Wage Determination in Russia: An Econometric Investigation

By: Peter J. Luke and Mark E. Schaffer

Working Paper Number 295
March 2000

Wage Determination in Russia:
An Econometric Investigation

Peter J. Luke
P.J.Luke@hw.ac.uk

and
Mark E. Schaffer
M.E.Schaffer@hw.ac.uk

Centre for Economic Reform and Transformation
Department of Economics
School of Management
Heriot-Watt University
Edinburgh EH14 4AS, UK

March 2000

Abstract

Using a firm level dataset from four regions of Russia covering 1996/97, an investigation was carried out into how the surplus created within the firm is divided between profits and wages. An efficient bargaining framework based on the work of Svejnar (1986) is employed which takes into account the alternative wage or outside option available to employees in the firm as well as the value added per employee. Statistical differences in the share of the surplus taken by employees employed in state, private and mixed forms of firms are found. In addition, the results prove sensitive to the presence of outliers and influential observations. A variety of diagnostic methods are employed to identify these influential observations and robust methods are employed to lessen the influence of them. Whereas in practice some of the diagnostic and robust methods utilised proved incapable of identifying or accommodating the gross outlier(s) in the data, the more successful methods included robust regression, Winsorising, the Hadi and Siminoff algorithm, Cook's Distance and Covratio.

JEL Classification: C21, J300.

Keywords: Russian labour markets, efficient bargaining, outliers, regression diagnostics, robust regression.

Acknowledgements

1. Introduction

This paper investigates wage determination in Russia. Wage equations derived from an efficient bargaining model are estimated econometrically on a new firm-level dataset for 1996–97 covering several thousand industrial and construction firms from four regions of Russia. The central economic question investigated in the paper is how wage determination is related to form of ownership. In particular, we ask whether the degree of state ownership – full state ownership, fully privatised, or mixed state/private ownership – is related to the share of the firm's surplus taken by workers. Do workers in state-owned firms take a larger share of the surplus than in private firms, as our *ex ante* expectation of the nature of corporate governance would suggest? Is wage determination in partially privatised firms more akin to that in state-owned or in private firms?

The 'outside option' or 'alternative wage' – the expected income faced by workers who leave a firm – plays an important role in wage bargaining models of this type. The paper makes two contributions in this regard. First, we relax the usual modeling assumption that the alternative wage is exogenous, and derive from the model the implications of what happens when the wage set by a (large) firm has an impact on wages set by other firms – a not unrealistic possibility for Russia, where labour markets are highly segmented regionally, and wage setting by individual firms may have a general impact on wage setting in the region. Second, in our empirical work we calculate and work with a firm-specific alternative wage that is intended to capture alternative employment possibilities within the same sector and region as the firm. The motivation for this is again the regional segmentation of Russian labour markets.

Finally, our empirical investigation directly addresses a practical problem facing many researchers conducting econometric analyses using data from transition countries – the problem of 'dirty data' or outliers. Rather than deal with this problem in an *ad hoc* way, as most investigators do, we employ a wide range of econometric methods of outlier detection and robust estimation. We find that the econometric estimates of the model can indeed be highly sensitive to the treatment of outliers. The general lesson from this is that when investigators suspect the data they are using are 'dirty', they should conduct and report a range of outlier detection and robust estimation techniques.

The paper is organised as follows. Section 2 presents the theoretical and estimating frameworks used in the paper. Section 3 discusses various econometric techniques for outliers and robust estimation. The dataset used and the estimation results are described in Sections 4 and 5, respectively. Section 6 concludes.

2. Theoretical and Estimating Frameworks

2.1. Right-to-Manage vs. Efficient Bargaining

In attempting to model how insiders may raise their wage above the outside option or 'alternative wage' – the wage that can be expected to be earned by a worker were he/she to leave his/ her current job – various approaches have been taken by different researchers.

The question boils down to whether the firm is on its labour demand curve or on the contract curve. This is often expressed as the right-to-manage model versus the efficient bargaining model.[1] In the former model, the two sides within the firm – management and workers – bargain over the wage but not employment, manangement then sets the level of employment given the wage, and the outcome lies somewhere on the labour demand curve. In the latter model, both parties bargain over wages and employment and employment will be higher than if they bargain only over wages; efficient from the perspective of the firm but not of society as a whole.[2]

In its general form the right-to-manage model has a generalised Nash bargaining solution which is the product of the weighted net gains to each party: the gain to employees from expected income (from either employment in the firm or alternative employment) over certain alternative employment, and the gain to the firm from profits earned vs. shut-down or zero profits.

$$
\max_{w} B = \left\{ \left[\frac{L}{T} u(w) + \left(1 - \frac{L}{T} \right) u(w_A) \right] - u(w_A) \right\}^{g} \left\{ pQ(L) - wL - 0 \right\}^{1-g}
$$

$$
= \left\{ \frac{L}{T} [u(w) - u(w_A)] \right\}^{g} \left\{ pQ(L) - wL \right\}^{1-g}
$$

(1)

where L is employment, w is the wage, w_A is the alternative wage, u is utility from wage income, Q is output, $pQ - wL$ is firm profits, g is the bargaining strength of the employees, and T can be interpreted as the number of all workers (in the Western context, this would be union membership) including both those who end up being employed by the firm and those who are not employed by the firm and take alternative employment instead. The weights L/T and $(1 - L/T)$ can be interpreted as the probabilities faced by a representative worker of employment in the firm, and of having to take alternative employment (including possibly being unemployed), respectively. Bargaining (maximisation) is over the wage w, and the employment level L is chosen by the firm to maximise profits. The efficient bargaining model has an identical setup, with the simple difference that bargaining (maximisation) is over both the wage w and the employment level L, i.e.,

---

[1] See, in particular, for a very good summary of the issues, Booth (1995).
[2] See, however, Layard and Nickell (1990) who show that this is only true under limited assumptions.

$$\max_{w,L} B = \left\{ \frac{L}{T} [u(w) - u(w_A)] \right\}^g \{pQ(L) - wL\}^{1-g} \qquad (2)$$

Which of the two models would be more appropriate for modeling wage and employment determination in the Russian transition economy is a difficult question, as there is little micro empirical evidence on whether or how far bargaining within Russian firms extends beyond wage bargaining to include bargaining over employment. The labour hoarding that still continues today within Russian enterprises can be interpreted either within an efficient bargaining model, or, in a right-to-manage framework in which wages are extremely flexible, where employees have simply traded off real wages for relative employment security and in the process moved (a long way) down the labour demand curve.[3]

In practice, however, the two models yield very similar estimating equations. As Booth points out,[4] there is as yet no clear test to discriminate between the two models. For our purposes, the choice of model is mostly one of convenience. We use in this paper an efficient bargaining model by Svejnar (1986) that has also been used to estimate wage equations in a transition context (Prasnikar *et al.*, 1994 and 1997).

2.2. Estimating Framework

Svejnar (1986) sets out a Nash bargaining model of wage determination in which unions have fixed memberships and maximise their members' expected utility which is taken as a function of income and employment; the firm's utility is taken as equal to their profit. In the context of a transition economy such as the Russian Federation it is not labour unions but firm insiders that are seeking to maximise utility.

In Svejnar's model, Y is total labour income and the union utility function V reflects constant relative risk aversion:

$$V(Y) = \frac{g^d}{d} \qquad (3)$$

and the coefficient of relative risk aversion is given by:

$$r = \frac{-V''(Y)Y}{V'(Y)} \qquad (4)$$

Note $d < 1$ ($d > 1$) implies risk aversion (risk loving). Svejnar's final estimating model, conditional on $d = 1$, i.e., risk neutrality, is:

---

[3] See, e.g., Standing (1996) who argues that wages in Russia have been extremely flexible and the main reason for low observed rates of unemployment.

[4] *Op. cit.* p. 135.

$$Y = g\frac{\left(R - Y_a L\ - H\right)}{L} + Y_a \tag{5}$$

or

$$Y = Y_a + g(\ valad\ ) \tag{6}$$

where Y is the firm wage, Y_a is the alternative wage or outside option, R is firm revenue, H is non-labour costs, L is employment,[5] $valad = (\ R - Y_a L\ - H\ )/L$ is the value added per worker evaluated at Y_a, and g as before is a measure of workers' bargaining power. Value added per worker, *valad*, can be interpreted as the surplus per worker generated by the firm above the amount workers would obtain if they took up their outside option. Note that the main implication of the above equation is that the measure of insider power, g will also be the share of value added that workers take over and above the outside option. A value of one indicates that all added value is taken by the workers and so insiders have complete power. A value of zero indicates that the insiders have no power and receive only the outside option as a wage.

We are particularly interested in the impact of ownership form on wage formation, and therefore in the estimations we allow both for ownership-specific coefficients on *valad* and ownership intercept dummies. The data allow us to distinguish between state-owned firms, privatised firms, and 'mixed' firms with both state and private ownership shares. Our prior expectations are that the share of value-added appropriated by insiders will be larger in state-owned firms than in private firms because of the superior corporate governance in the latter, with mixed firms being an intermediate case.

Currie (1991) in her study of teachers' wages in Ontario uses an efficient contracts model and a standard labour demand model. The alternative wage is assumed to be a log-linear function of wages in surrounding school districts, average income in the district, local manufacturing wages, and a local employment index. Ashenfelter and Brown (1986) in their particular study of union branches belonging to the International Typographical Union actually use eleven different measures of the alternative wage. MaCurdy and Pencavel (1986) again using data on branches of the International Typographical Union use a wage index representing the real hourly earnings received by production workers in durable goods manufacturing. The point to note from this is that there does not appear to be a universal method of calculating *the* definitive alternative wage. It is very much a case of the particular circumstances determining the approach to be taken.

---

[5] Under the assumption of risk neutrality, i.e., $d = 1$, Svejnar's model leads to a vertical contract curve and hence a unique value of employment which coincides with the socially efficient level of employment, just as if the firm were on its labour demand curve paying its workers the outside option or the alternative wage. This is a strong assumption especially given the situation within Russia. Svejnar later swaps this assumption for one of a Cobb-Douglas production function which allows the value of d to vary.

If we wish to take into consideration unemployment and unemployment benefits then a further complication arises. In the unemployment benefit system in the Russian Federation that has been in place since the early 1990s, there is no single measure of unemployment benefit. Benefit is calculated as a percentage of the worker's average pay over the final three months of employment. At the time of writing the worker receives 0.75 of this figure for a total of three months; a further 0.6 of their previous average pay for a maximum of four months after the first three months have expired and a final 0.45 for a remaining five months with the possibility of a further six months on top of this after a review. This is further complicated by the fact that there is a minimum and a maximum which the claimant can receive. If in the first three months their calculated benefit exceeds the maximum then it is this maximum figure that they receive and not the 0.75 of their previous average wage. The final complication is that this maximum and minimum varies from *oblast* to *oblast* since it is calculated on a monthly basis using a basket of local consumer goods and hence local prices as opposed to the average price level for the country as a whole.

The Russian labour market is highly segmented, with very little labour mobility between regions. Labour skills are also sector-specific. This implies that the alternative wage faced by a worker in a firm will be the wage in alternative employment in a similar firm in the same region of Russia. Our strategy is to estimate the wage equation in which the alternative wage is calculated for each firm as the average wage in all other firms in the same sector and region; in other words, we make the alternative wage firm-specific:

$$Y_{a,i} = \frac{\sum\limits_{j,j \neq i}^{n} wage\ fund\ firm_j}{\sum\limits_{j,j \neq i}^{n} employment\ firm_j} \tag{7}$$

where n is the number of firms in a particular sector in a particular region.

We also recognise that recorded unemployment rates are likely to understate the difficulty a departing worker will have in finding a new job; although observed unemployment rates in Russia are low, outflow rates from unemployment are also low. We therefore do not use observed unemployment rates to weight the alternative wage but instead the coefficient on the alternative wage to vary from unity. Our prior expectation here is that the coefficient will be positive but significantly less than one.

Our estimating equation is therefore:

$$Y_i = a + a_{private} + a_{mixed} + g\,valad_i + g_{private}\,valad_i$$
$$+ g_{mixed}\,valad_i + bY_{a,i} \tag{8}$$

7

where *i* is used to index observations on individual firms and the coefficients denoted 'private' and 'mixed' are intercept and slope coefficients relating to the relevant ownership dummy variables. We also include regional and industry dummy variables in the estimation.

2.3. The Alternative Wage - Exogenous or Endogenous

Finally, there is a question mark on the endogeneity of the alternative wage. Lockwood and Manning (1989) have shown that in a dynamic framework all variables that affect firm profits will influence employment determination. However, for the purposes of the regressions, we have assumed, as Svejnar himself does, that the alternative wage is exogenous. This issue is, however, taken up when we allow the firm wage to vary with the alternative wage.

In Svejnar (1986), as in the literature generally, the alternative wage is taken as exogenous and the change in the firm's own wage is not seen to have any influence on the alternative wage. However, it can be argued that if the individual firm is large enough relative to the size of the branch in which it is operating, then any change in the firm wage will be noticed by workers in other firms who will recalculate *their* alternative wage on the basis of the new wage that they now see in effect in the other (large) firm. A change in the alternative wage for these workers will, as shown by McDonald and Solow (1981), lead to a leftward movement of their contract curve and so to a higher wage. The workers who originally began the process see that their alternative wage has increased and so their contract curve will move to the left and a higher wage will result.

The above mirrors, to a degree, the annual wage round within Britain in the sixties and seventies in particular when individual unions were keen 'to maintain differentials' and maintain their position within a pay league relative to other groups of workers. We have reworked Svejnar's basic model for his estimating equation but allow Y_a, the alternative wage, to be a function of Y, the firm wage.

$$U_L^g U_M^{1-g} = \left[\frac{L(Y_i - Y_a)}{\bar{\bar{L}}} \right]^g [R - Y_i L - H]^{1-g} \tag{9}$$

In a two-party framework the bargaining process is the maximisation of the above equation. This leads to the following steps:

$$\P(\bullet)/\P Y_i = \begin{aligned} &g\left[\frac{L(Y_i - Y_a)}{\bar{\bar{L}}} \right]^{g-1} [R - Y_i L - H]^{1-g} \left[\frac{L}{\bar{\bar{L}}} - A\frac{L}{\bar{\bar{L}}} \right] + \\ &\left[\frac{L(Y_i - Y_a)}{\bar{\bar{L}}} \right]^g (1-g)[R - Y_i L - H]^{-g}(-L) = 0 \end{aligned} \tag{10}$$

As is well known, there is a general problem with the reliability of data reporting within Russia. This arises for two reasons. Firstly, there is a problem related to the recording of the data by the relevant state statistical bodies.[8] Outliers arising from this fall within the realm of what are described as 'non-statistical factors'.

The second problem is potentially more serious and arises from contamination from another distribution other than the one under consideration. Here we are concerned with statistical factors that might help us explain why outliers occur, how they can be

---

[6] Grubbs (1969) quoted by Barnett and Lewis (1994).

[7] Rousseeuw and Leeroy (1987).

[8] To give but a few examples from the data we use in this paper. Two firms apparently paid negative wages to their workforces. Another two recorded negative fixed capital. A file containing observations on trade related firms within the Moscow area had several hundreds of duplicate observations masquerading as unique observations. In the original data that were obtained, the area or regional Goskomstat offices were using different units of account to measure the financial totals of some variables; for example, the fixed capital figures from the Chuvash file had to be divided by 1000 to ensure the same units were used throughout.

detected and if necessary whether they should be removed. For example, we may wish to examine the distribution of sales of Russian enterprises. If 60 per cent of enterprises report truthfully, but 40 per cent do not, then in effect we have two distributions masquerading as one.

Even if we assume there is no contamination, extreme values may well occur which have been truthfully reported and recorded by the relevant statistical body. Should these values be deleted, ignored or somehow accommodated? As Thomas (1997) writes, '... most econometricians prefer to retain unusual observations in their datasets'. But he continues, 'However, because OLS is so sensitive to such observations, other methods of observation are sometimes employed'. Reiss (1990) comments, 'When an observation is both an outlier and influential, investigators usually report regression results both with and without the observation'. Using a variety of econometric techniques that take on board the presence of outliers within the data, we present results based on the model outlined above. We focus in particular on the estimated relationship between ownership structure and the distribution of surplus (value added) and how it is affected by the removal of outliers; in other words, how robust are the findings?

There are two types of outliers. One is where the residual formed from the difference between the fitted $\hat{Y}_i$ and the actual Y_i shows up as particularly large. The second is where the X_i is far from the mass of X_i's. This means, for example, that an outlier will cause the direction of the fitted line to change but this will not show up in a high residual. This is a leverage point and the observation is usually referred to as influential. While outliers and influential observations are conceptually different it is not unusual to see influential observations simply referred to as outliers.

Identifying outliers when there is only one contained within a dataset is relatively straightforward. If there is more than one then problems can arise from what is known as masking and swamping effects. Masking occurs when a subset of outliers is not detected because of the presence of another subset, usually close to the former. Swamping, on the other hand, occurs when 'normal' observations are taken as outliers by the detection process because of the presence of another subset of observations.

One approach to outlier identification is to make use of the residuals from the model estimation. However, given the *normal* error model the difficulty arises that the estimated residuals do not have constant variance (Behnken and Draper, 1972). The variance of the OLS residuals is given by:

$$var(\boldsymbol{e}_i) = \boldsymbol{s}^2 \left\{ \left(1 - \frac{1}{n}\right) - \left(x_i - \bar{x}\right)^2 \middle/ \sum_{1}^{n} \left(x_i - \bar{x}\right)^2 \right\} \qquad (12)$$

As can be seen the ith individual residual is not homoskedastic and is more variable the closer the x_i explanatory variable is to the overall mean. This 'ballooning' effect, as it

is known, would have to be taken into account if it is pronounced when examining the size of residuals as an indication that it is an outlier.

In addition it can also be shown that the OLS residuals have different y's in common and so are correlated, i.e.,

$$e_i = y_i(1 - h_{ii}) + \sum_{j \neq i} h_{ij} y_j \qquad (13)$$

where h represents the hat matrix. It is partly for the above reasons that some researchers prefer to use other types of residuals when constructing diagnostic tests. The description of these and the outlier detection/robust methods are described more fully in Appendix 2.

As Rousseeuw and Leroy (1987) point out, approachs to outlier detection can be placed in one of two camps. They distinguish between robust regression and outlier diagnostics. Regression diagnostics attempts to identify observations that are potentially anomalous in the dataset after applying a regression and then examining the residuals (of whatever hue). One may then decide to keep or delete the outliers having studied them. One of the criticisms thrown at regression diagnostics is that it employs the residuals from the regression. However, the size of the residual is not necessarily any guide to whether the observation is an outlier or not since leverage points will have small residuals.

Robust regression, on the other hand, tries to lessen or to accommodate the influence of any outlying observations in the estimators employed so that it is not necessary to remove them at all as their influence is downweighted. Some diagnostic methods attempt to take this on board (Cook's Distance, for example) by calculating the leverage of the residual as well as the size of the residual. The methods employed in our study are summarised below.

Diagnostic Techniques	Robust Estimation
Dffits	Median Least Squares (MLS)
Cook's Distance	Reweighted Least Squares (RLS)
Welsch Distance	Winsorising
Covratio	Non-Linear Least Squares
Hadi and Simonoff Procedure (HSP)	Robust Regression

This is not an exhaustive list, nor is this the first time a comparison has been made of different techniques on a dataset (see e.g., Cook and Weisberg, 1982, and Belsley *et al.*, 1980). This paper is to our knowledge, however, the first time that such techniques have been applied and compared using a micro dataset from a transition economy such as the Russian Federation.

4. The Dataset and Summary Statistics

The data cover a three year period from 1995 to 1997 and come from Goskomstat, the Russian Federation state statistical body. The information is collected from all medium-sized and large firms, of all categories of ownership – state-owned, privatised, and 'mixed' firms with both private and state ownership shares. The four regions covered are Moscow City, Krasnoyarsk Krai, Chuvashia Republic and Chelyabinsk Oblast.[9] The data cover all medium and large industrial and construction firms in these regions.[10]

As argued above, it is appropriate in the Russian case to treat the alternative wage as region- and sector-specific. Alternative wages at the two-digit branch level were calculated as follows.[11] First, the 'branch wage fund' and 'branch employment' were calculated for each of the four regions in the data. Second, for each firm, that firm's wage fund and employment level were subtracted from the branch wage fund and branch employment, respectively, to generate an alternative branch wage fund and an alternative employment level (given the same branch and location). Finally, the former divided by the latter gives the alternative wage for each firm. The alternative wage thus calculated can be interpreted as the wage a worker would expect, on average, if leaving his current job and taking a job in a different firm in the same branch in the same region.

The wage fund is made up of monetary payments for labour activities performed *and* of payments in kind. The fund takes in incentive payments, bonuses and special payments connected with the conditions of work (dangerous or harmful conditions, night working, overtime payments etc.). Also included in the wage fund are regular payments on food, fuel and living accommodation that are received by the workforce either as a result of agreements between the workforce and management or due to legislation. Paid holidays, leave for training and professional retraining are also included in this. The average wage for each firm was calculated as the firm wage fund divided by the average listed number of workers.

Value added per worker is calculated as the sum of the wage fund and reported profits, less the alternative wage (see above). Profit is calculated according to the standard Russian definition except that it is gross of depreciation. Depreciation is not reported directly in the data but is instead estimated for each firm using reported fixed capital and the average depreciation rate for Russian industries reported in the annual statistical yearbook published by Goskomstat.

---

[9] The reasons for choosing them as representative economic regions within the Russian Federation are discussed in Lehmann, Wadsworth and Acquisti (1998).

[10] The raw data also cover wholesale and retail trade firms, but missing data meant there was an insufficient number of firms from these sectors to be included in the study.

[11] The two-digit levels means, for example, the fuel sector (11), ferrous and non-ferrous metals (12), machine construction (14) and so on.

Tables 1–2 present summary statistics for the data that relate to those observations actually used in the regression analysis, i.e., excluding firms with incomplete data. Output and output per worker, while not used in the estimating equation, are also presented.[12] All financial data are current prices in millions of rubles.

5. Estimation Results

Table 3 below gives the results obtained from using the above methods on the dataset when applied to the basic model estimated in levels for the year 1997. The first column reports the results of estimating the model on the full sample of data using ordinary least squares (OLS). Subsequent columns report estimation after deletion of outliers and/or using the named robust estimation technique. The number of observations dropped as outliers in each estimation is reported in the last row of the table. The omitted regional intercept dummy variable is Chelyabinsk.

Starting first with outlier detection methods, several results are worth stressing. In general the largest reduction in the residual sum of squares caused by the removal of a subset of n observations from a dataset are seen as the most likely candidates to be outliers (Gentleman and Wilk, 1975). Put another way the greatest increase in the R^2 is one way to discriminate between different outlier detection methods. On that basis, the Hadi and Simonoff procedure (HSP) performs best – the R^2 increases from 0.29 with the full sample to 0.66 when the outliers are deleted. Cook's distance and covratio are next with R^2's of 0.47 and 0.48, respectively. Welsch's distance and Dffits, however, actually decrease the R^2 by the deletion of their outliers.

Turning to the robust estimation techniques two methods stand out – winsorising and robust regression. In the latter case, an R^2 is not applicable because of the weighting procedure. The coefficents generated by robust regression are close to Cook's distance, which follows from the fact that this robust regression technique uses Cook's distance initially to remove 'gross outliers' (see Appendix 2). The R^2 from the winsorised regression is of the same magnitude as the HSP and the coefficients produced between the two methods are, relatively speaking, very similar.

In the course of the study it was found that there was one very influential observation. Its deletion results in a dramatic change in the coefficients and in the value R^2 – see the final column in Table 3, 'Minus One'. The coefficient on value added per employee, for example, jumps from 0.001 to 0.27; on the interactive dummies for value added per employee they change from 0.071 and –0.001 to –0.198 and –0.270 for private and mixed property firms, respectively.

Further examination of the results reported in Table 3 show that the estimation results can be separated into two groups: those in which the estimated coefficient on value

---

[12] A small number of firms do not report output data but report profit, wage and employment data and hence can be included in the regression analysis.

added per worker is extremely small, about 0.001 – OLS applied to the full sample, Dffits, Welsch's distance, non-linear least squares, and median least squares – and those in which the estimated coefficient is between 0.27 and 0.33, values that are relatively reasonable and consistent with our prior expectations – HSP, Cook's distance, Covratio, robust regression, Winsorising, and 'Minus One'. What is happening is that the HSP, Cook's distance, covratio, robust regression and Winsorising methods pick up this gross outlier while the other methods do not.

With one dependent variable and one regressor a graph of *y* against *x* can be displayed along with the fitted line of a regression. Outlying observations then stand out for the eye to see. The added variable plot is a diagnostic tool that is an attempt to do the same but for models where there is more than one regressor. In our case, we can see graphically whether value added per employee has any outliers by creating an added variable plot with the residuals from a regression of value added per employee on all other explanatory variables on the X-axis, and the residuals from a regression of the firm wage on all the right-hand side variables except value added per employee on the Y-axis. Figure 1 shows such a plot for our basic regression for 1997, with each observation marked by a '1' if the HSP outlier detection process deemed it an outlier and a zero otherwise. The added variable plot in Figure 1 is quite dramatic and shows the value of this diagnostic tool. The outlier is clearly a classic leverage point that because of the extreme distance that it lies from the other points gives the mistaken impression of the rest of the observations forming a vertical line.

Figure 1: Added-variable plot using value-added per worker in 1997
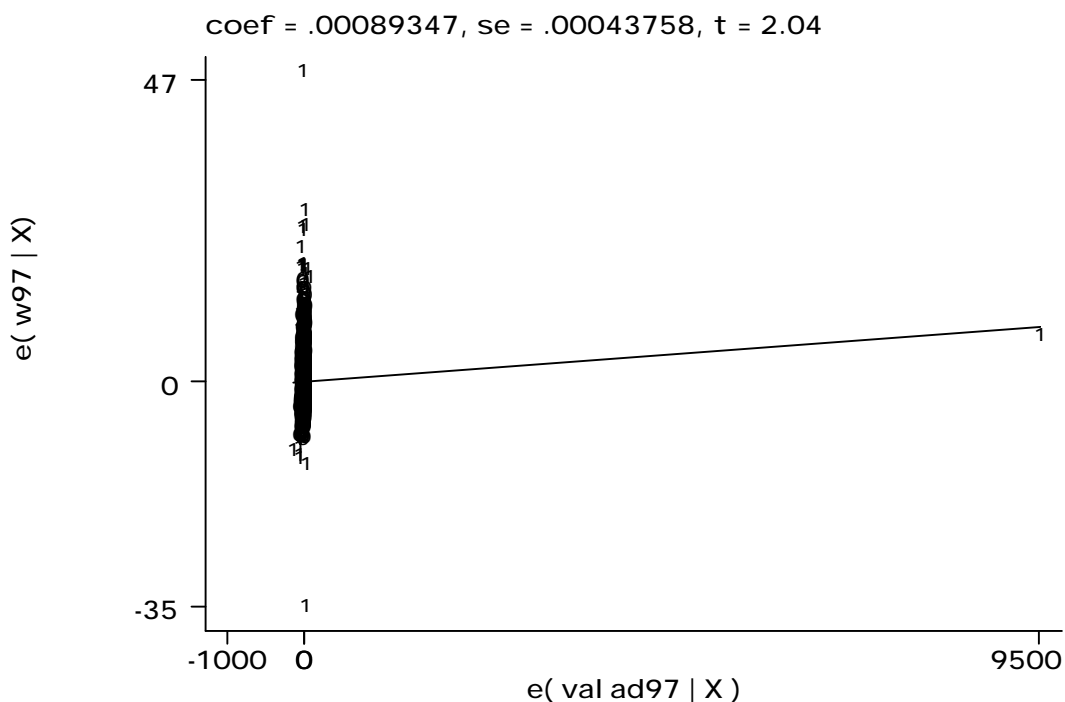
This gross outlier provides an acid test for the regression diagnostics employed along with the robust estimating techniques. The observation is contained within the Krasnoyarsk region, is a state-owned firm and is in the construction sector. Statistical information on it and the rest of the data are presented separately in Table 4. It can be seen that this firm has an extremely high value of profits per worker, with unremarkable values for the other variables. We note that this outlier is not the result of a simple reporting error – the same firm reports a very large value for profit in the preceding year, 1996, as well.

Our first conclusion, therefore, is that even sophisticated outlier detection methods and robust estimation techniques can fail to detect a gross outlier whose presence in the dataset can have an overwhelming influence on the regression results.

Compared to the coefficients on value added per employee the coefficients on the alternative wage are relatively stable regardless of which technique is employed. The value of the coefficient varies from 0.538 to 0.720. This is significantly less than unity, and is consistent with our prior expectations (see above).

We turn now to the estimated coefficients on value added per worker for the different ownership categories, considering only the results from the six methods that excluded the gross outlier: HSP, Cook's distance, Covratio, robust regression, Winsorising, and simple exclusion of the gross outlier ('Minus One'). The benchmark ownership category is state-owned firms; the interactive coefficients on *valad97* reported for private and mixed firms are deviations from this benchmark category. The value of the share of the surplus taken by insiders for these ownership categories is obtained from the sum of the benchmark (state) coefficient and the estimated coefficient. For example, the coefficient on *valad97* in the HSP results is 0.339, indicating that workers in state-owned firms appropriate 33.9 per cent of the available surplus. The coefficient on *private*valad97* is –0.045, indicating that workers in private firms appropriated – 4.5 per cent less than workers in state-owned firms, or 29.4 (33.9 – 4.5) per cent of the available surplus.

All six methods show that the share of the surplus taken by workers in private firms is significantly less than that taken by workers in state-owned firms; the range of the estimated difference is from 4.2 percentage points less (Covratio) to 19.8 percentage points less ('Minus One'). Note that the value of the coefficient for 'Minus One' is significantly larger that those for the other methods, suggesting that the other methods are detecting and excluding other, less gross, outliers.

The results for mixed ownership firms are themselves rather mixed. The estimations for the HSP and winsorising show that the shares of surplus taken by workers in mixed ownership firms are close to the shares taken by those in private firms, whereas the results for Cook's distance, covratio, robust regression, and 'Minus One' indicate that workers in mixed ownership firms are appropriating much smaller shares of the surplus than in either state-owned or private firms. Indeed, the coefficient on *mixed*valad97*

in these latter results is approximately the same magnitude as, but opposite in sign from the coefficient on *valad97* – indicating, in other words, that workers in mixed firms have zero bargaining strength and are appropriating *none* of the firms' surplus. The former set of results – bargaining power in mixed firms similar to that of private firms – is more or less consistent with our prior expectations, the interpretation being that private ownership leads to lower bargaining power for workers regardless of whether it is full or partial private ownership. The latter set of results – workers in mixed firms have no bargaining power at all – is, however, harder to explain. One possibility, of course, is that the latter results are generated by outliers remaining in the sample but that are removed or attenuated by the HSP or Winsorising procedures; this is consistent with the fact noted above that the latter two procedures score highly in terms of the increase in R^2. In any case, the results for mixed firms demonstrate again the sensitivity of results to the treatment of outliers.

The next step in our investigation is to consider a wider range of estimations – covering 1996 as well as 1997, estimating using instrumental variables as well as OLS, and estimating in first-differences to allow for firm-specific fixed effects – but employing a smaller range of robust/outlier estimation methods. Based on the results above, using as our criteria the improvement in R^2 and consistency with prior expectations, we choose the HSP and Winsorising methods. We also report, for reference, the estimation results for the full sample, and for the full sample minus the gross outlier ('Minus One'). Value added per worker is instrumented with its lagged value from the previous year; all other variables are treated as exogenous. The results are reported in Tables 5–9. Regional and industry (construction) dummy variables were included in the estimations but are not reported for reasons of brevity.

OLS results for 1996 (Table 5) and 1997 (Table 6) are very similar for both the HSP and Winsorising procedures: in both years the share of the surplus taken by workers in state-owned firms is estimated to be approximately 31 to 34 per cent, with shares taken by workers in private and mixed firms about 4 to 9 percentage points less than this; as noted earlier, the estimated shares taken by workers in private and mixed firms are very similar using these estimation methods. The estimation results when simply the gross outlier is excluded are, however, rather unstable, with the results for 1996 and 1997 changing quite sharply.

Tables 7 and 8 report results for instrumental variable estimation. The results reported for 1996 are generally poor, with few significant coefficients. The reason for this is unclear and may be related to the fact that we did not check for outliers amongst the instruments; further investigation is needed here. The 1997 IV estimation does not have these problems; the results are quite good and qualitatively similar to the OLS results for that year. Compared to the previous 1997 estimations, the coefficient on *valad97* is somewhat higher at 0.39 (Winsorising) or 0.43 (HSP), indicating that, when account is taken of simultaneity, the estimated share of the surplus taken by workers in state-owned firms is somewhat higher, at about 40 per cent. The results for private and

mixed firms are similar to each other, indicating workers take a share of the surplus that is about 10 percentage points lower than that taken by workers in state firms. The IV coefficients for private and mixed firms using the HSP and Winsorising methods are therefore a bit higher than the OLS coefficients using these methods – workers in these firms are taking about 30 per cent of the surplus.

Finally, Table 9 reports the results of estimation in first differences. The results here, by contrast with those reported above, are somewhat inconclusive. The HSP and Winsorising results indicate that workers in state-owned firms are taking a smaller share of the surplus than when estimated in levels – 14 per cent (Winsorising) or 16 per cent (HSP). All but one of the ownership coefficients are insignificant, meaning we are unable to detect significant differences between the process of wage determination in state firms and private or mixed firms. Finally, even the alternative wage, highly significant in all the previous estimations, is insignificant in our first-differences results. A possible explanation for these poor results, and for the insignificant coefficient on the alternative wage in particular, is that measurement error is overwhelming the true changes in the underlying variables.

6. Conclusions

In this paper we have analysed the relationship of ownership form to wage determination in Russia using a firm-level dataset that includes several thousand industrial and construction firms from four representative regions of Russia. Our empirical results can be summarised as follows.

We find that wage determination is indeed related to ownership form in Russia. Workers in state-owned firms take a larger share of their firm's surplus than workers in private (privatised) firms. Interestingly, the share of the surplus taken by workers in firms with mixed state/private ownership is similar to that in fully private firms. These results can be interpreted as suggesting that corporate governance (wage setting) is more disciplined in private firms, and that partial privatisation is enough to generate this improved wage discipline.

We also find that our results are highly sensitive to the treatment of outliers in the data, and the conclusions above must be taken with this caveat in mind. There is no easy answer to this problem – remarkably, we found that several outlier detection and robust estimation techniques were unable to detect a single gross outlier, the inclusion/exclusion of which had a huge impact on our estimation results. Our strategy was to use a wide range of such techniques, report the results of all of them, and then investigate further using a subset of these techniques that generated results that were reasonable on both statistical grounds and on the grounds of consistency with prior expectations based on theory and other evidence. We would recommend that other researchers do the same when employing problematic data – as data from transition countries often are.

Biometrics, 31, pp. 387–410.

Greene, W. H. (1993), *Econometric Analysis*, 2nd. Edition, Prentice-Hall International Inc., p. 288.

Grubs, F. E. (1969), 'Procedures for detecting outlying observations', *Technometrics*, 11, pp. 1–21.

Hadi, A. S. and J. S. Simonoff (1993), 'Procedures for the Identification of Multiple *Journal of the American Statistical Association*, 88(424), pp. 1,265–72.

Layard, R. and S. Nickell (1990), 'Is Unemployment Lower if Unions Bargain Over *Quarterly Journal of Economics*, August, 105(3), pp. 773–87.

Lehmann, H., J. Wadsworth and A. Acquisti (1998), 'Grime and Punishment: Job Insecurity and Wage Arrears in the Russian Federation', William Davidson Institute Working Paper Series, Ann Arbor, Michigan.

Journal of Econometrics, 43, pp. 293–315.

Rousseeuw, P. J. and A. M. Leeroy (1987), '*Robust Regression and Outlier Detection*', New York: John Wiley & Sons, Inc.

Standing, G. (1996), '*Russian Unemployment and Enterprise Restructuring – Reviving Dead Souls',* London: Macmillian Press, Chapter 5: The Paradox of Wage Flexibility.

Svejnar, J. (1986), 'Bargaining Power, Fear of Disagreement, and Wage Settlements: Theory and Evidence from U.S. Industry', *Econometrica*, 54(5), pp. 1,055–78.

Thomas, R. L. (1997), *Modern Econometrics – An Introduction*, Harlow: Addison Wesley Longman.

Welsch, R. E. and E. Kuh (1977), *Technical Report 923-77: Linear Regression Diagnostics,* Cambridge, MA: Sloan School of Management, Massachusetts Institute of Technology.

Tables

Table 1: Summary Statistics by Ownership Category

1996	obs.	Total Sample Mean (St. Dev)	obs.	State Firms Mean (St. Dev)	obs.	Privatised Firms Mean (St. Dev)	obs.	Mixed Firms Mean (St. Dev)
Employment	2331	612.81 (2497.06)	705	324.75 (1018.85)	605	508.01 (1099.96)	1021	873.83 (3560.68)
Average Wage	2331	10.52 (8.12)	705	9.19 (6.06)	605	11.82 (10.97)	1021	10.66 (7.19)
Alternative wage	2331	11.84 (4.68)	705	11.97 (4.56)	605	12.01 (4.42)	1021	11.65 (4.90)
Output	2201	59866.98 (424332.6)	664	18821.24 (81146.91)	564	44635.67 (203249 1)	973	96706.47 (613576.3)
Output per Employee	2201	61.14 (105.69)	664	40.31 (49.47)	564	69.09 (82.86)	973	70.76 (138.60)
Profit	2331	8766.08 (130374.4)	705	4317.56 (39129.45)	605	4252.32 (12240.87)	1021	14512.44 (193967.2)
Profit per Employee	2331	64.04 (1928.052)	705	23.99 (487.17)	605	8.57 (16.35)	1021	124.57 (2884.63)

1997	obs.	Total Sample Mean (St. Dev)	obs.	State Firms Mean (St. Dev)	obs.	Privatised Firms Mean (St. Dev)	obs.	Mixed Firms Mean (St. Dev)
Employment	2355	560.10 (2319.89)	707	299.41 (944.07)	615	463.50 (1024.93)	1033	796.03 (3306.60)
Average Wage	2355	12.50 (8.56)	707	11.06 (7.62)	615	13.39 (8.64)	1033	12.96 (8.99)
Alternative wage	2355	14.47 (5.51)	707	14.67 (5.39)	615	14.61 (5.22)	1033	14.24 (5.74)
Output	2226	61045.46 (428830.1)	661	20822.06 (98035.58)	577	48959.52 (227310.1)	988	95014.34 (612904.2)
Output per Employee	2226	68.16 (109.04)	661	45.48 957.32)	577	79.71 (119.91)	988	76.59 (125.38)
Profit	2355	11957 (232431.7)	707	4549.91 (48360.52)	615	5066.77 (27904.09)	1033	21128.63 (347875)
Profit per Employee	2355	42.29 (864.89)	707	27.56 (624.83)	615	9.31 (41.12)	1033	72.01 (1198.54)

Table 2: Summary Statistics by Industry Category

	obs.	Industry Mean 1996 (St. Dev)	obs.	Industry Mean 1997 (St. Dev)	obs.	Construction Mean 1996 (St. Dev)	obs.	Construction Mean 1997 (St. Dev)
Employment	1768	712.26 (2817.66)	1789	651.49 (2618.76)	562	300.52 (872.73)	565	271.12 (781.90)
Average Wage	1768	9.74 (6.36)	1789	11.71 (7.75)	562	12.97 (11.76)	565	15.02 (10.35)
Alternative wage	1768	10.82 (3.97)	1789	13.29 (4.91)	562	15.06 (5.26)	565	18.20 (5.63)
Output	1730	71120.83 (476957.8)	1751	72516.11 (482223.9)	498	17515.4 (59928.69)	500	17807.37 (48028.62)
Output per Employee	1730	62.39 (114.18)	1751	67.36 (101.29)	498	53.45 (65.06)	500	67.53 (131.41)
Profit	1768	7000.49 (93495.98)	1789	12223.78 (255011.7)	562	14142.3 (207377.9)	565	10883.02 (138896.1)
Profit per Employee	1768	6.54 (18.47)	1789	5.98 (29.59)	562	244.39 (3923.67)	565	156.59 (1761.20)

Table 3: Regression Diagnostics and Robust Estimation for 1997

	Full Sample	Hadi & Simonoff	Dffits	Cook's Distance	Welsch's Distance	Covratio	Non Linear Least Squares	Median Least Squares	RLS	Robust Regression	Winsor	Minus One
valad 97	0.001	0.339	0.001	0.277	0.001	0.248	0.001	0.001	0.001	0.346	0.340	0.270
	(0.000)*	(0.012)**	(0.000)**	(0.013)**	(0.000)*	(0.013)**	(0.000)*	(0.000)**	(0.000)*	(0.010)**	(0.012)**	(0.015)**
private*valad97	0.071	–0.045	0.047	–0.095	0.055	–0.042	0.070	0.202	0.215	–0.111	–0.073	–0.198
	(0.007)**	(0.016)**	(0.006)**	(0.017)**	(0.007)**	(0.016)**	(0.007)**	(0.005)**	(0.010)**	(0.013)**	(0.015)**	(0.016)**
mixed*valad97	–0.001	–0.061	–0.001	–0.263	–0.001	–0.235	–0.001	0.000	–0.001	–0.346	–0.087	–0.270
	(0.000)	(0.014)**	(0.000)**	(0.013)**	(0.000)**	(0.013)**	(0.000)	(0.000)**	(0.000)	(0.010)**	(0.014)**	(0.015)**
Private	0.908	0.167	0.770	0.303	0.783	0.274	1.049	0.498	0.374	0.196	0.143	0.857
	(0.242)**	(0.146)	(0.203)**	(0.177)	(0.216)**	(0.169)	(0.233)**	(0.179)**	(0.217)	(0.161)	(0.142)	(0.227)**
Mixed	1.249	0.381	1.064	0.947	1.021	0.819	1.275	1.045	1.129	0.596	0.420	1.233
	(0.205)**	(0.124)**	(0.171)**	(0.149)**	(0.182)**	(0.143)**	(0.205)**	(0.152)**	(0.182)**	(0.135)**	(0.120)**	(0.192)**
alt wage97	0.680	0.576	0.562	0.614	0.585	0.584	0.720	0.682	0.623	0.573	0.538	0.665
	(0.039)**	(0.027)**	(0.033)**	(0.032)**	(0.035)**	(0.029)**	(0.027)**	(0.028)**	(0.034)**	(0.026)**	(0.027)**	(0.036)**
Moscow	–0.308	–0.106	–0.359	–0.378	–0.260	–0.446	–0.308	–0.472	–0.300	–0.254	–0.182	–0.200
	(0.254)	(0.156)	(0.213)	(0.186)*	(0.227)	(0.178)*	(0.254)	(0.186)*	(0.216)	(0.168)	(0.152)	(0.238)
Chuvashia	–1.026	–1.036	–1.391	–1.198	–1.319	–1.191	–1.026	–0.998	–1.150	–1.060	–1.108	–1.021
	(0.293)**	(0.176)**	(0.244)**	(0.211)**	(0.260)**	(0.204)**	(0.293)**	(0.214)**	(0.248)**	(0.194)**	(0.171)**	(0.274)**
Krasnoyarsk	0.118	1.056	–0.290	0.302	–0.149	0.286	0.118	–0.081	–0.009	0.548	0.938	0.620
	(0.314)	(0.189)**	(0.262)	(0.227)	(0.279)	(0.219)	(0.314)	(0.230)	(0.267)	(0.209)**	(0.181)**	(0.295)*
Construction	–0.119	0.169	–0.326	–0.226	–0.163	–0.136	–0.119	0.049	–0.172	0.120	0.086	0.201
	(0.235)	(0.146)	(0.198)	(0.175)	(0.210)	(0.167)	(0.235)	(0.172)	(0.200)	(0.156)	(0.141)	(0.220)
Constant	0.999	1.652	1.971	1.575	1.834	1.800	0.999	0.679	1.610	1.686	1.968	0.950
	(0.364)**	(0.240)**	(0.308)**	(0.283)**	(0.328)**	(0.265)**	(0.364)**	(0.266)*	(0.310)**	(0.241)**	(0.234)**	(0.340)**
N	2354	2256	2294	2282	2326	2254	2354	2354	2347	2352	2354	2353
R^2	0.29	0.66	0.27	0.47	0.27	0.48	0.29	0.21[1]	0.38	n.a.	0.66	0.37
Outliers	0	98	60	72	28	100	0	0	7	2	0	1

Notes: ** significant at the 1 per cent level; * significant at the 5 per cent level; [1]Refers to a pseudo R^2.

Table 4: Gross Outlier and Rest of Sample Compared

	Gross Outlier	Full Sample Less Gross Outlier, i.e., 2,353 observations			
		Mean	Standard Deviation	Min	Max
Wage	17.341	7.206	4.933	0. 029	58.393
Alternative wage	14.380	8.343	3.176	2.335	23.916
Profit	507694.7	6651.679	133706	–233775.2	6072081
Profit per worker	9579.145	20.238	458.465	–203.189	17585.85
Value added per worker	9582.105	19.102	458.570	–215.458	17591.6
Employment	53	560.413	2320.851	1	73930

Table 5: Estimation using OLS 1996

1996	Full Sample	Hadi & Simonoff	Winsorising	Minus One
Valad 96	0.001	0.312	0.320	0.115
	(0.0004)*	(0.011)**	(0.011)**	(0.009)**
*Private*valad96*	0.336	–0.069	–0.076	0.222
	(0.012)**	(0.015)**	(0.015)**	(0.015)**
*Mixed*valad96*	–0.001	–0.055	–0.076	–0.115
	(0.0004)	(0.014)**	(0.013)**	(0.009)**
Private	–0.159	0.321	0.279	–0.058
	(0.246)	(0.144)*	(0.137)*	(0.239)
Mixed	1.096	0.498	0.395	1.207
	(0.203)**	(0.121)**	(0.114)**	(0.197)**
alt wage96	0.602	0.482	0.443	0.595
	(0.039)**	(0.027)**	(0.024)**	(0.038)**
Constant	1.571	2.053	2.435	1.462
	(0.347)**	(0.224)**	(0.203)**	(0.337)**
N	2330	2235	2330	2329
R^2	0.41	0.63	0.65	0.45

Notes: **significant at 1 per cent level, * 5 per cent level.

Table 6: Estimation using OLS 1997

1997	Full Sample	Hadi & Simonoff	Winsorising	Minus One
valad 97	0.001	0.339	0.340	0.270
	(0.0004)*	(0.012)**	(0.012)**	(0.015)**
private*valad97	0.071	–0.045	–0.073	–0.198
	(0.007)**	(0.016)**	(0.015)**	(0.016)**
mixed*valad97	–0.001	–0.061	–0.087	–0.270
	(0.0004)	(0.014)**	(0.014)**	(0.015)**
Private	0.908	0.167	0.143	0.857
	(0.242)**	(0.146)	(0.142)	(0.227)**
Mixed	1.249	0.381	0.420	1.233
	(0.205)**	(0.124)**	(0.120)**	(0.192)**
alt wage97	0.680	0.576	0.538	0.665
	(0.039)**	(0.027)**	(0.027)**	(0.036)**
constant	0.999	1.652	1.975	0.950
	(0.364)**	(0.240)**	(0.234)**	(0.340)**
N	2354	2255	2354	2353
R^2	0.29	0.66	0.66	0.37

Table 7: Estimation using instrumental variables, 1996

1996	Full Sample	Hadi & Simonoff	Winsorising	Minus One
valad 96	0.008	0.313	0.151	0.120
	(0.005)	(2.013)	(0.072)*	(0.083)
private*valad96	0.314	0.123	0.148	0.183
	(0.168)	(1.496)	(0.141)	(0.183)
mixed*valad96	–0.007	3.481	0.130	–0.119
	(0.005)	(34.792)	(0.146)	(0.083)
private	0.095	0.139	0.007	0.183
	(0.969)	(4.120)	(0.485)	(0.911)
mixed	1.275	–7.757	0.174	1.310
	(0.244)**	(84.714)	(0.388)	(0.236)**
alt wage96	0.607	–0.775	0.458	0.624
	(0.054)**	(13.489)	(0.059)**	(0.048)**
constant	1.232	7.918	2.359	1.019
	(0.452)**	(64.018)	(0.409)**	(0.427)*
N	2022	1938	2022	2021
R^2	0.36	.	0.62	0.45

Note: A full stop in the R^2 box indicates that the calculated R^2 was negative and hence is not reported.

Table 8: Estimation using instrumental variables, 1997

1997	Full Sample	Hadi & Simonoff	Winsorising	Minus One
valad 97	0.001	0.429	0.388	0.349
	(0.0004)*	(0.025)**	(0.014)**	(0.031)**
private*valad97	0.226	–0.096	–0.100	–0.122
	(0.016)**	(0.029)**	(0.019)**	(0.034)**
mixed*valad97	0.000	–0.120	–0.097	–0.349
	(0.001)	(0.027)**	(0.017)**	(0.031)**
private	0.115	0.012	0.059	0.051
	(0.278)	(0.152)	(0.145)	(0.265)
mixed	1.216	0.267	0.313	1.198
	(0.227)**	(0.129)*	(0.122)*	(0.216)**
alt wage97	0.613	0.555	0.510	0.594
	(0.043)**	(0.028)**	(0.027)**	(0.041)**
constant	1.572	1.770	2.132	1.505
	(0.407)**	(0.247)**	(0.236)**	(0.387)**
N	2322	2226	2322	2321
R^2	0.13	0.65	0.66	0.21

Table 9: Estimation in First Differences, 1996–97

	Full Sample	Hadi & Simonoff	Winsorising	Minus One
diff valad 96	0.002	0.157	0.140	0.021
	(0.003)	(0.013)**	(0.013)**	(0.009)*
private*diff valad96	0.068	–0.013	0.015	0.048
	(0.007)**	(0.017)	(0.017)	(0.011)**
mixed*diff valad96	–0.002	–0.031	–0.033	–0.021
	(0.003)	(0.016)	(0.015)*	(0.009)*
private	–0.254	–0.078	–0.001	–0.282
	(0.193)	(0.097)	(0.085)	(0.193)
mixed	0.107	–0.027	0.050	0.081
	(0.164)	(0.083)	(0.072)	(0.165)
diff alt wage96	0.034	0.026	0.112	0.021
	(0.153)	(0.078)	(0.074)	(0.153)
constant	0.257	0.322	0.271	0.282
	(0.221)	(0.111)**	(0.097)**	(0.221)
N	2323	2218	2323	2322
R^2	0.07	0.20	0.20	0.07

Appendix 1

Allowing the Alternative Wage to Vary with the Firm Wage

Below we rework Svejnar's basic model for his estimating equation but allow Y_a, the alternative wage, to be a function of Y, the firm wage.

$$U_L^g U_M^{1-g} = \left[\frac{L(Y_i - Y_a)}{\overline{L}}\right]^g [R - Y_i L - H]^{1-g} = (\bullet)$$

In a two party framework the bargaining process is the maximisation of the above equation. This leads to the following steps:

$$\P(\bullet)/\P Y_i = \begin{array}{l} g\left[\dfrac{L(Y_i - Y_a)}{\overline{L}}\right]^{g-1}[R - Y_i L - H]^{1-g}\left[\dfrac{L}{\overline{L}} - A\dfrac{L}{\overline{L}}\right] + \\[2em] \left[\dfrac{L(Y_i - Y_a)}{\overline{L}}\right]^g (1-g)[R - Y_i L - H]^{-g}(-L) = 0 \end{array}$$

where $A = \dfrac{\partial Y_a}{\partial Y_i}$

In Svejnar (1986) this conjectural variation has a value of zero, but here we allow it to vary. One would expect *a priori* that it would take a value between zero and one. If groups of workers are primarily concerned with keeping their relative positions in some kind of pay league then a value of one would imply unchanged positions in any ranking and a value of less than one that 'distance' between groups of workers is widening.

$$\left(\frac{g}{1-g}\right)\left[\frac{\overline{L}}{L(Y_i - Y_a)}\right]\left[\frac{(R - Y_i L - H)}{L}\right] = \frac{\overline{L}}{L - AL}$$

$$[g(R - H) - gY_i L](L - AL) = (1-g)L^2 Y_i - (1-g)L^2 Y_a$$

$$g(R - H)(L - AL) - (gY_i L)(L - AL) = Y_i L^2 - gL^2 Y_i - (1-g)L^2 Y_a$$

$$Y_i[L^2 + gL(L - AL) - gL^2] = g(R - H)(L - AL) + (1-g)L^2 Y_a$$

$$Y_i = \frac{g(R - H)(L - AL)}{L^2(1 - gA)} + \frac{(1-g)L^2 Y_a}{L^2(1 - gA)}$$

$$Y_i = \frac{g(R-H)}{L(1-gA)} - \frac{g(R-H)A}{L(1-gA)} + \frac{(1-g)Y_a}{(1-gA)}$$

$$Y_i = \frac{g(R-H)}{L(1-gA)} - \frac{g(R-H)A}{L(1-gA)} + \frac{Y_a}{1-gA} - \frac{gY_a}{1-gA}$$

$$Y_i = \frac{g(R-H)-gY_aL}{L(1-gA)} - \frac{g(R-H-Y_aL+Y_aL)A}{L(1-gA)} + \frac{Y_a}{1-gA}$$

$$Y_i = \frac{g(R-H-Y_aL)}{L(1-gA)} - \frac{g(R-H-Y_aL)A}{L(1-gA)} + \frac{Y_a}{1-gA} - \frac{gY_aA}{1-gA}$$

$$Y_i = \frac{g(R-H-Y_aL)}{L} * \left[\frac{1-A}{1-gA}\right] + Y_a$$

This should be compared with the original Svejnar model below.

$$Y_i = \frac{g(R-H-Y_aL)}{L} + Y_a$$

The effect of assuming that the wage set by firm i has an impact on wages paid elsewhere in the industry/region is to decrease the share of the surplus taken by the insiders of firm i.

Appendix 2

Residual Classification and Diagnostic/Robust Techniques

Standardisation of the Residuals

In the brief descriptions that follow there are several measures of residuals used. As Greene (1993) points out, if one wishes to identify which residuals are significantly large they should be standardised by dividing by the appropriate standard error for that particular residual.

Studentisation refers to the division of a scale-dependent statistic by a scale estimate that results in a statistic which is free of the nuisance scale parameters.[13] A further distinction is between *internal* and *external* studentisation. In the former, the scale statistic and the scale estimate are derived from the same data; in the latter they are independent but the necessary information for the construction of the residual comes from the fitting of the model using all the data to hand. For predicted residuals, calculations are based on a fit that has taken place without including the ith observation. The same division between internal and external can then be made.

The Standardised Residual

One residual used in outlier identification is the standardised residual. (This is the term the Stata Corporation use.) Hadi and Simonoff (1993) refer to this as the 'internally studentised residual' and Cook and Weisberg (1982), while acknowledging it as an internally studentised residual when referring to (1.1), simply adopt the term 'studentised residual'.) It is defined as:

$$\hat{e}_i = \frac{e_i}{\left(s\sqrt{1-h_{ii}}\right)}$$
(A2.1)

h_{ii} represents the ith diagonal element of the hat matrix **H**; s is the variance from the sample residuals and e_i is the ordinary residual.

Studentised Residual

A second residual used is the studentised residual. (This is what Hadi and Simonoff refer to as the externally studentised residual.) It is defined as:

$$r_i = \frac{e_i}{\left(s(i)\sqrt{1-h_{ii}}\right)}$$
(A2.2)

The *(i)* here refers to the fact that the standard error is calculated without using the ith observation. Hence the residual is independent from the standard error. $s_{(i)}$ is obtained from:

---

[13] Cook and Weisberg (*op. cit.*, p. 18).

$$s_{(i)}^2 = \frac{(n-k)s^2 - e_i^2/(1-h_{ii})}{n-k-1}$$

$$= s^2\left(\frac{n-k-\hat{e}_i^2}{n-k-1}\right)$$

(A2.3)

where n is the number of observations and k the number of explanatory variables.

Predicted Residuals

The ith predicted residual is found by estimating the given model without the ith observation and then using the least squares estimator of $\boldsymbol{b}$ to find the residual, i.e.,

$$e_{(i)} = y_{(i)} - \mathbf{x}_i^T \hat{\boldsymbol{b}}_{(i)} \qquad i = 1, 2, \ldots, n$$

(A2.4)

The predicted error can then be divided by the standard error of the prediction to obtain:

$$\hat{\boldsymbol{y}}_{(i)} = \frac{e_{(i)}}{s_{(i)}\left[1 + x_i^T\left(X_{(i)}^T X_{(i)}\right)^{-1} x_i\right]^{1/2}}$$

(A2.5)

where x_i is a row vector of explanatory variables on the ith observation and $X_{(i)}$ is obtained from the matrix X by deleting the ith row x_i^T. Moreover, the above scaled predicted residual is equal to the externally studentised residual, i.e., $r_{(i)}$. As mentioned in the main body of the text the coefficient on a dummy variable that has a one for the ith observation and zeros elsewhere will equal the predicted residual and the t statistic is equal to the externally studentised residual.

We can link the internally and externally studentised residuals with the predicted residuals. From the mean shift outlier model we know that:

$$e_{(i)} = \frac{e_i}{1 - h_{ii}}$$

(A2.6)

where $e_{(i)}$ is the predicted residual, e_i the ordinary residual. The externally studentised residual is given by:

$$r_i = \frac{e_i}{\left(s_{(i)}\sqrt{1 - h_{ii}}\right)}$$

(A2.7)

(A2.6) can be rearranged to give:

$$e_i = r_i \times s_{(i)}\sqrt{1 - h_{ii}}$$

(A2.8)

Substituting and rearranging gives:

30

$$r_{(i)} = \frac{e_{(i)}}{s_{(i)} / \sqrt{1 - h_{ii}}} \qquad (A2.9)$$

If we require the standard error with the ith observation then we have

$$\hat{e}_{(i)} = \frac{e_{(i)}}{s / \sqrt{1 - h_{ii}}} \qquad (A2.10)$$

where $\hat{e}_{(i)}$ is the internally studentised residual or the standardised residual.

Formulae for Outlier Detection

The basic formulae used in each of the outlier detection methods are given below along with possible connections between the various methods. As will be seen, they make use mainly of the studentised and the standardised residuals.

1) DFFITS (Welsch and Kuh, 1977 and Belsley, Kuh and Welsch, 1980)
This measures the difference between the fitted value $\hat{y}$ of y that results from dropping a particular observation. If $\hat{y}_{(i)}$ is the fitted value of y once the ith observation is deleted then the quantity $(\hat{y} - \hat{y}_{(i)})$ divided by the scaling factor $h_{ii} s_{(i)}$, where $s_{(i)}$ is the estimator of s^2 from a regression with the ith observation omitted is called DFFITS. It has been shown by Belsley *et al.*(1980) that:

$$DFFITS_i = r_i \sqrt{h_{ii} / (1 - h_{ii})} \qquad (A2.11)$$

where r_i are the studentised residuals. It should be noted that Belsley, Kuh and Welsch do not actually suggest dropping outliers even if upon examination they prove to be anomalous. Rather they suggest a bounded influence estimation where (to simplify) Welsch (1980) suggests that the regression is run; DFFITS is calculated and on the basis of the absolute value of DFFITS weights are attached by the following rule:
Minimize $\sum w_i (y_i - \boldsymbol{b} x_i)^2$, where $w_i = 1$ if $|DFFITS| \leq 0.34$ or $w_i = \frac{0.34}{|DFFITS|}$ if $|DFFITS| > 0.34$. If we had pursued this approached then this would have placed this in the robust category since it is clearly an attempt to accommodate outliers and influential observations. In our study we merely adopt the suggested cutoff value $2\sqrt{(k/n)}$.

2) Cook's Distance (Cook, 1977)
Cook's Distance is given by:

$$D_i = \frac{h_i e_i^2}{k s^2 (1 - h_i)^2} = \frac{1}{k} \frac{s_{(i)}^2}{s^2} DFFITS^2_i \qquad (A2.12)$$

where k is the number of variables including the constant in the regression, s is the root mean square error of the regression and $s_{(i)}$ is the root mean square error of the regression with the ith observation omitted.

3) Welsch's Distance
Welsch's Distance is given by:

$$W_i = \frac{r_i \sqrt{h_{ii}(n-1)}}{(1-h_{ii})} = DFFITS_i \sqrt{\frac{n-1}{1-h_{ii}}} \qquad (A2.13)$$

Values of Cook's distance greater than $4/n$ should be examined (Bollen and Jackman, 1990). Following similar logic, the cut-off for Welsch's Distance is approximately $3\sqrt{k}$ (Chatterjee and Hadi, 1988).

4) Covratio

$$covratio = \left(\frac{n-k-\hat{e}_i}{n-k-1}\right)^k \frac{1}{1-h_{ii}} \qquad (A2.14)$$

This measure is the ratio of the determinants of the covariance matrix, with and without the ith observation. Belsley, Kuh and Welsch (1980) suggest that observations for which

$$|covratio_i - 1| \geq \frac{3k}{n} \qquad (A2.15)$$

should be investigated further. In the calculation of outliers using the above procedures the recommended cut-off points were used.

5) Hadi Process (Hadi and Simonoff, 1993)
The Hadi and Simonoff procedure (HSP) firstly involves identifying an initial clean subset M, initially of size h = integer part of $(n + k - 1)/2$. Hadi and Simonoff (1993) propose two methods of identifying the initial subset. With the first method, M1, the dataset is divided into a basic subset that contains the first $k + 1$ observations and a non-basic subset that contains the remaining $(n - k - 1)$ observations. This is done after fitting a regression model to the whole dataset and then ordering the n observations according to some regression diagnostic, say, the absolute value of the adjusted residual, which is defined as $\hat{e}_i = e_i / \sqrt{1 - h_{ii}}$. Run a regression on the subset B, i.e., the basic subset. Compute $\hat{e}_i = e_i / \sqrt{1 - h_{ii}}$ if $i \in B$ and $\hat{e}_i = e_i / \sqrt{1 + h_{ii}}$ if $i \notin B$ and arrange the observations in ascending order. The size of the basic subset, s, is increased by one observation to $s + 1$. Repeat the above. When the size of the basic subset equals h substitute the first h observations for M and go to step 2 of their main algorithm (see below). M2, the second method is left for the cited reference. M1 is described in detail here since Hadi and Simonoff found that it was more powerful in the presence of high

leverage outliers compared to M2 but there was little difference in the presence of low leverage outliers.

In the second step of their algorithm the following is calculated:

$$d_i = \frac{y_i - x_i' \hat{b}_M}{\hat{s}_M \sqrt{1 - x_i' (X_M' X_M)^{-1} x_i}} \quad , \text{if } i \in B \text{ and} \qquad (A2.16)$$

$$d_i = \frac{y_i - x_i' \hat{b}_M}{\hat{s}_M \sqrt{1 + x_i' (X_M' X_M)^{-1} x_i}} \quad \text{if } i \notin B \qquad (A2.17)$$

Following this the observations are arranged in ascending order according to $|d_i|$. Under the normality assumption and if $\hat{b}_M$ and $\hat{s}_M$ are independent then the d_i's follow a t-distribution with $s - k$ degrees of freedom for each subset M of size s and for $i \notin B$.[14]

a) If $d_{(s+1)} \geq t_{a/2(s+1, s-k)}$, then declare all observations satisfying $|d_i| \geq t_{a/2(s+1, s-k)}$ as outliers and stop.

b) Otherwise, form a new subset M by taking the first $s + 1$ ordered observations and go to step two of the algorithm. If $n = s + 1$, then declare no outliers in the data and stop.

Using the above process Hadi and Simonoff demonstrate that their process is effective in detecting multiple outliers and maintain that their method is less affected by masking and swamping problems compared to other methods

Robust Estimation

There are various robust estimators that may be employed. Our selection is based as much on convenience as anything else in terms of those procedures that are already 'canned' in econometric software packages (with the exception of the Rousseeuw and Leroy procedure). A brief, general introduction to robust regression with a good list of books and articles at the end of the relevant chapter can be found in Draper and Smith (1998).

1) Median Least Squares (MLS)
Whereas in ordinary regression the object is to estimate the mean of the dependent variable in MLS the object is to estimate the median of the dependent variable. The median regression finds a line that minimises the sum of the absolute residuals rather

---

[14] As Hadi and Simonoff point out $\hat{b}_M$ and $\hat{s}_M$ are dependent and since both are determined from a preliminary examination of the data the estimate of $\hat{s}_M$ is biased downwards. That said the authors feel that the t-distribution provides a useful benchmark for establishing cut-off values.

than the sum of the squares of the residuals as in OLS, i.e., minimise, with respect to the elements of $\mathbf{b}$.

$$Median\left(Y_i - \mathbf{x}_i^T \mathbf{b}\right)$$

It is usual to see a pseudo R^2 reported with MLS. This is calculated as

1 – *(the sum of the absolute deviations about the predicted Y)/ (the raw sum of absolute deviations about the unconditional median)*
A more rigorous procedure advocated by Rousseeuw and Leroy (1987) is explained below. While MLS results are reported, results based on this procedure are also reported below.

2) Reweighted Least Squares (Rousseeuw and Leroy Procedure)
The following procedure does not appear explicitly in full within the book by Rousseeuw and Leroy (1987) but is taken from various parts of the book.

Step 1. Run MLS.[15]
Step 2. Calculate a preliminary scale estimate S^0

$$s^0 = 1.4826\left(1 + \frac{5}{n-p}\right)\sqrt{\underset{i}{med}\ r_i^2}$$

where r_i is the residual, n the sample size and p the number of explanatory variables.

Step 3. Calculate the standardised residuals r_i / s^0.

Step 4. Determine the weights w_i for the ith observation based on
$w_i = \left\{1\ if\ \left|r_i / s^0\right| \leq 2.5\ or\ 0\ otherwise\right.$.

Step 5. Calculate the scale estimate for LMS regression using $\mathbf{s}^* = \sqrt{\dfrac{\sum\limits_{i=1}^{n} w_i r_i^2}{\sum\limits_{i=1}^{n} w_i - p}}$.

Step 6. Standardise the residuals using σ^*. If $\left|r_i / \mathbf{s}^*\right| > 2.5$ investigate the observation as a possible outlier.

---

[15] Strictly speaking the dataset should be standardised before the above procedure is carried out. This makes the variables dimensionless and so helps avoid numerical inaccuracies caused by different units of measurement. Rousseeuw and Leroy recommend that the median of the jth variable is subtracted from the ith observation on the jth variable. This in turn is divided by the median of the absolute deviations from the median multiplied by the correction factor $1/f^{-1}(0.75) \approx 1.4826$. After MLS the results, of course, have to be transformed back. The Stata command qreg with its default setting is MLS but qreg does not transform the variables to be dimensionless (personal e-mail from Stata's technical support 22/4/99). Whether by transforming the data as suggested by Rousseeuw and Leeroy would have made a great deal of difference to the final results reported is open to question.

Step 7. Carry out reweighted least squares (RLS). In effect OLS except on the reduced dataset formed on the basis of deleting observations deemed outliers in step 6 above.

3) Winsorising

This is similar to the so-called α-trimmed mean, which is simply the mean of the sample after the proportion α of the largest and smallest observations have been deleted from the data. However, the *Winsorised mean*, rather than discarding the proportion α, 'accumulates' them at either end of the truncation point. In the regressions run using the Winsorising technique the largest and smallest 5 per cent of observations were 'Winsorised'.

4) Non Linear Least Squares

If transforming variables by taking natural logarithms can lessen the impact of outliers it might be worth exploring whether estimating the model by non-linear least squares assuming a *k* shifted log normal errors might lessen the impact of influential observations within the dataset. The specification is as follows.

$$ln(wage - k) = ln([\boldsymbol{b_0} + \boldsymbol{b_1}valad97 + \boldsymbol{b_2}alternative\ wage] - k) + \boldsymbol{n}_i$$

In running the regression k is given a value of zero, which is the case of proportion errors since the standard error of the response variable, here the log of the firm wage, is proportional to its expected value.

5) Robust Regression

In this procedure implemented by the statistical package Stata, Cook's Distance, D, is first calculated (see above) and then any observations from the regression which have a D larger than 1 are excluded. An iterative procedure then follows where weights are calculated on the basis of the residuals and then the regression is repeated. The process stops when the maximum change in weights drops below a certain level. There are two weight functions used: 1) Huber weights and 2) biweights. Both are used since Huber weights have problems dealing with severe outliers while biweights sometimes fail to converge or have multiple solutions.[16]

A. Huber weights

The ith scaled residual is calculated from $u_i = \dfrac{e_i}{s}$ where e_i is the ordinary residual and

$s = \dfrac{M}{0.6745}$ is the residual scale estimate and $M = med(|e_i - med(e_i)|)$ is the median absolute deviation from the median residual. The case weights are given by $w_i = 1\ if\ |u_i| \leq c_h$ if not then $w_i = c_h/|u_i|$ where $c_h = 1.345$. So downweighting starts when the absolute value of a residual exceeds $(1.345/\ 0.6745)M$ which is about $2M$.

---

[16] This brief summary is taken from the Stata Reference Manual, volume 3, Release 4, p. 135.

B. Biweights

All observations with non-zero residuals receive some downweighting using the following smoothly decreasing biweight function.

$$w_i = \left[1 - (u_i/c_b)^2\right]^2 \; if \; |u_i| \leq c_h \; \text{ or 0 otherwise}$$

where $c_b = 4.685$ (tuning value)/7. The tuning value can be altered with higher tuning constants producing milder downweighting and lower ones giving more drastic downweighting to the residuals. In the robust regression results reported the default setting of 7 was used.