Envelope Wages, Underreporting and Tax Evasion: The Case of Turkey *

Selin Pelek[†] Gokce Uysal[‡]

April 30, 2016

Abstract

Even though informality in labor markets have been studied widely, envelope wages, a hybrid form of informal employment has only recently attracted more attention. In this form, some employees receive two wages from their employers, an official wage that is reported to the tax authorities and an envelope wage that is undeclared. This underreporting of actual wages by the firms reduce the tax burden of the firms, but result in unfair competition, and limits the access of employees to social protection systems. Research on envelope wages is gaining momentum recently as more data becomes available. The institution set-up in Turkey provides a unique setting to study envelope wages as the wages are taxed at the source. Therefore the firms have an incentive to underreport and the households have no incentive to lie about their wages. In this paper, we use data from two different data sources, one collected at the firm level and the other at the household level, to identify individuals who may be receiving envelope wages. Following a Heckman sample selection estimation strategy, we estimate the coefficients in a Mincerian wage equation, which are later used to construct estimated wages for individuals whose wages may have been underreported. This paper is the first in the relevant literature to provide estimates of underreporting and the associated tax losses when data on underreporting is not directly available.

Keywords: Informal employment, envelope wages, underreporting, tax evasion
JEL codes: J33, J38, H26

* We would like to thank Jeffrey M. Wooldridge and Seyfettin Gursel for valuable comments. All remaining errors are our own.

[†] Galatasaray University, GIAM, Michigan State University, selinp@msu.edu
[‡] Bahcesehir University, Betam, IZA, ERF, gokce.uysal@eas.bau.edu.tr

1 Introduction

Informality in labor markets has been widely studied by researchers in both high- and low-income countries across the world. As a result, many forms of formal and informal employment have been defined. Some definitions concentrate on the type of production (informal firms) and some on the type of job (informal jobs). Even though there is continuous effort to define different types of informal employment and build a conceptual framework, as informal economies grow, various new forms of informal employment emerge (Hussmanns, Hussmanns). An example that is only recently attracting more attention is a hybrid form of informality, envelope wages. In Williams (2009), OECD (2003) and Woolfson (2007), this form is characterized by "envelope wages" referring to two wages paid by employers: an official wage declared and a residual undeclared cash-in-hand wage.

The main motivation of for underreporting of wages is to reduce the tax burden placed on labor in the form of social security contributions (Merikull and Staehr, 2010). Furthermore, unreported wage income may avoid paying other taxes, such as value added or environmental taxes. In this way, underreporting should be considered as a component of informality and expected to be more common where the taxes and state intervention are higher (Becker, 2004). The structuralist perspective argue that the under-reporting of wages could be a product of a weak regulation or inadequate levels of intervention (Williams, 2012). According to the this approach, the employers could make more profit by not declaring a part of wage bill and the employees unwillingly agree to receive an envelope wage (Ghezzi, 2009). Therefore, the underreporting of wages is expected to be more prevalent in countries with lower levels of labor market controls.

The firms who resort to underreporting do so to lessen the bite of employment taxes and severance pay. Thus, underreporting may help firms survive and sustain employment which they otherwise cannot. However, economy-wide, it breeds an environment of unfair competition among firms. On the employee side, underreporting of official wages constrains the opportunities available to the employees. When firms underreport wages, employees gain only limited access to some types of social protection that are a function of wages, e.g. retirement benefits. Their access to other types of financial services such as mortgage and loans are restricted as well.

Recent applied research on this area seems to be consistent with this conjecture. The European Commission report on undeclared work conducted in 2007, states that 5% of employees in European Union countries received all or part of their salary as envelope wages within the past 12 months of their survey (European Commission, 2007). However, this rate varies considerably by location. Central and Eastern European countries are reported with the highest levels of envelope wages (23% of wages by 2006) while in Germany, France, Luxembourg, Malta and United Kingdom the underreporting of wages has only a marginal share (less than 1%). Another comprehensive study by Williams and Padmore (2013) uses a multinational survey conducted in 27 EU Member States, by Eurobarometer in 2007 in order to investigate the prevalence and distribution of envelope wage payments. This specific survey includes 26,659 face-to-face interviews. Their results indicate that one in 18 formal employees receives an envelope wage amounting to 25% of their gross pay on average in EU. The authors find a negative relationship between GDP and the envelope wage practice. This phenomenon is also common in many other developing countries outside Europe, such as Russia, Ukraine, Argentina (Tonin, 2011).

Just like in other forms of informal employment, data on envelope wages is hard to come by, and when it exists, it is usually limited in scope either in terms of the number of observations or in terms of coverage. Except for the data sets within the Eurobarometer used by Williams and Padmore (2013) contain direct questions on envelope wages, there are no other nationally representative data sets providing information on underreporting. Turkey is no exception. Even though anectodal evidence points to a wide-spread usage of envelope wages, there are no data sets that directly ask about underreporting. However, the institutional set-up allows us to build an estimate of envelope wages under minimal assumptions. To measure the prevalence of envelope wages, we use a novel approach and two different data sets. There are two different types of micro data available, one collected at the household level, the other at the firm level. In Turkey, earnings are taxed at the source, i.e. the firm pays both its and the worker's contributions of the social security taxes, the unemployment taxes, the income tax as well as some other deductions. In other words, the firms are responsible for reporting earnings and paying taxes. Households receive and report net earnings. Since taxes are not the household's responsibility, they have no incentives to lie about their earnings. To sum up, if the firm is paying envelope wages, the firm will only report the official wage whereas the individual (surveyed at home) will report the actual wage, the sum of the official and the envelope wages. Thus, if there is underreporting, the wage distribution of the firm level data and that of the household level data will differ.

If and when the firm chooses to underreport the earnings of a given worker, the optimal decision would be to register the worker at the minimum wage, under the condition that the probability of getting caught is not a function of how much the wage is being underreported. Since the firms have an incentive to report earnings at the minimum wage, we expect relatively more employees registered at the minimum wage at the firm level data (Structure of Earnings Survey, SES) than at the household level data (Household Labor Force Survey, HLFS and Household Budget Survey, HBS). This is indeed the case. Figure 1 provides the data from both types of data. Clearly, the firm level data has a much higher spike at the minimum wage than household level data, consistent with our conjecture.

We estimate the coefficients in a Mincerian wage regression using household level data under the assumption that the true wages are observed in this data set, and then use these estimates to construct a hypothetical wage distribution for individuals in the firm level data under the assumption that these may be underreported. The difference between the hypothetical distribution and the actual distribution provides our measure of underreporting. The data sets and the method used are described in detail below.

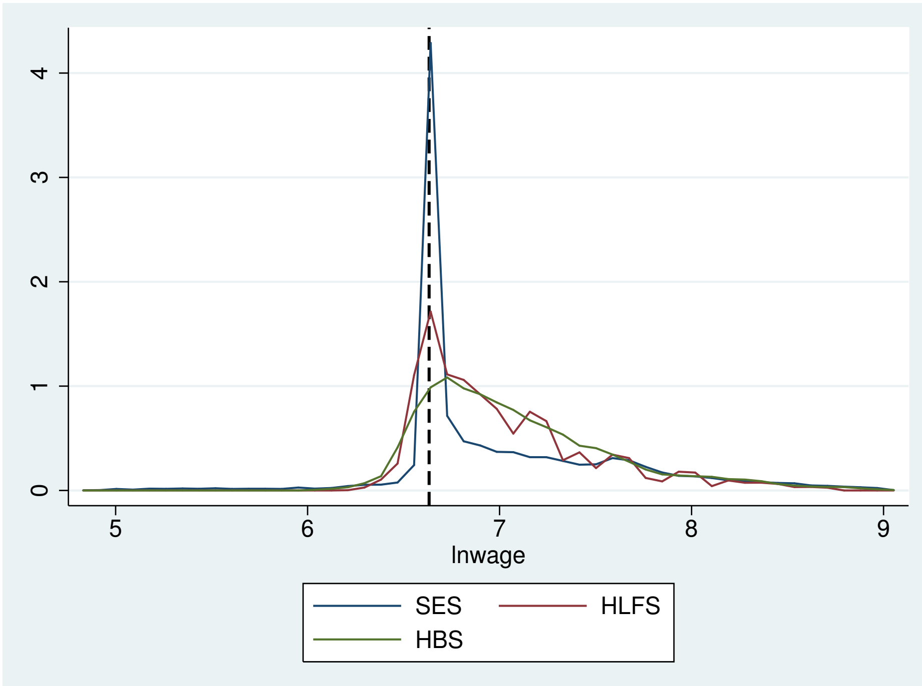

Figure 1: Kernel density plots of wages

Note: Distribution of wages of full-time formal workers in enterprises employing at least 10 workers. The vertical line represents the gross minimum wage.

Despite this clear evidence of envelope wages in many countries, the econometric studies addressed the issue is limited. Previous studies have mainly focused on qualitative, rather than quantitative, data and methods (Sedlenieks, 2003; Woolfson, 2007). Their results are in line with the European Commission report as well as Williams and Padmore (2013) highlighting that central and eastern Europe has been marked by a high incidence of underreporting while in the west of Europe the envelope wages are not common.

Kriz et al. (2007) examined the socio-economic characteristics of workers who receive underreported wages by employing three cross sectional data sets from Estonia. However, the samples used in this study are not randomly drawn. Their results suggest that the envelope wages and tax evasion are more frequent among the individuals who work part-time, are of non-Estonian ethnicity, have relatively short education, are younger, earn a low income, and are men. Another research carried out by Merikull and Staehr (2010), addressed the underreporting of wages in three Baltic countries; Estonia, Latvia, and Lithuania. In this comparative study, 900 face to face interviews between 1998 and 2002 used to seek the characteristics of individuals receiving envelope wages. They found different trends across countries: the prevalence of unreported wages decreased in Estonia from 1998 to 2002, while the opposite occurred in Latvia and Lithuania (Merikull and Staehr, 2010).

In Turkey, the labor market has a dual structure where the informal labor market is sizeable. Even though the labor law requires employers to pay at least minimum wage as a lower limit, it is a well known fact that a non-neglible share of employees, particularly those in informal employment, are paid less than the minimum wage (Pelek and Calavrezo, 2011). Besides the low-paid informal work, the underreporting of wages in the formal employment is a prevalent phenomenon in Turkish labor market. According to Ministry of Labor and Social Security, approximately 5 million workers, 40% of the wage earners in the private sector, earn the minimum wage officially. However, Mehmet Simsek, the former minister of Labor and Social Security, had asserted that only 2.8 million employees were *true* minimum wage earners while the wages of the remaining 2.2 million workers were underreported to Social Security Institution (Kivanc, 2015). The former minister pointed to the tax loss due to underreporting. Besides, as mentioned above, reporting less than the true amount of wage earnings, causes a reduction in the retirement pensions of employees. The anecdotal data suggest that the underreporting problem is not intrinsic to a particular sector; but rather it exists in all sectors. Erdogdu (2009) highlights the evidence on the underreporting of wages based on the confidential interviews with three employers from the manufacturing, construction and private education sectors. The employers confirmed the underreporting practices and stated that their basic motive was to evade high labor costs through paying lower taxes and social security contributions (Erdogdu, 2009) Notwithstanding the anecdotal evidence of envelope wages in Turkey, there are no studies that assess the extent to which firms underreport. In other words, the difference between the reported wages and the real wages and the extent of tax evasion due to underreporting have not been investigated using a nationally representative data set yet. In this study, we aim to address this particular gap by using a novel estimation strategy and two different data sets to construct estimates of underreported wages and associated tax losses.

2 Data and Descriptive Analysis

The data used in this paper come from three different surveys conducted by TURKSTAT:

i. *Structure of Earnings Survey (SES)*: This survey is conducted every four years and aims to provide detailed information on earnings, the demographic and job characteristics of workers such as age, gender, education level, tenure, industry, occupation. We use the 2010 wave which provides most up-to-date data. This survey is conducted on a total of 20,155 establishments covering the enterprises employing 10 and more employees. Since the data is collected at the firm level and that it consists of formally employed workers, it provides adequate data for the measurement of underreporting. In the data, there are 198,375 employees who work as wage earners, interns or apprentices. Employers and partners who only receive profit shares and unpaid family workers are excluded. The reference month of the survey is November 2010. A major drawback of SES is that it covers only formal employment in firms with more than 10 employees. This represents 63% of all formal employment in Turkey for the year 2010. So, the estimates provided here are a lower limit of the extent of underreporting in Turkey. It is easy to imagine that envelope wages would be more common among small firms where productivity levels are lower and the firms have a lower probability of getting caught.

ii. *Household Labor Force Survey (HLFS)*: This annual cross sectional data set provides information on the socioeconomic characteristics of the labor force in Turkey. HLFS data is designed to be representative of the whole non institutional population of Turkey. The respondents of HLFS are asked to report their net wages. The wage information reported directly by workers enables us to explore *real wages*. Thus, under the assumption that the wage information in HLFS do not represent the underreported ones, the comparison of these two dataset could give an idea about the underreporting of wages. Therefore, HLFS 2010 with a sample size of 522 171 is employed in this paper to estimate the envelope wages.

iii. *Household Budget Survey (HBS)*: This cross sectional survey is designed to provide information on household income and consumption, e.g. total income, main items of expenditure, and socio economic charecteristics at household and indiviudal levels as well. HBS is also representative of the whole non institutional population of Turkey. For the sake of robustness, alongside of HLFS, HBS 2010 is employed in the descriptive and econometric parts of this paper. 38 206 individuals from 10 082 households are observed in HBS 2010.

In order to harmonize sampling, we keep the observations for the employees working in the enterprises with 10 or more employees in the formal sector and reporting a monthly wage greater than zero from HLFS and HBS. Note that, the wages reported in these two datasets are the net wages. We calculate the gross wages from the net ones by using tax rates on labor provided by OECD tax database. Thereby, we obtain three different sample with the same scope coming from three data sources. The numbers of observation are 198 360, 32 039 and 2 420 for SES, HLFS and HBS respectively. By taking into account the sample size and the scope of survey, we prefer to use HLFS microdata rather than HBS in the econometric part. The descriptive statistics are provided in Table 1.

We begin with the specification of minimum wage workers in these three sub-samples. The amount of the gross minimum wage in the second half of 2010 was 760.5 TL (Approximately, it is rounded to 761 TL in this paper). We define the minimum wage workers as the ones whose gross wage is between 0.95 and 1.05 monthly minimum wage in 2010. These bounds have been used in previous research in order to account for rounding approximations (Gindling and Terrell, 2005; Pelek and Calavrezo, 2011). The share of workers earning the minimum wage highly varies depending on data source: 43.2% of wages reported in SES corresponds to minimum wage level while it is only 16.3% and 13.1% in the subsamples of HLFS and HBS respectively. Note that, these shares are consistent with the kernel density

Table 1: Descriptive statisrtics

	SES	HLFS (all)	HLFS (restricted sample)
Individual characteristics			
Gender			
Male	75	78.1	77.8
Average age	33.9	33.2	33.3
Education			
Primary or less	22.8	40.2	34.4
Secondary	16.1	21	16.6
High school	25.5	12.9	13.9
Vocational high sch.	9	13.3	16.8
University	26.7	12.6	18.3
Job characteristics			
Sector			
Industry	37.4	37.1	46.7
Construction	4.5	9.5	6
Services	58.1	53.4	46.3
Occupation			
Skilled	16	7	9.7
Semi-skilled	70.8	73.5	73
Unskilled	13.2	19.5	17.3
Firm size			
Less than 10	N.A.	41.7	-
10-49	51	28.9	45.4
50-249	17.5	19.5	35.5
250-499	10.9	4.5	8.6
500 or more	20.7	5.4	10.5
Average tenure	3.5	4.1	4.8

estimates provided in Figure 1 and further confirm the anecdotal data of underreporting of wages. The density belonging to SES is characterized by a sharp spike around the minimum wage and a narrower distribution. The shapes of HLFS and HBS are clearly broader and the peaks around the minimum wage are less pointy.

In order to clarify the differences of the wage distributions extracted from the three data sets, we rely on the decomposition method introduced by DiNardo, Fortin and Lemieux (1996), DFL hereafter. DFL is a semiparametric decomposition approach which is an extended version of standard Oaxaca Blinder method. Its extension is the evaluation of counterfactual differences on the whole distribution instead of the mean. Following DiNardo et al. (1996), we estimate a counterfactual distribution that should be called "the density that would have prevailed if individual attributes in HLFS and HBS were identical to those in SES and workers had been paid according to the wage schedule observed in HLFS and HBS respectively." Therefore, the difference between the actual density of wages obtained from HLFS and HBS and the counterfactual density estimated by DFL methodology reveals the potential effects of underreporting.

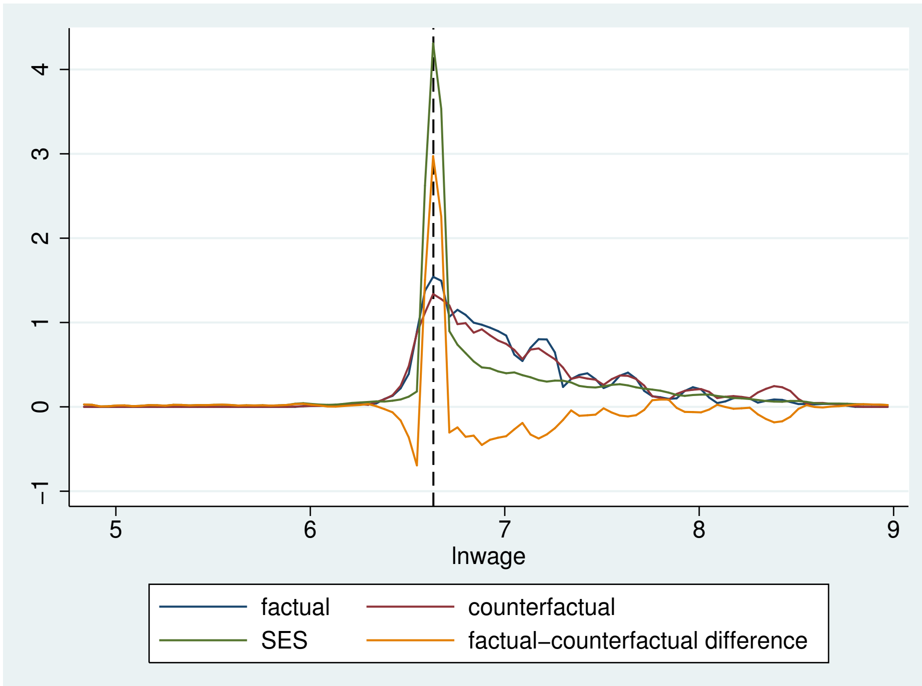

Figure 2: DFL estimation using HLFS

Note: The vertical line is the gross minimum wage.

To decompose the effects of the individual characteristics in the wage distribution comes from HLFS, we obtain a counterfactual distribution by keeping them the same individual characteristics constant as in SES. The individual attributes used are gender, age, age-squared, educational level, tenure, tenure-squared, occupation, sector, and firm size. Figure 2 plots actual kernel density estimations of wages from HLFS and SES samples and counter-factual kernel density estimations from HLFS if the individual characteristics are the same as in SES sample.Note that, we obtain a slightly smoother counterfactual distribution with the SES characteristics, but the overall picture does not change significantly. The sizeable difference between factual and the counterfactual distributions of still needs to be explained. Our assumption is that this difference is due to the underreporting practices.

We employ the DFL method to HBS sample too. As shown in Figure 3 the counterfactual kernel density estimates obtained with SES' characteristics almost overlaps the factual distribution. Once again, the difference between the factual distribution and the counterfactual distribution of HBS strengthens our conjecture of envelope wages.

Unfortunately the DFL strategy allows the computation of neither the extent nor the

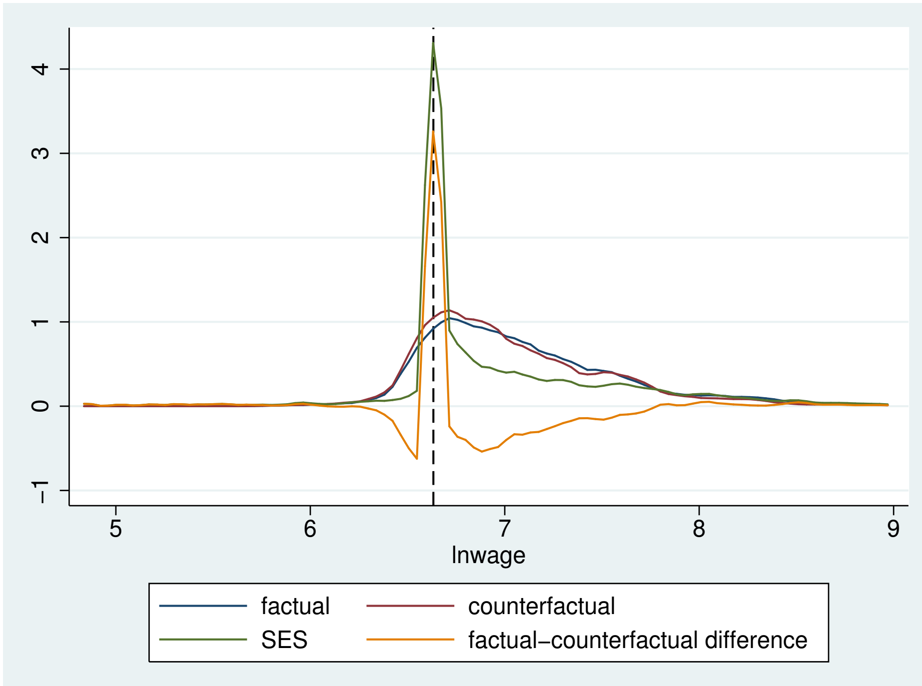

Figure 3: DFL estimation using HBS

Note: The vertical line is the gross minimum wage.

magnitude of envelope wages. In the next section, we discuss the econometric methodology used to estimate the magnitude of underreporting and the associated tax losses.

3 Estimation Strategy

The main assumption here is that the households report their true wages, that is, the sum of their official wages and their envelope wages, whereas the firms underreport. In other words, the true data generating process can be estimated by using household level data. Therefore, any discrepancy between the household data and the firm data stems from underreporting of true wages by the firms. Under this identifying assumption, we follow the steps summarized below to estimate the true wages of the individuals in the firm level data. Let the subscript i denote an individual in the household data and the subscript j denote an individual in the firm data. Then X_i^H denotes the labor market characteristics of individual i in household data H and X_j^F those of individual j in firm data F.

$$\ln w_i^H = X_i^H \beta + \epsilon_i \quad (1)$$
$$\ln \hat{w}_j^F = X_j^F \hat{\beta}^H \quad (2)$$
$$\hat{u}_j = \ln \hat{w}_j^F - \ln w_j^F \quad (3)$$

1. Use the household data of individuals i to estimate the coefficients of a Mincerian wage equation. Regression summarized in Equation (1) yields the estimated coefficients $\hat{\beta}^H$.

2. Use $\hat{\beta}^H$ to construct an estimated wage, $\ln \hat{w}_j^F$, for each observation j in the firm level data using their characteristics, X_j^F. Equation (2) represents our estimate of the real wages of individuals j in the firm level data.

3. The difference between the estimated wage $\ln \hat{w}_j^F$ and the reported wage $\ln w_j^F$ of individual j in the firm level data Fconstitute our estimate of the envelope wages shown in Equation (3). We will be using these estimates of envelope wages $\hat{u}_j$ to estimate the total losses in tax revenue.

4. We will also compute $\ln \hat{w}_j^H = X_j^H \hat{\beta}^H$ and compare the estimated wage $\ln \hat{w}_j^H$ to the observed wage $\ln w_j^H$ to assess the reliability of our estimates.

The next question is the regression method to be used in Equation (1).[1] Note that the firm level data is representative of only a subsample of the entire population, i.e. employees who are registered at the social security institution and working in private firms of at least 10 employees. Therefore, we use the Heckman sample selection model to consistently estimate the coefficients. As is widely known, the first stage of this estimation is based on the selection into the sample as in Equation (4). Here, w_k are the relevant characteristics of individual k whether they are from the household data H or the firm data F. Our equation of interest is still Equation (5). Note that we are in a slightly different setting than in the usual Heckman selection model. We observe the wages without any selection in the household data whereas we only observe the wages of the selected sample in the firm data.

$$z_k^* = w_k \gamma + \eta_k \quad (4)$$
$$\ln w_k = X_k \beta + \epsilon_k \quad (5)$$

In a standard set-up, Equation (6) is estimated to tackle sample selection. In the first stage, the selection equation provides an estimate for λ. Then the second step uses X_k and

[1]The OLS regressions provide biased estimates of the wage distribution. We provide the plain OLS regression results in the Appendix. Clearly, the estimated wages and the actual wages are systematically different, even in the household data where no such discrepancy is justified. We conclude that the OLS is not an appropriate estimation strategy in this setting and use the Heckman sample selection model. See HLFS **??** and HBS **??**

A as regressors and the coefficients β and β_λ are estimated. In our case, our ultimate goal is to construct an estimated wage for individuals in the firm data. To construct a better estimate of the actual wage than Equation (2), we would like to take possible selection into account.

$$E[\ln w_k|z_k^*] = X_k\beta + \underbrace{\beta_\lambda \underbrace{\frac{\phi\left(w_k\gamma/\sigma_\eta\right)}{\Phi\left(w_k\gamma/\sigma_\eta\right)}}_{\lambda}}_{B} \tag{6}$$

To that end, we use the two stages of this estimation strategy to construct an estimate for the selection term, B. In the first stage, we need an estimate for λ, but we only observe the selected sample in the firm level data. Therefore, we turn to the household data and estimate the first stage to construct $\hat{\lambda}$ using $\hat{\gamma}$ and $\hat{\sigma}_\eta$ that were estimated using household data only, hence the superscript H. However, this is not enough to construct an estimated wage that corrects for selection. We also need to have an estimate for β_λ. Still using household data only, we use X_i^H and $\hat{\lambda}^H$, we run the second stage regression of the Heckman sample selection model to get an estimate of $\hat{\beta}_\lambda^H$ using only the household data.

$$\ln \hat{w}_j^F = X_j^F \hat{\beta}^H + \hat{\beta}_\lambda^H \underbrace{\frac{\phi\left(w_j^F \hat{\gamma}^H/\hat{\sigma}_\eta^H\right)}{\Phi\left(w_j^F \hat{\gamma}^H/\hat{\sigma}_\eta^H\right)}}_{\hat{\lambda}^H} \tag{7}$$

To clarify, we assume that the true data generating process is the one that generates the household level data. Under this assumption, we estimate all necessary parameters using the household data only and then use these estimates to construct an estimated wage for all individuals in the firm data. The regression output is provided in Appendix **??**.

4 Results

The estimation strategy detailed above provides an estimated wage $\ln \hat{w}_j^F$ for all observations in the firm level data set. The resulting Kernel density is provided along with the observed kernel density in Figure 4. Clearly, the estimated distribution has a much lower peak than the actual distribution, implying that underreporting may be a widespread problem in the labor market.

Of course, since we lack actual data on envelope wages, it is difficult to run diagnostics to assess the performance of the estimation strategy. One potential robustness check is to construct wage estimates for the individuals in the household level data and to compare the estimated wages to the actual wages. The results are provided in Figure 5. It looks like the estimated wages and the actual wages of HLFS almost overlap. As expected, the distribution of estimated wages are smoother given that there is no error term.

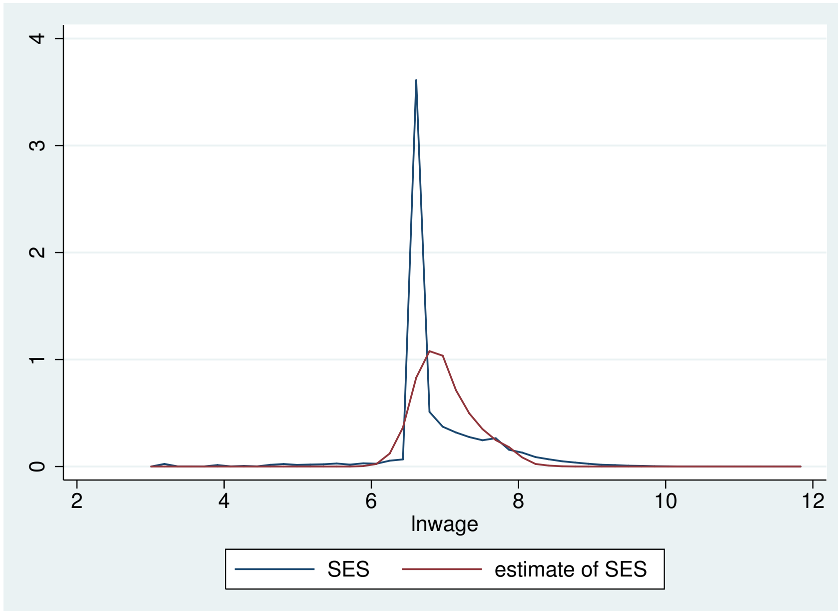

Figure 4: SES estimates with Heckman

4.1 Estimates of underreporting

As informative as the Kernel density estimations are, they are difficult to interpret. We provide the actual wages and the estimated wages of the firm level data in Table 2. First, our results show that at the lower end of the wage distribution, the minimum wage is binding. At the 10th percentile, our estimation results reveal positive envelope wages. In other words, the estimated wages are clearly above the reported wages in the firm level data. We estimate some negative envelope wages at the highest percentiles of the wage distribution.

Table 2: Heckman estimated wages

percentiles	observed wages	estimated wages	envelope wages
10%	760	875	115
20%	760	875	115
30%	761	883	122
40%	761	890	129
50%	768	899	131
60%	840	1036	196
70%	1049	1196	147
80%	1415	1124	-291
90%	2064	1604	–460

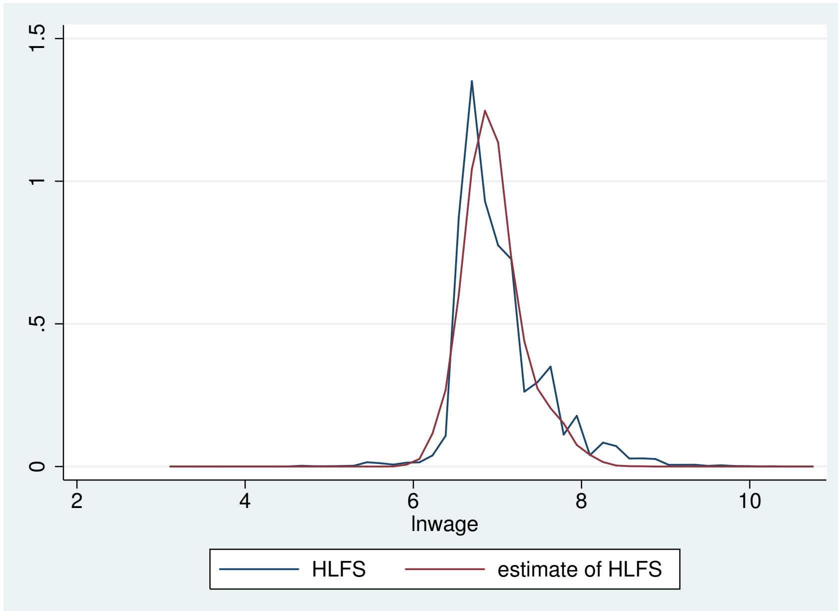

Figure 5: HLFS estimates with Heckman

4.2 Estimates of Tax Evasion

Using the estimated wages, we calculate an estimate for the tax losses. The tax loss is calculated as the difference between the social security contributions amounts calculated from the gross reported wages and the gross estimated wages. We take two different approaches when calculating tax losses. In one approach, we do not allow for negative envelope wages. In the other approach we assume that underreporting when the reported wage is twice the minimum wage is negligible.

In the first approach, negative envelope wages are not admissible. If the reported wage is higher than the estimated wage our model predicts, we update our estimate and impose a zero envelope wage. Then the tax losses are calculated for individuals who earn positive envelope wages. The underlying assumption is that the idiosyncratic component of the wages, embedded in the error term are small enough to ignore. If the error term is normally distributed, the mean is zero and this assumption is not restrictive. In this case, the sample size decreases from 198 360 to 110 642 observations. The monthly total tax paid on the reported wages for individuals. The monthly total tax paid on the reported wages for individuals who receive an positif or zero envelope wages is 35 829 008 TL (12 816 450 USD). If we assume that these individuals receive the estimated wage, the difference paid in envelopes, the estimated tax is 55 374 556 TL (19 808 118 USD). In other words, there is a 35.3% tax loss for this sub-sample. It corresponds to 177 TL (63 USD) tax loss *per capita* and 19.5% tax loss in the all SES sample including the individuals who are estimated to receive a negative envelope wages.

However, we are actually ignoring only negative error terms, which introduces a bias. So the second approach refrains from making this assumption. We assume that underreporting is non-existent at the upper end of the wage distribution, i.e. for individuals who earn at least twice the minimum wage, and restrict our attention to the lower end.

The monthly total tax paid on the reported wages for individuals who earn less than twice the minimum wage is 47 276 292 TL (16 911 275 USD). If we assume that these individuals receive the estimated wage, the difference paid in envelopes, the estimated tax is 61 530 140 TL (22 010 042 USD). Given that the observations number in this case is 152 678 individuals, per capita tax loss is estimated as 93.4 TL (33.4 USD) In other words, there is a 23.2 percent tax loss if the estimated wages provide a better estimate of the actual wages. It corresponds to 17.9 % tax loss in total sample.

5 Conclusion

Informal forms of employment are always a concern for labor economists as well as for policymakers. The usual classification of informal employment allows for individuals who hold formal jobs and informal jobs, and also for firms who operate formally and informally. However, as economies develop, new forms of informal employment emerge. Underreporting or envelope wages, are a relatively recent form of informality where the worker may formally be employed in a formal firm, but may be receiving some part of their wages in an "envelope", off-the-books. This hybrid form has been recently attracting attention, particularly since it is becoming widespread in especially in Eastern European countries. There are only two studies, European Commission (2007) and Williams and Padmore (2013) that use nationally representative data sets to study underreporting as the data available is extremely limited.

We use a novel approach to estimating the extent of underreporting and the associated tax losses in absence of a data set that asks directly about envelope wages. Our identification strategy is unique to Turkey where the firms are required to pay taxes as the wages get taxed at the source. In this set up, the firms have an incentive to underreport and employees do not. We use two different data sets, one at the firm level, the other at the household level to compare the wage distribution and to construct an estimate for underreporting and tax evasion. To this end, we estimate a Mincerian wage regression with a Heckman sample selection correction using the household data, then use the coefficients and the parameters estimated to calculate an estimated wage for the employees in the firm data. Our results indicate that underreporting is widespread and tax losses sizeable. The difference between the distribution of wages coming from individual and firm based data could be explained by underreporting practice imposed by firms. According to our estimations, we can suggest that underreporting of wages lead to a tax loss about 20% in the formal sector.

Williams and Padmore (2013) also draw attention to the trade-off between the minimum wage level and envelope wages. They suggest that a higher minimum wage could reduce

reduce the proportion of the wage that could be paid as an envelope wage. However it could increase the shift from formal employment to an informal one (Williams and Padmore, 2013). The minimum wages have been at the center of a heated public discussion in high-income countries such as the US and the UK. Turkey, a middle-income country, has acted faster and has increased the minimum wage by one third.

The minimum wage has became a topic of popular discussion in Turkey by governors, politicians, and economists prior to the last general elections in November 2015. All of the political parties gave the minimum wage central importance in their electoral campaigns and promised to increase the minimum wage by differing percentages. Consequently, the new government raised the net minimum wage by 30 %, from 1000 to 1300 TL, as of January 1, 2016. The wage and employment effects of this sudden and sizeable increase in the wages remains to be seen. We also expect this increase to have an effect on envelope wages and the related tax losses in Turkey. The employment effects of the minimum wage increase may be limited insofar as the envelope wages are common at the lower end of the wage distribution. In that case, increasing the minimum wage will curb tax evasion through underreporting.

References

Becker, K. F. (2004). The informal economy: a fact finding study prepared for the swedish international development agency. *Stockholm, Sweden, SIDA*.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044.

Erdogdu, S. (2009). Minimum wage in tension between economic and social concerns. In D. Vaughan-Whitehead (Ed.), *The minimum wage in the enlarged EU*. Edward Elgar.

European Commission, E. C. (2007). Undeclared work in the european union. Technical report, Special Eurobarometer 284.

Ghezzi, S. (2009). The fallacy of the formal and informal divide: lessons from a post-fordist regional economy. In J. Marcelli, C. C. Williams, and P. Joassart (Eds.), *Informal work in developed nations*, pp. 114–131. Routledge.

Gindling, T. and K. Terrell (2005). The effect of minimum wages on actual wages in formal and informal sectors in costa rica. *World Development 33*(11), 1905–1921.

Hussmanns, R. Defining and measuring informal employment. *Bureau of Statistics Paper, ILO, Geneva, February year= 2004*.

Kivanc, A. (2015, May). Resmi asgari secmen 5 milyon asli 2.8 milyon.

Kriz, K. A., J. Merikull, A. Paulus, and K. Staehr (2007). Why do individuals evade payroll and income taxation in estonia? *U of Tartu Economics and Business Administration Working Paper* (49-2007).

Merikull, J. and K. Staehr (2010). Unreported employment and envelope wages in mid-transition: comparing developments and causes in the baltic countries. *Comparative Economic Studies 52*(4), 637–670.

Pelek, S. and O. Calavrezo (2011). Les emplois du bas de lechelle salariale en turquie: une description des salaries des secteurs formels et informels, hors agriculture. *Travail et emploi 126*, 45–60.

Sedlenieks, K. (2003). Cash in an envelope: corruption and tax avoidance as an economic strategy in contemporary riga. *Everyday Economy in Russia, Poland and Latvia*, 37–52.

Tonin, M. (2011). Minimum wage and tax evasion: Theory and evidence. *Journal of Public Economics 95*(11), 1635–1651.

Williams, C. (2012). Cross-national variations in the under-reporting of wages in south-east europe: A result of over-regulation or under-regulation? *South East European Journal of Economics and Business 7*(1), 53–61.

Williams, C. C. (2009). The prevalence of envelope wages in the baltic sea region. *Baltic Journal of Management 4*(3), 288–300.

Williams, C. C. and J. Padmore (2013). Envelope wages in the european union. *Int'l Lab. Rev. 152*, 411.

Woolfson, C. (2007). Pushing the envelope: the 'informalization' of labour in post-communist new eu member states. *Wrok Employment Society 21*(3), 551–564.

A Quantile Regression

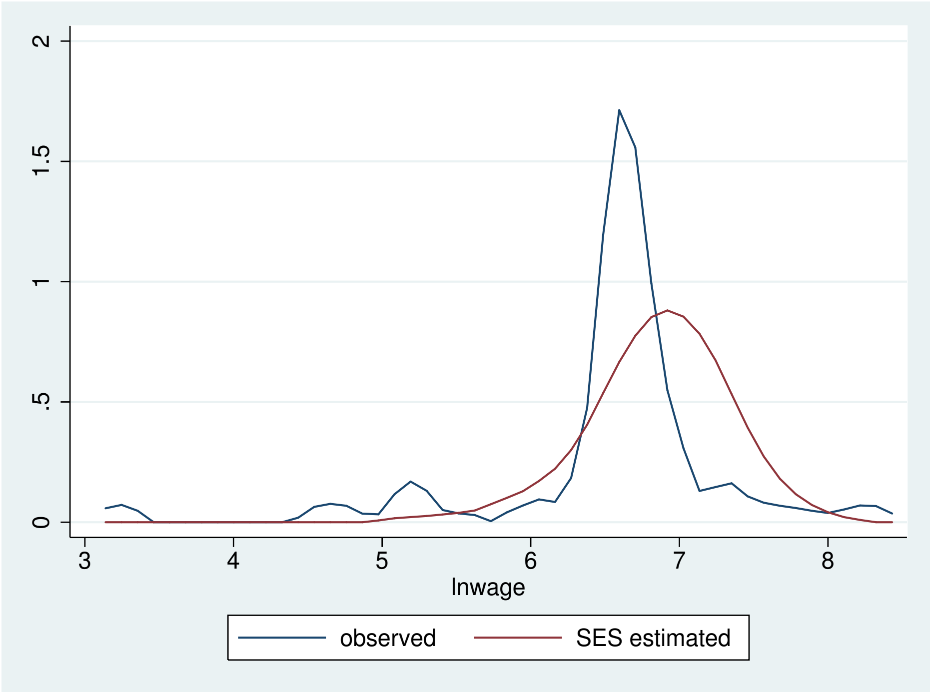

Figure 6: Quantile estimates of SES with HLFS

Note: observed vs estimated of SES sample (with HLFS)

Table 3: Quantile estimated wages HLFS

percentiles	observed wages	estimated wages
10%	760	515
25%	761	732
50%	796	987
75%	1415	1304
90%	2540	1673

B Regression Output

Table 4: Probit regression of HLFS

Variable	Coefficient	(Std. Err.)
gender	0.167	(0.024)
_Ieducation_2	0.072	(0.028)
_Ieducation_3	0.227	(0.035)
_Ieducation_4	0.310	(0.035)
_Ieducation_5	0.463	(0.047)
age	0.199	(0.005)
agesquared	-0.003	(0.000)
tenure	0.111	(0.005)
tenuresq	-0.003	(0.000)
_Ifirmsize_2	11.382	(155.871)
_Ifirmsize_3	11.809	(155.871)
_Ifirmsize_4	11.964	(155.871)
_Ifirmsize_5	12.141	(155.871)
_Ioccupatio_1	0.177	(0.068)
_Ioccupatio_2	0.290	(0.075)
_Ioccupatio_3	0.238	(0.047)
_Ioccupatio_4	0.358	(0.047)
_Ioccupatio_5	0.092	(0.042)
_Ioccupatio_6	0.056	(0.122)
_Ioccupatio_8	0.196	(0.032)
_Ioccupatio_9	-0.072	(0.032)
_Iactivity_1	-1.558	(0.077)
_Iactivity_2	-2.242	(0.374)
_Iactivity_3	-0.848	(0.295)
_Iactivity_5	0.207	(0.145)
_Iactivity_6	4.717	(360.852)
_Iactivity_7	0.151	(0.334)
_Iactivity_8	0.105	(0.170)
_Iactivity_9	-0.466	(0.242)
_Iactivity_10	-0.011	(0.059)
_Iactivity_11	4.851	(116.163)
_Iactivity_12	5.017	(656.451)
_Iactivity_13	-0.018	(0.060)
_Iactivity_14	-0.636	(0.050)
_Iactivity_15	-0.787	(0.086)
_Iactivity_16	-0.130	(0.121)
_Iactivity_17	0.148	(0.139)
_Iactivity_18	-0.103	(0.135)
_Iactivity_19	-0.016	(0.219)
_Iactivity_20	0.016	(0.112)
_Iactivity_21	0.930	(0.501)
_Iactivity_22	0.088	(0.095)
_Iactivity_23	-0.056	(0.066)
_Iactivity_24	0.251	(0.094)
_Iactivity_25	0.103	(0.086)
_Iactivity_26	0.413	(0.246)
_Iactivity_27	0.390	(0.112)
_Iactivity_28	0.052	(0.087)
_Iactivity_29	0.265	(0.105)
_Iactivity_30	0.166	(0.191)
_Iactivity_31	-0.079	(0.083)
_Iactivity_32	-0.180	(0.127)
_Iactivity_33	0.433	(0.192)
_Iactivity_35	0.327	(0.156)

Table 5: Coefficient estimation results with Heckman selection term

Variable	**Coefficient**	**(Std. Err.)**
gender	0.200	(0.005)
_Ieducation_2	0.024	(0.006)
_Ieducation_3	0.117	(0.007)
_Ieducation_4	0.110	(0.007)
_Ieducation_5	0.340	(0.009)
age	0.056	(0.002)
agesquared	-0.001	(0.000)
tenure	0.021	(0.001)
tenuresq	0.000	(0.000)
_Ifirmsize_2	-0.637	(0.078)
_Ifirmsize_3	-0.584	(0.079)
_Ifirmsize_4	-0.552	(0.079)
_Ifirmsize_5	-0.473	(0.079)
_Ioccupatio_1	0.587	(0.012)
_Ioccupatio_2	0.524	(0.014)
_Ioccupatio_3	0.216	(0.009)
_Ioccupatio_4	0.038	(0.009)
_Ioccupatio_5	0.001	(0.008)
_Ioccupatio_6	0.016	(0.021)
_Ioccupatio_8	0.049	(0.007)
_Ioccupatio_9	-0.116	(0.007)
_Iactivity_1	-0.112	(0.018)
_Iactivity_2	0.150	(0.060)
_Iactivity_3	0.076	(0.064)
_Iactivity_5	0.060	(0.032)
_Iactivity_6	-0.188	(0.138)
_Iactivity_7	0.202	(0.075)
_Iactivity_8	-0.004	(0.038)
_Iactivity_9	0.159	(0.067)
_Iactivity_10	-0.014	(0.011)
_Iactivity_11	0.019	(0.050)
_Iactivity_12	0.183	(0.073)
_Iactivity_13	-0.234	(0.012)
_Iactivity_14	0.032	(0.012)
_Iactivity_15	0.056	(0.023)
_Iactivity_16	-0.093	(0.023)
_Iactivity_17	-0.030	(0.027)
_Iactivity_18	0.047	(0.028)
_Iactivity_19	0.098	(0.049)
_Iactivity_20	0.020	(0.023)
_Iactivity_21	0.268	(0.041)
_Iactivity_22	-0.018	(0.019)
_Iactivity_23	-0.034	(0.014)
_Iactivity_24	0.033	(0.017)
_Iactivity_25	-0.029	(0.016)
_Iactivity_26	0.160	(0.036)
_Iactivity_27	-0.003	(0.019)
_Iactivity_28	-0.013	(0.018)
_Iactivity_29	0.003	(0.017)
_Iactivity_30	0.261	(0.039)
_Iactivity_31	0.019	(0.016)
_Iactivity_32	0.078	(0.027)
_Iactivity_33	0.064	(0.029)
_Iactivity_35	0.074	(0.028)