MODELING AND FORECASTING INCOME TAX REVENUE: THE CASE OF UZBEKISTAN

Marat Ibragimov

Center for Economic Research, Tashkent, Uzbekistan

Rustam Ibragimov*

Department of Economics, Harvard University, Cambridge, MA, USA

Nishanbay Sirajiddinov

Center for Economic Research, Tashkent, Uzbekistan

* Corresponding author. Department of Economics, Harvard University, Littauer Center, 1805 Cambridge St., Cambridge, MA 02138. Phone: (617) 496-4795; Fax: (617) 495-7730.

E-mail address: ribragim@fas.harvard.edu (Rustam Ibragimov)

Abstract

Income tax revenue crucially depends on the wage distribution across and within the industries. However, many transition economies present a challenge for a sound econometric analysis due to data unavailability. The paper presents an approach to modeling and forecasting income tax revenues in an economy under missing data on individual wages within the industries. We consider the situations where only the aggregate industry-level data and sample observations for a few industries are available. Using the example of the Uzbek economy in 1995-2005, we show how the econometric analysis of wage distributions and the implied tax revenues can be conducted in such settings. One of the main conclusions of the paper is that the distributions of wages and the implied tax revenues in the economy are well approximated by Gamma distributions with semi-heavy tails that decay slower than those of Gaussian variables.

JEL classification: C13, D31, H24, P20

Keywords: Income tax; Wage distribution; Inequality

1. Objectives of the paper

This paper discusses an approach to econometric analysis and forecasting for income tax revenue in situations where the data on wages and income are largely unavailable. Naturally, income tax revenue crucially depends on the wage distribution in the economy. However, many transition economies present a challenge for a sound econometric analysis due to data unavailability. This is typically the case for employee-level data, including wages and income.

We show that, in certain cases, one can make inference on wage distribution and the implied distribution of income tax revenues even in absence of microeconomic level data for individual households. We discuss an approach that allows one to estimate the form and parameters of the wage distribution in the economy using the official data on average wages across the industries. We then use these results to obtain estimates for the distribution of tax revenues. One of the main conclusions of the paper is that Gamma densities provide appropriate models for wage distributions. These approximations imply that the distributions of wages and the implied tax revenues have semi-heavy tails that decay slower than those of Gaussian variables.

The analysis presented in the paper is based on the official data on average wages across the industries and a sample of data for employees' wages in one of the industries, construction. The analysis thus uses the existing minimum of the available information. The estimates presented are, therefore, some of the only results that are possible to obtain in such settings.

2. Inference on wage distribution under missing household income data

Tables A1 and A2 in the appendix provide government data on the number of employees and the average wages across the industries of the Uzbek economy. Table 1 below provides a summary of the implied inequality measures for the wage distribution in the Uzbek economy. The inequality measures summarized in the table and their properties are discussed in detail, among others, in Ch. 13 in Marshall and Olkin (1979).

Table 1. Statistical characteristics of wage distribution in the Uzbek economy in 1995, 2000 and 2005

	1995	**2000**	**2005**
Minimal wage, soum	250	1320	9400
Average wage, $\overline{w}$, soum	1057	11225.1	85865.0
Gini coefficient[1]	0.235	0.233	0.252
The number of employees with wages less than $\overline{w}$	60%	43%	42%
Minimal majority[2]	68%	67,4%	67,5%

Table 2 presents Gini coefficients for the wage distribution in the Uzbek economy computed under the assumption that wages are uniformly distributed within each industry.

Table 2. Gini coefficients for wages in Uzbekistan under the assumption of uniform wage distribution within the industries

	1995	**2000**	**2005**
Gini coefficient	0.161	0.253	0.262

The Gini coefficient values reported in Tables 1 and 2 are considerably small comparing to other Newly Independent States. These values do not reflect the true inequality in wage distribution in the total economy.

The inequality in the total economy reflects both the disparities among the industries and inequalities in distributions within them. Thus, taking into account the wage inequality within the industries would lead to an increase in the calculated Gini coefficients and other inequality measures considered in Table 1.

[1] The Gini coefficient is computed under the assumption that each industry of the economy is considered as an individual unit regardless of the number of employees in the industry. In other words, here, the Gini coefficient characterizes the inequality in distribution of the average wage across the industries.

[2] In the context of political science, the minimal majority is the smallest number of individuals controlling a majority of legislature. This index can also be used as an inequality measure and has a simple expression in terms of the Lorenz curve characterizing the inequality among subjects: if wages $w_1,..., w_n$ determine the Lorenz curve *h*, then the minimal majority is $h^{-1}(0.5)$ (see Ch. 13 in Marshall and Olkin, 1979). Thus, in Table 1, minimal majority equals to the proportion of the employees who have 50% of the total income (wage fund).

As an example, we estimate the wage distribution within the construction industry. This wage distribution is estimated using a sample of wages in 1995 for 256 employees in the industry available to us (see Table A3 and Figure A1 in the appendix). The Gini coefficient for the construction industry estimated using the sample is 0.29.

Consider a random variable (r.v.) X with truncated normal density

$$f(x) = \begin{cases} \dfrac{1}{A\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, & if \;\; x > a; \\ 0, \;\; if \;\; x \le a. \end{cases} \tag{1}$$

where the truncation level a=250 (soum) is the minimal wage in 1995 and $A = \dfrac{1}{\sigma\sqrt{2\pi}} \int_a^{\infty} e^{-(x-\mu)^2/(2\sigma^2)} dx = 1 - \Phi\left(\dfrac{a-\mu}{\sigma}\right)$ is a normalizing constant.

If a is known, one may use the relations for the mean $E[X]$ and the variance $Var[X]$ of X given by

$$\begin{cases} E[X] = \mu + \sigma B \\ Var[X] = \sigma^2 (1 - B^2 + \dfrac{a-\mu}{\sigma} B) \end{cases} \tag{2}$$

where $B = \dfrac{\varphi\left(\dfrac{a-\mu}{\sigma}\right)}{A}$, to estimate μ and σ by the method of moments (see Section 19.3-4 in Korn and Korn, 1968, and Section 10.1 in Johnson, Kotz and Balakrishnan, 1994).

The p-value for the Kolmogorov-Smirnov test of the null hypothesis that wage in the construction industry follows the above distribution equals to α=0.21. Thus, the null hypothesis is not rejected at the $\alpha \le 0.21$ significance level (see Figure A2 in the appendix).

The coefficient of variation for the fitted truncated normal distribution is $CV = \dfrac{\sqrt{Var[X]}}{|E[X]|} = 0.51$.

To estimate the wage distribution in the whole economy we assume that, in each industry, wage has a truncated normal distribution with the same coefficient of variation CV=0.51 as in construction, that is, the parameters $E[X]$ and $Var[X]$ satisfy $\sqrt{Var[X]} = 0.51E[X]$. The wage distribution for the whole economy is a mixture of wage distributions within the industries. The cdf of wage in the whole economy is determined as

$$F(x) = \frac{N_1}{N} F_1(x) + ... + \frac{N_m}{N} F_m(x) , \qquad (3)$$

where $N=N_1+...+N_m$ is the total number of employees in the economy, N_j is the number of employees in the *i*th industry, and $F_j(x)$ is the cdf of wage in the *j*th industry, $j=1, ..., m$.

Figures A3-A5 in the appendix provide the cdf's for wages in the Uzbek economy in 1995, 2000 and 2005. The calculations are based on the above assumption of equal coefficients of variation for wage distributions within the industries. Evidently, this assumption may be lifted if, in addition to the data on the average wages in each of the industries, one also has the data on the wage variances within the industries.

Table 3 provides the means, variances and Gini coefficients for the fitted wage distribution in the Uzbek economy in 1995, 2000 and 2005. Columns 5 and 6 of Table A1 and columns 5, 6, 10 and 11 of Table A2 in the appendix provide the parameters μ and σ for truncated normal distributions (1) and the distributions' means $E[X]$ and variances $Var[X]$ calculated using (2) for each of the industries in 1995, 2000 and 2005.

Table 3**.** Statistical characteristics of wage distribution in the Uzbek economy in 1995, 2000 and 2005

	1995	2000	2005
Mean	1307	19973	136667
Standard deviation	808	12968	89396
Gini coefficient	0.321	0.356	0.359

The interpretation of Gini coefficients presented in Tables 1-3 can be described as follows.

1. The coefficients in Table 1 correspond to the case where one measures the inequality among the industries using known values of average wages in them. This equals to the true inequality index for the whole economy if wage is equally distributed within each of the industries and, in addition, the number of employees in the industries is the same.

2. The estimates in Table 2 correspond to the case where the average wages and the number of employees in the industries are known. Wage distributions within the industries are assumed to be uniform.
3. Table 3 provides the estimates of the wage inequality in the whole economy where the inequality in wage distributions within the industries is taken into account.

One should note that the Gini coefficient values in Table 1 provide lower bounds for the true Gini coefficients for the whole economy. The inequality indices in Table 3 may be regarded as the most appropriate ones. In addition, they are the closest to the true inequality measures under some natural general assumptions (see the discussion and related results in Ibragimov and Walden, 2007).

According to the results in this section, the inequality in wage distribution over the whole economy is relatively small, even when the inequalities within the industries are taken into account. The situation is similar to that in Russia in the first half of 90's where one observed "the power of institutional features in the wage settings that tended to dominate the redistributive effects transmitted through high inflation and decentralization in wage settings" (Commander, McHale and Yemtsov, 1995, p. 165) and relative stability of the low inequality among the industries.

3. Income tax revenue distribution and forecasting

We follow the following two criteria in approximating the wage distribution in the economy:

1. The distribution law should be widely known and parsimonious with a relatively small number of parameters;
2. The measure of approximation should have a clear interpretation.

Among two parameter distributions, the best fit appeared to be provided by gamma distribution with cdf

$$F_{gamma}(x;k;\theta) = \int_0^x f(u;k;\theta)du,$$

where $f(x;k;\theta) = x^{k-1}\dfrac{e^{-x/\theta}}{\theta^k\Gamma(k)}$, $x > 0$, $k,\theta > 0$ is the pdf of the distribution. The gamma cdf $F_{gamma}(x;k;\theta)$ has tails that asymptotically decay as $x^{\alpha}\exp(-\beta x)$ as $x \to \infty$. Thus, the tails of the distribution belong to the class of semi-heavy tails that decay slower than those of Gaussian distribution but faster than any power law tails.[3]

The shape and scale parameters k and θ are estimated by the method of moments using the relations

$$\begin{cases} E[X] = k\theta \\ Var[X] = k\theta^2 \end{cases}$$

for the mean $E[X]$ and variance $Var[X]$ of a gamma r.v. X with the pdf $f(x;k;\theta)$.

Table 4 provides a summary of the implied approximation for the wage cdf F estimated in (3) in the case where the measure $\Delta_{gamma} = \max|F(x) - F_{gamma}(x;k;\theta)|$ is used as a measure of approximation. For comparison, Table 4 also presents the values $\Delta_{normal} = \max|F(x) - F_{normal}(x;k;\theta)|$ for the normal distribution with the parameters $m=E[X]$ and $\sigma = \sqrt{Var[X]}$ (see also Figures A3-A5 in the appendix).

Table 4. Gamma distribution approximation results.

	1995	**2000**	**2005**
$E[X]$	1307.43	19973.28	136666.91
$Var[X]$	653450.74	168164070.26	7991717405.72
k	2.615916	2.372278605	2.337150
θ	499.798148	8419.450581	58475.876093
Δ_{gamma}	0.0496	0.0237	0.02533
Δ_{normal}	0.0977	0.0808	0.0792

As is seen from the table, the gamma distribution provides better approximation to $F(x)$ than does the normal distribution.

[3] Such tails were considered among others, by Barndorff-Nielsen (1997) and Barndorff-Nielsen and Shephard (2001) in the context of applications of Normal Inverse Gaussian distributions and their extensions and by Malevergne, Pisarenko and Sornette (2005) who consider the fitting by stretched exponential and related distributions. Together with widely used power law distributions applied in many works (see, for instance, the discussion in Loretan and Phillips, 1994, Gabaix, Gopikrishnan, Plerou and Stanley, 2003, Ibragimov, 2005, Rachev, Menn and Fabozzi, 2005, and references therein), semi-heavy tailed distributions were reported to provide good fit for a number of economic and financial time series.

The results obtained can be applied to forecast the income tax revenue using the tax rate and wage data. We describe the approach to forecasting using the parameters $E[X]$ and $Var[X]$ for 2005.

Denote by T_{total} the yearly income tax revenue in the economy, in million soum, and by w_{min} the minimal monthly wage, in soum. Let $T(x)$ denote the monthly income tax revenue from one employee with wage x.

The tax rates in Uzbekistan in 2005 are provided in Table A4 in the appendix.

The function relating the income tax $T(x)$ to the wage x has the following form:

$$T(x) = \begin{cases} t_1 x, \ \ 0 < x \le w_1, \\ t_1 w_1 + t_2(x - w_1), \ \ w_1 < x \le w_2, \\ \dots \\ t_1 w_1 + t_2(w_2 - w_1) + \dots + t_n(x - w_{n-1}), \ \ x > w_{n-1} \end{cases}, \tag{4}$$

where x is the wage, t_i are the tax rates for different wage levels, i=1, ..., n; $[w_{\min}; w_1]$, $[w_1; w_2]$, ..., $[w_{n-1}; \infty)$ is the partition of the set of possible wage values $[w_{min};\infty)$ into strata that correspond to the different tax rates (see Table A4 in the appendix). Thus function (4) has the following form for the Uzbek economy in 2005:

$$T(x) = \begin{cases} 0.13x, \ \ 0 < x \le 5w_{\min}, \\ -0.4w_{\min} + 0.21x, \ \ 5w_{\min} < x \le 10w_{\min}, \\ -1.3w_{\min} + 0.3x, \ \ x > 10w_{\min}, \end{cases}$$

where w_{min}=9400 soum.

As before, we assume that the wage level X has the gamma distribution with the cdf $F_{gamma}(x; k; \theta)$ and the pdf $f(x; k; \theta)$. Therefore, the cdf of the r.v. $t=T(x)$ is given by $F_{gamma}(T^{-1}(t); k; \theta)$ and the pdf of $t=T(x)$ is $f(T^{-1}(t); k; \theta)$ (see Figure A6 in the appendix).

According to the official site of the Ministry of Finance of the Republic of Uzbekistan[4], the income tax revenue in the Uzbek economy in 2005 was equal to T_{actual}=465641.1 million soum.

To compare income tax revenues predicted by the above results with the actual values, we make the following assumptions:

[4] http://www.mf.uz/

1. The monthly income tax revenue equals to $1/12$ of yearly tax revenue;
2. The income tax revenue from each employee equals to $1/N$ of the income tax revenue over the whole economy, where N is the number of employees included in calculation of the tax base. The monthly income tax revenue from each employee is the r.v. $T(x)$ with the cdf $F_{gamma}(T^{-1}(t);k;\theta)$.

Table 5 presents the estimates for the tail probabilities of yearly income tax revenues over the whole economy that hold under these assumptions.

Starting with 2000, the average wage for agriculture has not been calculated or reported by the State Committee for Statistics (Goskomstat) of the Republic of Uzbekistan due to the absence of data for the private sector (see Tables A1 and A2 in the appendix). Therefore, the income tax revenues from agriculture seem to be negligible in 2000 and 2005. Nevertheless, we present the estimates for the tail probabilities of the income tax revenue for both the cases where N equals to the total number of employees in the economy and where N does not include the employees in agriculture.

Table 5. Tail probabilities of the total income tax revenues in 2005

N, тыс.чел.	$P(T_{total} > T_{actual}) = P(T_{total} > 465641.1$ million soum)
10196.3: Total number of employees in all the industries of the economy	0.9496
7110.6: Total number of employees in all industries except agriculture	0.8988

The results in Table 5 suggest very high probability of the total tax revenue in the model being not less than the actual value of 465641.1 million soum in 2005. This probability is estimated to be about 0.9 if agriculture is excluded in calculation of the tax base, and to be about 0.95 if the estimate is calculated using the total number of employees in the economy.

4. Conclusion

The approach discussed in the paper allows one to obtain estimates of the income tax revenues via gamma distribution approximations using data on average wages in the economy and wage variances. At the same time, the calculations are based on several strict assumptions, in particular, those listed below

1. Within each industry, wage distribution is truncated normal;
2. Wage variances in all the industries are the same;
3. Wage in the whole economy follows a gamma distribution.

Each of these assumptions is very strong. The assumptions are motivated only by the absence of sample data across all the industries. Even if such studies are conducted in transition economies, their results are often inaccessible. On the other hand, the approach to wage distribution modeling and forecasting the implied income tax revenues discussed in the paper is based on the existing minimum of the available data. The estimates presented are, therefore, some of the only results that are possible to obtain in such circumstances.

References

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and the modeling of stock returns. *Scandinavian Journal of Statistics* **24**, 1-13.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **63**, 167-241.

Commander, S., McHale, J. and Yemtsov, R. (1995). *Russia.* In S. Commander and F. Coricelli, eds "*Unemployment, Structural Transformation and Labor Market in the Eastern Europe and Russia",* EDI Development Studies, World Bank, Washington, DC. , pp. 147-191.

Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature* **423**, 267-270.

Ibragimov R. and Walden, J. (2007). Value at risk under dependence and heavy-tailedness: Models with common shocks. *Working paper, Harvard University and University of California at Berkeley*.

Ibragimov, R. (2005), *New majorization theory in economics and martingale convergence results in econometrics*, Ph.D. dissertation, Yale University.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions*. Vol. 1, Second edition. John Wiley, New York.

Korn, G. A. and Korn, T. M. (1968). *Mathematical handbook for scientists and engineers.* Second edition, McGraw-Hill, New York.

Loretan, M. and Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* **1**, 211-248.

Malevergne, Y., Pisarenko V. and Sornette D**.** (2005). Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance* **5**, 379-401.

Marshall, A. W. & Olkin, I. (1979), *Inequalities: theory of majorization and its applications*, Academic Press, New York.

Rachev, S. T., Menn, C. and Fabozzi, F. J. (2005). *Fat-tailed and skewed asset return distributions: Implications for risk management, portfolio selection, and option pricing*, Wiley, Hoboken, NJ.

Appendix

Table A1. Wage distribution in the industries of Uzbek economy in 1995

Industries	Number of employees, N, thousands	Average wage, $E[X]$, soum	Variance, $Var[X]$	μ	σ
1	**2**	**3**	**4**	**5**	**6**
1.Heavy industry	854.9	1529	608072.4	1264	973.14
2. Horticulture	378.1	808	169809.9	342.32	655.49
3. Animal husbandry	84.3	620	99982.44	-522.74	723.05
4. Forestry	5.3	694	125273.5	-3.9443	659.67
5.Transportation	259.5	1427	529649.2	1158	919.95
6.Communications	43.4	1521	601726	1255.8	968.94
7.Construction	320.5	1655	712420.4	1392.4	1039.9
8. Construction-related services	13.9	1915	953845.2	1651.4	1180.1
9. Trade	194.4	556	80406.27	-1726.4	888.4
10. Public catering	46.9	528	72511.72	-1733.8	860.71
11.Computer services	2.7	1716	765905	1453.7	1072.6
12. Housing and utilities	116.9	1064	294458.2	747.27	743.15
13. Health	469.6	665	115022.7	-153.27	674.25
14.Education	858.5	577	86594.83	-1317.8	840.35
15.Arts and culture	62.4	770	154213.3	251.76	650.92
16.Science	41.6	1209	380183.2	920.57	810.42
17.Insurance and pensions	38	1789	832455.5	1526.7	1111.8
18.Administration	87	1041	281865.4	717.77	733.17
19.Other industries	252.1	949	234246.3	590.78	696.16

The values μ and σ are the parameters of the truncated normal distribution (1) fitted to the industry data, and $Var[X]$ is the variance of the distribution calculated using (2).

Source: The State Committee for Statistics (Goskomstat) of the Republic of Uzbekistan (Columns 1-3) and the authors' calculations (Columns 4-6).

Table A2. Wage distribution in the industries of Uzbek economy in 2000 and 2005

Industries	**2000**					**2005**				
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
	N	$E[X]$	$Var[X]$	μ	σ	N	$E[X]$	$Var[X]$	μ	σ
1.Heavy industry	1145	21861.5	124308074	19970	12774	1145	145364.2	5496104461	132430.0	85171.0
2. Construction	676	18796.6	91896342	17064	11054	676	140114.8	5106320438	127460.0	82226.0
3. Transportation	313.7	18569.3	89687023	16848	10926	367.97	134430.6	4700419575	122060.0	79038.0
4. Communications	68.3	25081.1	163619392	23016	14584	80.1	135686.9	4788680620	123250.0	79742.0
5. Trade and public catering	754	9815.6	25059452	8427.6	6070.5	754	70209.0	1282111887	60321.0	43398.0
6. Health	587	8177.2	17392138	6780.9	5193	587	48411.0	609576842	37913.0	31924.0
7. Education	1054	9317.3	22579624	7932.6	5801.2	1274	55454.2	799851274	45420.0	35524.0
8. Arts and culture	61.02	9591.9	23930226	8205.8	5949.4	73.75	57977.2	874287115	48027.0	36846.0
9. Science	30.98	16053.3	67029991	14453	9518.6	37.445	90070.1	2110090877	83956.0	53187.0
10. Housing and utilities	251	13265.4	45769964	11784	7966.4	251	84703.9	1866150447	74494.0	51332.0
11. Credits and insurance	52	24834.8	160420567	22783	14446	52	198770.0	1.0276E+10	182950.0	115210.0
12. Other industries	904.3	30779.5	246412885	28397	17793	1812.4	200156.0	1.042E+10	184260.0	115990.0

The notations used are the same as in Table A1.
Source: The State Committee for Statistics (Goskomstat) of the Republic of Uzbekistan (Columns 1-3, 7, 8) and the authors' calculations (Columns 4-6, 9-11).
Note: Starting with 2000, the average wage for agriculture has not been calculated by the Goskomstat due to the absence of data for the private sector.

Table A3. Wage distribution in the construction industry

W, %	*n*, %		*W*, %	*n*, %		*W*, %	*n*, %
0-4	1,56		74-79	3,13		148-153	1,56
4-9	1,17		79-83	2,73		153-157	0,39
9-13	3,13		83-87	1,95		157-161	1,17
13-17	1,56		87-92	2,73		161-166	0,78
17-22	1,56		92-96	3,52		166-170	1,56
22-26	1,17		96-100	4,69		170175	0,78
26-31	1,17		100-105	3,13		175-179	0,78
31-35	1,56		105-109	4,30		179-183	1,56
35-39	2,34		109-113	2,34		183-188	1,17
39-44	3,13		113-118	3,52		188-192	1,17
44-48	0,39		118-122	5,08		192-196	0,39
48-52	1,56		122-127	3,13		196-205	0,39
52-57	3,52		127-131	3,13		205-218	0,78
57-61	1,95		131-135	2,73		218-236	0,39
61-65	2,34		135-140	3,52		236-246	0,39
65-70	1,95		140-144	1,95		>246	0,39
70-74	1,95		144-148	2,73			

Based on a sample of 256 employees. *W*,% denotes the average wage in percent to the average wage in the sample; *n*,% is the percentage of employees in the sample who receive the wage within the indicated ranges.

Table A4. Income tax rates in Uzbekistan in 2005

Range of the income, *I*	Corresponding tax amount, *T(I)*
$I \le 5w_{\min}$	$T(I) = 0.13I$
$5w_{\min} < I \le 10w_{\min}$	$T(5w_{\min}) + 0.21(I - 5w_{\min})$
$I > 10w_{\min}$	$T(10w_{\min}) + 0.3(I - 10w_{\min})$

The value *I* denotes the income to date from the beginning of 2005; $w_{\min}$ denotes the minimal wage in 2005. Source: State Tax Committee of the Republic of Uzbekistan

Figure A1. Wage frequency distribution in the construction industry

Based on a sample of 256 employees. *Wage,%* denotes the average wage in percent to the average wage in the sample. *Number of workers,%* is the percentage of employees in the sample who receive the wage within the indicated ranges.

Figure A2. The cdf of wage in the construction industry and the cdf of the fitted truncated normal distribution

Figure A3. The cdf $F(x)$ for the wage distribution in the Uzbek economy in 1995 and the cdf's of the fitted gamma and normal distributions

Figure A4. The cdf $F(x)$ for the wage distribution in the Uzbek economy in 2000 and the cdf's of the fitted gamma and truncated normal distributions

Figure A5. The cdf $F(x)$ for the wage distribution in the Uzbek economy in 2005 and the cdf's of the fitted gamma and truncated normal distributions

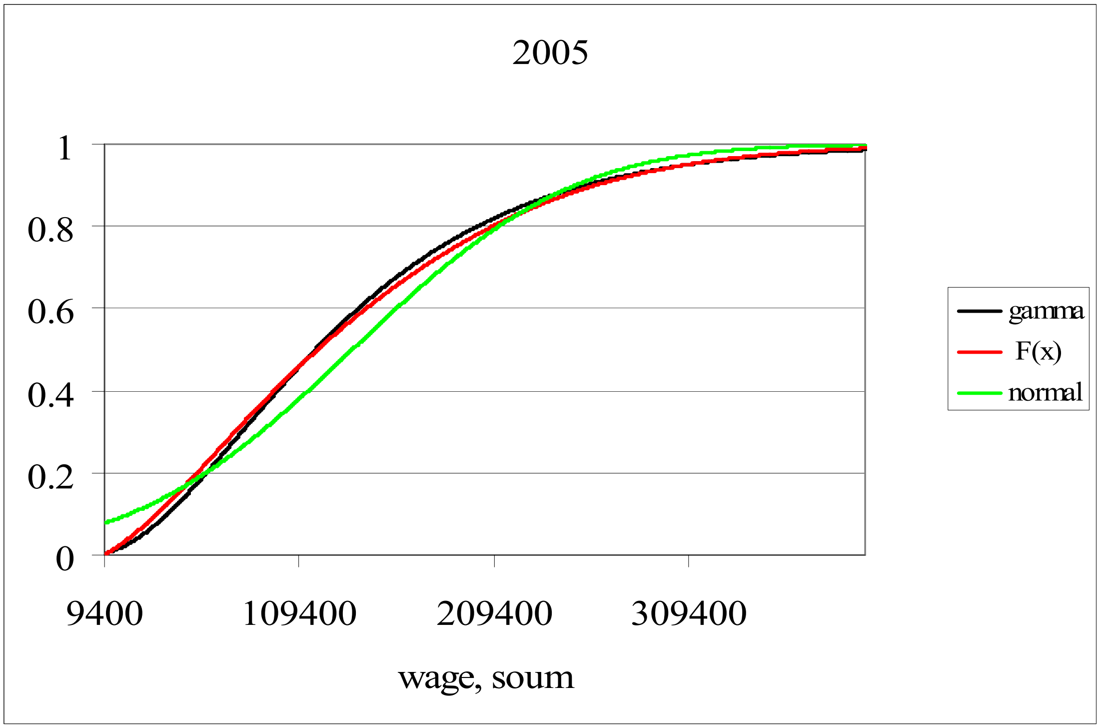

Figure A6. The pdf of the estimated income tax revenue from one employee in 2005