to obtain the highest payoff. Therefore, omitting the subscripts for individual workers to save on notation, and denoting $u_k + \zeta_k = \eta_k$, a worker will be employed in the formal sector if $\omega_f + \pi_f > \eta_n - \eta_f$ and $\omega_f + \pi_f - \omega_i - \pi_i > \eta_i - \eta_f$, but will be employed in the informal sector if $\omega_f + \pi_f - \omega_i - \pi_i < \eta_i - \eta_f$ and $\omega_i + \pi_i > \eta_n - \eta_i$ and will not be employed if $\omega_f + \pi_f < \eta_n - \eta_f$ and $\omega_i + \pi_i < \eta_n - \eta_i$.

Since wages in sector k are given by $w_k = X\beta_k + u_k$ while the self-selection of workers into the individual sectors is driven by the unobserved variable y_k , determined by:

$$y_k = X\beta_k + Z\gamma_k + \eta_{jk} \quad (1)$$

average wages in sector k will only be observed if $y_k > \max_{k' \neq k} y_{k'}$ and are given by

$$E(w_k | y_k > \max_{k' \neq k} y_{k'}) = X\beta_k + E(u_k | y_k > \max_{k' \neq k} y_{k'}) \quad (2)$$

2 Estimation

In this theoretical set up therefore estimating sector wage equations by ordinary least squares will suffer from selection bias which arises from the choice between different states. As shown by Lee (1983) models such as those formulated in equations (1) and (2) can be estimated by a Heckman-type two-step selection model based on multinomial choice. In this in the first step the “reduced form” multinomial logit selection equation (1) is estimated. In our case in this regression the dependent variable takes on a value of zero if the person under consideration is not employed, one if the person is employed in the informal sector and two if the person is employed in the formal sector. This step allows for an analysis of selection into different sectors based on observables and is also used to construct predicted probabilities of the choice of a particular alternative.

In the second step a linear regression (as in equation 2) on the wage level for each of the sectors including a selection correction for the choice derived from the first step estimates is estimated by OLS with log observed wages (either in the formal or informal sector) as the dependent variable. This step results in set of unbiased estimates of the parameter β_k and an estimate of the signs of the correlations driving selectivity in equation (2). Furthermore, as demonstrated by Lee (1979) based on the estimates for equation (2) unbiased predictions of the unobserved expected wages of individuals in other sectors than those in which they are employed can be derived. These can then be used to estimate a structural (multi-nomial) selection equation which measures the impact of wages in different sectors on the choice of sector of employment.

Since Lee’s (1983) article a number of other authors have provided alternative methods based on the same idea but differing in the details for correcting for selection bias (e.g. Dubin and McFadden 1984, Dahl 2002) to estimate equation (2). In a recent Monte Carlo study Bourguignon et al. (2007) show that a slightly modified version of Dubin and McFadden’s (1984) method leads to the most reliable results. In this method - under the assumption that u_k is linearly associated to n_k (which is obviously the case in the model above since $u_k = \eta_k - \zeta_k$) and that the η_k are independently and identically Gumbel distributed, equation (3) can be estimated by:

$$w_k = X\beta_k + \sigma_k \rho_k m(P_k) + \sigma_k \sum_{k' \neq k} \rho_{k'} m(P_{k'}) \frac{P_{k'}}{P_{k'} - 1} + v_k \quad (3)$$

with $v_k = \log(P_k) + \eta_{jk}$, P_k the probabilities of choosing state k and $m(P_k)$ a complicated function of these probabilities, which, however, can be calculated from the first stage estimates.

Furthermore as noticed by Dimova and Gang (2007) the estimated coefficient on the selection terms ($\sigma_k \rho_{k'}$), can be interpreted as an indication of both the direction of the selection bias in estimating equation (3) as well as an indication of how this bias is impacted on by the choice of individuals between any two sectors. The estimates of the coefficients ($\tau_{kk'} = \sigma_k \rho_{k'}$) on the selection term ($m(P_{k'}) \frac{P_{k'}}{P_{k'}-1}$) in equation (3) can thus be used to assess whether self-selection is due to comparative advantages: For instance a positive coefficient on the selection term on non-employment in either the informal or formal sector (i.e. $\tau_{fn} > 0$ or $\tau_{in} > 0$) wage equation would suggest that average wages in the formal (informal) sector are increased on account of self-selection of more able workers from non-employment to the formal (informal) sector. Similarly with respect to the self-selection between sectors four possible situations can arise: The first one applies if $\tau_{f,i} > 0$ and $\tau_{i,f} > 0$. In this case average informal sector wages are increased due to workers with high informal sector skills self-selecting into the informal sector and formal sector wages are increased due to workers with high formal sector skills self-selecting into the formal sector. In this situation therefore workers self-select into sectors according to unobserved comparative advantages. In the second case $\tau_{f,i} > 0$ and $\tau_{i,f} < 0$ formal sector wages are increased due to workers with high formal sector skills self-selecting into the formal sector but informal sector wages are reduced due to workers with low informal sector skills self-selecting into the informal sector, while in the third case ($\tau_{f,i} < 0$ and $\tau_{i,f} > 0$) the opposite applies. In these cases, therefore, either workers in the formal or informal sector select to this sector based on comparative advantages while for workers in the other sector the labor market is segmented. Finally, a fourth possibility is that $\tau_{f,i} < 0$ and $\tau_{i,f} < 0$. In this case both formal and informal sector wages are reduced because of the self-selection of workers to the respective sector. This would, however, imply a misallocation of resources because wages are reduced in both sectors on account of the self-selection of workers between sectors.

3. Data

We use data from the Tajikistan Standards of Living Survey (SLS) of 2007 to estimate equations (1) and (3). This questionnaire is one of many SLS that have become standard data sets for labor market analysis in developing countries (Nguyen et al. 2013) and contains a representative sample of about 1500 Tajik households. This data has been previously used in several papers on the Tajik labor market (see: Justino and Shemyakina, 2012; Arabsheibani and Staneva 2012 and Abdulloev et al. 2011). We use the individual level data of this survey and restrict attention to the main job held by working age

(15 to 60 years old) persons in the last 14 days.¹ Setting these restrictions and omitting persons with missing variables for one of the explanatory variables we obtain 7750 usable observations of which 5904 are employed either in the formal or the informal sector and 1646 are not employed.

3.1 Dependent Variables

We follow the literature (Marcoulier et al. 1997, Henley et al. 2006, Nguyen et al. 2013) and use two definitions of informal sector employment. In the first (the social security definition) a person is considered to be formally employed if the respondent stated that he or she was affiliated to the Tajik social security scheme in their main job. In the second (the contract definition) a person is encoded as employed in the formal sector if respondents signed a contract or written agreement with their employer and informally employed if no such agreement existed.² Finally, in both instances persons are considered not employed if they do not have a primary job which generates income.

The top panel of table 1 shows the distribution of non-employed, informal and formal sector workers according to these two definitions by different demographic groups. The overall share of non-employed 15 to 60 year olds is almost 22 percent. These levels are markedly higher for females and young as well as less qualified persons in both definitions of formal employment. Also for both definitions of formality the share informal sector employment exceeds that of formal sector employment in aggregate as well as for most demographic groups. This accords with the results of Arabsheibani and Staneva (2012). The only exceptions (in both definitions) are females, persons aged 45 or older and persons with tertiary education. For the youngest and persons with basic education (when considering the social security definition) or persons with primary education (when considering the contract definition) the share of informally employed exceeds that of the formally employed by most. Furthermore, the share of informal sector employment is somewhat higher when considering the social security definition than when considering the contract definition both in aggregate and for almost all demographic groups.

{Table 1: Around Here}

For the second stage equation, the dependent variables are (log) hourly wages in the formal and informal sector. These are calculated from two questions in the questionnaire in which respondents were first asked on their wages in their main job in the previous month and weekly working hours in

¹ This selection ensures omission of a large number of employment relationships with very low income and is also dictated by data since the Tajik SLS only collects information on wages, hours worked and social security entitlement as well as contractual basis in the main job held by the respondent.

² Arabsheibani and Staneva (2012) and others also define informal sector employees as persons working in firms with less than 5 employees. We do not use this definition as Henley et al. (2006) in their review suggest that this definition should only be used if others are unavailable.

the main job.³ The bottom panel of table 1 shows the average formal and informal sector log hourly wages in Tajikistan together with a t-test for difference in means between sectors. According to these tests informal sector wages were significantly higher than formal sector wages in almost all segments of the Tajik labor market in both definitions of formality. The only exceptions are female and very young workers, when considering the contract definition of the formal sector. Furthermore, wage differentials between the formal and informal sector are slightly smaller in the contract definition of the formal sector than in the social security definition.

3.2 Explanatory variables

As explanatory variables in the second step, wage level equation we include the variables usually included in a Mincerian wage regression - (the log of) age and age squared and controls for human capital. The later are dummy variables for the highest completed education of the individual (primary, basic, secondary and tertiary education) and a dummy variable measuring whether the respondent can read. In addition, we also control for whether the person is not Tajik, female or lives in the capital city of Dunshabe through dummy variables. These variables are included because literature (Karnite, 2010) suggests that non-Tajik workers earn lower wages than Tajiks and to control for potential gender differences in wages as well as for the high urban-rural wage differentials found in many studies (for instance Brühlhart and Koenig, 2006) on transition countries. Furthermore, in the first stage regressions, as our *Z* variables, we include a dummy variable for married persons, a dummy for children under the age of six living in the household, and a set of dummy variables for the self-reported health status of respondents (which may be very good, good or fair or bad) because we expect these variables to influence non-pecuniary benefits of (and capability to) work in a sector without them having a (direct) impact on wages levels.⁴

{Table 2: around here}

The descriptive statistics for these variables (in Table 2) suggest that on average the non-employed are significantly younger and more often female and have more children than those in formal employment. They are, however, also more often Tajik and also more often reside in Dunshabe, but are married less often than those working in the formal sector. Individuals not employed are also significantly less well educated than the employed in the formal sector and have a worse health status (that is a statistically significantly larger share of persons with a bad or fair health status).

³ We also conducted the same analysis as below on an additional wage measure in which includes proxies for hourly in-kind and bonus payments as a robustness check. Since this analysis led to very similar results to those presented below, we do not report them here, but make them available from the authors upon request.

⁴ This assumption was also tested by including these variables directly in the wage level equations, where they remained insignificant.

Those working in the informal sector, by contrast, are usually significantly younger, more often Tajik but less often female, less often reside in capital cities, more often have good health and also can read more often than those in the formal sector. They also more often have children under the age of six. Furthermore, these differences are often larger when considering the social security definition of the formal sector than when considering the contract definition.

4 Results

4.1 Results for the selection regressions

Table 3 presents the marginal effects for the first stage (reduced form) multinomial logit regressions for the selection of workers into one of the three labor market states (non-employment, informal sector and formal sector employment) for the two definitions of the formal sector used. These results suggest that older workers have a higher probability of formal sector employment, while younger ones are both more likely to be employed in the informal sector and to not be employed. A one percent age increase increases the probability of formal sector employment by around 0.16 percentage points, but reduces the probability of non-employment by 0.06 percentage points and the probability of informal sector employment by between 0.09 to 0.10 percentage points. Furthermore, the non-employment probability reduces with education while informal sector workers primarily come from the lower and medium educated population, in particular when the social security definition of the informal sector is considered. Workers with a basic or a secondary education have an around 0.07 to 0.08 percentage points higher probability of informal sector employment than persons with only compulsory education and persons with a tertiary education have a 0.15 percentage points lower probability of informal sector employment. When considering the contract definition, by contrast, coefficients for workers with basic and secondary education remain insignificant. This suggests that these workers have a similar probability to work in the informal sector as those with compulsory education: Tertiary educated workers, however, have a 0.17 percentage points lower probability of informal sector employment. The probability of formal sector employment, by contrast, increases with educational attainment. This applies in particular to tertiary educated workers. Their probability of working in the formal sector is (all else equal) over 0.30 percentage points higher than of workers with compulsory education in both definitions of the formal sector. After controlling for education, however, being able to read significantly influences the selection into formal and informal sector employment as well as non-employment only if the social security definition of the formal sector is considered. In this case persons, who can read, are 0.10 percentage points less likely to work in the formal sector but 0.10 respectively 0.01 percent more likely to work in the informal sector or to not be employed at all than people who cannot read.

{Table 3: Around Here}

There are some indications of selection by gender, marital status and ethnicity. Females have a significantly higher probability of formal sector employment or not to work at all. Their probability to work in the informal sector is (depending on the definition of the informal sector) between 0.12 to 0.19 percentage points lower than that of the males. Similarly, non-Tajiks are significantly more often employed in the formal sector and less often non-employed in both definitions of the formal sector, and weakly significantly more often employed in the informal sector when considering the social security definition, than the Tajik population. Marriage, by contrast, mostly affects the probability of non-employment and formal sector employment, while its impact on the probability of informal sector employment remains insignificant. Married persons are by 0.05 percentage points more likely to not be employed and by 0.05 percentage points less likely to be formally employed. Furthermore, informal sector employment is also strongly associated with residence in Dunshabe and thus also an urban phenomenon. Persons residing in Dunshabe have an around 0.12 to 0.13 percentage points lower probability of non-employment than persons living outside Dunshabe, while the probability of employment in the formal sector in Dunshabe is either comparable (in the social security definition) or only slightly higher (in the contract definition) than outside Dunshabe. As a consequence residents of Dunshabe are by between 0.09 to 0.10 percentage points more likely to be informally employed.

Finally, health conditions mainly predict the probability of formal and informal sector employment: Persons who state that their health conditions are fair or bad have a (depending on definition) 0.08 to 0.10 percentage points lower probability of formal sector and a 0.07 to 0.09 percentage point higher probability of informal sector employment. In addition, when considering the social security definition of the formal sector, persons stating to have good health have a significantly (0.04 percentage points) lower probability of formal sector employment but a significantly (0.06 percentage points) higher probability of informal sector employment than persons stating to have very good health, while self-reported health conditions have no significant impact on the probability of non-employment.⁵

In sum therefore the results of the first stage multinomial regressions suggest that informal sector workers are selected on age and education (with younger and less educated workers having a higher probability to be employed in the informal sector) but also on gender, ethnicity and health and that informal sector employment in Tajikistan is also more preponderant in Dunshabe.

⁵ One explanation for the better health conditions of the workers in the formal sector is that working conditions in this sector are better than in the informal sector, another one is that the formal sector also comprises a number of branches (e.g. police, military service) in which good health conditions are a precondition for employment.

4.2 Results for the wage regressions

Comparing these results to the (selectivity corrected) wage equation estimates (in table 4), suggests that this self-selection to a large degree accords with the differences in wage setting in the formal and informal sector. For instance, the pattern of self-selection in terms of education accords with the substantially higher returns to medium levels of education in the informal sector (when considering the social security definition of informal work) than in the formal sector and a penalty for tertiary education in the informal sector in both definitions of the informal sector. When considering the social security definition of the formal sector, informal sector workers with a basic education earn 0.5 percent and informal sector workers with a secondary education 0.3 percent more than informal sector workers with only compulsory education. Formal sector workers with basic and secondary education have no significantly different wages than workers with only compulsory education. When considering the contract definition of the informal sector, returns to basic and secondary education are insignificant in both the formal and the informal sector. Tertiary educated informal sector workers, however, earn wages that are substantially (by 0.8 percent to 1.6 percent) lower than those of persons with compulsory education, while tertiary educated formal sector workers get a significant wage premium of between 0.3 percent to 0.6 percent in both definitions of formality. Being able to read, by contrast, only has a positive impact on wages of informal sector workers when considering the social security definition, while it remains an insignificant wage determinant in all other wage regressions.

Similarly, returns to age are in line with the self-selection of older workers into the formal sector. Although in both definitions of the formal sector the significant coefficients on the linear age term are substantially higher in the informal sector than in the formal sector, the equally larger coefficient on the squared term on age in the informal sector imply that marginal returns to age are higher in the informal sector as of an age of 23 when the social security definition is considered and of 26 when the contract definition is considered. The differences in gender and ethnic wage differentials in the formal and informal sector are also consistent with the self-selection of workers into these two sectors. Non-Tajik minorities all else equal earn significantly lower wages than Tajiks in the informal sector irrespectively of how this is defined, while in the formal sector there are no significant differences between these two groups. Women earn significantly less than men both in the formal and informal sector. The female wage penalty, however, is much larger in the informal sector. Location in Dunshabe, by contrast, leads to rather similar wage increases in both the formal and the informal sector according to both definitions of informality, so that here selection of residents in the capital city into the informal sector cannot be so easily justified by higher returns of such a decision in the capital city.

{Table 4: Around Here}

Furthermore the selection terms in the bottom of table 4 (indicated by the variables m_0 to m_3) suggest that first of all both formal and informal sector workers are weakly positively selected on unobservables relative to the non-employed, although this positive selection is much stronger and statistically significant only for informal sector workers. Second of all, these results also suggest that workers also select into the formal and informal sector on the basis of unobserved comparative advantages. The coefficients on the selection terms m_1 and m_2 are both positive in the case of formal sector employment, although this term is insignificant for the m_1 variable when considering the contract definition only. This suggests that wages in the formal sector are increased by the self-selection of workers with a high probability to either work in the formal or informal sector. Similarly, these terms are also positive for the informal sector wage regressions, although the m_2 term remains insignificant. Therefore also informal sector workers' wages are increased by the self-selection of these workers relative to the formal sector, although this effect is rather weak and remains insignificant in the case of selection from the informal to the formal sector.

4.3 Oaxaca decompositions

The results of the wage regression in table 4 can also be further analyzed by using Oaxaca-Blinder decomposition. In these the informal-formal sector wage differential can be decomposed into four components: a difference in parameters effect, a difference in coefficients effect, a selection into work and a selection between sectors effect.⁶ The results of these decompositions (table 5) suggest that the self-selection of workers from non-employment to employment into the formal or the informal sector is the main factor contributing to higher informal sector wages in Tajikistan. In both definitions of the formal sector the selection into work effect contributes significantly to the higher informal sector wages, with the size of this effect exceeding the total log wage differences between the two sectors. Similarly, also the selection between sectors effect contributes positively to these wage differences, while the differences in returns to observed characteristics (differences in parameter effect) works to reduce these wage differences in both definitions of the formal sector. These effects, however, - while sizeable in terms of coefficient size, remain statistically insignificant throughout. The contribution of the differences in characteristics of the workers in the formal and

⁶ These decompositions use the fact that by equation (3) informal-formal sector wage differentials can be written as $\bar{w}_l - \bar{w}_f = [(\bar{X}_l - \bar{X}_f)\hat{\beta}_i] + [(\hat{\beta}_l - \hat{\beta}_f)\bar{X}_f] + [\hat{\gamma}_{0i}\bar{\mu}_{0i} - \hat{\gamma}_{0f}\bar{\mu}_{0f}] + [\hat{\gamma}_{1i}\bar{\mu}_{1i} - \hat{\gamma}_{1f}\bar{\mu}_{1f} + \hat{\gamma}_{2i}\bar{\mu}_{2i} - \hat{\gamma}_{2f}\bar{\mu}_{2f}]$ where bars over variables represent means, hats indicate estimated parameters, the $\bar{\mu}_{kk}$ are the various selectivity terms and the terms in the first, second, third and fourth square brackets are the difference in characteristics, difference in parameters, selection into work and selection between sector effects respectively (see Yun 2005 and Madden, 2000 for details and Jann 2005 for standard errors).

informal sector to total wage differences, by contrast, is both small in terms of absolute size as well as remaining statistically insignificant.

{Table 5: Around Here}

4.4 Results for the structural form multinomial logit regressions

Finally, table 6 reports the results of the structural form multinomial logit regression for choosing a particular sector. In this regression we excluded all variables in the wage equation to avoid collinearity with wages. These structural form regressions suggest a rather strong impact of relative wages on the labor supply in the formal and informal sector. An increase of informal sector wages by 1 percent significantly reduces the probability to be inactive by 0.09 percentage points and the probability to be employed in the formal sector by 0.21 percentage points but increases the probability to be employed in the formal sector by 0.29 percentage points, when considering the social security definition of the informal sector, while such an increase significantly reduces the probability to be inactive (by 0.07 percentage points) and employed in the formal sector (by 0.21 percentage points), but increases the probability to be in informal employment by 0.25 percentage points when considering the contract definition. An increase in the formal sector wage by one percent, by contrast, statistically significantly reduces the probability of inactivity (by 0.38 percentage points) and of informal sector employment (by 0.26 percentage points). It also increases the probability of working in the formal sector by 0.64 percentage points when considering the social security definition of the informal sector, while it statistically significantly reduces the probability of informal sector employment (by 0.13 percentage points) and the probability of non-employment (by 0.28 percentage points) but increases the probability of that formal sector employment (by 0.41 percentage points) in the contract definition of the formal sector.

{Table 6: Around Here}

Furthermore the control variables reconfirm the selection based on health also found in the reduced form regressions. As in the reduced form multinomial logit results the healthiest persons usually work in the formal sector, since persons who state that their health status is good as well as persons stating to have fair or bad health have a lower probability to work in the formal sector than persons stating to have very good health. By contrast, workers with good and fair or bad health conditions have a higher probability to work in the informal sector than those with very good health conditions, while self-reported health conditions – as in the reduced form results - have no significant impact on the probability of non-employment except for when the contract definition of informality is used.

Similarly the results with respect to married persons suggest that these significantly less often work in the informal sector but more often do not work at all when considering both definitions of the formal sector (and also work significantly more often in the formal sector when considering the contract definition of the formal sector). Finally, persons living in households with children under the age of 6 significantly less often work in the formal sector, but are significantly more often employed in the informal sector in both definitions of the formal sector and significantly more often do not work at all than persons living in households with no children under the age of 6 when considering the contract definition of the formal sector.

Conclusions

This paper uses Tajikistan as an example of a poor transition economy, where, informal sector wages are higher than formal ones, and in which the informal sector, according to World Bank (2007) estimates, accounts for more than half of total employment, to analyze whether the self-selection of workers into formal and informal sector employment accords with the competitive advantage or segmentation hypothesis of informal work. Our results suggest that even in this rather untypical country stylized facts accord with self-selection on comparative advantages both with respect to observable and unobservable characteristics of workers rather than labor market segmentation and also indicate a rather strong impact of relative wages in the two sectors on the decision of a worker to work in the formal or informal sector. An increase in formal sector wages by 1 percent reduces the probability of working in the informal sector by 0.13 to 0.26 percentage points but increases the probability to work in the formal sector by 0.41 to 0.69 percentage points. An increase in informal sector wages by 1 percent, by contrast, increases the probability of working in the informal sector by 0.25 to 0.29 percentage points but reduces both the probability of working in the formal sector by around 0.32 percentage points. This suggests that relative wages have an important impact on the decision of a worker to work in the formal or informal sector. On the policy side this implies that policies that increase wages in the formal sector (or reduce those in the informal sector) would in all likelihood also be effective in reducing the extremely high share of the informal sector employment in Tajikistan.

We also ask to what degree differences in returns to education and other productive characteristics of workers between these sectors contribute to explaining wage differential between these two sectors. Although we find a number of differences between formal and informal sector wage setting in Tajikistan, such as that returns to education as well as to experience are higher in the formal than in the informal sector and that wage penalties for women and non-Tajik workers are higher in the informal than the formal sector, Oaxaca-Blinder decompositions of the formal-informal sector wage

differential suggest that the self-selection of workers on unobservables is the main reason for higher wages in the informal than the formal sector. This thus strengthens our conclusions of self-selection based on comparative advantages rather than labor market segmentation. We, however, also find that a large share of the workers in the formal sector could expect to attain higher wages if they moved to the informal sector. This may indicate that – despite selection on the base of comparative advantages – workers still consider non-pecuniary aspects of work in the formal sector more attractive than workers in the informal sector and are thus willing to accept the wage discounts in the formal sector. These results are largely robust to using alternative definitions of the formal sector and also to using different wage definitions.

References

- AMARAL, P., QUINTIN, E. (2006): A competitive model of the informal sector, *Journal of Monetary Economics*, 53(7), pp. 1541–1553.
- ARABSHEIBANI, G.R., STANEVA A. (2012): Is There an Informal Employment Wage Premium? Evidence from Tajikistan, IZA Discussion Paper 6727, Institute for the Study of Labor (IZA).
- ABDULLOEV I., GANG I.N , LONDON-LANE J. (2011): Migration as a Substitute for Informal Activities: Evidence from Tajikistan, IZA Discussion Papers 6236, Institute for the Study of Labor (IZA).
- BATINI, N., KIM, Y.-B. , LEVINE, P., LOTTI, E. (2010): Informal Labour and Credit Markets: A Survey”, IMF Working Paper 10/42.
- BARGAIN, O., KWENDA, P., (2009): The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data, IZA Discussion Papers 4286.
- BENNETT, J. (2008): Formality, informality, and social welfare”, IZA Discussion Paper No.3550.
- BENNETT, J., ESTRIN, S. (2007): Informality as a stepping stone: entrepreneurial entry in a developing economy, IZA Discussion Paper No.2950.
- BERNABÈ, S. (2002): *Informal Employment in Countries in Transition: A Conceptual Framework*, CASE paper 56, London: Centre for Analysis of Social Exclusion.
- BOURGUIGNON, F., FOURNIER, M., GURGAND M. (2007): Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons, *Journal of Economic Surveys*, 21(1):174-205.
- CUNNINGHAM, W., MALONEY, W. (2001): Heterogeneity in the Mexican micro-enterprise sector: an application of factor and cluster analysis, *Economic Development and Cultural Change*, 50(1), pp. 131–156.
- DAHL, G.B. (2002): Mobility and the returns to education: testing a Roy model with multiple markets, *Econometrica*, 70(6):2367-2420
- DICKENS, W., LANG, K. (1985): A test of dual labor market theory, *The American Economic Review*, 67(2):792–805.
- DIMOVA, R., GANG, I.N., (2007): Self-selection and wages during volatile transition, *Journal of Comparative Economics*, 35(3):612-629.
- DUBIN, J.A., MCFADDEN, D.L. (1984): An econometric analysis of residential electric appliance holdings and consumption, *Econometrica*, 52(2):345-362.

- FIELDS, G. (2005): A guide to multisector labor market models, Paper prepared for the World Bank Labor Market Conference, URL: http://www.ilr.cornell.edu/directory/gsf2/downloads/Multisector_mar05.pdf.
- GINDLING, T. (1991): Labor market segmentation and the determination of wages in the public, private-formal and informal sectors in San-Jose, Costa-Rica, *Economic Development and Cultural Change*, 39(3): 585–603.
- GÜNTHER, I., LAUNOV, A. (2012): Informal employment in developing countries: Opportunity or last resort?, *Journal of Development Economics*, 97(1): 88-98.
- HARRIS, J., TODARO, M. (1970): Migration, unemployment and development: a two-sector analysis", *The American Economic Review*, 60(1):126–142.
- HECKMAN, J., SEDLACEK, G. (1985): Heterogeneity, aggregation and market wage functions: an empirical model of self-selection in the labor market, *Journal of Political Economy*, 93(6): 1077–1125.
- HENLEY, A., ARABSHEIBANI, R., CARNEIRO, F.G. (2006): On Defining and Measuring the Informal Sector, IZA Discussion Papers 2473, Institute for the Study of Labor (IZA).
- JANN, B. (2005): Standard Errors for the Blinder-Oaxaca Decomposition:" [http://repec.org=dsug2005=oaxaca se handout.pdf](http://repec.org=dsug2005=oaxaca%20se%20handout.pdf).
- LEE F. (1979): Identification and Estimation of Binary Choice Modes with Limited (Censored) Dependent Variables, *Econometrica*, 47:977-996.
- LEE L-F. (1983): Generalized econometric models with selectivity, *Econometrica*, 51(2):507-512.
- LOAYZA, N., RIGOLINI, J. (2006): Informality trends and cycles, The World Bank Policy Research Working Paper No.4078.
- MADDEN, D. (2000): Towards a Broader Explanation of Male-Female Wage Differences," *Applied Economics Letters*, 7(12):765-70
- MAGNAC, T. (1991): Segmented or competitive labor markets, *Econometrica*, 59(1): 165–187.
- MALONEY, W. (1999): Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico, *The World Bank Economic Review*, 13(2):275–302.
- MALONEY W. (2004): Informality revisited, *World Development*, 32(7):1159–1178.
- MARCOUILLER, D., RUIZ DE CASTILLA, V., WOODRUFF, C. (1997): Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru, *Economic Development and Cultural Change*, 45(2): 367-92.
- NGUYEN, H. C., NORDMAN, C. J., ROUBAUD, F. (2013): Who Suffers the Penalty? A Panel Data Analysis of Earnings Gaps in Vietnam," IZA Discussion Papers 7149, Institute for the Study of Labor (IZA).
- PAULSON, A., TOWNSEND, R. (2005): Financial constraints and entrepreneurship: evidence from the Thai financial crisis, *Economic Perspectives*, 3Q/2005: 34–48.
- PRATAP, S., QUINTIN, E. (2006): Are labor markets segmented in developing countries? A semiparametric approach, *European Economic Review*, 50(7):1817–1841.
- TANSEL, A., KAN, E. O. (2012): The Formal/Informal Employment Earnings Gap: Evidence from Turkey, IZA Discussion Papers 6556, Institute for the Study of Labor (IZA)
- THE WORLD BANK (2007): World Bank, Informality: Exit and Exclusion, The World Bank, Washington D.C (2007).
- YUN, M.-S. (2005): A Simple Solution to the Identification Problem in Detailed Wage Decompositions", *Economic Inquiry*, 43(4):766-72

Table 1: Distribution of formal and informal sector employment as well as non-employment by demographic groups and average log wages in the two sectors according to different definitions of the formal sector

	Definition 1: Social security entitlement			Definition 2: Written contract		
	Not employed	Informally employed	Formally employed	Not employed	Informally employed	Formally employed
Share of observations						
Total	21.80	42.21	35.99	21.80	40.29	37.91
Gender						
Male	18.08	48.92	33.00	18.08	44.20	37.72
Female	27.68	31.61	40.70	27.68	34.11	38.21
Age						
15-24	26.36	44.41	29.23	26.36	43.60	30.04
25-44	21.28	43.90	34.82	21.28	40.71	38.01
45-60	19.18	37.10	43.72	19.18	36.80	44.03
Education						
primary	33.76	35.02	31.22	33.76	41.77	24.47
basic	31.32	42.17	26.51	31.32	41.23	27.45
secondary	22.58	46.10	31.32	22.58	42.48	34.94
tertiary	8.38	28.09	63.53	8.38	30.49	61.13
log wage excluding bonuses and in kind payments						
Total		1.49***	1.00		1.36***	1.16
Gender						
Male		1.77***	1.25		1.67***	1.43
Female		0.80***	0.68		0.73	0.73
Age						
15-24		1.13***	0.63		0.95	0.90
25-44		1.62***	1.07		1.51***	1.22
45-60		1.55***	1.07		1.42***	1.18
Education						
primary		0.82***	0.58		0.74**	0.65
basic		1.16***	0.45		1.00***	0.73
secondary		1.51***	0.87		1.40***	1.07
tertiary		1.92***	1.46		1.71***	1.55

Source: TLSS 2007, Sample 15 to 60 year olds. ***, (**), (*) show significance of a t-test for difference in mean log wages between the formal and informal sector at the 1%, (5%), (10%) level.

Table 2: Descriptive Statistics of dependent variables

	Overall		Not Employed		Social security definition				Contract definition			
	Mean	S.D.	Mean	S.D.	informally employed		formally employed		informally employed		formally employed	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Lnage	3.52	0.35	3.48***	0.36	3.50***	0.34	3.57	0.34	3.50***	0.35	3.57	0.33
Lnagesq	12.51	2.40	12.23***	2.50	12.34***	2.35	12.88	2.36	12.35***	2.40	12.85	2.31
Not Tajik	0.21	0.41	0.16***	0.37	0.21***	0.41	0.23	0.42	0.22	0.42	0.22	0.41
Female	0.39	0.49	0.49***	0.50	0.29***	0.45	0.44	0.50	0.33***	0.47	0.39	0.49
reside in capital	0.14	0.35	0.06***	0.24	0.15***	0.36	0.18	0.38	0.16**	0.36	0.18	0.38
Read	0.98	0.14	0.97	0.16	0.99***	0.12	0.98	0.15	0.98	0.14	0.98	0.12
primary education	0.03	0.17	0.05***	0.22	0.03	0.16	0.03	0.16	0.03***	0.18	0.02	0.14
basic education	0.14	0.35	0.20***	0.40	0.14***	0.35	0.10	0.30	0.14***	0.35	0.10	0.30
secondary education	0.66	0.47	0.69***	0.46	0.72***	0.45	0.58	0.49	0.70***	0.46	0.61	0.49
tertiary education	0.17	0.37	0.06***	0.24	0.11***	0.31	0.29	0.46	0.13***	0.33	0.27	0.44
very good health	0.11	0.31	0.12	0.32	0.10***	0.30	0.12	0.33	0.10**	0.30	0.12	0.32
good health	0.75	0.43	0.73	0.45	0.76***	0.42	0.75	0.44	0.74**	0.44	0.77	0.42
fair or bad health	0.14	0.35	0.15***	0.36	0.14	0.35	0.13	0.34	0.15***	0.36	0.11	0.32
Married	0.74	0.44	0.72**	0.45	0.73*	0.44	0.75	0.43	0.73***	0.45	0.76	0.43
Children under 6	0.53	0.50	0.54***	0.50	0.55***	0.50	0.51	0.50	0.54**	0.50	0.52	0.50
Observations	7550		1646		3187		2717		3042		2862	

Source: TLSS 2007, Sample 15 to 60 year olds. ***, (**), (*) show significance of a t-test for difference in means of the respective subaggregate relative between the formal sector at the 1%, (5%), (10%) level.

Table 3: Results of first step selection equation (marginal effects)

	P(inactive)	P(informal)	P(formal)	P(inactive)	P(informal)	P(formal)
	Social security definition			Contract definition		
ln(age)	-0.064*** (0.017)	-0.097*** (0.022)	0.161*** (0.021)	-0.064*** (0.017)	-0.093*** (0.022)	0.156*** (0.021)
primary education	base category					
basic education	-0.012 (0.026)	0.067* (0.040)	-0.055 (0.036)	-0.014 (0.025)	-0.010 (0.037)	0.024 (0.041)
secondary education	-0.069** (0.027)	0.072** (0.036)	-0.003 (0.036)	-0.072*** (0.027)	-0.013 (0.035)	0.084** (0.036)
tertiary education	-0.165*** (0.016)	-0.155*** (0.037)	0.320*** (0.039)	-0.169*** (0.016)	-0.165*** (0.034)	0.334*** (0.038)
Not Tajik	-0.079*** (0.010)	-0.003 (0.015)	0.083*** (0.015)	-0.078*** (0.010)	0.027* (0.015)	0.051*** (0.015)
Female	0.074*** (0.010)	-0.192*** (0.012)	0.117*** (0.012)	0.077*** (0.010)	-0.116*** (0.012)	0.039*** (0.012)
can read	0.007*** (0.031)	0.095** (0.044)	-0.102** (0.046)	0.007 (0.031)	-0.032 (0.044)	0.025 (0.044)
capital city	-0.126*** (0.012)	0.098*** (0.018)	0.028 (0.017)	-0.124*** (0.011)	0.089*** (0.018)	0.035** (0.017)
very good health	base category					
good health	-0.020 (0.015)	0.064*** (0.019)	-0.044** (0.019)	-0.021 (0.015)	0.029 (0.019)	-0.008 (0.018)
fair or bad health	0.009 (0.019)	0.072** (0.025)	-0.08*** (0.022)	0.011 (0.019)	0.092** (0.024)	-0.103*** (0.021)
Married	0.045*** (0.013)	0.005 (0.018)	-0.050*** (0.017)	0.045*** (0.012)	0.008 (0.017)	-0.053*** (0.017)
children under 6	-0.004 (0.010)	0.013 (0.012)	-0.010 (0.012)	-0.004 (0.010)	0.006 (0.012)	-0.002 (0.012)
No. Observations	7750			7750		
Log likelihood	-7511.499			-7666.920		

Source: TLSS 2007, Sample 15 to 60 year olds, values in brackets are (heteroskedasticity robust) standard errors of the estimate ***, (**), (*) show significance at the 1%, (5%), (10%) level.

Table 4 Wage Regression (dependent variable log hourly wages)

	Social security definition		Contract definition	
	Informal Sector	formal sector	Informal Sector	Formal Sector
ln(age)	8.90*** (1.21)	5.07*** (1.17)	8.40*** (1.19)	7.00*** (1.28)
ln(age) ²	-1.33*** (0.18)	-0.72*** (0.17)	-1.22*** (0.17)	-0.96*** (0.18)
primary education				
basic education	0.45*** (0.13)	-0.1 (0.11)	-0.051 (0.12)	0.095 (0.13)
secondary education	0.29** (0.14)	0.12 (0.12)	-0.11 (0.13)	0.17 (0.13)
tertiary education	-1.58*** (0.27)	0.26** (0.13)	-0.83*** (0.23)	0.51** (0.22)
Not Tajik	-0.54*** (0.09)	-0.069 (0.07)	-0.27*** (0.09)	-0.024 (0.08)
Female	-1.72*** (0.16)	-0.59*** (0.13)	-1.10*** (0.11)	-0.24*** (0.09)
can read	1.07*** (0.19)	0.089 (0.15)	0.16 (0.16)	0.4 (0.25)
capital city	0.25** (0.13)	0.33*** (0.09)	0.32** (0.15)	0.24** (0.12)
m_0 (inactivity)	4.49*** (0.98)	1.12 (0.75)	3.86*** (1.69)	1.59 (1.37)
m_1 (informal employment)	3.47*** (0.53)	1.69* (0.95)	2.23*** (0.62)	1.76 (1.28)
m_2 (formal employment)	0.63 (1.07)	1.32*** (0.37)	2.17 (1.69)	2.51*** (0.63)
Constant	-14.9*** (2.09)	-7.07*** (2.07)	-14.1*** (2.00)	-11.4*** (2.16)
Number of obdervations	3187	2717	3042	2862
R squared	0.202	0.188	0.211	0.161

Source: TLSS 2007, Sample 15 to 60 year olds, values in brackets are (heteroskedasticity robust) standard errors of the estimate ***, (**), (*) show significance at the 1%, (5%), (10%) level. R-sq= R squared values

Table 5: Oaxaca decompositions of informal-formal sector wage differentials (hourly wages)

	Social security definition		Contract definition	
	Difference	S.E.	Difference	S.E.
Total difference	0.495***	0.029	0.2046***	0.030
selection to work	0.980***	0.355	1.2698***	0.625
selection between sectors	0.989	0.747	0.8317	1.182
Diff in Characteristics	0.046	0.147	-0.0650	0.157
Age	-0.003	0.143	-0.0267	0.156
education	-0.033***	0.011	-0.0540***	0.024
minorities	0.001	0.001	-0.0002	0.002
Gender	0.088***	0.015	0.0153***	0.005
Read	0.001	0.001	-0.0021**	0.001
Capital	-0.008***	0.003	0.0027***	0.002
Diff in parameters	-1.520	7.242	-1.8319	7.471
Age	5.853	6.608	1.7342	6.856
education	0.265	0.178	0.0570	0.057
minorities	0.137**	0.046	0.097**	0.047
Gender	0.237***	0.080	0.1477***	0.055
Read	0.479***	0.120	-0.1242	0.106
Capital	0.063	0.067	-0.1561*	0.083
Constant	-8.554***	2.957	-3.4983	2.964

Source: TLSS 2007, Sample 15 to 60 year olds. Table reports contributions of individual variables to total informal-formal sector wage differential based on Oaxaca decompositions based on results in table 5, ***, (**), (*) show significance at the 1%, (5%), (10%) level. S.E. = heteroskedasticity robust standard error of the estimate

Table 6: Marginal effects of the structural multinomial selection equation (hourly wages)

	P(non-employed)		P(informal sector)		P(formal sector)	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
	Social security definition					
Predicted wage in informal sector	-0.088***	0.022	0.293***	0.027	-0.205***	0.026
Predicted wage in formal sector	-0.384***	0.034	-0.259***	0.037	0.643***	0.036
very good health	Base category					
good health	0.001	0.016	0.144***	0.019	-0.144***	0.020
fair or bad health	0.020	0.022	0.177***	0.026	-0.197***	0.019
Married	0.055***	0.011	-0.034**	0.014	-0.021	0.014
children under 6	0.012	0.010	0.050***	0.012	-0.062***	0.012
Observations	7750					
Log Likelihood	-7719.402					
	Contract definition					
Predicted wage in informal sector	-0.072**	0.032	0.248***	0.039	-0.205***	0.039
Predicted wage in formal sector	-0.377***	0.037	-0.132***	0.045	0.409***	0.046
very good health	Base category					
good health	0.023	0.015	0.041**	0.019	-0.064***	0.020
fair or bad health	0.180***	0.038	0.075**	0.038	-0.256***	0.023
Married	0.019*	0.011	-0.056***	0.014	0.037***	0.014
children under 6	0.028***	0.010	0.024**	0.012	-0.052***	0.012
Observations	7750					
Log Likelihood	-7864.192					

Source: TLSS 2007, 15 to 60 year olds. ***, (**), (*) show significance at the 1%, (5%), (10%) level. Coeff. = marginal effects, S.E.= heteroskedasticity robust standard error of the estimate..

Appendix

To check for robustness of results, we conducted the same analysis as above on an additional wage measure in which we include proxies for in-kind and bonus payments defined from a set of questions in which persons are asked whether their main job included bonus payments or other payments in the last twelve months, how high these payments were as well as how often these payments were made. From this we calculate the average hourly bonus and in-kind payments (by dividing the annual amount by 12 and the average working time) and add this amount to the average hourly wage. This appendix reports the results of this analysis

Table A1: Distribution of formal and informal average log wages (including benefits and in-kind income) in the two sectors according to different definitions

	Definition 1: Social security entitlement			Definition 2: Written contract		
	Not employed	Informally employed	Formally employed	Not employed	Informally employed	Formally employed
log wage including bonuses and in kind payments						
Total		1.50***	1.03		1.37**	1.19
Gender						
Male		1.79***	1.28		1.68***	1.46
Female		0.81**	0.71		0.74	0.77
Age						
15-24		1.14***	0.65		0.96	0.93
25-44		1.63***	1.10		1.53***	1.26
45-60		1.56***	1.11		1.43***	1.22
Education						
primary		0.84	0.59		0.75	0.68
basic	0.39	1.17***	0.48	0.39	1.01**	0.75
secondary	-0.75	1.53***	0.90	-0.75	1.41***	1.11
tertiary	-0.63	1.93***	1.50	-0.63	1.74***	1.58

Source: Tajikistan Living Standards Survey of 2007, Sample 15 to 60 year old persons. ***, (**), (*) show significance of a t-test for difference in mean log wages between the formal and informal sector at the 1%, (5%), (10%) level.

Table A2: Wage regression (dependent variable log hourly wages including bonuses and in-kind payments)

	Social security definition		Contract definition	
	Informal Sector	Formal sector	Informal Sector	Formal Sector
ln(age)	8.97*** (1.21)	5.17*** (1.16)	8.41*** (1.19)	7.04*** (1.27)
ln(age) ²	-1.34*** (0.18)	-0.73*** (0.17)	-1.22*** (0.17)	-0.96*** (0.18)
primary education				
basic education	0.43*** (0.13)	-0.085 (0.12)	-0.052 (0.12)	0.088 (0.13)
secondary education	0.25* (0.14)	0.14 (0.12)	-0.11 (0.13)	0.17 (0.13)
tertiary education	-1.59*** (0.27)	0.27 (0.23)	-0.83*** (0.23)	0.52** (0.22)
Not-Tajik	-0.54*** (0.09)	-0.075 (0.07)	-0.27*** (0.09)	-0.23*** (0.08)
Female	-1.71*** (0.16)	-0.59*** (0.13)	-1.11*** (0.11)	-0.25*** (0.09)
can read	1.09*** (0.19)	0.099 (0.15)	0.17 (0.16)	0.44*** (0.14)
capital city	0.13 (0.13)	0.35*** (0.09)	0.34** (0.15)	-0.12 (0.12)
m_0 (inactivity)	4.54*** (0.98)	1.01 (0.75)	3.92** (1.67)	2.25 (1.37)
m_1 (informal employment)	3.46*** (0.53)	1.55* (0.85)	2.32** (0.81)	1.46 (1.28)
m_2 (formal employment)	0.61 (1.06)	1.23** (0.37)	2.11 (1.67)	2.34*** (0.63)
_cons	-15.0*** (2.08)	-7.30*** (2.06)	-14.1*** (1.99)	-11.6*** (2.15)
N	3187	2717	3042	2862
R-sq	0.231	0.185	0.241	0.154

Source: Tajikistan Living Standards Survey of 2007, Sample 15 to 60 year old persons. Table reports marginal effects, values in brackets are (heteroskedasticity robust) standard errors of the estimate ***, (**), (*) show significance at the 1%, (5%), (10%) level. R-sq= R squared values

Table A3: Oaxaca decompositions of informal-formal sector wage differentials (hourly wages including bonuses and in kind transfers)

	Social security definition		Contract definition	
	Difference	S.E.	Difference	S.E.
Total difference	0.477 ***	0.029	0.183 ***	0.030
selection to work	0.933 ***	0.354	1.160 ***	0.620
selection between sectors	0.971	0.745	0.848	1.175
Diff in Characteristics	0.045	0.147	-0.068	0.157
Age	-0.004	0.143	-0.027	0.155
Education	-0.033	0.031	-0.056 ***	0.020
Minorities	0.002	0.001	0.000	0.002
Gender	0.088 ***	0.015	0.016 ***	0.005
Read	0.001	0.001	-0.002	0.001
Capital	-0.009 ***	0.003	0.002 ***	0.002
Diff in parameters	-1.472	7.218	-1.756	7.435
Age	5.785	6.586	1.669	6.823
Education	0.257 ***	0.078	0.058	0.057
Minorities	0.135 ***	0.045	0.011	0.046
Gender	0.234 ***	0.080	0.148 ***	0.055
Read	0.481 ***	0.120	-0.129	0.106
Capital	0.077	0.067	-0.156 *	0.082
Constant	-8.441 ***	2.948	-3.356	2.949

Source: Tajikistan Living Standards Survey of 2007, Sample 15 to 60 year old persons. Table reports contributions of the individual variables to total informal-formal sector wage differential based on Oaxaca decompositions based on results in table 5, ***, (**), (*) show significance at the 1%, (5%), (10%) level. S.E.= heteroskedasticity robust standard error of the estimate

Table A4: Marginal effects of the structural multinomial selection equation (hourly wages including bonuses and in-kind payments)

	P(non-employed)		P(informal sector)		P(formal sector)	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
	Social security definition					
Predicted wage in informal sector	-0.090 ***	0.022	0.385 ***	0.026	-0.295 ***	0.025
Predicted wage in formal sector	-0.377 ***	0.033	-0.243 ***	0.036	0.621 ***	0.035
very good health						
good health	0.000	0.016	0.141 ***	0.019	-0.141 ***	0.020
fair or bad health	0.020	0.022	0.174 ***	0.026	-0.194 ***	0.019
Married	0.056 ***	0.011	-0.034 **	0.014	-0.022	0.014
children under 6	0.011	0.009	0.049 ***	0.012	-0.059 ***	0.012
Observations	7750					
Log Likelihood	-7716.568					
	Contract definition					
Predicted wage in informal sector	-0.085 **	0.033	0.191 **	0.041	-0.106 ***	0.041
Predicted wage in formal sector	-0.297 ***	0.039	-0.126	0.047	0.422 ***	0.048
very good health						
good health	0.025 *	0.015	0.040 **	0.020	-0.066 ***	0.020
fair or bad health	0.194 ***	0.040	0.066 *	0.039	-0.261 ***	0.024
Married	0.022 *	0.011	-0.055 ***	0.014	0.033 **	0.014
children under 6	0.028 ***	0.010	0.024 **	0.012	-0.052 ***	0.012
Observations	7750					
Log Likelihood	7861.333					

Source: Tajikistan Living Standards Survey of 2007, Sample 15 to 60 year old persons. Table reports marginal effects, ***, (**), (*) show significance at the 1%, (5%), (10%) level. Coeff. = marginal effects, S.E.= heteroskedasticity robust standard error of the estimate