DISCUSSION PAPER SERIES

IZA DP No. 410

Wage Arrears and the Distribution of Earnings in Russia

Hartmut Lehmann
Jonathan Wadsworth

December 2001

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Wage Arrears and the Distribution of Earnings in Russia

Hartmut Lehmann
Herriot-Watt University, Edinburgh, WDI, University of Michigan Business School, Russian Center for Labor Studies, Moscow Higher School of Economics and IZA, Bonn

Jonathan Wadsworth
Royal Holloway College, University of London, Centre for Economic Performance, LSE, WDI, University of Michigan Business School and IZA, Bonn

Discussion Paper No. 410
December 2001

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

IZA Discussion Paper No. 410
December 2001

ABSTRACT

Wage Arrears and the Distribution of Earnings in Russia*

The increase in wage inequality in Russia during its transition process has far exceeded the increase in wage dispersion observed in other European countries undergoing transition. Russia also has an extremely large incidence of wage arrears. We analyse to what extent wage arrears affect the wage distribution and measures of wage inequality in Russia. We present counterfactual distributions, derived from a variety of different methods, which suggest that conventional measures of earnings dispersion would be some 20 to 30 per cent lower in the absence of arrears. We then go on to show how wage gaps at various points in the pay distribution across gender, education, region and industry are influenced by a failure to allow for wage arrears. Using our counterfactual estimates we show, for example, that the median gender wage gap would be around twenty-five points higher than the actual gap that we observe. Similarly, the counterfactual ratio of mean graduate pay to mean pay of those with primary education is around twenty points lower than observed. We show that the parameters of the counterfactual wage distributions are very similar to the parameters of the observed wage distributions of those not in arrears. This means that for those wishing to study aspects of wage differentials and inequality in Russia, it may be feasible to use the subset of those not in arrears and still get close to the true population parameters.

JEL Classification: J6

Keywords: Wage arrears, earnings inequality, counterfactuals, transition economies

Hartmut Lehmann
Heriot-Watt University
Edinburgh EH14 4AS
United Kingdom
Tel.: 44-131-451 3626
Fax: 44-131-451 3008
E-mail: H.Lehmann@hw.ac.uk

* This paper took its inspiration from previous joint work with Ruslan Yemtsov, who has moved on to more important work at the World Bank.
The authors are grateful to John DiNardo and Jochen Kluve for making their STATA routines available; to Boris Augurzky, Todd Idson, Alan Manning, Mark Schaffer, Christoph Schmidt, Jan Svejnar, Paul Walsh, participants at the IZA-WDI Conference on Labour Markets in Transition Countries in May 2000 in Bonn, at the CEPR-WDI Conference on Transition in Moscow and at an ACES/Econometric Society session of the ASSA Conference in January 2001 in New Orleans as well as participants in seminars at the Universities of Heidelberg, Heriot-Watt University, the London School of Economics, the University of Trento and the WDI, University of Michigan for valuable comments. The Fritz Thyssen Foundation provided financial support within the project "Economic Reform and Labour Market Adjustment in the Russian Federation." The authors are also grateful to IZA for providing the facilities for the authors to undertake research on this paper. The opinions expressed are those of the authors and do not necessarily reflect those of the institutions with which the authors are affiliated.

Wage Arrears and the Distribution of Earnings in Russia

Hartmut Lehmann and Jonathan Wadsworth

I. Introduction.

Wage inequality in Russia following the end of central planning has risen far more than in Central and Eastern European (CEE) countries undergoing transition. According to estimates based on official statistics, the Gini coefficient for wages in Russia rose from 0.22 before transition to around 0.5 in 1996 and the 90:10 income decile ratio tripled from 3.3 before transition to 10 in 1995 (Flemming and Micklewright, 1997). In contrast, over the same period, the estimated Gini index for wages in CEE grew from levels in the range of 0.2 to 0.25 to levels in the range 0.3 to 0.35. The level of wage inequality in Russia is now also very high by international standards.[1] Rising earnings dispersion seems to have been the major factor behind rising inequality in personal incomes.

All these trends are now well documented in the literature. But the reasons for (a) the sharp increase in earnings inequality in Russia and (b) the divergence between Russia and Central and Eastern Europe, are not entirely clear. Why was the rise less pronounced in the advanced reformer-countries compared to a country lagging in economic reforms, and not the opposite, as the logic of emerging returns to market oriented skills would suggest? There is little evidence of any large earnings discrepancies in aggregate data between industries. A majority of Russian workers are still employed in state-owned enterprises (SOEs) or the government, and therefore still subject, at least in theory, to regulated normative wages, under the "tariff ladder". The evidence presented in Brainerd (1998) also suggests that whilst returns to education, if not experience, have grown over the period, they remain low by Western standards. As such these

1 In Chile, the Gini coefficient is around 0.45 and in Turkey around 0.37.

factors cannot explain the extent of inequality observed in Russia.

One simple explanation of growing inequality in the wage distribution could be the presence of wage arrears. If in any given month a substantial subset of workers, receive only a part of the normal wage, or even no wage at all, then inequality in wages in any given month will be extremely high. However, the timing of the dramatic rise in inequality during the first years of transition documented in Brainerd (1998) indicates to us that most of the rise in inequality occurred before the problem of wage arrears really began. Hyperinflation at the onset of reforms is probably the major contributing factor to the rise in inequality at this time, however, as inflation subsided inequality has not fallen back. It, therefore, seems important to try to analyse to what extent wage arrears have affected the earnings distribution since payment problems began.

Wage arrears have been a pervasive feature of Russian economic life since 1994 affecting large sections of the workforce (Lehmann, Wadsworth and Acquisti, 1999, show that this affects between 40 and 70 percent of the workforce). The withholding of wage payments has been systematic and concentrated heavily on sub-sections of the working population (see e.g. Earle and Sabirianova, 1999, and Lehmann, Wadsworth and Acquisti, 1999). An explicit treatment of distributional effects of wage arrears has, however, not yet been undertaken.[2] Most studies of wages in Russia tend to ignore the presence of wage arrears without considering the possible consequences. Using Russian Longitudinal Monitoring Survey (RLMS) data, covering the years 1994 through 1996 and 1998, we explore the issue of how wage arrears have affected the wage distribution and the level of wage inequality in Russia.

In order to demonstrate the effects of wage arrears on the wage distribution, Table 1 and Figure 1 give summary measures of the changes in real monthly wage distribution across our

2 Gimpelson (1998) discusses distributional issues connected to wage arrears from a political rather than an economic perspective.

sample period. We also provide estimates from Poland and Britain as benchmark comparisons[3]. The Russian data suggest that, by 1998, around 70 percent of employees did not receive a wage complete and on time, and half of these received nothing in the preceding month. Whilst disturbing in itself, this finding of a large number of zero wage observations among the working population means that any conventional measures of inequality based around logarithmic transformations will be of little use here. In what follows therefore, we focus on the real monthly wage distributions and eschew any techniques that rely on logarithmic transformations.

Real average earnings fall markedly over the sample period prompted by a series of national economic crises which left inflation soaring and nominal wages failing to keep pace. The earnings distribution also widens over the first half of the sample period, while the evidence for the second half of the sample period is mixed. The coefficient of variation continues to increase, albeit more gently, but the Gini coefficient and the 90:50 ratio fall back. By 1996, the Gini coefficient in Russia was more than twice that observed in Poland and 60 percent higher than in Britain.

It is apparent, however, that the Russian results are strongly influenced by wage arrears. Figure 2 tracks the increased skewness of the real monthly wage distribution as the incidence of arrears builds up. The bottom panel of Table 1 confirms that inequality rises by much less amongst those who receive wages in full during the sample period. The Gini coefficient, for example, is roughly one third lower for those without wage arrears.

In some sense it is difficult to analyse wage distributions in Russia for this period since wage arrears scramble the distribution. Persons appear in low deciles solely because they are not

3 The figures for Poland are for full-time workers only, though, as in Russia, part-time working amounts to less than 3% of the Polish workforce.

paid at all or paid only part of their wages. So to analyse distributional issues seriously counterfactual wage distributions for years in which wage arrears are a problem seem to be required.

In what follows, we try to estimate what the wage distribution would have looked like during this period if all workers had been paid in full and on time in order to establish the "true" parameters of the distribution and any between-group differences. As we argue in the next section there are several reasons for undertaking such an exercise. We use seven methods to construct counterfactual distributions. The first is a simple least squares prediction and the second is least squares with the addition of a random residual, both of which use parameters from a wage equation estimated on the sample without wage arrears to predict wages for those in arrears. The third is a Tobit II extension of the second method, which corrects for the incidental truncation of the wage distribution (Heckman correction). We then apply a different residual according to the method proposed by Juhn, Murphy Pierce (1993). We also provide counterfactual estimates of what the wage distribution would look like if everyone were paid on time following the Kernel density approach pioneered by DiNardo, Fortin and Lemieux (1996). Our sixth method employs a variation of the exact matching techniques used by, among others, Heckman, Ishimura and Todd (1997), and Kluve, Lehmann and Schmidt (1999), to assign wages to those in arrears by matching their characteristics to the sub-sample of those who continue to be paid in full but who had a similar labour market pre-treatment history. The last method matches on the propensity score (Lechner, 2000). These matching estimators, we suggest, may take account of unobserved heterogeneity that could be missed by the other approaches. Our results, similar across the various methods, suggest that most of the earnings dispersion in Russia occurs amongst the stock of workers affected by wage arrears, and that earnings dispersion may have been some 30 percent lower in the absence of arrears.

Having estimated these counterfactual distributions we then examine the implications for estimates of between-group wage differentials commonly addressed in the literature. One interesting application is a reassessment of the analysis of the gender wage gap. Much of the existing work on the Russian wages has ignored the presence of wage arrears.[4]
However, given that, on average, women seem to be less affected by wage arrears (cf. Lehmann, Wadsworth and Acquisti, 1999) we would expect the mean gender gap to be larger with any of the counterfactual distributions than with the actual observed wages. This prior is confirmed by our analysis. We also can look at the gender gap sweeping through the entire earnings distribution, something we cannot do when wage arrears are present on a massive scale. In addition, we look at how wage arrears might affect returns to education and relative wage distributions by region and industry.

In the next section we look at the rationale for constructing counterfactual wage distributions in the Russian case. The subsequent section presents the various methods employed to construct counterfactual wage distributions, while section IV discusses data issues. Section V analyses earnings inequality in Russia and the decomposition of its change over time, followed by a presentation of the counterfactual results. Section VII then concludes.

II. Economic Reality in Russia and the Construction of Counterfactual Wage Distributions

One might legitimately ask why one would like to construct counterfactual wage distributions in Russia that assume payment of wages in full and on time for all employed members of the workforce. One could argue that during transition as labour hoarding continues the Russian

4 Oglobin (2000) is an exception to this, using a selection equation in his analysis of the mean gender pay gap.

economy is confronted with a macro constraint that makes it impossible to pay the contracted wage. Since mass layoffs are in the short-run politically and economically too costly[5], the Russian economy may then be intrinsically unable to honour all contractual wage claims in any given month. The practice of wage arrears then, is an economic policy tool that is consciously chosen by policy makers and managers to deal with output contraction in Russia. Equally, instead of, for example, imposing an inflationary tax on the entire workforce, the major costs of transition could be put on the shoulders of weak sub-groups of the workforce by withholding regular wage payments from them. Wage arrears are, therefore, an integral part of the labour market experience of many Russian workers. The upshot of these considerations then would be that the actual wage distribution is what matters and not some elusive counterfactual.

The above lines of reasoning do not preclude, in our opinion, the construction of counterfactual wage distributions, for the following reasons. First, if wage arrears are brought about because of a conscious policy of avoiding mass layoffs or a reluctance to use a general inflationary tax, then we can think of counterfactual wage distributions as reflecting a counterfactual economic policy that encourages the release of labour from unproductive, declining sectors. Such a policy, which has been used in most countries of Central and Eastern Europe despite large initial falls in output, seems to avoid inflationary bottlenecks and reverses the output decline. If such a counterfactual economic policy had been chosen those in work would almost certainly get paid in full and on time[6].

Secondly, even if there is no conscious attempt by policy makers and managers to

5 Russian labour market legislation stipulates severance pay of three monthly salaries for workers laid off in a mass layoff.

6 It may be that there would be differential unemployment levels across the two scenarios. The manner in which unemployment affects the parameters of the wage distribution in Russia is however unclear. The evidence on wage arrears emerging from the analysis in this paper suggests that wage arrears are distributed rather randomly across the wage distribution.

concentrate the costs of transition on some sub-groups of the workforce, there is no reason to assume that the "non-payment equilibrium" (Earle and Sabirianova, 2000) is the only natural, rational outcome that had to arise in the Russian labour market during transition.[7] In this case wage arrears occur because of some constraints that are not exogenous. The political constraint that does not allow mass layoffs could have been relaxed as could have been labour market legislation that imposes too large costs on firms in connection with these layoffs. One could even envisage the counterfactual as that which would emerge in the absence of large shocks. In summary, as long as we can think of sensible counterfactual policy regimes or scenarios it seems legitimate to construct counterfactual wage distributions.

Finally, we believe that the dynamic nature of the arrears process provides the strongest rationale for the use of counterfactual densities. Aggregate data from Russian Statistical Office (Goskomstat) tell us that since 1996 the stock of wage arrears has been approximately in a steady state, equivalent to two monthly wage bills. This means that the amount of contractual wages not paid to (some) workers in month t roughly equals to the amount of wage debts paid back to (some) workers in month t.[8] This implies that even though wage arrears are not a purely stochastic phenomenon in the sense that incidence is not random, most of the workers affected by them do get paid the owed wages eventually. The RLMS provides a monthly data window as far as wage payments are concerned. This monthly window might then be too narrow to obtain an estimate of the "permanent" earnings of those workers affected by the irregularity of pay.[9]

---

7 Desai and Idson (2000) seem to assume this non-payment equilibrium as the natural outcome in the Russian transition period

8 Payroll data from the city of Ryazan recently collected by one of us also seem to confirm this pattern.

9 Consider a simple thought experiment. Assume an economy where all workers get paid monthly. Let us make the additional assumption that the data window on earnings is the third week of the month in which we undertake the survey. So, we ask: "How much did you get paid in the third week of month x?" Some workers will have

However, issues dealing with the distribution of earnings like the gender gap or returns to human capital should be investigated using estimates of "permanent" earnings. The counterfactual distributions constructed by us provide estimates of such "permanent" earnings, albeit imperfect ones.[10]

III. Building Counterfactual Estimates of the Effects of Wage Arrears

Counterfactual wage distributions have been applied to a variety of economic and statistical issues, e.g. minimum wages (DiNardo, Fortin and Lemieux ,1996), item non-response (Biewen, 1999) and international differences in wage inequality (Blau and Kahn, 1996). The literature suggests at least 7 ways of building counterfactuals.

OLS methods

Following Oaxaca (1973) we can estimate a wage equation using the sample of those without wage arrears. Using the vector of estimated parameters from the no arrears equation and the observed characteristics of those in arrears we then predict wages, which those in arrears would have received if they had been paid in full. More formally, let B_{NW} be the vector of parameter estimates from the wage equation of the sample without wage arrears and let $X_{i,WA}$ be a vector of characteristics of the i-th person who experiences arrears. Then the predicted wage of this individual, $Y_{i, WA}$, will simply be:

---

been paid their monthly salary in this third week, but many will have been paid in another week of month x. Estimation of monthly earnings on this weekly window will be certainly inefficient, or even misleading. If, in the Russian case, we had a window of, say, two, three or four months, we could obtain better estimates of "permanent" earnings of Russian workers. The construction of counterfactuals is a good substitute for such estimates.

10 They give imperfect estimates of "permanent" income since the counterfactuals ignore the losses in earnings over time due to inflation. One should recall, though, that wage arrears are particularly virulent in times of low inflation (Gimpelson, 1998).

$$Y_{i,\,WA} = B'_{NW}\, X_{i,WA} \qquad (1)$$

Since this method gives only a mean prediction, we can add a residual so as to proxy wage dispersion in full, since the actual wage equals the sum of the predicted wage and a residual, $w = \hat{w} + \hat{u}$. We do this by first taking the standard error of the regression from the no arrears equation, σ_{NW}, and multiplying by a, randomly assigned, standard normal random variable z_i. It follows that a random residual which can be added to the predicted wage for the arrears sub-group then is given by

$$\varepsilon_{iWA} = z_i * \sigma_{NW} \qquad (2)$$

Heckman Selection Model

If there is any incidental truncation of the wage distribution in case of wage arrears then the coefficients used in equation (1) are not consistent if unobserved factors that determine the level of wages are correlated with unobserved forces driving the incidence of wage arrears. In order to achieve consistent estimates of the B_{NW} vector we estimate a Tobit II variant of Heckman's selection correction model, where the parameters of the selection equation are assumed to be different from the parameters entering the wage equation. Such an assumption seems reasonable as the effect of most regressors on the level of wages should be different from their effect on the probability of experiencing wage arrears. The Tobit II model is estimated using Maximum Likelihood. The main difficulty in this estimation is to find a regressor that identifies the model, as it is hard to think of factors that determine the probability of experiencing wage arrears but not the level of wages. Drawing on previous results from our research on the Russian labour market we use a rural/urban dummy as an identifier in the selection equations , (Lehmann, Wadsworth and Acquisti, 1999). We take the estimated coefficients of the no arrears group in the presence of the selectivity term, apply them to the characteristics of the arrears group and add a random residual. However, it seems clear to us that the results of the Tobit II model are extremely sensitive to specification and that alternative methods of constructing counterfactual wage distributions need to be also explored.

Juhn, Murphy and Pierce

Juhn, Murphy and Pierce (1993) and Blau and Kahn (1996) have suggested that it may be worthwhile trying to take into account unobserved heterogeneity as measured by the percentile ranking of each individual in the residual wage distribution. With a simple transformation of the residual into the product of a standard normal residual, θ, and the residual standard deviation from the wage equation, σ, the predicted wage can be written as

$$Y_{i,\,WA} = B'_{NW}\, X_{i,WA} + \sigma_{NW}\, \theta_{WA} \tag{3}$$

So that the counterfactual is the set of wages that would result if the no arrears wage coefficients and residual standard deviation were given to those currently in arrears. The estimates from the equations used to construct these estimates are given in Table A1 in the Appendix. It is apparent that the estimated coefficients vary widely between the arrears and no-arrears groups. Since many of the observations on the dependent variable in the arrears sample are zero, this technique relies on the assumption of normality in the residuals estimated from this subset.[11]

Kernel Density Counterfactuals

DiNardo, Fortin and Lemieux (1996), (hereafter DFL), have suggested that a broader insight may be obtained by taking into account the entire wage structure, allowing the returns to observables and unobservables to vary across the distribution of wages. The principal remains the same, to estimate the wages that those in arrears would receive had they been paid as those paid in full. Given the joint distribution of wages, w, and characteristics, x, the marginal distribution of wages

11 This is not always the case in our data.

conditional on x can be written g(w) = $\int f(w/x)h(x)dx$. The conditional expectation, f( ) is similar to an estimated regression line and the marginal density of x, h( ) is analogous to the vector of characteristics. Following DFL, using Bayes' law, it can be shown that the counterfactual wage distribution if everybody were paid in full can be obtained by taking the observed wage distribution of the subset of those paid in full and reweighting by a parameter Φ(x), where Φ(x) reflects the relative incidence of arrears conditional on characteristics x, Φ(x) = Pr(No Arrears) / Pr(No Arrears/x). The weights are normalised to sum to one. So,

$$g(w) = \int \Phi(x) f^{NoArrears}(w/x)h(x/i = NoArrears)dx$$

The integral is approximated using Kernel density estimation, which means that we do not get predictions of individual wages, only the quantiles of the distribution. The numerator in Φ(x) is the sample proportion of those not in arrears in any year and the denominator is estimated by a logit regression conditional on a set of observed characteristics. The estimates from the logit equations used to construct these estimates (Table A2) confirm the dominance of location and firm characteristics in explaining the incidence of arrears as found in Lehmann, Wadsworth and Acquisti (1999).

Matching Estimators

If there were unobserved heterogeneity amongst those in arrears, then the preceding techniques would fail to account for this. The JMP approach and the DFL density approach perhaps come closest, the latter using the non-parametric structure of the entire distribution. However they implicitly assume that heterogeneity amongst those not in arrears is duplicated amongst those in arrears. If this is not the case, those not in arrears are different from those in arrears, the counterfactual estimates could be biased in some way. Moreover, the JMP method uses the standard residuals from the arrears regression to calculate counterfactuals. This standardised residual is usually interpreted as an individual's ranking in the residual wage distribution and as such a measure of unobserved relative skill. However, the outcome we analyse in equation (3) gives an individual's relative ranking in the residual arrears distribution, which is hard to interpret as a measure of unobserved skill. This, together with our wish to construct counterfactuals untainted by arrears leaves this method open to question.

We therefore experiment with alternative approaches based on the matching estimator literature. The first technique follows Heckman, Ishimura and Todd (1997) in that we also condition, non-parametrically, on "pre-treatment history" in order to minimise biases arising from unobserved heterogeneity. In our case this means conditioning on events **before** wage arrears began, together with a set of current observable, exogenous characteristics, in order to try and capture heterogeneity in the arrears population, i.e. to ensure that the treatment and the control group do not differ systematically. Conditioning on a set of pre-treatment covariates is assumed to be sufficient to allow the assumption of assignment to the treatment group as random, such that unobservables may be ignored. If Y_{i1} is the outcome with treatment and Y_{i0} is the outcome without treatment for individual i and X and H are sets of controls for observable characteristics and "pre-treatment history", then the identification assumption becomes,

$E(Y_{i0} / T = 1, X, H) = E(Y_{i0} / T = 0, X, H)$. Heckman, Ishimura and Todd (1997) find that for this type of matching estimators to work well the same data set should be used for the control and treatment group, the groups should be in the same local labour markets and the data set should contain a rich set of variables relevant to the treatment decision.

Treatment in our study is the experience of wage arrears and the labour market history we condition on, using the panel element of the RLMS, is labour market status one year earlier and if employed, the ranking in the wage distribution of those paid in full. If the individual was out of work one year earlier we create unemployed and inactive categories. If the individual was in arrears one year earlier we create a separate sub-category. We divide last year's wage distribution, excluding arrears, into deciles. Matching proceeds for those sub-groups of the treated and the non-treated who have the same "pre-treatment history", and in addition we match according to age (with a maximum allowed difference of ten years), gender, region (3 groups) and qualifications (6 groups) in the current year. This strategy conforms broadly to the criteria set out by Heckman et al. (1997) required for a good performance of a matching estimator. Also, the assumption here is that the variables used for matching are not affected by the treatment (arrears).[12]

We assign the wages of those currently paid in full to those in the treatment group, who were placed in the same decile a year ago when both treatment and control groups were paid in full. Those in arrears now who were also in arrears last year or non-employed are given the wages of those currently paid in full who were in the same category one year earlier. In this way, we hope to reduce the difference in unobserved skills and other characteristics that might exist between the individuals experiencing wage arrears and those who are unaffected by them. If more than one person can be matched with the individual we assign the average wage of the matched controls. With this direct matching procedure the set of variables used is much smaller than can be afforded by a regression based technique which is unaffected by empty cells. The matching algorithm is shown in Box A1 in the appendix.

The approach assumes that individuals do not move rapidly through the earnings distribution. As a check, Table A3 in the appendix presents one and four-year earnings transition

12 Whilst within region mobility may be affected by arrears, the regions in the RLMS are so large as to make mobility between regions as a result of arrears unlikely.

matrices using quintiles of the wage distribution. It is apparent that, whilst there is a degree of mobility across earnings quintiles, there is considerably less mobility amongst those not subject to wage arrears. Figure 3 also suggests that those in arrears are drawn from across the entire wage distribution. Since this approach can only be used when there are at least two consecutive years of longitudinal data, we confine our estimates using this approach to 1996 and provide comparable estimates using the other counterfactual techniques.

Propensity Scores

When performing non-parametric matching we lose around 10 per cent of potential matches due to empty cells. To avoid this, we also employ propensity score matching, where individuals are matched according to the closeness in the estimated probability of experiencing wage arrears. We use the matching algorithm suggested by Dehejia and Wahba (1998).[13] We estimate probit regressions, conditional on the same co-variates as used in the matching approach, take the predicted probability – the propensity score - and match, with replacement, those in arrears with those not in arrears with the nearest propensity score. It can be shown that if Y_{i1} and Y_{i0} are independent of treatment, T, given X and H (that is, given sufficient disaggregation by age, sex and region, for example, as well as by "pre-treatment history"), then the two groups may be treated as the same. In other words, T is *ignorable* given X and H, so that

$E(Y_{i0} / T = 1, P(X,H)) = E(Y_{i0} / T = 0, P(X,H)) = E(Y_{i0} / P(X,H))$. We estimate two variants of the propensity score, one where pre-treatment variables are included in the set of co-

---

13 As Kluve, Lehmann and Schmidt (2001) state, "the reduced dimension comes at a cost, however. The propensity score is not known and has to be estimated. Also, in samples of limited size, for some *i* and *j* it may occur that $p(X_i)=p(X_j)$ even if $X_i \neq X_j$, resulting in imperfect balancing of the distributions of covariates." The literature stresses that there seems to be a bias vs. efficiency trade-off between non-parametric and propensity score matching. Smith and Todd (2001) show that estimates from different propensity score matching methods do not vary much as long as the conditioning variables satisfy the requirements set out by Heckman et al. (1997).

variates and one without them. In the latter case the identification assumption becomes,

$$E(Y_{i0} / T = 1, P(X)) = E(Y_{i0} / T = 0, P(X)) = E(Y_{i0} / P(X)).$$

IV. Data.

Our main data source is the second phase of the Russian Longitudinal Monitor Survey, (RLMS), a longitudinal panel of around 4000 households across the Russian federation conducted in the autumn of 1994, 1995, 1996 and 1998. The data contains a set of demographic and establishment characteristics, together with information on the labour market activities of its sample. Despite its relatively small size, the advantage of this source for our purposes, is that we can track individual wages and the incidence of wage arrears over time. We restrict our sample to employees of working age and exclude the military.[14] The survey design does not follow individuals if they move, but does sample new occupants of the same address. There are around 10,000 individual observations in each wave, of which around 4000 are in work in any wave and around 3,500 give wage related information.

The survey questions dealing with wage arrears ask whether, conditional on being in work, whether an individual was owed money by the firm in the past month or was paid "in kind" with goods produced by the firm. This constitutes our sample of those in arrears in any wave. Some of those in arrears are paid a certain amount of money, whilst others, around one half of those in arrears, receive nothing.[15] Respondents, both those paid in full and those in arrears, are asked to state the amount of money received from their employers after tax in the past month.

---

14 The RLMS is ambiguous on the nature of self-employment, referring instead to the extent of self-ownership in the enterprise where the individual works. We exclude only those who say they own between 51 and 100% of the enterprise.

15 The RLMS also asks for the total amount owed, together with the number of months since the worker was paid last.

These are total wage receipts and not contractual wages. There is no distinction made between basic wages and bonus. These wage responses are then deflated by a national price deflator indexed to 100 at January 1998.[16] There is no indication whether wage arrears are estimated before or after tax. We remove outliers from that data, namely those earning in excess of 4000 roubles a month, or less than 50 roubles if the respondents are not in arrears.

We also provide some data from a smaller, household survey data set, VCIOM, undertaken in 1993 in order to provide summary evidence on pay from an earlier period when wage arrears were less prevalent.

V. Earnings Distributions and Inequality in Russia

Table 2 provides a formal decomposition of changes in earnings inequality over the period into its between and within-group components. [17] Following Cowell (1995) we can decompose any generalised entropy measure of inequality[18]

$$I_a = I_{between} + I_{within}$$

where

$$I_{between} = \frac{1}{\theta^2 - \theta}\left[\sum_{j=1}^{k} f_j \left[\frac{\overline{y_j}}{\overline{y}}\right]^{\theta} - 1\right]$$

and

16 There are no population weights in the data sets.

17 Fields (1999) decomposition of the sources of wage inequality relies on a decomposition of the log variance of earnings, which is inappropriate here given the large number of zero wage observations.

18 Note that this approach calculates the between group component assuming that everyone within a group receives mean income, which is clearly not the case in Russia

$$I_{within} = \sum_{j=1}^{k} w_j I_j \qquad \text{where} \quad w_j = g_j^{\theta} f_j^{1-\theta}$$

and f_j is the population share of group j and g_j is the share of group j in total income = $f_i y_i / \bar{y}$. So total within group inequality is a weighted average of inequality in each sub group, though the weights do not add to one unless θ=1 or 0. This decomposition is sensitive to the choice of parameter θ, so in Table 2 we present estimates based on two different θ values. The results suggest that differences between those in arrears and those not accounts for around 20 to 30 percent of the rise in inequality between 1994 and 1996. The majority of the rise in inequality however comes from within the group in arrears. The results are more ambiguous over the second half of the sample period. Inequality rises or falls depending on the value of θ used, as do the within and between group components. The entropy estimate based on the low θ value falls between 1996 and 1998 most likely because low values put more weight on distances between wages in lower parts of the distribution and the share of those paid zero wages falls. This is not reflected in the other entropy estimate, which rises as it gives greater weight to wage changes in the upper tail. It remains true however that the majority of earnings inequality in any one period comes from amongst those in arrears.

Table 3 gives the results using the first four estimation approaches and the propensity score matching estimation without conditioning on pre-treatment history for the years 1994, 1996 and 1998. Figure 4 graphs the counterfactual Kernel densities. Not surprisingly the mean and various centiles of the distributions are all higher using the counterfactual estimates. Mean earnings rise by around 30% in 1994 and between 50% and 70% when wage arrears were highest in 1998. Similarly estimated overall dispersion, as measured by the coefficient of variation, is between 23% and 36% lower in 1994 and between 29% and 46% lower in 1998. The Gini coefficients are now in the same range as those of Britain. The estimates of the various wage centiles show a narrower distribution when based on simple OLS predictions (OLS I) and the propensity score matching estimator than the estimates based on other methods. Since simple OLS estimation performs a regression to the mean, it comes as no surprise that the OLS I based earnings counterfactual distribution is narrower; it this feature of OLS estimation that provides actually a justification for the addition of a random error term. The larger 9-to-1 and 5-to-1 decile ratios for the OLS II and Heckit estimates can be explained by the skewness of the wage distribution. Adding a random normal residual to the predicted value will then generate too many negative values for predicted wages, giving a very low value for the 10-th percentile.

Table 4 uses the panel element of the data in order to add the exact matching estimator and a second propensity score estimator with "pre-treatment history" included as an additional regressor. We compare the results with those using the other methods for the year 1996. We also show the distribution of those in the sample who get paid in full and on time (second column). Apart from the estimates based on simple OLS prediction (OLS I) all other counterfactual distributions have a very similar spread as can be seen from the close coefficients of variation and the GINI coefficients. The distribution of the propensity score estimates seems to depend on the set of covariates used to generate the propensity scores. It also noteworthy that the no arrears distribution differs little from the counterfactuals, a point to which we will return later.

We now examine the implications of these counterfactual estimates for pay gaps between various sub-groups of the workforce. In Table 5 we compare levels and ratios of pay across gender using the actual distribution, the no arrears distribution and the counterfactual distributions for the year 1996. If everyone were paid in full, then there would be more dispersion in pay between men and women. If we exclude the propensity score estimates, the mean gender wage gap rises from the observed 19 percentage points to between 28 and 32 percentage points when counterfactual distributions are used. Also noteworthy is the fact that five of the counterfactual distributions show the largest wage gap at the tenth percentile, while in Western economies the widest divergence between male and female earnings usually occurs at the ninth decile. The no-arrears distribution gives mean and median ratios that are very similar to the levels of all but the propensity score based counterfactuals. On the other hand, the no-arrears distribution shows no variation of the ratios across deciles, which does not hold for any of the counterfactuals.

Table 6 shows mean and medium wages of three educational categories (graduate, intermediate and primary) and presents mean and medium ratios relative to the low educational category using the actual, the no arrears and all counterfactual distributions. It is striking that the actual distribution suggests a higher relative return to graduate education than the counterfactual estimates, while there is little difference in the relative returns for the intermediate group. It is also noteworthy, that the ratios from the no arrears distribution are again quite similar to those from the counterfactual distributions.

We now turn to two dimensions that impact strongly on the incidence of wage arrears, region and industry. We divide the sample into three areas: those living in Moscow and St. Petersburg (Metro), where the incidence of wage arrears is low and wages are high; those living in the Urals region, where wage arrears are massive, but wages are highest; and those living in the rest of the country, where wages are lower and the incidence of wage arrears is high. The actual distribution gives a 25% higher mean wage gain from living in one of the metropolitan regions compared with the counterfactuals and, ignoring the propensity score estimates, understates the mean wage difference between Urals and other regions by 20% on average. For the median ratios these biases are even more pronounced (Table 7).

In Table 8 we aggregate industries into two sectors, production and services, with workers in the former more likely to experience wage arrears than in the latter. There are discernible differences between the ratios of the actual and the other distributions at the median and the 90th percentile. The counterfactual distributions suggest that if everyone were paid in full, then there would be more dispersion in pay between production and service sectors. The production sector seems to be more affected by wage arrears than services resulting in an increase of roughly 30% points as one goes from the actual to the no arrears or the counterfactual distributions.

VII. Conclusions

Russia now has one of the highest levels of wage inequality in the world. While wage arrears were not responsible for the large increase in inequality, the estimates in our paper suggest that they may have been partly responsible for the failure of inequality to fall back following the unanticipated price shocks in the first half of the nineties. The majority of earnings inequality is experienced within the population experiencing wage arrears at any point in time. The large share of employees who receive no wages in any one month renders many conventional estimates of inequality inoperable. Counterfactual estimates of the wage distribution in the absence of arrears indicate that average earnings would be some twenty to fifty percent higher, depending on the extent of arrears and that earnings dispersion would be lower by similar amounts if everyone were paid in full. This puts earnings inequality back towards levels of inequality currently experienced in Western countries like Britain and the United States.

On the basis of the counterfactual distributions we find higher gender wage gaps through most of the distributions, with the mean gap taking on values approximately 10 percentage points higher than the actual gap in the year 1996. In contrast, our estimates suggest that the relative return to graduate education would be compressed by around 15 percent if everybody were paid in full. Regional pay differentials would become more compressed and sectoral differentials would be widened in the absence of wage arrears.

One striking feature of our exercise is that the parameters of the counterfactual wage distributions are very similar to the parameters of the observed wage distributions of those not in arrears. While this does not mean that experience of wage arrears is a random event as confirmed by evidence in Earle and Sabirianova (2001) and Lehmann, Wadsworth and Acquisti (1999), it does suggest that those in wage arrears are drawn reasonably uniformly from throughout the wage distribution. For those wishing to study aspects of wage differentials and inequality in Russia, it may, therefore, be feasible to use the subset of those not in arrears and still get close to the true population parameters, subject to an efficiency loss.

References.

Atkinson A., Micklewright J., (1992) Economic Transformation in Eastern Europe and the Distribution of Income., Cambridge University Press.

Biewen, M., (1999) 'Item Non-Response and Inequality Measurement: Evidence from the German Earnings Distribution', University of Heidelberg, Economics Department Discussion Paper No. 295, March.

Blau, F. and Kahn, L. (1996), 'International Differences in Male Wage Inequality', Journal of Political Economy, vol. 104, no.4, 791-836.

Brainerd, E., (1998) "Winners and Losers in Russia's Economic Transition" American Economic Review, December.

Commander, McHale and Yemtsov. ' Russia' in Unemployment, Restructuring and the Labour Market in Eastern Europe and Russia. World bank, 1995

Cowell, F., (1995), Measuring Inequality, Phillip Allen, London.

Desai, P. and Idson, T. (2000), Work Without Wages: Russia's Nonpayment Crisis, MIT Press, Cambridge, Mass. and London.

DiNardo, J., Fortin, N. and Lemieux, T., (1996), 'Labour Market Institutions and the Distribution of Wages', 1979-1992, Econometrica

Dehejia, R. and Wahba, S., (1998) , 'Propensity Score Matching Methods for Non-Experimental Causal Studies', NBER Working Paper No. 6829

Heckman, James J., Hidehiko Ishimura and Petra E. Todd (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme' Review of Economic Studies 64, 605-654.

Earle J. and Sabirianova K. (2001), 'Understanding wage arrears in Russia. Journal of Labor Economics, forthcoming.

Flemming and Micklewright . Income distribution, economic systems and transition (EBRD 1997

Garner T. and Terell K. 'A Gini decomposition analysis of inequality in the Czech and Slovak Republics during the transition' Economics of Transition, Vol6, 1, 1998.

Gimpelson, V. (1998), The Political Economy of Wage Arrears, Budapest, mimeo.

Juhn, C., Murphy, K. and Pierce, B., (1993), 'Wage Inequality and the Rise in Returns to Skill', Journal of Political Economy, Vol. 101, No.3, pp. 410-42.

Katz K. (1994) "Gender Differentiation and Discrimination. A Study of Soviet Wages." PhD. Dissertation. Goteborgs Universitet.

Kluve, J, Lehmann H. and Schmidt C., (1999), 'Active Labour Market Policies in Poland: Human Capital Enhancement, Stigmatization or Benefit Churning?', Journal of Comparative Economics, January, vol. 27, 61-89

Kluve, J, Lehmann H. and Schmidt C., (2001), 'Disentangling Treatment Effects of Polish Active Labour Market Policies: Evidence from Matched Samples', IZA Discussion Paper No. 355, September.

Lehmann, H. Wadsworth J. and Aquisti A. (1999). 'Grime and punishment: Job insecurity and wage arrears in the Russian Federation' Journal of Comparative Economics, December, vol. 27, 595-617

Oaxaca, R., (1973), 'Male-Female Earnings Differentials in Urban Labor Markets', International Economic Review, 14, 693-709.

Ogloblin C. G. (1999), 'The Gender Earnings Differential in the Russian Transition Economy', Industrial and Labor Relations Review, 52, 4, 602-627.

Reilly B. The Gender Pay Gap in Russia during the Transition (1992-1996), Economics of Transition (1998).

Smith, J.A. and Todd P.E. (2001), 'Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods', American Economic Review, vol. 91, no.2, 114-118.

World Bank. Poverty and income distribution in a high growth economy: Chile, 1998.

Figure 1. Real Wage Distribution 1994-98 (RLMS)

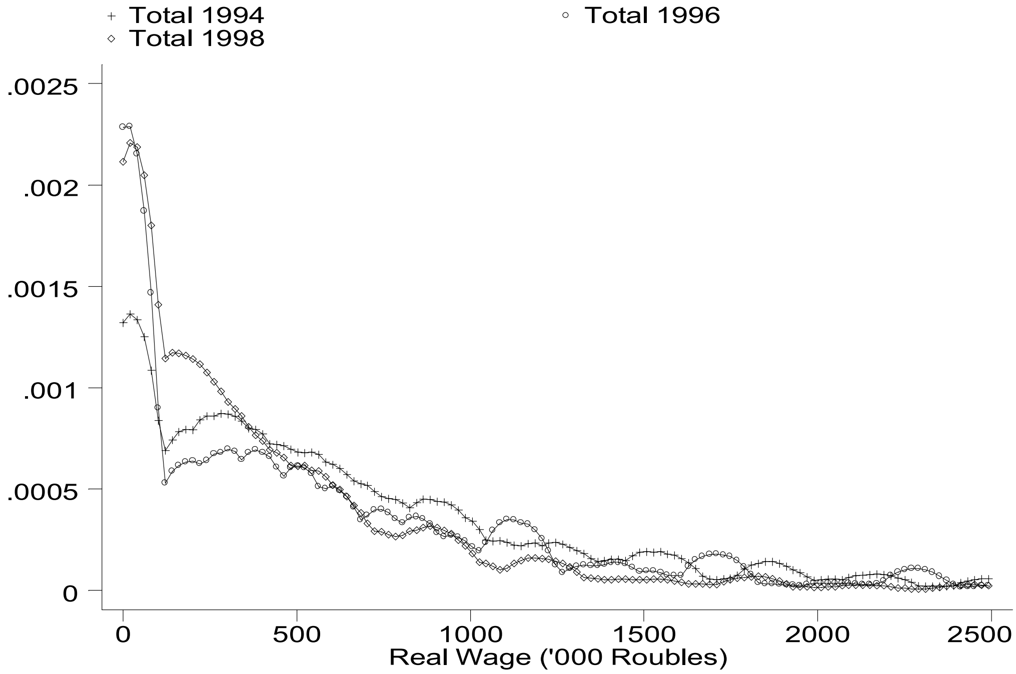


Figure 2. Distribution of Real Wages in Russia

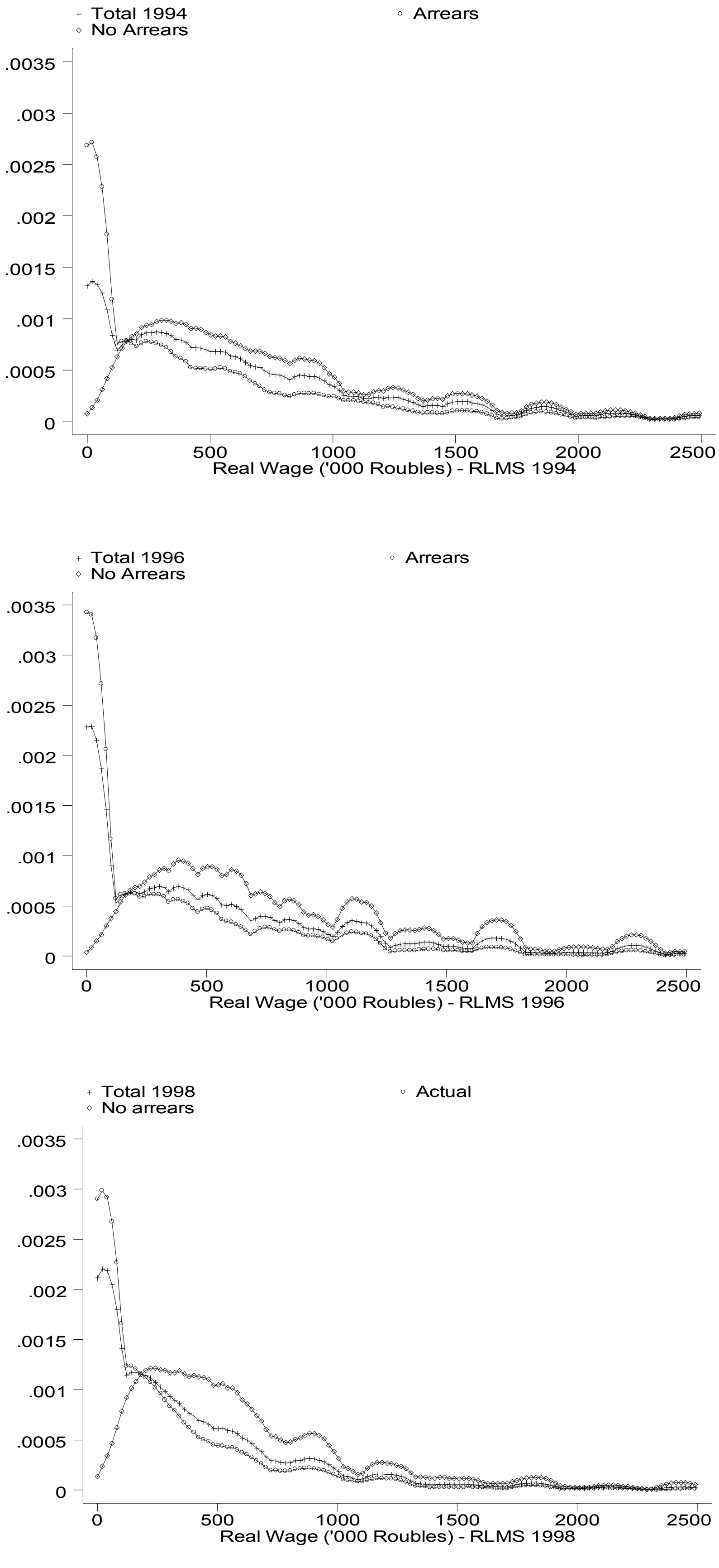


Figure 3. Decile Origin of Those in Wage Arrears

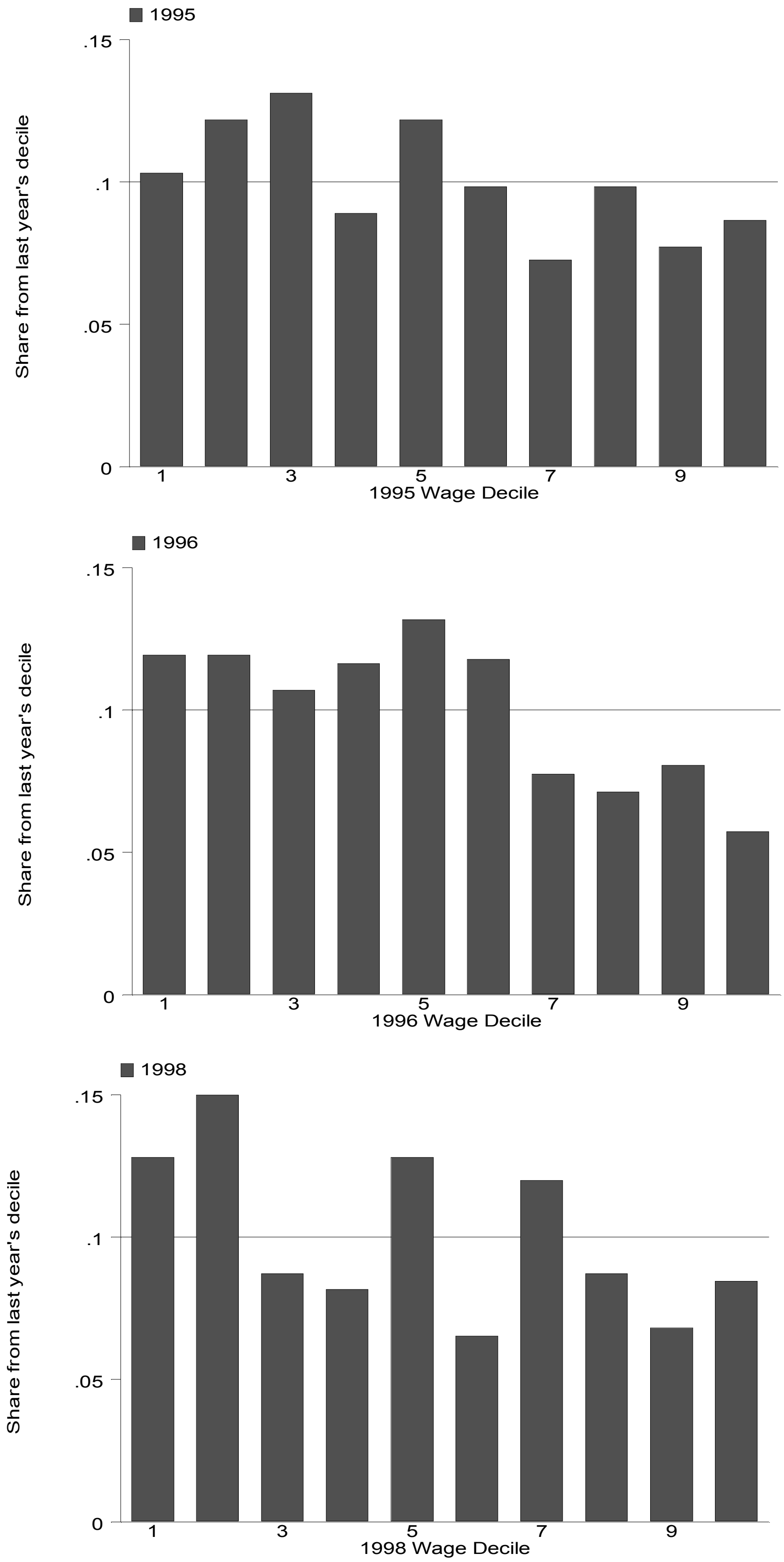
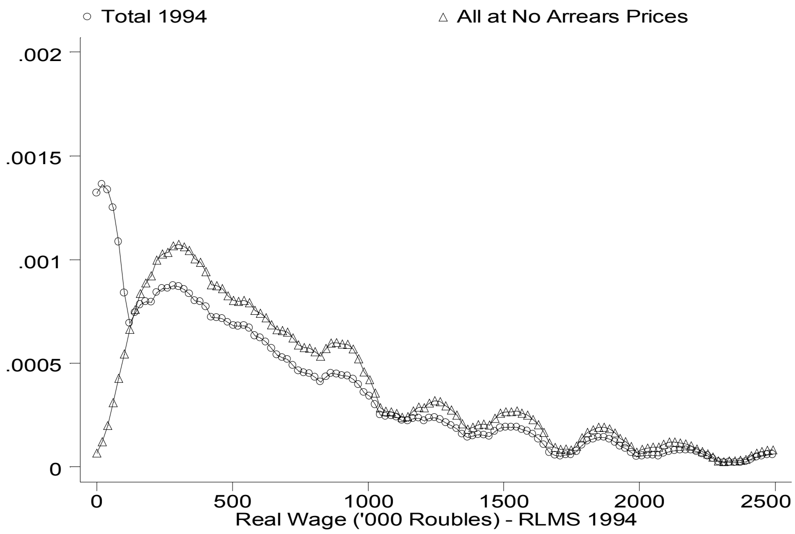
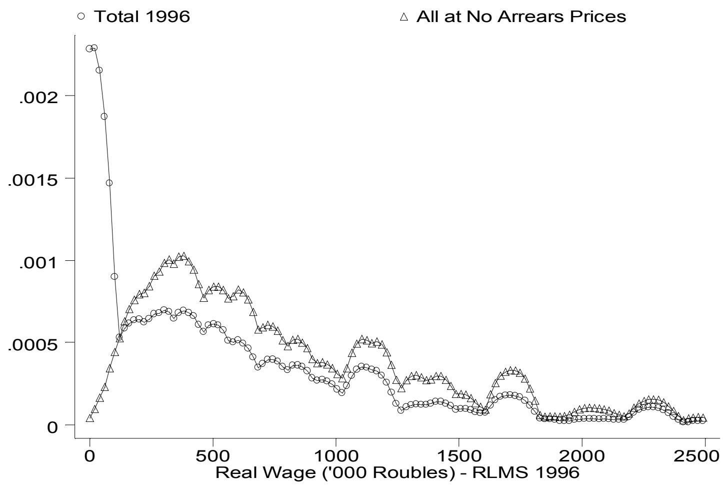


Figure 4. Counterfactual Estimates of Wage Distribution in Absence of Arrears





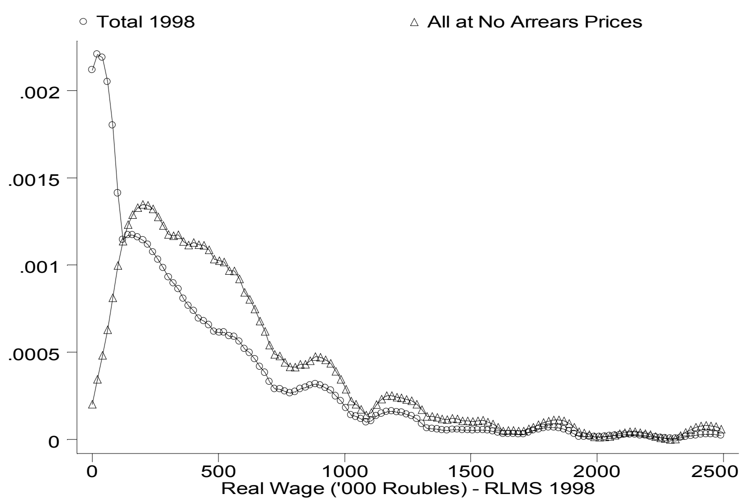


Table 1. Summary Measures of Real Monthly Wage Distribution

	1993 VCIOM	1994 RLMS	1996 RLMS	1998 RLMS	1996 Poland	1998 Britain
Total						
Mean	930	867	547	396	140	1247
90th	1724	1563	1464	968	219	2316
50th	690	469	338	242	120	1054
10th	276	0	0	0	81	271
% no pay	0.4	18	32	26	0	0
90/10	6.25	n/a	n/a	n/a	2.70	8.55
90/50	2.5	3.33	4.33	4.00	1.83	2.20
50/10	2.5	n/a	n/a	n/a	1.48	3.89
Coef. Var	1.12	1.05	1.26	1.31	0.62	0.80
Gini	0.410	0.535	0.622	0.613	0.239	0.387
% arrears	9	49	66	72	0	0
No Arrears						
Mean	944	845	948	660		
90^{th}	1724	1818	1989	1355		
50th	690	625	688	491		
10th	276	200	229	166		
90/10	6.25	9.09	8.69	8.16		
90/50	2.5	2.91	2.89	2.75		
50/10	2.5	3.13	3.00	2.96		
Coef. Var	1.12	0.80	0.79	0.85		
Gini	0.407	0.414	0.409	0.419		

Table 2. Between and Within Group Real Wage Inequality by Arrears

	1993 VCIOM	1994 RLMS	1996 RLMS	1998 RLMS
Entropy (θ=0.5)	0.298	0.660	0.979	0.900
within arrears (w_i)	0.003	0.435 (.39)	0.731 (.52)	0.715 (.61)
within no arrears	0.264	0.171 (.60)	0.123 (.44)	0.107 (.37)
between group	0.031	0.054	0.125	0.078
%Share Between Group Inequality (θ=0.5)	0.3	8.1	12.8	8.1
% Change in inequality accounted for by between group	N/a	N/a	22.3	59.5 [10.0]
Entropy (θ=2.0)	0.622	0.546	0.798	0.856
within arrears	0.037	0.208 (.21)	0.345 (.26)	0.481 (.39)
within no arrears	0.584	0.286 (.89)	0.316 (1.01)	0.286 (.79)
between group	0.001	0.052	0.137	0.089
%Share Between Group Inequality (θ=2.0)	0.2	9.4	17.2	10.4
% Change in inequality accounted for by between group	N/a	N/a	34.1	-82.8 [11.9]
% in Arrears	9.1	48.7	66.3	71.7

Note. Figure in square brackets in column 5 give the change of between-group shares from 1994 to 1998. Figures in round brackets give within-group weights.

Table 3. Counterfactual Real Wage Distributions

	Mean	90th Pctile	Median	10th pctile	90/10	90/50	50/10	Coef. Var.	Gini
1994									
Actual	609	1500	422	0	N/a	3.55	N/a	1.08	0.546
OLS I	793	1500	665	210	7.14	2.25	3.17	0.73	0.380
OLS II	800	1689	644	150	11.26	2.62	4.29	0.83	0.447
Heckit	818	1719	662	156	11.01	2.59	4.24	0.81	0.438
JMP	794	1655	618	186	8.89	2.68	3.32	0.84	0.424
DFL	788	1687	582	178	9.48	2.89	3.27	0.84	0.426
PS I	802	1562	631	253	6.17	2.48	2.49	0.69	0.359
1996									
Actual	500	1376	287	0	N/a	4.79	N/a	1.32	0.636
OLS I	830	1452	739	262	5.54	1.96	2.82	0.66	0.342
OLS II	841	1743	727	124	14.05	2.39	5.86	0.84	0.457
Heckit	919	1880	803	172	10.93	2.34	4.67	0.77	0.423
JMP	830	1720	619	229	7.51	2.78	2.70	0.85	0.410
DFL	817	1720	585	184	9.35	2.94	3.18	0.85	0.419
PS I	749	1465	583	223	6.57	2.51	2.61	0.79	0.395
1998									
Actual	371	907	206	0	N/a	4.40	N/a	1.33	0.618
OLS I	580	1030	500	167	6.17	2.06	2.99	0.75	0.367
OLS II	590	1291	484	66	19.56	2.66	7.33	0.94	0.503
Heckit	634	1356	533	91	14.90	2.54	5.85	0.88	0.471
JMP	580	1209	423	136	8.89	2.85	3.11	0.95	0.442
DFL	571	1210	417	121	10.00	2.94	3.40	0.95	0.442
PS I	577	1124	434	158	7.11	2.59	2.75	0.85	0.405

Source: RLMS. Note: OLS I is OLS estimate without residuals, OLS II includes residuals, PS I is estimate based on propensity score without conditioning on pre-treatment history.

Table 4. Counterfactual Real Wage Distributions, 1996

	Actual	No Arrears	OLS I	OLS II	Heckit	JMP	DFL	Match.	PS I	PS II
Mean	510	839	820	814	886	819	813	798	825	807
90th	1284	1720	1405	1720	1740	1720	1720	1641	1720	1720
50th	322	635	732	688	782	609	608	596	648	596
10th	0	225	268	122	172	229	195	225	229	206
90/10	n/a	7.64	5.24	14.09	10.11	7.51	8.82	7.29	7.51	8.35
90/50	3.99	2.71	1.92	2.50	2.22	2.82	2.83	2.75	2.65	2.88
50/10	n/a	2.82	2.73	5.64	4.54	2.66	3.11	2.64	2.83	2.89
Coef. Var	1.26	0.81	0.65	0.83	0.76	0.83	0.83	0.83	0.79	0.83
Gini	0.621	0.411	0.336	0.449	0.416	0.403	0.411	0.407	0.402	0.417

Source: RLMS. Notes. See Table 3. PS II is estimate based on propensity score conditioning on pre-treatment history. Sample size = 2538, of which 1351 are in arrears and 1187 are paid in full and on time.

Table 5. Comparing the Gender Wage Ratio, (1996)

	Actual	No Arrears	OLS I	OLS II	Heckit	JMP	DFL	Match.	PS I	PS II
Men										
Mean	569	1013	972	967	1051	972	949	970	918	887
Median	337	803	917	898	980	737	788	745	749	683
90th	1490	2199	1562	1865	1984	1973	1950	2178	1950	1950
10th	0	287	344	225	268	350	241	305	229	229
Women										
Mean	462	716	693	698	761	693	697	662	748	741
Median	310	539	605	575	642	526	516	520	563	563
90th	1147	1464	1212	1515	1620	1376	1456	1311	1548	1577
10th	0	195	229	108	145	191	178	185	212	195
Ratio										
Mean	0.81	0.71	0.71	0.72	0.72	0.71	0.73	0.68	0.81	0.84
50th	0.92	0.67	0.66	0.64	0.66	0.71	0.65	0.70	0.75	0.82
90th	0.77	0.67	0.78	0.81	0.82	0.70	0.75	0.60	0.79	0.80
10th	n/a	0.68	0.66	0.48	0.54	0.55	0.74	0.61	0.93	0.85

Source: RLMS. Sample size=2538, of which 1153 are male and 1385 female.

Table 6. Comparing Education Wage Ratios, (1996)

	Actual	No Arrears	OLS I	OLS II	Heckit	JMP	DFL	Match.	PS I	PS II
Upper										
Mean	600	903	900	910	969	900	902	886	860	853
Median	401	688	792	802	859	688	688	688	675	631
Intermed										
Mean	436	779	749	786	870	749	760	729	807	780
Median	229	597	687	688	767	573	573	563	642	573
Low										
Mean	434	760	757	722	804	757	709	709	776	751
Median	229	470	676	581	675	573	459	458	573	563
Ratio: Low										
Mean upper	1.38	1.19	1.19	1.26	1.21	1.19	1.27	1.25	1.11	1.14
Inter	1.01	1.03	0.99	1.09	1.08	0.99	1.07	1.03	1.04	1.04
Median upper	1.75	1.46	1.17	1.38	1.27	1.20	1.49	1.50	1.18	1.12
Inter	1.00	1.27	1.02	1.18	0.54	1.00	1.25	1.23	1.12	1.02

Source: RLMS. Sample size=2538, of which 1157 are upper, 888 intermediate and 493 lower.

Table 7. Comparing Regional Wage Ratios, (1996)

	Actual	No Arrears	OLS I	OLS II	Heckit	JMP	DFL	Match.	PS I	PS II
Metro.										
Mean	758	1027	1036	1042	1094	1036	906	1002	923	924
Median	563	803	945	917	962	788	802	803	788	688
Urals										
Mean	630	1305	1213	1253	1346	1213	1270	1176	1132	982
Median	189	1032	1130	1146	1296	940	963	705	844	642
Other										
Mean	439	734	719	727	802	719	716	715	763	758
Median	275	573	655	642	688	563	513	570	573	573
Ratio: Other										
Mean Metro	1.73	1.40	1.44	1.43	1.36	1.44	1.27	1.40	1.11	1.22
Urals	1.44	1.80	1.69	1.72	1.68	1.69	1.77	1.64	1.04	1.30
Median Metro	2.05	1.40	1.44	1.43	1.40	1.40	1.56	1.41	1.18	1.20
Urals	0.69	1.80	1.73	1.79	1.88	1.67	1.69	1.24	1.12	1.12

Source: RLMS. Sample size=2538, of which 427 are metropolitan, 241Urals and 1871 other.

Table 8. Comparing Industry Wage Ratios, (1996)

	Actual	No Arrears	OLS I	OLS II	Heckit	JMP	DFL	Match.	PS I	PS II
Production										
Mean	527	886	853	874	965	852	839	845	843	829
Median	271	681	788	756	863	624	596	653	676	608
10^{th}	0	229	275	130	212	251	206	229	229	229
90^{th}	1261	1834	1452	1945	1970	1720	1720	1720	1720	1834
Services										
Mean	462	805	788	789	844	788	788	755	808	787
Median	344	605	688	688	745	596	614	573	630	573
10^{th}	0	206	258	149	169	229	194	225	217	200
90th	1305	1689	1351	1605	1689	1605	1605	1463	1720	1720
Ratio										
Mean	1.14	1.10	1.08	1.11	1.14	1.08	1.06	1.12	1.04	1.05
50^{th}	0.79	1.13	1.15	1.10	1.16	1.05	0.97	1.14	1.07	1.06
10^{th}	n/a	1.11	1.07	0.87	1.25	1.10	1.06	1.02	1.06	1.15
90th	0.97	1.09	1.07	1.21	1.17	1.07	1.07	1.18	1.00	1.07

Sample size=2538, of which 1227 are production and 1312 services.

Appendix

Box A1
Exact matching – algorithm and scheme of conditioning on pre-treatment history

Exact matching algorithm

I. Condition on following possible pre-treatment labour market history:
- **employed and fully paid and in xth decile of wage distribution**
- **unemployed**
- **inactive**
- **employed and experiencing wage arrears (WA)**

II. Match treated individuals to individuals with same pre-treatment history using following observable characteristics:
- **gender**
- **region (4 categories)**
- **qualifications (6 categories)**
- **age (maximum allowed difference of 10 years – choose those controls that have the minimum age difference)**

Assumption: these variables are not affected by the treatment (WA).
Because treated are more than potential controls, matching is done with replacement.

III. Assign wage of matched control to treated individual, or assign average of wages of matched controls

Scheme of Conditioning on pre-treatment history by example

Pre-treatment period	**Treatment period**
Potential Control 1 in 95 **Employed and fully paid and in 2nd decile of wage distribution**	***Potential Control 1 in 96*** **Employed and fully paid**
Treated 1 in 95 **Employed and fully paid and in 2nd decile of wage distribution**	***Treated 1 in 96*** **In wage arrears**
Potential Control 2 in 95 **Unemployed**	***Potential Control 2 in 96*** **Employed and fully paid**
Treated 2 in 95 **Unemployed**	***Treated 2 in 96*** **In wage arrears**

Table A1. OLS Real Weekly Wage Estimates

	1994		1996		1998	
	No Arrears	Arrears	No Arrears	Arrears	No Arrears	Arrears
Female	-328.844	-82.797	-333.160	-40.009	-233.198	-66.771
	(24.318)**	(28.675)**	(36.521)**	(23.981)	(25.946)**	(16.803)**
Age	27.947	5.922	24.714	4.414	35.168	4.891
	(6.544)**	(8.201)	(9.800)*	(6.724)	(6.817)**	(4.694)
Age2	-0.389	-0.092	-0.367	-0.069	-0.449	-0.093
	(0.077)**	(0.098)	(0.116)**	(0.080)	(0.081)**	(0.056)
University	379.099	112.386	243.810	91.789	266.344	140.187
	(37.974)**	(46.869)*	(54.541)**	(36.136)*	(41.377)**	(25.664)**
Technical	190.469	2.117	69.194	54.203	57.383	47.081
	(36.510)**	(42.481)	(52.169)	(33.435)	(40.152)	(24.389)
PTU 1	49.374	-43.377	-117.411	-24.110	-38.326	2.641
	(40.950)	(45.926)	(61.551)	(37.010)	(44.640)	(26.754)
PTU 2	29.637	-66.157	-34.488	45.818	-105.285	-1.476
	(47.106)	(53.921)	(72.830)	(46.796)	(53.254)*	(30.662)
Other Quals.	42.720	-82.750	-144.592	-48.117	-14.694	45.585
	(42.341)	(47.680)	(66.053)*	(36.181)	(48.155)	(28.830)
North West	83.355	-248.990	-74.875	-283.137	118.066	-116.845
	(51.778)	(73.535)**	(78.128)	(65.712)**	(57.282)*	(48.022)*
Central	-265.516	-306.965	-311.085	-318.211	-179.586	-218.597
	(39.265)**	(63.264)**	(57.314)**	(60.239)**	(42.953)**	(43.774)**
Volga	-368.290	-311.739	-474.672	-440.988	-215.499	-251.302
	(40.676)**	(62.459)**	(61.497)**	(59.326)**	(45.960)**	(43.270)**
Caucasus	-308.003	-339.224	-313.739	-399.983	-205.358	-260.768
	(45.716)**	(65.725)**	(69.570)**	(61.104)**	(50.731)**	(46.063)**
Urals	-190.510	-185.588	-273.558	-228.785	-189.485	-181.967
	(41.643)**	(64.530)**	(62.013)**	(59.346)**	(46.324)**	(44.090)**
Western Siberia	198.770	-247.740	189.993	-337.197	230.365	-193.010
	(48.672)**	(68.635)**	(73.810)*	(62.005)**	(56.277)**	(46.780)**
East	82.152	-190.736	-133.817	-363.395	-22.999	-191.335
	(48.010)	(66.243)**	(83.308)	(61.224)**	(54.385)	(45.698)**
State	-86.539	-34.385	-178.204	-27.526	-117.536	-58.777
	(25.128)**	(29.600)	(39.320)**	(24.746)	(27.499)**	(17.230)**
Agriculture	-250.863	-97.303	-251.971	-60.974	-169.516	-111.494
	(59.234)**	(54.806)	(95.687)**	(52.865)	(57.422)**	(32.569)**
Manufacturing	-11.362	15.018	46.962	46.869	-93.825	-35.587
	(45.083)	(52.705)	(71.786)	(51.275)	(44.513)*	(29.455)
Construction	257.273	282.506	291.500	174.120	75.914	-23.680
	(57.922)**	(62.508)**	(98.849)**	(62.138)**	(68.167)	(38.384)
Energy	260.745	1.952	441.998	244.459	146.231	92.060
	(52.590)**	(70.162)	(85.853)**	(58.332)**	(52.746)**	(34.916)**
Transport	232.231	38.041	300.775	91.000	71.933	105.861
	(52.218)**	(63.508)	(81.381)**	(60.297)	(50.453)	(36.463)**
Retail	41.109	234.521	30.394	152.660	76.558	58.465
	(51.443)	(72.558)**	(78.304)	(67.108)*	(49.225)	(44.084)
Finance	383.194	729.191	457.864	427.141	131.729	111.311
	(94.452)**	(187.211)**	(122.168)**	(215.898)*	(84.225)	(100.764)
Health/Education	-65.742	79.929	28.073	-19.806	-92.157	-20.960
	(42.608)	(55.645)	(69.499)	(49.456)	(41.488)*	(28.996)
Firm size 11-50	-9.312	11.026	27.490	-0.066	83.296	-69.379
	(47.472)	(57.127)	(73.808)	(53.442)	(55.176)	(37.761)
Firm size 51-100	13.184	54.215	-12.727	46.439	40.027	-28.619
	(53.774)	(64.810)	(83.160)	(58.747)	(60.579)	(41.245)
Firm size 101-500	78.130	109.737	47.045	62.613	66.823	-19.890
	(48.484)	(57.951)	(76.932)	(53.216)	(58.393)	(37.207)
Firm size 501-1000	177.653	189.683	86.523	59.353	293.726	-6.018

	(51.093)**	(60.095)**	(80.915)	(57.817)	(60.541)**	(39.212)
Firm size missing	55.855	37.957	2.156	-27.662	93.407	-22.801
	(48.708)	(58.531)	(70.627)	(52.176)	(55.107)	(36.204)
Job Tenure 1-2 yrs	50.552	-15.643	99.058	53.044	61.998	33.746
	(40.085)	(50.518)	(63.362)	(43.535)	(43.738)	(28.735)
2-5 yrs	-44.714	12.982	184.181	66.827	72.079	32.913
	(35.527)	(43.460)	(54.085)**	(37.186)	(38.963)	(24.867)
5-10 yrs	-14.406	-15.183	125.330	25.428	73.904	62.436
	(39.176)	(45.520)	(60.085)*	(38.875)	(42.646)	(26.842)*
10-20 yrs	25.448	-3.080	120.659	13.702	99.791	56.950
	(38.021)	(45.063)	(60.814)*	(37.612)	(45.074)*	(27.370)*
20 yrs+	112.283	-32.667	159.724	62.463	102.873	113.582
	(44.758)*	(52.423)	(68.120)*	(42.388)	(50.534)*	(30.381)**
Constant	470.968	412.070	727.343	401.153	33.770	350.394
	(142.358)**	(181.035)*	(206.173)**	(155.882)*	(152.983)	(107.224)**
N	2645	1332	1368	1532	1669	1674
Adjusted R-squared	0.28	0.12	0.25	0.11	0.23	0.11

Table A2. Logit Estimates of Probability of Not Being in Arrears

	1994	1996	1998
Female	0.273	0.148	0.100
	(0.078)**	(0.088)	(0.080)
Age	-0.028	-0.025	-0.043
	(0.022)	(0.024)	(0.022)*
Age2	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
University	-0.110	0.358	0.266
	(0.125)	(0.131)**	(0.125)*
Technical	-0.164	0.310	0.216
	(0.116)	(0.124)*	(0.120)
PTU 1	-0.056	-0.004	0.056
	(0.127)	(0.142)	(0.133)
PTU 2	0.022	0.276	0.068
	(0.147)	(0.173)	(0.155)
Other Quals.	0.030	-0.149	0.221
	(0.132)	(0.147)	(0.142)
North West	-0.810	-1.205	-1.206
	(0.184)**	(0.208)**	(0.201)**
Central	-0.446	-0.616	-0.745
	(0.153)**	(0.176)**	(0.170)**
Volga	-0.659	-1.067	-1.252
	(0.153)**	(0.177)**	(0.173)**
Caucasus	-0.695	-1.136	-0.944
	(0.167)**	(0.194)**	(0.191)**
Urals	-0.561	-1.015	-1.095
	(0.157)**	(0.178)**	(0.176)**
Western Siberia	-0.826	-1.264	-1.270
	(0.173)**	(0.196)**	(0.196)**
East	-0.884	-1.777	-1.335
	(0.169)**	(0.208)**	(0.192)**
State	-0.271	-0.182	-0.270
	(0.080)**	(0.093)*	(0.083)**
Agriculture	-0.760	-0.716	-0.383
	(0.168)**	(0.221)**	(0.172)*
Manufacturing	-0.382	-0.401	-0.582
	(0.143)**	(0.180)*	(0.138)**
Construction	-0.530	-0.462	-0.617
	(0.173)**	(0.232)*	(0.194)**
Energy	0.319	-0.230	-0.202
	(0.178)	(0.208)	(0.161)
Transport	-0.010	0.265	0.099
	(0.168)	(0.207)	(0.162)
Retail	0.355	0.482	0.622
	(0.184)	(0.216)*	(0.182)**
Finance	0.792	1.955	1.039
	(0.434)	(0.564)**	(0.377)**
Health/Education	0.396	-0.235	-0.311
	(0.143)**	(0.174)	(0.131)*
Firm size 11-50	0.045	-0.048	0.297
	(0.156)	(0.191)	(0.180)
Firm size 51-100	0.155	-0.178	0.335
	(0.178)	(0.212)	(0.196)
Firm size 101-500	0.161	-0.373	-0.038
	(0.159)	(0.196)	(0.183)
Firm size 501-1000	0.002	-0.192	0.150
	(0.167)	(0.209)	(0.193)

Firm size missing	-0.044	0.076	0.169
	(0.160)	(0.186)	(0.177)
Job Tenure 1-2 yrs	0.080	0.075	0.294
	(0.134)	(0.158)	(0.137)*
2-5 yrs	-0.004	0.182	0.283
	(0.117)	(0.134)	(0.120)*
5-10 yrs	-0.180	0.026	0.270
	(0.125)	(0.145)	(0.130)*
10-20 yrs	-0.108	-0.136	0.231
	(0.122)	(0.143)	(0.135)
20 yrs+	-0.190	-0.118	0.205
	(0.143)	(0.160)	(0.151)
loc3	-0.592	-0.483	-0.508
	(0.100)**	(0.119)**	(0.106)**
Constant	2.065	1.710	1.911
	(0.480)**	(0.535)**	(0.493)**
N	3977	2899	3341
Log L	-2364.0	-1818.6	-2158.9

Table A3. Earnings Mobility in Russia, 1994-98

a) 1994/95

Total			1995		
1994	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
1	46.5	21.1	12.2	10.8	9.5
2	20.9	37.8	25.3	10.6	5.1
3	17.1	22.5	31.5	21.9	7.0
4	11.1	8.2	22.2	34.8	23.9
5	11.0	5.2	8.2	20.7	54.8
No Arrears			1995		
1994	1	2	3	4	5
1		62.5	25.0		12.5
2		49.8	33.5	10.9	5.0
3		21.6	41.6	29.2	7.6
4		4.9	24.4	43.2	27.4
5		1.5	6.3	23.6	68.5

b) 1995/96

Total			1996		
1995	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
1	58.9	8.0	14.6	11.5	7.1
2	36.8	17.5	29.0	11.5	5.2
3	25.1	7.9	33.8	24.2	9.0
4	21.1	2.5	20.2	32.3	24.0
5	17.2	1.8	8.7	20.2	52.1
No Arrears			1996		
1995	1	2	3	4	5
1					
2		21.3	54.6	19.4	4.6
3		3.0	49.4	38.1	9.5
4		1.0	15.4	48.0	35.6
5		1.4	3.2	20.7	74.7

c) 1995/98

Total			1998		
1995	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
1	41.2	17.7	18.2	13.7	8.9
2	30.8	24.6	26.0	13.3	5.3
3	24.9	10.8	29.7	23.9	10.8
4	19.9	7.1	17.9	31.9	23.1
5	14.9	3.4	12.3	24.9	44.6
No Arrears			1998		
1995	1	2	3	4	5
1		50.0	16.7	33.3	
2		21.1	43.7	29.6	5.6
3		3.5	41.4	41.4	13.8
4		2.6	15.7	44.4	37.8
5		0.9	8.9	20.5	69.6

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
395	P. Manzini C. Ponsatí	Stakeholders, Bargaining and Strikes	6	11/01
396	M. A. Shields S. Wheatley Price	Exploring the Economic and Social Determinants of Psychological and Psychosocial Health	5	11/01
397	M. Frondel C. M. Schmidt	Evaluating Environmental Programs: The Perspective of Modern Evaluation Research	6	11/01
398	M. Lindeboom F. Portrait G. J. van den Berg	An Econometric Analysis of the Mental-Health Effects of Major Events in the Life of Elderly Individuals	5	11/01
399	J. W. Albrecht J. C. van Ours	Using Employer Hiring Behavior to Test the Educational Signaling Hypothesis	1	11/01
400	R. Euwals	The Predictive Value of Subjective Labour Supply Data: A Dynamic Panel Data Model with Measurement Error	5	11/01
401	J. Boone P. Fredriksson B. Holmlund J. C. van Ours	Optimal Unemployment Insurance with Monitoring and Sanctions	3	11/01
402	O. Ashenfelter D. Card	Did the Elimination of Mandatory Retirement Affect Faculty Retirement Flows?	5	11/01
403	L. Ljungqvist	How Do Layoff Costs Affect Employment?	1	11/01
404	H. Battu C. R. Belfield P. J. Sloane	Human Capital Spill-Overs Within the Workplace	1	11/01
405	L. Locher	Testing for the Option Value of Migration	3	11/01
406	P. Garibaldi E. Wasmer	Labor Market Flows and Equilibrium Search Unemployment	1	11/01
407	R. Schettkat L. Yocarini	Education Driving the Rise in Dutch Female Employment: Explanations for the Increase in Part-time Work and Female Employment in the Netherlands; Contrasted with Germany	5	12/01
408	H. N. Mocan E. Tekin	Nonprofit Sector and Part-Time Work: An Analysis of Employer-Employee Matched Data of Child Care Workers	1	12/01
409	P. Apps R. Rees	Fertility, Female Labor Supply and Public Policy	6	12/01
410	H. Lehmann J. Wadsworth	Wage Arrears and the Distribution of Earnings in Russia	4	12/01

An updated list of IZA Discussion Papers is available on the center's homepage *www.iza.org.*