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

Selin Pelek (Galatasaray University & MSU)
Gokce Uysal (Bahcesehir University & Betam & IZA)

NTA Annual Conference

Nov. 12th, 2016

Introduction

- Envelope wages are a hybrid form of informal employment
- OECD (2003), Woolfson (2007), Williams (2009)
- Main motivation is tax evasion
  - More common where taxes/state intervention are higher (Becker, 2004)
  - May be due to weak regulation/inadequate level of intervention (Ghezzi, 2009; Williams, 2012)
- Why do we care?
  - Limits employees' access to benefits and to financial services
  - Unfair competition among firms
  - Tax losses on the government's side

Recent evidence on underreporting of wages

- Hard to measure, little data
- EU Commission (2007) says 23% in CEE, 5% in WE
- No evidence on US
- Tonin (2011) finds a negative relationship between GDP and envelope wages.
- Turkey provides a nice setting:
  - According to Ministry of Labor and Social Security 2.2 million employees (18 % of wage earners) underreported in 2015.
  - Underreporting of wages exists in all sectors (Erdogdu, 2009).
  - An estimation is possible under minimal assumptions.

Our Contribution

- Questions:
  - How common are envelope wages in Turkey?
  - What percentage of total pay is paid in an envelope?
  - How much tax is evaded through underreporting?
- Contributions:
  - First quantitative study using nationally representative data sets
  - First study to estimate tax losses
  - First study on underreporting in Turkey
  - When no direct measure exists!

The institutional set-up

- Earnings taxed at source: firm pays own and worker's contributions.
- Some firms declare lower official wages and pay envelope wages.
- *Assumption*: If the firm chooses to underreport the earnings of a worker, the optimal decision would be to register the worker at the minimum wage under the assumption that the probability of getting caught is not a function of the extent of underreporting.
- Households receive and thus report their net earnings.
- *Assumption*: Households have no incentives to lie about their earnings.

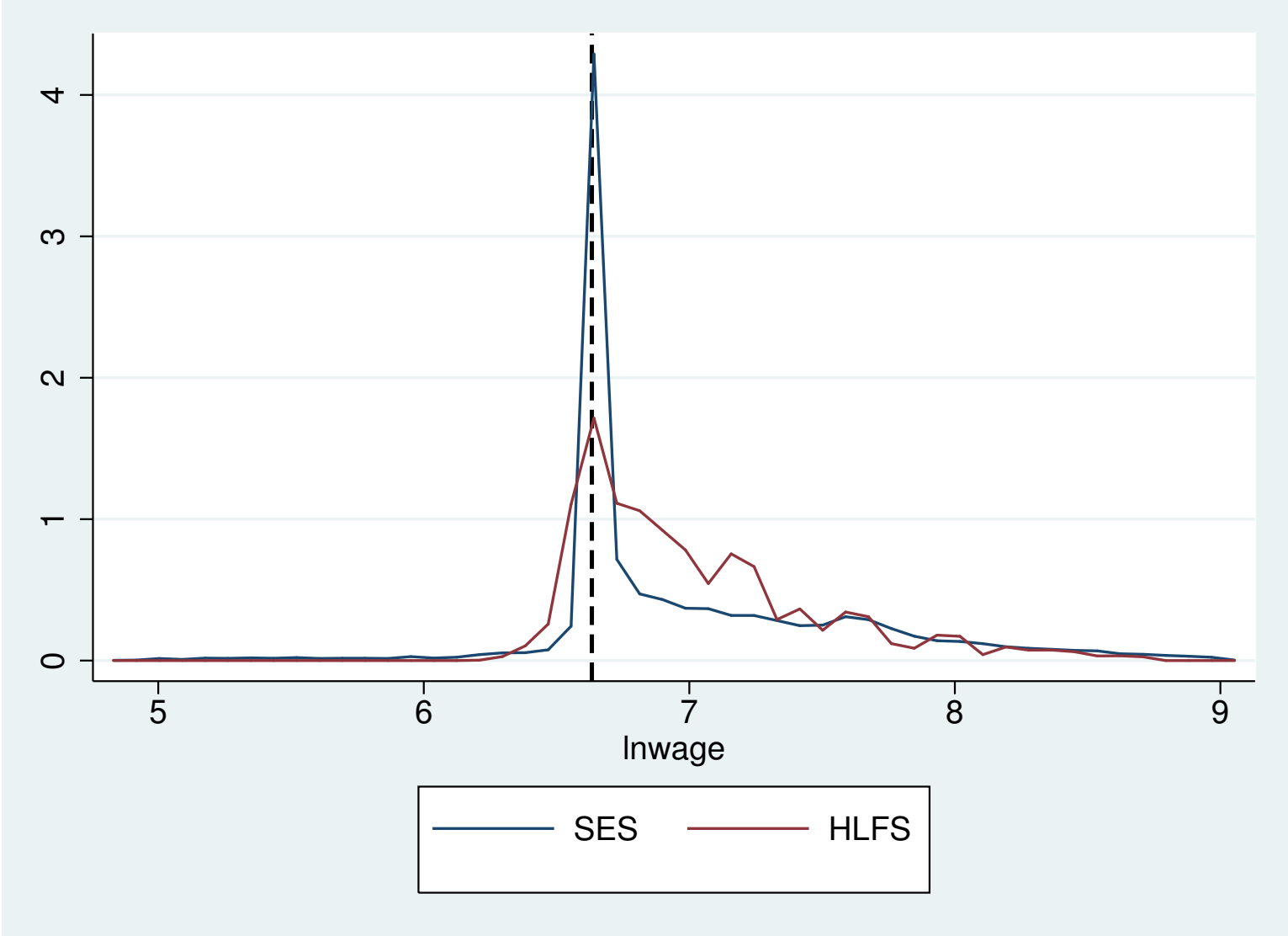
What do we expect to see in the data?

- If there is underreporting, the wage distributions of the firm level data and that of the household level data will differ significantly.
- **Structure of Earnings Survey (SES)**:
  - Covering only formal employment in firms with more than 10 employees
  - If underreporting is more severe among small firms, we underestimate the prevalence of underreporting
  - 2010 wave: 20 155 firms; 198 375 employees
- **Household Labor Force Survey (HLFS)**:
  - 2010 wave: 92 239 employees
- More employees registered at the minimum wage at the firm level (SES) than at the household level (HLFS).

A first look at the data: Harmonize

- Workers employed in firms with 10 or more employees
- Workers holding a formal job
- Transform net wages to gross wages using the OECD tax database
- Final samples: 198 360 in SES and 32 039 in HLFS
- The gross minimum wage in 2010: 760.5 TL, approximately 262 US$
- Share of workers with gross wages between 0.95 and 1.05 the monthly minimum wage.
- 43.2% in SES
- 16.3% in HLFS

A first look at the data: Kernel density plots



Estimation Strategy

$$\ln w_i^H = X_i^H \beta^H + \epsilon_i$$

- Under the assumption that the households report their actual wages, the true data generating process can be estimated by using household level data.
- Let the subscript i denote an individual in the household data (H) and the subscript j denote an individual in the firm data (F).
- A simple Mincer regression yields $\hat{\beta}^H$.

Estimation Strategy

$$\ln \hat{w}_j^F = X_j^F \hat{\beta}^H \quad (1)$$

$$\hat{u}_j = \ln \hat{w}_j^F - \ln w_j^F \quad (2)$$

1. 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.
2. The difference between the real and the estimated wage $\hat{u}_j$ is the estimate of the envelope wages.
3. Use $\hat{u}_j$ to estimate the total losses in tax revenue.
4. Compare $\ln \hat{w}_j^H = X_j^H \hat{\beta}^H$ and $\ln w_j^H$ to assess reliability.

Empirical Method

$$
\begin{aligned}
z_i^* &= y_i\gamma + \eta_i \\
\ln w_i &= X_i\beta + \epsilon_i \\
E[\ln w_i | z_i*] &= X_i\beta + \beta_\lambda \underbrace{\frac{\phi\left(y_i\gamma/\sigma_\eta\right)}{\Phi\left(y_i\gamma/\sigma_\eta\right)}}_{\lambda}
\end{aligned}
$$

- We use a Heckman sample selection model where the first stage is selection into the sample.
- y_i are the relevant characteristics of individual i.

Empirical Method

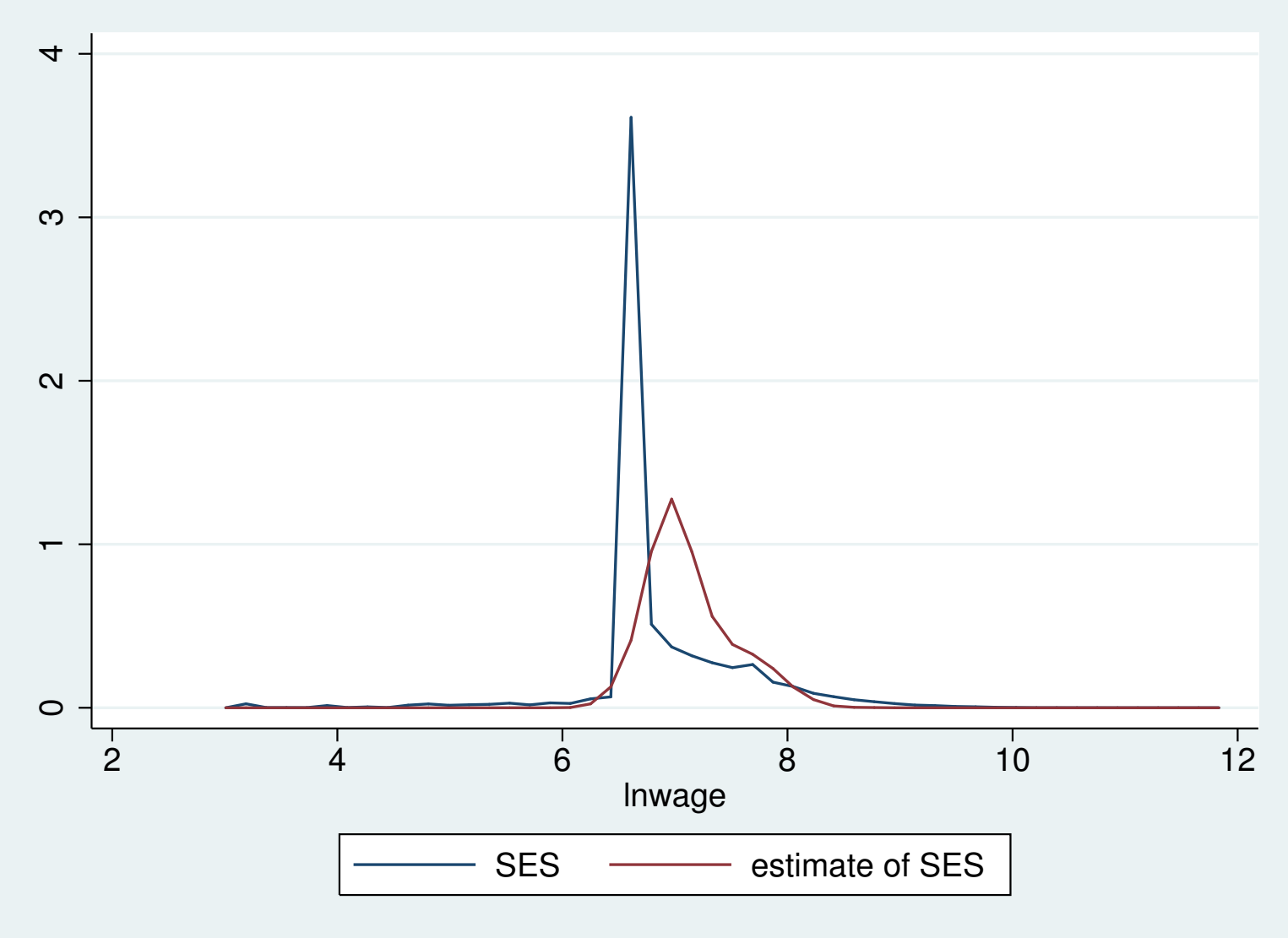
$$\ln \hat{w}_j^F = X_j^F \hat{\beta}^H + \hat{\beta}_\lambda^H \underbrace{\frac{\phi\left(y_j^F \hat{\gamma}^H / \hat{\sigma}_\eta^H\right)}{\Phi\left(y_j^F \hat{\gamma}^H / \hat{\sigma}_\eta^H\right)}}_{\hat{\lambda}^H}$$

- 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.
- 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.

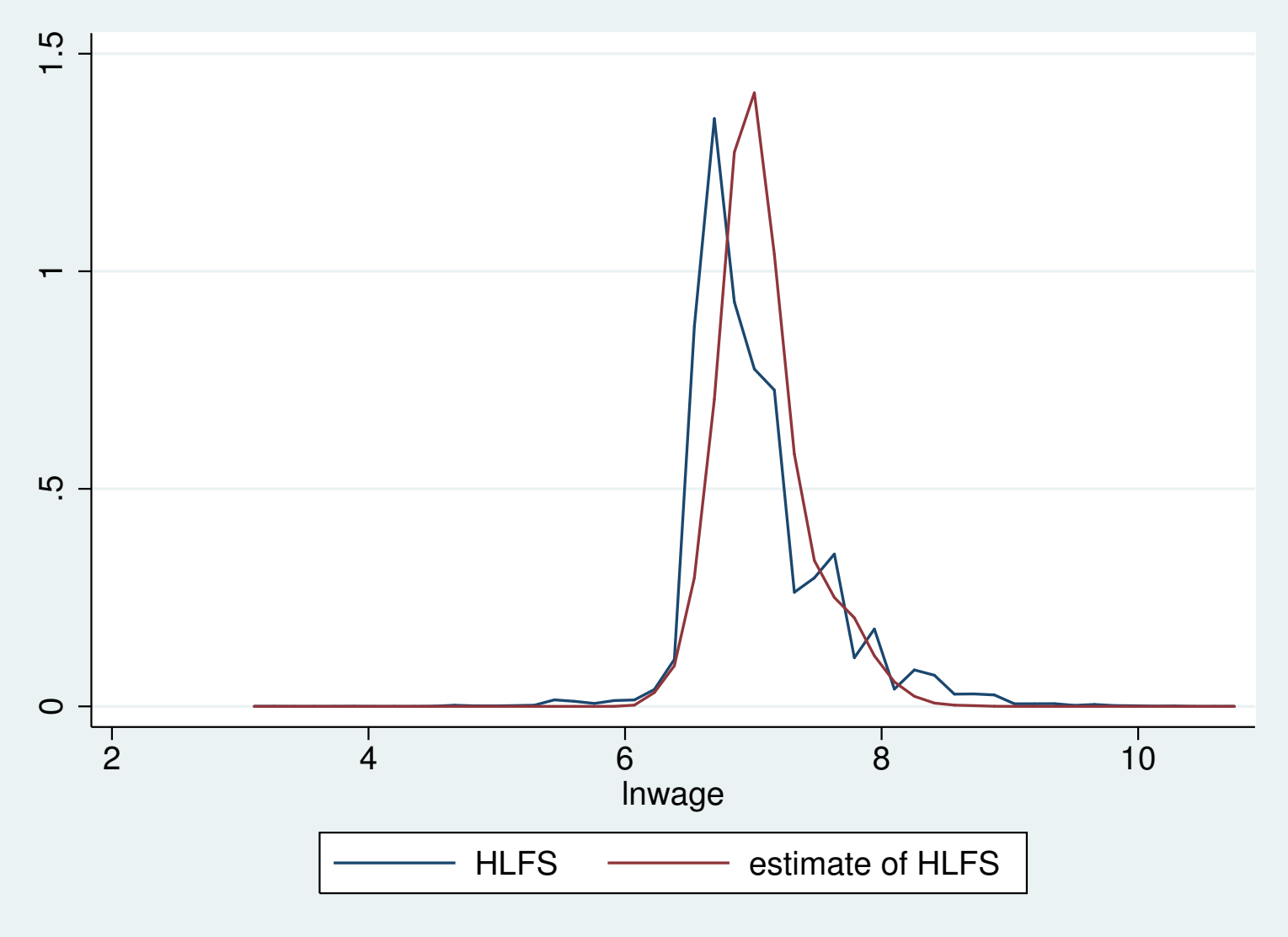
Results

- An estimated (hypothetical) wage ln $\hat{w}_j^F$ for all observations in the firm level data set.
- Drawing Kernel density estimations of these hypothetical wages

SES estimates with sample selection



Performance of the sample selection model



Estimates of Underreporting

percentiles	observed wages	estimated wages	envelope wages
10%	760	1095	335
20%	760	1095	335
30%	761	1098	337
40%	761	1109	348
50%	796	1127	331
60%	930	1264	334
70%	1207	1406	199
80%	1710	1409	-301
90%	2540	2297	-243

Estimates of Tax Evasion

- **Approach 1:**
- Assume negative envelope wages do not exist.
- Negative envelope wages are irrelevant for tax evasion
- High wage earners could have some unobservable characteristics
- 135 241 with positive envelope wages
- Estimated tax loss for the sub-sample : 37.4 %
- 208 TL (69.5 USD) *per capita*
- Estimated tax loss for the full-sample : 24.3 % of total collected tax

Estimates of Tax Evasion

- **Approach 2:**
- Assume that underreporting of wages is only a problem at the lower end of the wage distribution.
- An upper limit for two minimum wages.
- Estimated tax loss for the sub-sample : 35.1 %
- 163 TL (56 USD) *per capita*
- Estimated tax loss for the full-sample : 20.9 % of total collected tax

Correction à la Blackburn (2007)

	Observed wage	Log wage regression	Squared residuals regression	Absolute values of res. reg.	Exponentiated res. reg.
10	760	956.1	1140.6	1162.1	957.8
20	760	956.1	1140.6	1162.1	957.8
30	760.5	950.5	1132.6	1153.5	941.1
40	761	973.8	1170.1	1188.6	983.3
50	796	972.3	1141.9	1174.7	1056.6
60	930	1131.2	1309.3	1372.8	1178.1
70	1207	1188.8	1301.9	1383.9	1250.9
80	1710	1342.3	1709.3	1708.2	1467.9
90	2540	1906.3	2245.5	2407	2738.3

QMLE à la Blackburn (2007)

Percentiles	Observed wage	Nonlinear regression	Gamma (qml)	Poisson .(qml)	Gaussian (qml)
10	760	1090	1037.8	1052.9	1067.2
20	760	1090	1037.8	1052.9	1067.2
30	760.5	1075.6	1033	1048	1062.1
40	761	1120.3	1053.2	1069.3	1084.4
50	796	1114.4	1046.6	1061.8	1075.7
60	930	1308.1	1199.4	1223.1	1247.8
70	1207	1216.6	1247.5	1255.4	1254.7
80	1710	1647.5	1420.1	1440.8	1466.7
90	2540	2714.7	1885.5	1935.8	1974.9

Conclusion

- Underreporting of wages is widespread and tax losses are sizeable in Turkey
- The difference between the distribution of wages coming from individual and firm based data could be explained by underreporting practice of firms.
- Recent increase (30% as of January 2016) of the minimum wage could reduce envelope wages.
- That is, if the job is still a formal one!

Tax Evasion

APPROACH 1 (only positive envelope wages)

Linear Regression	Observed tax	Estimated tax	Tax evasion amount	Observed Tax/ Estimated tax (%)
Log wage	38 535 188	58 548 252	20 013 064	66
Squared res.	53 140 792	92 613 360	39 472 568	57
Absolute res.	55 622 116	96 423 272	40 801 156	58
Exponential res.	38 535 188	62 314 660	23 779 472	62
Exponential Reg.				
Nonlinear reg.	54 372 808	94 599 544	40 226 736	58
Gamma (qml)	43 595 444	67 371 304	23 775 860	65
Poisson (qml)	45 299 632	70 722 064	25 422 432	64
Gaussian (qml)	46 730 096	73 669 208	26 939 112	63

Tax Evasion

APPROACH 2 (UP TO 2 MW)

Linear Regression	Observed tax	Estimated tax	Tax evasion amount	Observed Tax/ Estimated tax (%)
Log wage	47 549 284	62 160 900	14 611 616	76
Squared res.	47 549 284	78 464 496	30 915 212	61
Absolute res.	47 549 284	79 654 088	32 104 804	60
Exponential res.	47 549 284	64 741 980	17 192 696	73
Exponential Reg.				
Nonlinear reg.	47 549 284	75 485 320	27 936 036	63
Gamma (qml)	47 549 284	67 466 952	19 917 668	70
Poisson (qml)	47 549 284	68 900 816	21 351 532	69
Gaussian (qml)	47 549 284	70 205 512	22 656 228	68