WAGE ARREARS AND INEQUALITY IN THE DISTRIBUTION OF PAY:

LESSONS FROM RUSSIA

HARTMUT LEHMANN [1] AND JONATHAN WADSWORTH [2 *]

January 2006

1. University of Bologna; IZA, Bonn; CERT, Heriot-Watt University, Edinburgh

2. Royal Holloway College, University of London; Centre for Economic Performance, LSE; and IZA, Bonn

* Corresponding author:

Jonathan Wadsworth
Economics Department
Royal Holloway College, University of London
Egham
United Kingdom
TW20 0EX

E:mail: j.wadsworth@rhul.ac.uk

Tel: (00 44 0)1784 443464

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, and two anonymous referees, participants at the CEPR-WDI Conference on Transition in Moscow as well as seminar participants 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 the Microeconomics of Labor 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.

WAGE ARREARS AND INEQUALITY IN THE DISTRIBUTION OF PAY:

LESSONS FROM RUSSIA

Abstract

Many developing and transition countries, and even some in the industrialized West, experience periods in which a substantial proportion of the workforce suffer wage arrears. We examine the implications for estimates of wage gaps and inequality using the Russian labor market as a test case. Wage inequality grew rapidly as did the incidence of wage arrears in Russia in the 1990s. Given data on wages and the incidence of wage arrears we construct counterfactual wage distributions, which give the distribution of pay were arrears not present. The results suggest that wage inequality could be some 30 per cent lower in the absence of arrears. If individuals in arrears are distributed across the underlying wage distribution, as appears to be the case in Russia, we show that it may be feasible to use the wage distribution for the subset of those not in arrears to estimate the underlying population wage distribution parameters.

JEL Classification No.: O1, J0

Key Words: Wage Arrears, Earnings Inequality, Pay Gaps, Counterfactual Estimates

I. Introduction.

Many countries in the developing world, those undergoing the transition from planned to market economic systems and even those in the industrialized West, experience periods in which a substantial proportion of the workforce suffer wage arrears[1]. For any research based on wage distributions, such as estimation of wage inequality, gender pay gaps or the returns to education, failure to account for wage arrears can have important implications, as we show below. Russia is particularly interesting in this regard, since it experienced well-documented increases in both the incidence of wage arrears and wage inequality over the first decade of the transition period. Moreover, the availability of data on both these issues facilitates exploration of the linkages between the two that is not always possible in other countries.

One contributory factor toward inequality in the wage distribution, in any country, could be the presence of wage arrears. If in any given month some workers receive only part of their normal wage, or no wage at all, then wage inequality will be higher, or in exceptional cases lower, than otherwise. Wage inequality in Russia following the end of central planning rose much more than in Central and Eastern European (CEE) countries undergoing transition. The Gini coefficient for wages in Russia rose from 0.22 before transition to around 0.5 in 1996, (Flemming and Micklewright, 1997) and has remained around this level ever since. Wage inequality in Russia is also very high by international standards.[2] Wage arrears were also a pervasive feature of Russian economic life over the 1990s. Lehmann, Wadsworth and Acquisti, (1999), show that around 65 percent of the workforce was owed money at the height of the problem in 1998. Moreover, the withholding of wage payments was systematic and concentrated

---

1 A glance at the BBC web site: www.bbc.co.uk contains reports on unpaid wages in Argentina, Azerbaijan, Belarus, Bulgaria, Central African Republic, China, Colombia, Honduras, Iran, Kazakhstan, Kenya, Kosovo, Mexico, Niger and the Ukraine as well as Russia over the last 5 years. Following the introduction of the national minimum wage in Britain in 1999, a recent report indicates that some 36% of firms were underpaying their minimum wage workers http://news.bbc.co.uk/1/hi/business/2255947.stm

2 Over the same period, the Gini indices 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. In Chile, the Gini coefficient is around 0.45 and in Turkey around 0.37.

heavily on sub-sections of the workforce in certain regions and industries (see e.g. Earle and Sabirianova, (2002), and Lehmann, Wadsworth and Acquisti, 1999). Despite their prevalence, most studies of wages, however, tend to ignore the effect of wage arrears on the earnings distribution.[3]

In what follows, we use data for Russia, to try to estimate what the wage distribution would have looked like if all workers had been paid the full contractual wage on time. By establishing the parameters of the underlying distribution it is then possible to adjust estimates of any between-group differences based on the observed wage distribution. Using Russian Longitudinal Monitoring Survey (RLMS) data, we apply several different imputation methods to generate predicted wages for those in arrears and construct counterfactual estimates of the underlying wage distributions.

We find similar results across the various imputation methods. Earnings dispersion may have been some 30 percent lower if workers had been paid in full. Since, on average, women seem to be less affected by wage arrears, (Lehmann, Wadsworth and Acquisti, 1999), the mean gender gap is larger in the counterfactual distributions compared with the observed gender pay gap. We also look at pay gaps across other quantiles of the earnings distribution, which cannot be done in the presence of large-scale wage arrears. We then look at how wage arrears affect estimates of the returns to education and relative wage distributions by region and industry.

In the next section we discuss the rationale for constructing counterfactual wage distributions. The subsequent section outlines the various methods used 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 the

3 Oglobin (1999) is an exception, using a selection equation in his analysis of the mean gender pay gap in Russia.

counterfactual results. Section VII then concludes.

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

Is a wage distribution that assumes payment of wages in full and on time for all employees a realistic counterfactual to pursue? Of course, economic welfare depends on the actual distribution of earnings, but a comparison of the actual and counterfactual will provide a means of estimating the cost of arrears. Also, if wage arrears are a problem of irregular pay and not of permanently withheld wages, then we have a strong rationale for constructing counterfactual wage distributions, since there is less concern over possible general equilibrium effects concerning any trade-off between the elimination of wage arrears and employment.[4] We believe that evidence garnered from various sources on the dynamic nature of the arrears process provides this rationale. Aggregate data from Russian Statistical Office (Goskomstat) indicate that since 1996 the stock of wage arrears has been approximately stable, equivalent to the aggregate wage bill for two months. At the same time, there is strong evidence in the RLMS data, the principal source of data in this paper, which supports the hypothesis of wage arrears as a problem of irregular pay rather than that of permanently withheld wages. Lehmann, Wadsworth and Acquisti (1999), use the RLMS to document the existence of simultaneous inflows into and outflows from wage arrears. In the data we analyze below, 10% of workers are in arrears at all four interview points and 20% never experience wage arrears.

These flow patterns allied to the fact that the stock of wage arrears is, approximately, in a steady state at the end of our sample period, suggest that the amount of contractual wages not paid to (some) workers is close to the amount of wage debts paid back to (some) workers in any month. Payroll data from a sample of 19 firms in a central Russian industrial city also seem to confirm this pattern in Figure A1. At times, the stock of arrears in some firms rises, while falling

4 With no trade-off between wage arrears and employment the counterfactual becomes the actual underlying

in others. Moreover, the Figure indicates that wage arrears are eventually paid off, and at different rates across firms. It seems that most workers do get paid the wages owed to them eventually.

Given this, because the RLMS elicits information on wages received in the month of the survey, this window might be too narrow to obtain an estimate of the contracted earnings of workers affected by pay irregularities. For example, in an economy where all workers get paid monthly but the data window on earnings is the third week of the month, if we ask: "How much did you get paid in the third week of this month?" some workers will have been paid their monthly salary in this week, but many will have been paid in another week of the month. 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, six months, we could obtain better estimates of the contracted monthly earnings of Russian workers. Since we do not have such a wide sample window in which to observe everyone paid in full at least once, then counterfactual distributions provide one way of estimating contracted monthly earnings.[5]

III. Building Counterfactual Estimates of the Effects of Wage Arrears

The literature suggests several methods of building counterfactual mean estimates, Y^0, given membership of a treatment group, $T_i \in \{0,1\}$ essentially built on the conditional independence assumption, (CIA), whereby assignment to the treatment group is ignorable conditional on a set of exogenous control variables, X, that are unaffected by the treatment and that $Y^0 \perp T/X$. Given the CIA and assuming there is overlap, or common support, in the X distributions of those in arrears and those not, if we take experience of wage arrears as the

---

wage distribution that would occur in the absence of arrears.

5 These are imperfect estimates, since the counterfactuals ignore the losses in earnings over time due to inflation, foregone interest and the costs of borrowing. However incidences of wage arrears in Russia were much higher after the hyper-inflations of the mid-90s when inflation rates were back to single figures (Gimpelson, 2000). The RLMS data do not give the dates of when arrears occurred so it is not possible to ascertain the dynamic history of the wage arrears process needed to infer inflation, interest and borrowing costs.

treatment and let the X variables influence the likelihood of being observed in arrears, then the counterfactual mean of the wage distribution for this treatment group equals the mean of the wage distribution for the no arrears control group, adjusted for differences in observable characteristics across the two groups.

In our case we are interested in not just the counterfactual mean but also the counterfactual distribution of wages, netting out the effect of 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, 2001) and international differences in wage inequality (Blau and Kahn, 1996). Given the CIA assumption Imbens (2004) shows that it is possible to identify different quantiles of a counterfactual distribution. Fröhlich (2003) shows that either matching on observables or propensity score estimation can be used to estimate counterfactual density functions consistently in addition to counterfactual means, since $E[\xi_{y/x,t}(X,t)/ X=r, T=t] = \xi_{y/p,t}(r,t)$ is satisfied both for $\xi_{y/x,t}(X,t)= E[Y/X=x, T=t]$ and the conditional density function $\xi_{y/x,t}(X,t)= f_{y/X, T=t}$. Once the counterfactual density is estimated the counterfactual quantiles can be recovered.[6]

If selection into the treatment group also depends on unobservables, then identification of the counterfactual densities, as with counterfactual means, requires data from before the treatment began in order to difference or net out any bias caused by unobservables. This generally requires the assumption that the bias caused by unobservables is constant over time. Whether researchers can ever be truly confident that treatment selection is observable, or that any bias from unobservables is constant, are moot points. We therefore produce a series of estimates that rely on the CIA, but which involve different sets of assumptions and look to

6 Firpo (2004) demonstrates that it is possible to estimate the quantiles directly without first estimating the counterfactual distribution. For estimates of the conditional variance or quantiles of the distribution, Fröhlich (2003) shows that while matching on a set of covariates X is consistent, propensity score matching is not.

compare the estimates based on the different methods.

We begin with a simple least squares prediction and then use 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[7]. We then apply a different residual according to the method proposed by Juhn, Murphy Pierce (1993). We next provide counterfactual estimates of the wage distribution following the Kernel density approach pioneered by DiNardo, Fortin and Lemieux (1996). We then employ a variation of the exact matching techniques used by, among others, Heckman, Ichimura 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 labor market pre-treatment history. The last method used matches on the propensity score rather than a vector of characteristics, (for example Lechner, (2002)).

OLS methods

Following Oaxaca (1973) we can estimate a wage equation using the sample of those without wage arrears. Using the vector of (consistently) estimated parameters from this equation and the observed characteristics of those in arrears we then predict wages, which those in arrears would receive 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 individual and job-related characteristics that determine whether the i-th person experiences arrears. The set of covariates is based on those used by Lehmann, Wadsworth and Acquisti, (1999) who used the same data set to examine the incidence of wage arrears[8]. The predicted

7 Imbens (2004) notes that the “debate concerning the practical advantage of the various estimators … is still ongoing with no firm conclusions yet reached.”

8 The set of controls include individual controls for age, gender, education and tenure job-level controls for 1 digit industry, firm size and region (see Appendix Table A1). Lehmann et al. (1999) and Earle and Sabirianova (2000, 2002) find that job and location rather than individual characteristics are the more relevant predictors of the incidence of arrears.

wage of this individual, $Y_{i,\,WA}$, will be:

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

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

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

Table A1 in the appendix gives the estimates from the OLS real wage equations for the no arrears group used to generate these estimates.

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

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

Applying this method in the context of wage arrears, the counterfactual is then the set of wages that would result if the no arrears wage coefficients and residual standard deviation were given to those currently in arrears. 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.[9] The method uses the standard residuals from the arrears regression to calculate counterfactuals. This standardized 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 analyze in equation (3) gives an individual's relative ranking in the residual arrears wage distribution, which is hard to interpret as a measure of unobserved skill, unless one is prepared to make the unlikely assumption that the size of non-payment reflects unobserved skill. The estimates from the equations for those not in arrears used to construct the counterfactuals are given in Table A1 in the Appendix.

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 principle 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 conditional on x can be written $g(w) = \int f(w/x)h(x)dx$. Following DFL, using Bayes' law, the counterfactual wage distribution if everyone 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 $\Phi(x)$, where $\Phi(x)$ reflects the relative incidence of arrears conditional on characteristics x, $\Phi(x)$ = Pr(No Arrears) / Pr(No Arrears/x). The weights are normalized 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, producing no predictions of individual wages, only the quantiles of the distribution. The numerator in $\Phi(x)$ is the sample

9 This is not always the case in our data.

proportion of those not in arrears in any year and the denominator is estimated by a logit regression conditional on a set of characteristics determining the incidence of arrears. The estimates from the logit equations used to construct these estimates (Table A2) confirm the dominance of location and firm characteristics in explaining arrears, as found in Lehmann et al, (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, however they implicitly assume that heterogeneity amongst those not in arrears is duplicated amongst those in arrears. If those not in arrears are different from those in arrears, the counterfactual estimates could be biased.

We therefore experiment with alternative approaches based on the matching estimator literature. The first technique follows Heckman, Ichimura and Todd (1997) in that we also condition, non-parametrically, on "pre-treatment history" in order to minimize any biases arising from unobserved heterogeneity. 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. 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. Heckman, Ichimura and Todd (1997) find that for this type of matching estimator to work well the same data set should be used for the control and treatment group, the groups should be in the same local labor markets and the data set should contain a rich set of variables relevant to the treatment decision.

Using the panel element of the RLMS we condition on labor 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. 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 in both years are given the current wages of those not in arrears now that were in arrears one year earlier. Those in arrears now but non-employed a year ago are given the current wages of those non-employed a year ago but paid in full now. In each case, if more than one person can be matched with the individual we assign the average wage of the matched controls. In addition we match according to age (with a maximum allowed difference of ten years), gender, region (3 groups, Metropolitan Moscow and St.Petersburg, East, and West) and qualifications (3 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[10]

The matching algorithm is shown in Box A1 in the appendix. 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 comparisons using the other counterfactual techniques estimated over the same sample. The approach assumes that individuals do not move much across the earnings distribution.[11] Figure 2, which for those currently in arrears, plots the share coming from each wage decile in the previous year, also suggests that those in arrears are drawn from across the entire wage distribution.

Propensity Scores

10 Sample size constraints prevent us from matching within all eight macro regions identified by the data and used in the OLS estimates. Also, whilst within regional 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.
11 The IZA discussion paper version of this paper presents one and four-year earnings transition matrices. Whilst there is a degree of mobility, there is considerably less amongst those not in arrears.

The non-parametric matching approach omits around 10 per cent of potential matches for whom a donor from the control group cannot be found. To avoid this lack of common support, we also employ propensity score matching, where all 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 (2002).[12] 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 to those not with the nearest propensity score. We estimate two variants of the propensity score, one with pre-treatment variables included in the set of co-variates and one without.

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. We use the surveys conducted in the autumn of 1994, 1995, 1996 and 1998, the period in which wage arrears first emerged and subsequently affected two-thirds of the workforce at the height of the problem in 1998. The data contains a set of demographic and establishment characteristics, together with information on the labor market activities of its sample. Despite its relatively small size, the advantage of this source 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.[13] 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

12 See also Kluve, Lehmann and Schmidt (2001). 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).

13 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.

which around 4000 are in work and around 3,500 give wage related information.

The survey questions dealing with wage arrears ask whether, conditional on being in work, 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 some money, whilst others, around one half of those in arrears, receive nothing. The RLMS also asks for the total amount owed, together with the number of months since the worker was paid last, but does not give the dates of when arrears occurred so it is not possible to ascertain the dynamic history of the wage arrears process. It may be that some of those not in arrears are paid more than their monthly wage if arrears are paid back. There may also be some in arrears who were paid in full in the current month. However there is no way of ascertaining these issues from the data. Respondents, both those paid in full and those in arrears, are also asked to state the amount of *money* received from their employers after tax in the past month. These are total wage receipts and not contractual wages, on which there is no reliable information.[14] There is no distinction made between basic wages and bonus. This constitutes the "true" wage for those paid on time.

These wage responses are then deflated by a national price deflator indexed to 100 at January 1998. We remove outliers from that data, namely those earning in excess of 4000 rubles a month, or less than 50 rubles if the respondents are not in arrears.15 Since we are interested in the impact of arrears on the aggregate distribution, we do not construct gender-specific counterfactual wage distributions.[16] Standard errors around the quantiles of the observed and

14 A question on the contractual wage appears for the first time in 1998, but the responses given for those in arrears unfortunately hardly differ from the actual wage responses. Therefore, we cannot use this information.
15 This comprises less than 1% of those at the bottom of the no treatment group and less than 1% of those at the top of the wage distribution.
16 Given a sufficiently large sample this would, of course, be possible. In what follows we capture gender effects through a simple intercept dummy variable.

counterfactual distributions are generated using the bootstrap method.[17] We also use a smaller, Russian household survey data set, VTsIOM[18], undertaken in 1993, in order to provide summary comparative evidence on pay from an earlier period when wage arrears were less prevalent, together with labor force survey data from Poland and Britain as benchmark comparisons. The former is a transition economy without wage arrears or a dominant oligarchy that followed a different restructuring process where more attention was given to sharing the costs of reform equally, (Hellman, 1998). The latter is a Western economy where wage inequality had risen sharply just prior to the sample period.[19]

V. Earnings Distributions and Inequality in Russia

The timing of the dramatic rise in inequality during the first years of transition, documented in Brainerd (1998), indicates that most of the rise in inequality occurred before the problem of wage arrears really began, though hyperinflation at the onset of reforms was probably not the sole contributing factor to the initial rise in inequality. However, as inflation subsided aggregate inequality remained high. The RLMS data indicate that inequality fell in regions with a low incidence of wage arrears, and rose most in regions with the largest increase in wage arrears. The Gini coefficient in the metropolitan areas, where arrears are lowest, fell from 0.39 to 0.35 between 1994 and 1998, but rose from 0.43 to 0.49 in the Far East, where arrears are highest. It seems important, therefore, to try to analyze to what extent wage arrears have affected the earnings distribution since payment problems began.

In order to demonstrate the effects of wage arrears on the wage distribution, Table 1 gives summary measures of the changes in real monthly wage distribution across our sample period.

---

17 Imbens (2004) questions the validity of bootstrap-based standard errors in the case of matching. In practice, the subsequent tables show that the standard errors of the matching estimates do not differ markedly from the standard errors of the estimates derived from the other methods.

18 VTsIOM is the Russian acronym for the All-Union Center for the Study of Public Opinion

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

The VTsIOM data show that wage inequality was already higher in Russia than in Poland before wage arrears took off, indicative of the different restructuring paths pursued by the two transition countries. By 1996, the Gini coefficient on Russian wages was more than twice that observed in Poland and 60 percent higher than in Britain. 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 slowly, but the Gini coefficient and the ratio of the 90^{th} to 50^{th} wage quantiles falls back. The Table also shows that real average earnings fell markedly over the sample period, as a series of national economic crises left inflation soaring and nominal wages failing to keep pace. By 1998, around two thirds of employees were not receiving a wage complete or on time, and around 40% of these received nothing in the preceding month. The large number of zero wage observations means that any conventional measures of inequality based around logarithmic transformations will be of little use.

The inequality estimates are influenced strongly by wage arrears. Figure 1 tracks the increasing skewness of the real monthly wage distribution as the incidence of arrears builds up. The bottom panel of Table 1 confirms that inequality is lower and rises by much less amongst those paid in full during the sample period. The Gini coefficient, for example, is around one third for the subset of those without wage arrears, in any period. Many individuals appear in low deciles solely because they are not paid at all or paid only part of their wages.

Counterfactual Estimates

We now present our counterfactual estimates of the underlying wage distribution for the years 1994, 1996 and 1998. Table 2 summarizes details of the estimated distributions for the different methods used.20 Figure 3 graphs the counterfactual Kernel densities, the sum of the

20 Other quantiles and moments of the distributions are available on request.

actual wage of those paid in full and the predicted wage of those in arrears. Table 2 confirms that the mean and various quantiles of the distributions are all higher using any of the counterfactual estimates. The bootstrapped standard errors indicate that all the imputed distributions lie within 2 standard errors of each other, with the exception of the OLSI estimates – though these do not contain a random residual and so would be expected to differ. The magnitudes of the estimated standard errors are also similar. In general then, the counterfactuals indicate that mean wages would have been around 30% higher in 1994 and around 60% higher in 1998 in the absence of wage arrears. Similarly the estimated overall dispersion, as measured by the coefficient of variation, would be around 20% lower in 1994 and some 40% lower in 1998 in the absence of arrears. The counterfactual Gini coefficients are now similar to that observed in Britain around the same time but much higher than for Poland. Interestingly, the counterfactual Gini coefficients are also similar to those of the No Arrears sub-group in Table 1.

Table 3 uses the panel element of the data in order to add estimates based on exact matching and a second propensity score estimator based on "pre-treatment history" included as additional regressors in the propensity score logit. We compare the results with those using the other methods for the year 1996, based on the sub-sample with valid pre-treatment histories. We also show the distribution of those in the sample who get paid in full and on time, (column 2). The pattern of results follows that of Table 2. Mean wages would be around 60% higher and the wage distribution narrower by around 40% in the absence of wage arrears. Apart from the estimates based on simple OLS prediction (OLS I) all other counterfactual distributions have a similar spread as can be seen from the coefficients of variation and Gini coefficients.[21] Conditioning on pre-treatment history for the propensity score estimates (PSII), results in

21 The OLS estimates without the added residual are used only as a benchmark to highlight the problem of distribution imputation based solely on predicted values from a wage equation and we do not recommend that this technique be used to estimate counterfactuals.

estimates within two standard errors of the propensity score estimates without pre-conditioning. This suggests that unobserved heterogeneity as captured by this method is not important for this sample. Note that the quantiles of the no arrears distribution again appear insignificantly different from the counterfactuals, a point to which we return later.

These counterfactual techniques can also be applied to wages observed over any combination of years to give estimates of the average wage distribution over a given interval. One advantage of pooling data across years is that we can net out the influence of unobservables in the prediction equations through random effects estimation of the wage of treatment equations. Rather than reveal the counterfactual distribution at a single year, it inevitably reveals a medium-run average wage distribution, which is based on the predictions of a much smaller proportion of the population who are never in arrears over successive years. Table A2 in the appendix, showing the results for pooling the years 1994 to 1996, suggests little difference between the pooled and the random effects estimates of the counterfactual distributions.

Gender, Region and Education Pay Gaps Revisited

We now examine the implications of these counterfactual estimates for pay gaps between various sub-groups of the workforce. If the incidence of wage arrears is concentrated on sub-groups of the population, then pay gaps estimated on the observed distribution may be misleading.[22] In Table 4 we compare gender pay ratios using the actual distribution, the no arrears distribution and the counterfactual distributions for the year 1996. The imputation methods are broadly in agreement with the exception of the propensity score based estimates which show a narrowing of the gender pay gap rather than the expected widening when the incidence of arrears across gender is taken into account.23 The observed distribution suggests a mean gender pay gap of

22 Table A3 in the appendix gives marginal effects from logit estimates of the probability of being in arrears. The same estimates are used to generate the counterfactual kernel density estimates.
23 The differences for PSI and PSII relative to the other methods are not caused by the chosen parametric

around 20%, (column 1). Since women are less likely to be observed with wage arrears, the counterfactual estimates, other than PSI and PSII, suggest that if everyone were paid in full there would be more dispersion in pay between men and women and the gender wage gap would be closer to 30%.

Table 5 gives mean and median wages of three educational categories (graduate, intermediate and primary) and median pay ratios of the first two groups relative to the primary educational category using the actual, the no arrears and all counterfactual distributions. This time all the imputation methods are in broad agreement. Since graduates are under-represented among the arrears group, the observed distribution suggests a higher relative return to graduate education than the counterfactual estimates. There is less difference in the estimates of the relative returns for the intermediate group, since the incidence of arrears does not vary much compared with the default group.

We now turn to two dimensions that have the largest explanatory power in the incidence of wage arrears estimates, namely region and industry. We divide the sample into two areas: those living in Moscow and St. Petersburg (Metro), where the incidence of wage arrears is low and wages are high and those living outside the major metropolitan areas where wages are lower and the incidence of wage arrears is high. In Table 6, the actual distribution suggests that there is a 100% median wage gain from living in the metropolitan areas. Accounting for the skewed incidence of wage arrears by region reduces this regional wage premium to around 30%.

In Table 7 we aggregate industries into two sectors, production and services. Table A2 suggests that workers in the former are more likely to experience wage arrears than workers in the latter. The actual distribution suggests a median pay penalty in production relative to services. However, since the production sector is affected more by wage arrears, if everyone

---

specification of the prediction equation, since the results are very similar across different specifications. Nor do were paid in full this would be sufficient to generate a small pay premium for the production sector. Again the different imputation methods are broadly in agreement.

One striking feature 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 (2002) and Lehmann, Wadsworth and Acquisti (1999), it does suggest that those in arrears are drawn from throughout the underlying wage distribution. Figure 2 seems to confirm this. 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 to estimate the population parameters, subject to an efficiency loss.

VII. Conclusions

It seems apparent that estimates of wage inequality, and pay gaps in general, can be affected strongly in countries that experience bouts of wage arrears. Studies that fail to account for wage arrears can over-estimate wage inequality substantially in countries where arrears are eventually paid back. In countries where wage arrears are never paid back the actual wage distribution is more relevant for measuring inequality, assessing welfare costs and formulating appropriate policy responses. In countries where arrears are paid back, pay gaps across sub-groups of the population could be mis-leading if no account is taken of the differing incidence of wage arrears across these sub-groups. Russia in the 1990's, having both one of the highest levels of wage inequality and a large incidence of wage arrears, is a particularly interesting case. The large share of employees who receive no wages in any month also renders many conventional estimates of inequality based on logarithmic transformations inoperable.

Using imputation techniques that could be applicable to any data set for any country with

---

the results vary significantly depending on the propensity score matching method. Results available on request.

information on wages and wage arrears, we show that in the absence of arrears 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 conclusion is broadly the same whatever imputation method is used. This would put Russian wage inequality back towards levels currently experienced in Western countries like Britain. In the absence of arrears, the gender pay gap could be around 10 percentage points higher than the observed gap, though the imputation methods are less in agreement in this regard. Regional pay differentials would become more compressed and sectoral differentials would narrow in the absence of wage arrears. In this particular study, it appears that those in arrears are drawn from throughout the underlying wage distribution. For those wishing to study wage differentials and inequality, it may for Russia, be feasible to use the subset of those not in arrears and get close to the true population parameters.

References

Biewen, M., (2001) 'Item Non-Response and Inequality Measurement: Evidence from the German Earnings Distribution', *Journal of the German Statistical Society (Allgemeines Statistisches Archiv)*, 85, 409 – 425.

Blau, F. and Kahn, L. (1996), 'International Differences in Male Wage Inequality', *Journal of Political Economy*, 104 (4), 791-836.

Brainerd, Elizabeth (1998), 'Winners and Losers in Russia's Economic Transition.' *American Economic Review* 88 **(**5), 1094-1116.

Dehejia, R. and Wahba, S., (2002) , 'Propensity Score Matching Methods for Non-Experimental Causal Studies', *Review of Economics and Statistics*, 84 (1), 151-161.

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, 64 (5), 1001-44.

Earle J. and Sabirianova K. (2000), 'Equilibrium Wage Arrears: A Theoretical and Empirical Analysis of Institutional Lock-In.' IZA Discussion Paper 196.

Earle J. and Sabirianova K. (2002), 'Understanding wage arrears in Russia. *Journal of Labor Economics,* 20 (3) 661-707.

Firpo, S., (2004), 'Efficient Semiparametric Estimation of Quantile Treatment Effects.' University of British Columbia, Economics Department Discussion Paper 04-01, January.

Flemming, J. and Micklewright , J., (1997) *Income distribution, economic systems and transition*, EBRD, London.

Fröhlich, M., (2003), 'Propensity Score Matching Without Conditional Independence Assumption', Department of Economics, University of St. Gallen Working Paper.

Gimpelson, V. (2000), "Politics of Labor Market Adjustment. The Case of Russia", in J. Kornai, S. Haggard, and R. Kaufman, *Reforming the State: Fiscal and Welfare Reform in Post-SocialistCountries*, New York: Cambridge University Press,

Heckman, J., Ichimura, H. and Todd, P. (1997), 'Matching as an Econometric Evaluation Estimator.' *Review of Economic Studies*, 65 (2), 261-294.

Hellman, J., (1998), 'Winners Take All: The Politics of Partial Reform in Postcommunist Transition World Politics', *World Politics*, 50 (2), 203-234.

Imbens G., (2004), 'Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A review', *Review of Economics and Statistics*, 86 (1), 4-29.

Juhn, C., Murphy, K. and Pierce, B., (1993), 'Wage Inequality and the Rise in Returns to Skill', *Journal of Political Economy*, 101 (3), 410-42.

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, 27 (1), 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 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, 27, (5), 595-617

Lechner, M., ‘Program Heterogeneity and Propensity Score Matching’, *Review of Economics and Statistics*, 84 (2), 205-220.

Oaxaca, R., (1973), ‘Male-Female Earnings Differentials in Urban Labor Markets’, *International Economic Review*, 14, (5), 693-709.

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

Smith, J.A. and Todd P.E. (2001), ‘Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods’, *American Economic Review*, 91 (2), 114-118.

Figure 1. Distribution of Real Wages in Russia

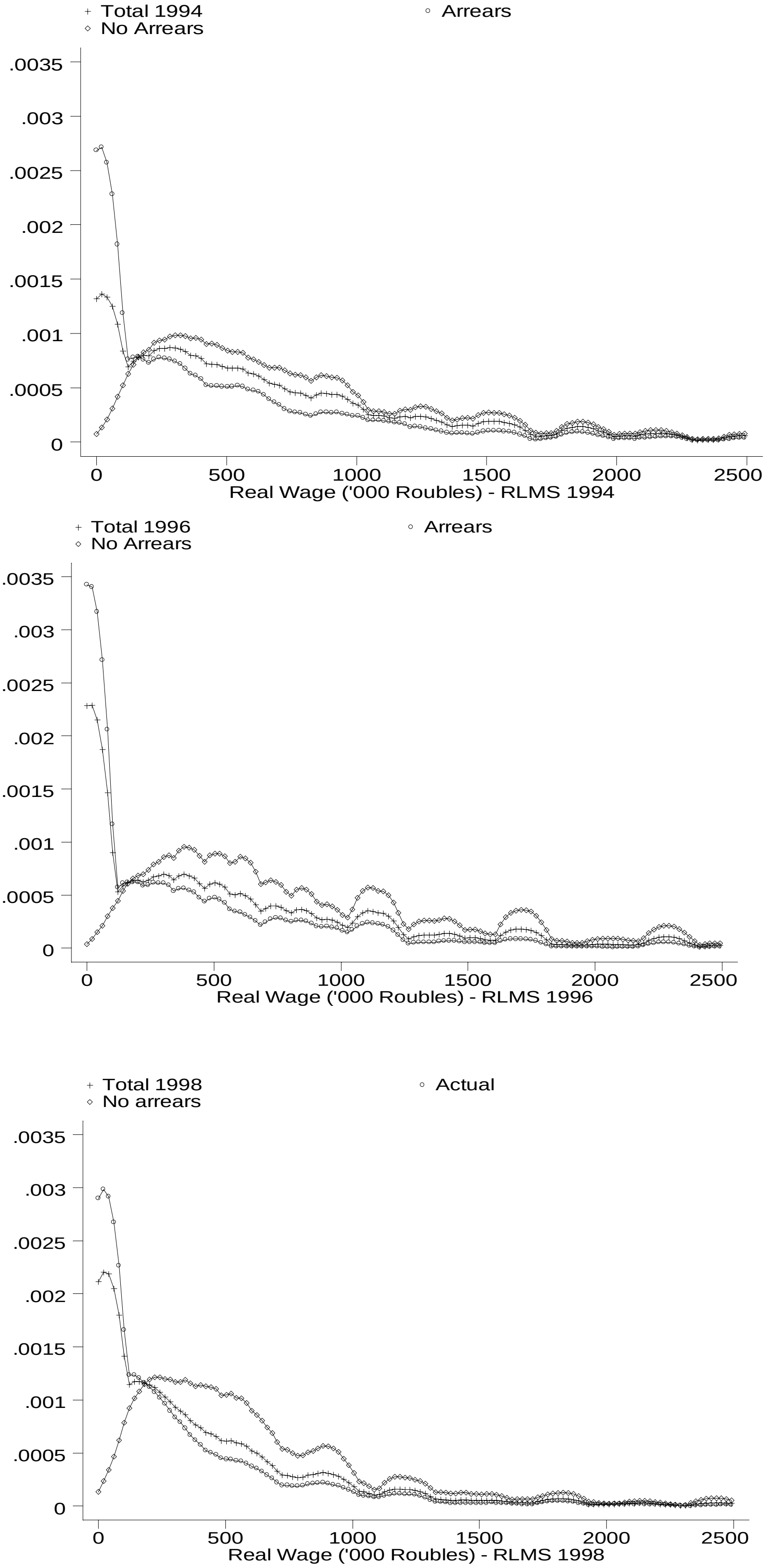


Figure 2. Previous Wage Decile of Those in Arrears

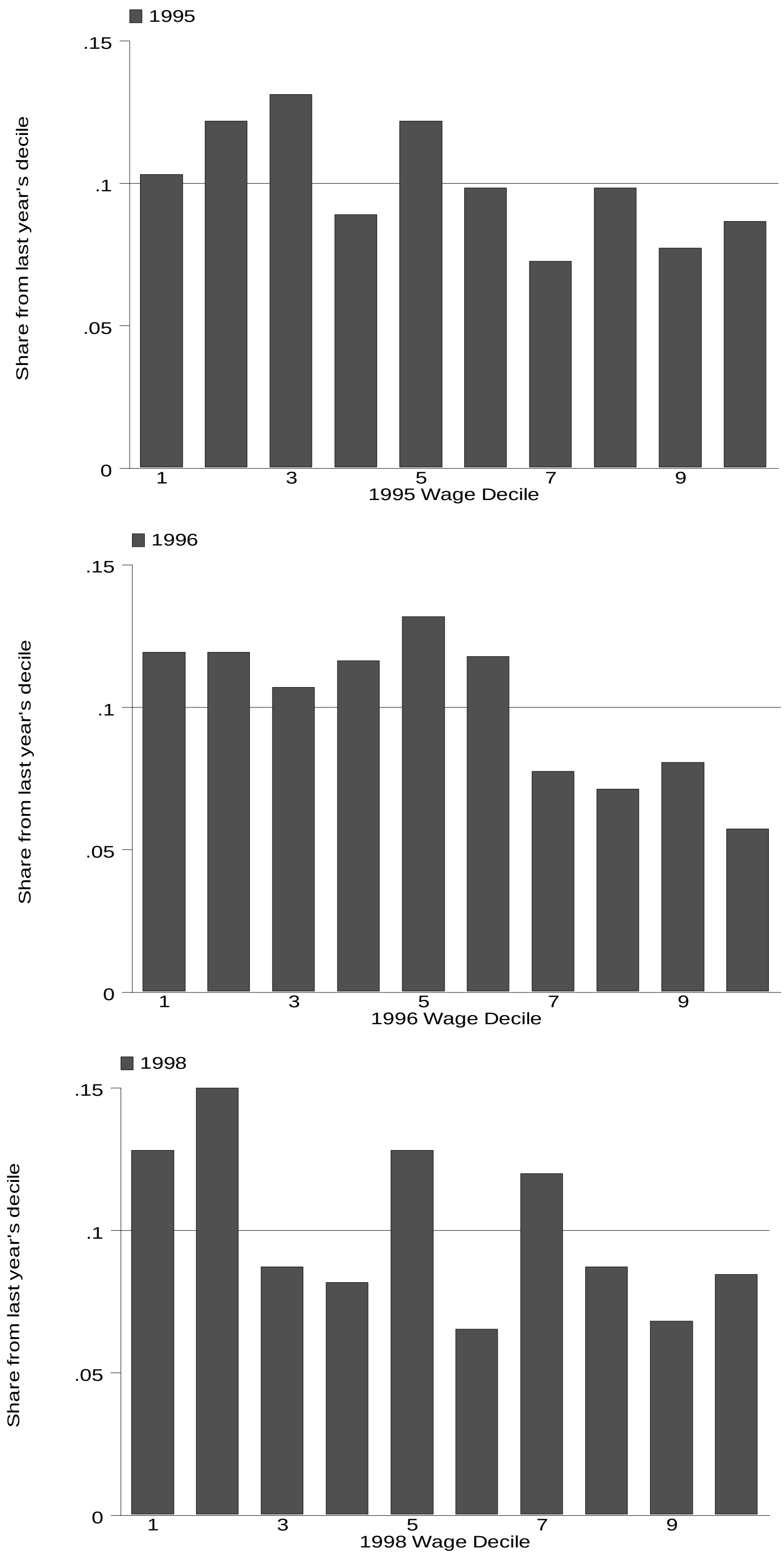
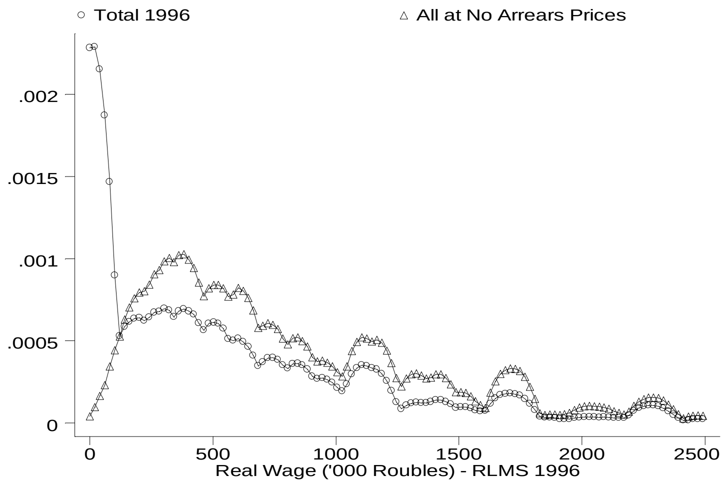


Figure 3. Counterfactual Estimates of Wage Distribution in Absence of Arrears


○ Total 1994
△ All at No Arrears Prices
.002
.0015
.001
.0005
0
0
500
1000
1500
2000
2500
Real Wage ('000 Roubles) - RLMS 1994


○ Total 1996
△ All at No Arrears Prices
.002
.0015
.001
.0005
0
0
500
1000
1500
2000
2500
Real Wage ('000 Roubles) - RLMS 1996

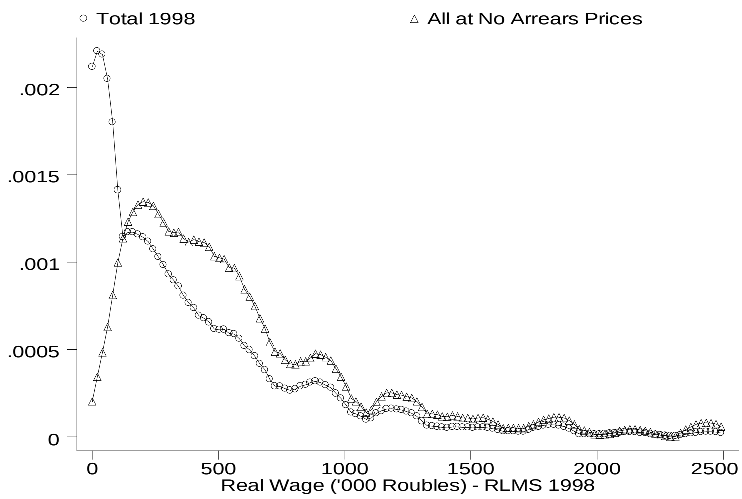

○ Total 1998
△ All at No Arrears Prices
.002
.0015
.001
.0005
0
0
500
1000
1500
2000
2500
Real Wage ('000 Roubles) - RLMS 1998

Table 1. Real Monthly Wage Distributions in Russia

	1993 VTsIOM	1994 RLMS	1996 RLMS	1998 RLMS	1996 Poland	1996 Britain
Total						
Mean	916	609	501	371		
	(1014)	(656)	(659)	(494)		
90th	1724	1500	1376	907		
	(23)	(19)	(14)	(11)		
50th	690	422	287	217		
	(28)	(13)	(34)	(11)		
10th	276	0	0	0		
	(20)					
90/10	6.25	n/a	n/a	n/a	2.70	8.55
90/50	2.5	3.55	4.79	4.18	1.83	2.20
50/10	2.5	n/a	n/a	n/a	1.48	3.89
Coef. Var.	1.11	1.11	1.32	1.33	0.62	0.80
Gini	0.407	0.547	0.637	0.619	0.239	0.387
	(.009)	(.005)	(.006)	(.006)		
% arrears	10	44.4	64.9	67.6	0	0
	(0.6)	(0.8)	(0.9)	(0.8)		
% no pay	0	19.3	34.6	28.1	0	0
		(0.6)	(0.9)	(0.8)		
No Arrears						
Mean	944	808	896	629		
	(1030)	(625)	(727)	(550)		
90th	1724	1718	1802	1273		
	(24)	(27)	(37)	(25)		
50th	690	625	677	484		
	(30)	(18)	(30)	(22)		
10th	276	188	229	187		
	(21)	(16)	(26)	(18)		
90/10	6.25	9.14	7.87	6.81		
90/50	2.5	2.75	2.66	2.63		
50/10	2.5	3.32	2.96	2.59		
Coef. Var	1.12	0.77	0.81	0.87		
Gini	0.407	0.420	0.415	0.428		
	(.011)	(.005)	(.008)	(.009)		

Note: wage data indexed to December 1997 prices. Wage observations for population of employees aged 18-69. Standard errors in brackets, bases on bootstrapping over 100 replications. Inequality measures use delta method approximation using standard normal distribution. Standard errors of proportions are used in percentage rows.

Table 2. Counterfactual Real Wage Distributions

	Mean	90th P'centile	Median	10th P'centile	90/10	90/50	50/10	Coef. Var.	Gini
1994									
Actual	629	1538	451	0	N/a	3.4	N/a	1.04	0.532
OLS I	743	1406	607	250	5.6	2.3	2.4	0.73	0.365
	(12)	(32)	(14)	(8)				(.01)	(.006)
OLS II	816	1672	613	190	8.8	2.7	3.2	0.88	0.429
	(17)	(60)	(15)	(8)				(.03)	(.006)
JMP	815	1688	625	203	8.3	2.7	3.1	0.81	0.411
	(15)	(57)	(13)	(11)				(.02)	(.007)
DFL	805	1719	625	188	9.1	2.8	3.3	0.82	0.417
	(16)	(80)	(15)	(5)				(.01)	(.005)
PS I	832	1818	625	188	9.7	2.9	3.3	0.81	0.420
	(17)	(74)	(10)	(7)				(.02)	(.006)
1998									
Actual	384	907	242	0	N/a	3.7	N/a	1.30	0.605
OLS I	517	907	422	212	4.3	2.1	2.0	0.73	0.337
	(14)	(23)	(11)	(10)				(.02)	(.008)
OLS II	594	1210	425	146	8.3	2.8	2.9	0.98	0.443
	(18)	(40)	(12)	(7)				(.05)	(.009)
JMP	607	1211	451	145	8.4	2.7	3.1	0.90	0.430
	(18)	(42)	(18)	(17)				(.03)	(.014)
DFL	588	1210	423	121	10.0	2.9	3.5	0.91	0.433
	(16)	(38)	(13)	(11)				(.03)	(.009)
PS I	609	1247	434	127	9.8	2.9	3.4	0.91	0.449
	(22)	(89)	(23)	(12)				(.03)	(.011)

Source: RLMS authors' calculations. Note: OLS I is OLS estimate without residuals, OLS II includes residuals, JMP is the Juhn-Murphy-Pierce decomposition, DFL is the DiNardo, Fortin, Lemieux technique, PS I is the estimate based on propensity score without conditioning on pre-treatment history. Actual values may vary from Table 1 due to missing observations on covariates used to construct counterfactuals. Bootstrapped standard errors in brackets based on 300 replications. Sample sizes: 3962 in 1994 and 3336 in 1998.

Table 3. Counterfactual Real Wage Distributions, 1996

	Actual	No Arrears	OLS I	OLS II	JMP	DFL	Match.	PS I	PS II
Mean	512	897	762	858	860	845	889	887	861
	(14)	(26)	(21)	(30)	(26)	(32)	(30)	(32)	(36)
90th	1261	1835	1351	1750	1776	1720	1720	1720	1802
	(66)	(125)	(65)	(79)	(64)	(71)	(114)	(130)	(129)
50th	339	688	635	630	674	630	688	688	631
	(18)	(14)	(21)	(25)	(19)	(37)	(37)	(23)	(31)
10th	0	229	304	227	194	221	229	229	225
		(10)	(20)	(10)	(24)	(22)	(9)	(12)	(18)
90/10	n/a	8.0	4.4	7.7	9.2	7.8	7.5	7.5	8.0
90/50	3.7	2.7	2.1	2.1	2.6	2.7	2.5	2.5	2.9
50/10	N/a	3.0	2.1	2.8	3.4	2.9	3.0	3.0	2.8
Coef. Var	1.26	0.79	0.68	0.88	0.83	0.83	0.77	0.81	0.84
	(.03)	(.02)	(.02)	(.05)	(.03)	(.03)	(.03)	(.03)	(.03)
Gini	0.617	0.405	0.332	0.423	0.411	0.411	0.392	0.409	0.423
	(.008)	(.009)	(.011)	(.011)	(.012)	(.012)	(.012)	(.012)	(.014)

Source: RLMS authors' calculations. 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. Bootstrapped standard errors in brackets.

Table 4. Counterfactual Gender Wage Ratio, (1996)

	Actual	No Arrears	OLS I	OLS II	JMP	DFL	Match	PS I	PS II
Men									
Mean	578	1113	922	1076	1043	1009	1078	929	910
Median	344	917	803	803	839	803	917	688	688
90th	1577	2294	1615	2231	2079	1950	2293	1835	1720
10th	0	344	351	279	322	252	321	229	203
Women									
Mean	459	752	633	723	714	704	687	847	821
Median	310	573	533	550	560	550	573	656	619
90th	1126	1605	1080	1425	1498	1456	1261	1720	1720
10th	0	221	262	201	145	184	216	229	203
GenderRatio									
Mean	0.79	0.68	0.69	0.67	0.68	0.70	0.64	0.91	0.90
50th	0.90	0.62	0.66	0.68	0.67	0.68	0.62	0.95	0.90
90th	0.71	0.70	0.67	0.64	0.72	0.75	0.55	0.94	1.00
10th	n/a	0.64	0.75	0.72	0.45	0.73	0.67	1.00	1.00

Source: RLMS. Sample size=2193, of which 976 are male and 1217 female.

Table 5. Actual and Counterfactual Education Wage Ratios, (1996)

	Actual	No Arrears	OLS I	OLS II	JMP	DFL	Match	PS I	PS II
Graduate									
Mean	594	944	823	923	917	907	902	904	852
Median	394	732	688	688	692	688	722	688	676
Intermed									
Mean	437	831	702	800	772	771	815	865	867
Median	248	631	573	573	573	563	642	653	630
Primary									
Mean	448	874	721	804	880	835	824	871	868
Median	229	581	585	569	688	607	574	676	631
Ratio:									
wrt primary									
Graduate	1.59	1.26	1.18	1.21	1.01	1.13	1.25	1.02	1.07
Intermed	1.08	1.09	0.98	1.01	0.83	0.93	1.12	0.96	1.00

Source: RLMS. Sample size=2193, of which 1059 are graduate, 759 intermediate and 415 primary. Ratios are based on median values in each group.

Table 6. Actual and Counterfactual Regional Wage Ratios, (1996)

	Actual	No Arrears	OLS I	OLS II	JMP	DFL	Match	PS I	PS II
Metro.									
Mean	758	1073	971	1072	1034	972	1074	1000	971
Median	573	845	821	803	802	803	917	803	788
Other									
Mean	462	847	720	815	824	819	817	861	838
Median	275	654	588	588	650	573	642	676	631
Ratio:									
wrt other									
Metro.	2.08	1.30	1.22	1.37	1.23	1.40	1.43	1.19	1.25

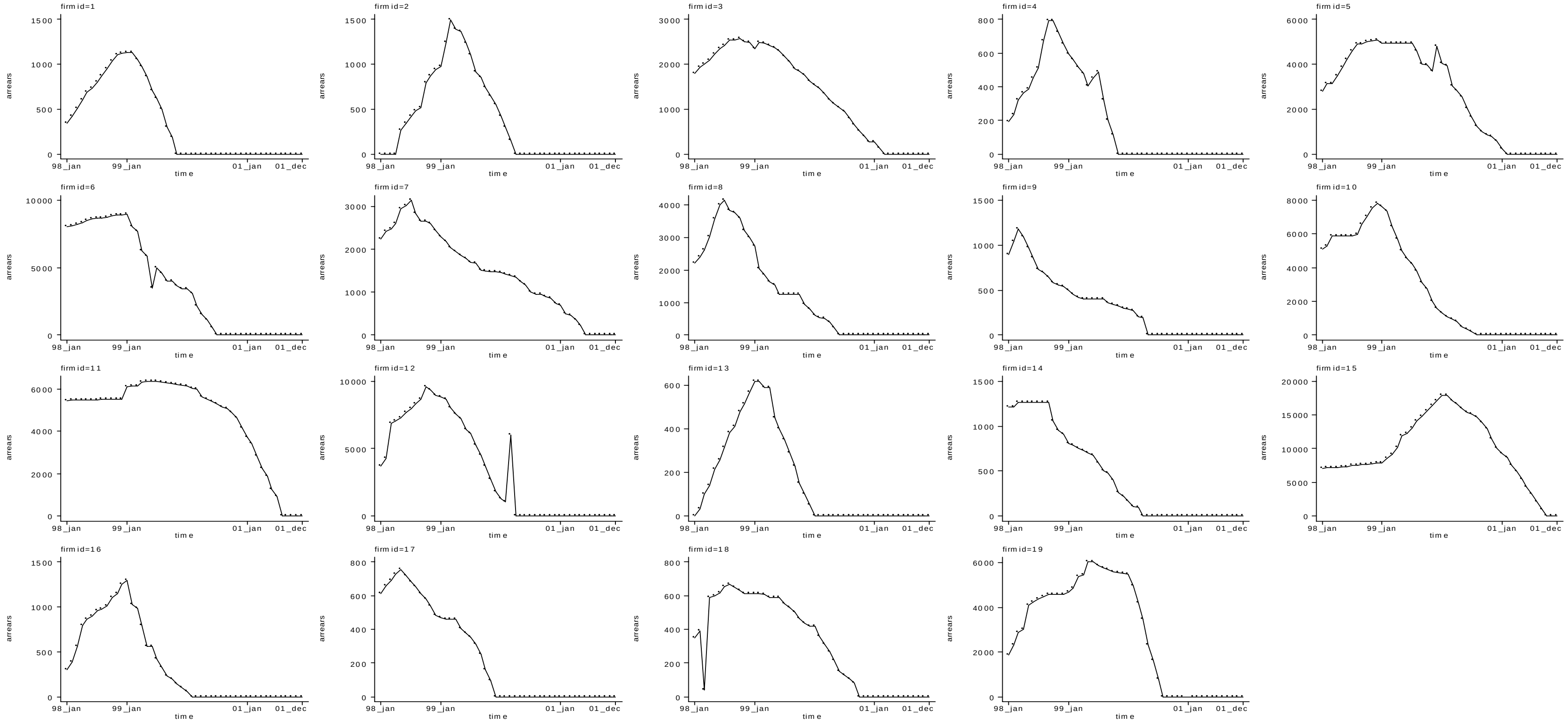
Source: RLMS. Sample size=2193, of which 332 are metropolitan, 1702 are elsewhere. Ratios are based on median values in each group.

Table 7. Actual and Counterfactual Industry Wage Ratios, (1996)

	Actual	No Arrears	OLS I	OLS II	JMP	DFL	Match	PS I	PS II
Production									
Mean	491	991	796	884	910	876	897	887	884
Median	281	784	675	642	739	596	748	688	654
Services									
Mean	531	843	730	835	814	813	824	882	839
Median	344	676	603	630	619	653	654	688	631
Ratio									
wrt services	0.82	1.16	1.12	1.02	1.19	0.91	1.14	1.00	1.04

Sample size=2193, of which 975 are production and 1059 services. Ratios are based on median values in each group.

Figure A1. Monthly stock of wage arrears within Russian firms (City of Ryazan –1998-2001)



Source: Authors' calculations based on CERT regional firm data.

Appendix

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

<u>Exact matching algorithm</u>

I. Condition on following possible pre-treatment labor market history:

- **employed and fully paid and in x-th 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

<u>Scheme of Conditioning on pre-treatment history by example</u>

<u>Pre-treatment period</u>	**<u>Treatment period</u>**
Potential Control 1 in 95 **Employed and fully paid and in 2^{nd} decile of wage distribution**	***Potential Control 1 in 96*** **Employed and fully paid**
Treated 1 in 95 **Employed and fully paid and in 2^{nd} 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 Log Real Weekly Wage Estimates for those Not in Arrears

	1994	1996	1998
Female	-0.430	-0.446	-0.417
	(0.033)**	(0.048)**	(0.047)**
Age	0.056	0.057	0.052
	(0.009)**	(0.012)**	(0.012)**
Age2	-0.001	-0.001	-0.001
	(0.000)**	(0.000)**	(0.000)**
University	0.512	0.251	0.456
	(0.051)**	(0.070)**	(0.076)**
Technical	0.302	0.084	0.193
	(0.049)**	(0.069)	(0.074)**
PTU 1	0.090	-0.094	-0.042
	(0.055)	(0.080)	(0.081)
PTU 2	0.052	0.004	-0.035
	(0.065)	(0.093)	(0.100)
Other Quals.	0.052	-0.125	-0.061
	(0.059)	(0.089)	(0.090)
North West	0.088	-0.063	-0.165
	(0.072)	(0.104)	(0.112)
Central	-0.349	-0.313	-0.311
	(0.052)**	(0.069)**	(0.070)**
Volga	-0.509	-0.528	-0.462
	(0.054)**	(0.078)**	(0.081)**
Caucasus	-0.479	-0.310	-0.438
	(0.060)**	(0.090)**	(0.086)**
Urals	-0.229	-0.232	-0.297
	(0.056)**	(0.078)**	(0.079)**
Western Siberia	0.119	0.278	0.281
	(0.065)	(0.098)**	(0.100)**
East	-0.014	-0.098	-0.178
	(0.068)	(0.112)	(0.101)
State	-0.115	-0.162	-0.229
	(0.034)**	(0.051)**	(0.050)**
Agriculture	-0.271	-0.352	-0.190
	(0.094)**	(0.143)*	(0.109)
Manufacturing	0.084	0.149	-0.028
	(0.062)	(0.091)	(0.079)
Construction	0.303	0.459	0.120
	(0.081)**	(0.131)**	(0.132)
Energy	0.331	0.423	0.313
	(0.072)**	(0.108)**	(0.096)**
Transport	0.287	0.373	0.196
	(0.070)**	(0.102)**	(0.088)*
Retail	0.073	0.162	0.163
	(0.069)	(0.095)	(0.081)*
Finance	0.411	0.634	0.248
	(0.121)**	(0.145)**	(0.130)
Health/Education	-0.098	0.052	-0.186
	(0.058)	(0.087)	(0.076)*
Firm size 11-50	0.040	0.038	0.044
	(0.063)	(0.094)	(0.093)
Firm size 51-100	0.093	0.048	0.110

	(0.072)	(0.109)	(0.105)
Firm size 101-500	0.176	0.127	0.117
	(0.064)**	(0.101)	(0.101)
Firm size 501-1000	0.277	0.171	0.403
	(0.068)**	(0.106)	(0.105)**
Firm size missing	0.109	-0.027	0.090
	(0.064)	(0.090)	(0.093)
Job Tenure 1-2 yrs	0.076	0.177	0.112
	(0.053)	(0.080)**	(0.076)
2-5 yrs	-0.026	0.252	0.117
	(0.048)	(0.068)**	(0.067)
5-10 yrs	0.021	0.107	0.183
	(0.052)	(0.077)	(0.076)**
10-20 yrs	0.081	0.201	0.292
	(0.051)	(0.077)**	(0.082)**
20 yrs+	0.224	0.243	0.215
	(0.060)**	(0.089)**	(0.092)**
Constant	5.470	5.635	5.373
	(0.190)**	(0.255)**	(0.268)**
N	2213	1019	1091
R^2	0.31	0.31	0.31

Standard errors in parentheses ** significant at 5%. Default region is metropolitan Moscow & St. Petersburg. Default industry is other services

Table A2. Counterfactual Average Real Wage Distributions, 1994-96

	Mean	90^{th} P'tile	Median	10^{th} P'tile	90/10	90/50	50/10	Coef. Var.	Gini
Actual	535	1250	375	0	N/a	3.3	N/a	1.11	0.555
OLS I	731	1215	625	360	3.4	1.9	1.7	0.58	0.284
	(22)	(58)	(19)	(12)				(.02)	(.009)
OLS I_ RE	746	1246	641	370	3.4	1.9	1.7	0.58	0.282
	(26)	(64)	(21)	(12)				(.02)	(.010)
OLS II	842	1674	633	248	6.8	2.6	2.6	0.85	0.401
	(24)	(55)	(19)	(11)				(.03)	(.008)
OLS II_RE	861	1670	654	252	6.6	2.6	2.6	0.85	0.400
	(30)	(63)	(20)	(10)				(.03)	(.009)
JMP	758	1536	581	213	7.2	2.6	2.7	0.80	0.400
	(13)	(36)	(11)	(12)				(.02)	(.007)
JMP_RE	750	1518	573	219	6.9	2.6	2.6	0.80	0.400
	(13)	(39)	(10)	(14)				(.02)	(.009)
DFL	753	1562	573	188	8.3	2.7	3.0	0.82	0.405
	(16)	(80)	(15)	(5)				(.01)	(.007)
DFL_RE	732	1562	530	181	8.6	2. 9	2.9	0.86	0.416
	(16)	(82)	(16)	(7)				(.01)	(.009)

Note: RE=counterfactual based on random effects regressions for prediction equations.

Table A3. Logit Estimates of Probability of Being in Arrears, (Marginal Effects)

	1994	1996	1998
Female	-0.070	-0.037	-0.018
	(0.019)**	(0.021)	(0.018)
Age	0.012	0.007	0.008
	(0.005)**	(0.006)	(0.005)
Age2	-0.0002	-0.0001	-0.0001
	(0.00006)**	(0.0001)	(0.0001)
University	0.030	-0.084	-0.086
	(0.029)	(0.031)**	(0.029)**
Technical	0.031	-0.030	-0.061
	(0.028)	(0.029)	(0.029)**
PTU 1	-0.007	0.001	-0.049
	(0.030)	(0.033)	(0.032)
PTU 2	0.018	-0.091	-0.029
	(0.036)	(0.043)**	(0.038)
Other Quals.	0.031	0.054	-0.044
	(0.032)	(0.033)	(0.035)
North West	0.204	0.326	0.382
	(0.042)**	(0.047)**	(0.046)**
Central	0.070	0.119	0.151
	(0.034)**	(0.037)**	(0.034)**
Volga	0.122	0.278	0.319
	(0.034)**	(0.039)**	(0.036)**
Caucasus	0.083	0.247	0.218
	(0.039)**	(0.044)**	(0.040)**
Urals	0.126	0.257	0.259
	(0.035)**	(0.039)**	(0.036)**
Western Siberia	0.145	0.333	0.299
	(0.039)**	(0.044)**	(0.042)**
East	0.252	0.429	0.358
	(0.039)**	(0.049)**	(0.043)**
State	0.079	0.051	0.109
	(0.019)**	(0.022)*	(0.019)**
Agriculture	0.262	0.216	0.074
	(0.045)**	(0.057)**	(0.042)
Manufacturing	0.071	0.156	0.162
	(0.034)**	(0.042)**	(0.031)**
Construction	0.152	0.142	0.183
	(0.042)**	(0.055)**	(0.048)**
Energy	-0.063	0.047	0.057
	(0.041)	(0.046)	(0.037)
Transport	-0.055	-0.067	-0.022
	(0.039)	(0.047)	(0.036)
Retail	-0.105	-0.143	-0.175
	(0.042)*	(0.048)**	(0.038)**
Finance	-0.254	-0.444	-0.338
	(0.098)**	(0.111)**	(0.078)**
Health/Education	-0.110	0.081	0.130
	(0.032)**	(0.040)**	(0.030)**
Firm size 11-50	0.062	-0.031	-0.031
	(0.038)	(0.045)	(0.041)
Firm size 51-100	0.021	0.056	-0.042

	(0.043)	(0.046)	(0.046)
Firm size 101-500	0.007	0.094	0.030
	(0.038)	(0.042)**	(0.041)
Firm size 501-1000	0.074	0.072	0.009
	(0.041)	(0.045)	(0.043)
Firm size missing	0.042	-0.040	-0.015
	(0.039)	(0.044)	(0.040)
Job Tenure 1-2 yrs	0.007	0.039	-0.027
	(0.032)	(0.037)	(0.032)
2-5 yrs	0.066	0.005	-0.024
	(0.028)**	(0.031)	(0.043)
5-10 yrs	0.069	0.053	-0.027
	(0.030)**	(0.034)	(0.027)
10-20 yrs	0.089	0.074	0.025
	(0.030)**	(0.034)**	(0.032)
20 yrs+	0.102	0.106	0.031
	(0.035)**	(0.038)**	(0.036)
Rural	0.207	0.197	0.182
	(0.025)**	(0.030)**	(0.027)**
N	3962	2884	3336
Log L	-2448	-1590	-1831

Standard errors in parentheses ** significant at 5%;