A Note on the Calculation of Firm–specific and Skill–specific Labor Costs from Firm–level Data

ULRICH KAISER*

Centre for European Economic Research (ZEW), Mannheim
and
Center for Finance and Econometrics at the University of Konstanz

January 2000

Abstract: Virtually all empirical firm–level studies on the demand for labor do not include labor cost in the econometric specification. This is due to the fact that business and innovation survey data usually lack information on labor cost. This paper shows how reliable skill–specific and firm–specific labor cost can be calculated from firm–level data on the basis of information on total labor cost and firms' skill mix only. The simple method proposed here is applied to German innovation survey data.

Keywords: labor cost, Kernel density estimation

JEL classification: J3, C14

Address:
Ulrich Kaiser
Centre for European Economic Research (ZEW)
Department of Industrial Economics and International Management
P.O. Box 103443
D–68163 Mannheim, Germany
phone: +49.621.1235.292
fax: +49.621.1235.333
eMail: `kaiser@zew.de`

*This paper owes much to discussions with Wolfgang Franz, François Laisney, Georg Licht and Winfried Pohlmeier. I am also indebted to the ZEW's Mannheim Innovation Panel team — namely to Günther Ebling, Sandra Gottschalk, Norbert Janz and Hiltrud Niggemann — for their ongoing effort to create the data set used in this paper, and to Jürgen Moka for maintaining the ZEW's basic data bases. Paul Townsend provided an excellent proof–reading job. This research was partially funded by the German Science Foundation (DFG) within the 'Industrial Economics and Input Markets Program'.

Non–technical summary:
Empirical firm–level studies on the demand for heterogenous labor usually do not include factor prices in the econometric specification. This is a severe drawback of these studies since relative labor costs are of course major determinants of the demand for heterogeneous labor. The reason for this deficiency is simple: firm–level data usually do not contain this information.
In this paper I suggest a method to calculate skill– and firm–specific labor costs from information on total labor cost, skill mix, and observable firm characteristics only.
The method proposed here provides an useful framework for empirical labor economists. It can be applied to almost any firm–level data set.
A comparison of the labor cost calculated on the basis of the method described here with actual labor cost averages from official statistics shows that the procedure described in this paper leads to reliable skill– and firm–specific estimated labor cost.

1 Introduction

The increasing availability of firm–level data has considerably broadened the pace and scope of empirical labor and industrial economics. Since the firm is the place where hiring and firing, investment and disinvestment decisions actually arise, the firm level potentially is the most interesting level of econometric analysis. In recent years, the focus of analysis of labor economists has shifted from an aggregated perspective to the firm level. Early studies of the impact of new technologies on the demand for heterogeneous labor have for example investigated this topic on a sectoral or on a country level.[2] The discussion nowadays focuses on the firm–level. Using microdata, Adams (1999), Doms et al. (1997), Haskel and Heden (1999), Blanchflower and Burgess (1998) and others discuss the relation between new technology and labor demand on the firm–level.[3] Most of these studies estimate either cost or employment share equations for different labor inputs derived from a quasi–fixed translog specification of the cost function. Though relative costs of factor inputs are arguments in such share equations, firm–level studies do not take them into account for an obvious reason: firm–level data do usually not provide information on labor cost for different types of labor. In this paper, I present a simple method for the disaggregation of total labor cost into skill–specific labor cost. This method requires information which is usually available from innovation or business survey data: total labor cost, skill mix and a set of observable firm characteristics. An empirical illustration for German innovation survey data shows that my method leads to results which compare well to the figures found in official statistics.

2 Labor cost decomposition

Assume that the data set at hand differentiates between three different skill groups. High skilled labor (university or polytechnic graduates, as in the empirical example of section 3) is the first, medium skilled labor (workers with completed vocational training) is the second and low skilled labor (workers with no formal qualification) is the third skill group. The running index of the skill groups is denoted by i, and $i = 1, 2, 3$.
Let C_{im} denote firm m's labor cost associated with skill group i and let L_{im} denote firm m's total number of workers of quality i. The average labor costs per employee for firm m can be written as the following identity:

$$\frac{\sum_{i=1}^{3} C_{im}}{L_m} = \sum_{i=1}^{3} p_{im} \frac{L_{im}}{L_m}, \tag{1}$$

where p_{im} denotes labor costs for labor of quality i for firm m and $L_m = \sum_{i=1}^{3} L_{im}$. I assume the skill–and firm–specific labor cost p_{im} are determined by the average labor costs for each skill group, p_i and a set of observable firm characteristics which are summarized in a vector $\boldsymbol{K_m}$. Thus, p_{im} is assumed to be given by

$$p_{im} = p_i + \boldsymbol{K_m} \, \boldsymbol{\theta_i} + \epsilon_{im}, \tag{2}$$

[2] E.g. Bartel and Lichtenberg (1987), Berman et al. (1994) and Berndt et al. (1992).

[3] Chennels and van Reenen (1999) provide a detailed survey on existing studies on skill–biased technological change.

where $\boldsymbol{\theta_i}$ is a vector which relates $\boldsymbol{K_m}$ to p_{im} and ϵ_{im} is an i.i.d. error term with variance σ_i^2, mean 0 and a covariance between ϵ_{im} and ϵ_{jm} $(i \neq j)$ of 0. Substitution of (2) into (1) leads to

$$\frac{\sum_{i=1}^{3} C_{im}}{L_m} = \sum_{i=1}^{3} p_i \frac{L_{im}}{L_m} + \sum_{i=1}^{3} \boldsymbol{K_m} \, \boldsymbol{\theta_i} \frac{L_{im}}{L_m} + \nu_m, \tag{3}$$

where the error term $\nu_m = \sum_{i=1}^{3} \frac{L_{im}}{L_m} \epsilon_{im}$ is heteroscedastic of known form. Equation (3) hence is estimated by FGLS. The term $\boldsymbol{K_m} \frac{L_{im}}{L_m}$ in equation (3) represents interactions between the elements of $\boldsymbol{K_m}$ with the shares of the three skill levels. Skill– and firm–specific labor cost are obtained by inserting the estimated parameters $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$ and $\hat{p}_i$ into equation (2).[4]
As it becomes apparent from equation (3), factor prices as calculated by the method proposed here have to be instrumented in labor demand equations since L_{im} occurs both on the right and the left–hand side of the labor demand equation.
Lastly, it has to be noted that the method to decompose labor costs proposed here of course is quite similar to estimating a reduced form for factor prices and labor demand. For a linear labor demand equation, e.g., $L_{im} = \gamma_{i0} + \sum_{i=1}^{3} \gamma_{i2} p_{im} + \xi_{im}$, insertion of (2), using the identity equation (1) and rearranging terms leads to a linear estimation equation from which the entire set of parameters can be identified. The error term evolving of such an equation, however, is heteroscedastic of known form which causes problems if panel data models are applied to estimate the labor demand equations.[5] If structural labor demand equations such as the Generalized Leontief or the Translog model are considered, an additional problem occurs: the reduced form equation cannot be estimated by a linear regression.

3 Empirical illustration

In order to give some insights on the accuracy of my method, I estimate equation (3) using innovation survey data for the German business–related services sector and compare the results with labor cost calculated from official statistics. The data set used here is the Mannheim Innovation Panel in the Service Sector (MIP–S), which is collected by the ZEW on behalf of the German Federal Ministry of Science and Education. This data set was originally collected in order to analyze the innovation behaviour of the German service sector. It is described in detail in Janz and Licht (1999). The MIP–S data used here are from the second wave conducted in 1997, the data refer to 1996. The survey design extends the traditional concept of innovation surveys in manufacturing industries as summarized in the OECD Oslo-Manual (OECD (1997)) to the service sector. Information collected in the questionnaire includes (1) general information about the firm (regional affiliation: East/West Germany, number of employees, total sales, exports), (2) innovative activity (introduction of product or process innovations, objectives and impacts of innovative activity, innovation expenditures, information sources used in the innovation process, innovation cooperation, factors hampering innovation), (3) personnel, qualification and training (skill structure, changes in the skill structure (on an ordinal scale), labor cost, expenditures for training) and (4) information technologies and customer relation.

[4] I used STATA6.0's `regress` option do run the estimation.

[5] Panel FGLS estimation requires that $T >> N$ which is unlikely to be the case for firm–level data (Baltagi, 1995, ch. 5.1) It is, however, possible to estimate the linear labor demand equation by GMM.

My specification of the vector of firm characteristics $\boldsymbol{K_m}$ contains six sector dummies: management consultancy (*MANAGEM*), architectural and engineering activities (*ARCENG*), advertising (*ADVER*), sewage and refuse disposal (*SEWAGE*), transport and storing (*TRANSPORT*), and computer and related activities (*SOFTWARE*). The base category is 'other' business–related services consisting of firms from real estate activities, renting of machinery and equipment, labor recruitment and industrial cleaning. I also include a dummy variable for East German firms and firm size (the natural logarithm of the number of employees both in linear and quadratic form. Clearly, this specification does not consider some potentially important variables such as the age distribution of workers (or, as a proxy variable, firm age) or the degree of unionization in a sector (Katz and Autor, 1999). My intention simply is to illustrate my method and to show that it leads to reasonable results even when some important variables are neglected. FGLS estimation results of equation (3) are displayed in Table 1.

TABLE 1

FGLS estimation results of equation (3): dep. var. $\frac{C_m}{L_m}$

Variable	Coeff.	Std. err.
S^{high}	27.4579	28.1060
S^{low}	-11.0817	14.7420
Management cons. $\cdot S^{high}$	7.1908	23.1012
Management cons. $\cdot S^{med}$	13.8845***	4.9918
Management cons. $\cdot S^{low}$	-6.1391	8.8141
Architecture $\cdot S^{high}$	-14.8632	22.4948
Architecture $\cdot S^{med}$	21.7455***	7.1806
Architecture $\cdot S^{low}$	0.3736	13.2164
Advertising $\cdot S^{high}$	-2.8447	28.6890
Advertising $\cdot S^{med}$	12.3963*	7.4067
Advertising $\cdot S^{low}$	-1.4082	9.0763
Waste removal $\cdot S^{high}$	8.6966	28.8968
Waste removal $\cdot S^{med}$	6.1498	7.2380
Waste removal $\cdot S^{low}$	5.8495	5.4703
Software $\cdot S^{high}$	1.4257	22.9532
Software $\cdot S^{med}$	7.6112	7.0249
Software $\cdot S^{low}$	4.3256	13.4917
East Germany $\cdot S^{high}$	-29.8135***	6.5171
East Germany $\cdot S^{med}$	-26.4026***	3.9325
East Germany $\cdot S^{low}$	-14.1415***	4.5226
$log(\#of empl.) \cdot S^{high}$	8.4429	6.9018
$log(\#of empl.) \cdot S^{med}$	0.2466	3.8577
$log(\#of empl.) \cdot S^{low}$	1.2202	4.9580
$log(\#of empl.)^2 \cdot S^{high}$	-0.2419	0.8948
$log(\#of empl.)^2 \cdot S^{med}$	0.2416	0.4593
$log(\#of empl.)^2 \cdot S^{low}$	-0.0402	0.5796
Constant	66.6039***	8.4093
F–Tests for joint significancy		
Sector dummies	$F(15, 929)$	1.66**
Skill shares	$F(2, 929)$	1.07
East Germany dummies	$F(3, 929)$	44.95***
Firm size	$F(6, 929)$	3.73***
R^2	0.2602	
# of obs.	956	

The shares of high, medium and low skilled labor are abbreviated by S^{high}, S^{med} and S^{low}, respectively.

My specification of equation (3) includes a constant term so that $\hat{p}_1$ ($\hat{p}_3$) is the sum of the coefficient of the constant term and the coefficient related to the share of high skilled labor, e.g., $\hat{p}_1 = 66.6 + 27.5$ ($\hat{p}_3 = 66.6 - 11.1$). The unit of measurement of average labor cost for high skilled labor are thousand German Marks (DM) which means that $\hat{p}_1 = 94.1$ and that $\hat{p}_3 = 55.7$ thousand DM p.a. The corresponding standard errors are 25.5 and 36.1 thousand DM p.a., respectively.[6] Average labor cost for low skilled labor are 66.6 thousand DM p.a. (standard error 8.4 thousand DM).

Signs and magnitude of the sectoral dummies interacted with the shares in total employment indicate that salaries generally increase with increasing skills. Not very surprisingly, East German labor cost are significantly lower than West German labor cost. Firm size has a U–shape impact on high and low skilled labor cost, and a linear and positive impact on low–skilled labor cost. The adjusted R^2 of this model is 0.2602 which is fairly high for linear regressions on firm–level, cross–sectional data.

Table 2 displays means, medians and 10 and 90 percent quantiles of the labor cost distribution for the three different skill groups.

TABLE 2

Means, quantiles and standard deviation of estimated annual labor cost (in thousands of DM)

		Quantile			
	mean	10%	50%	90%	std. dev.
High skilled	106.5	75.9	109.7	132.8	21.3
Medium skilled	69.7	47.9	70.6	87.8	14.6
Low skilled	52.6	41.6	52.5	63.1	8.3

Figure 1 presents Kernel density estimates for the wage distribution of high, medium and low skilled labor.[7] Labor cost for low skilled labor are concentrated more narrowly and around lower values than those for medium and high skilled labor.

Mean labor cost as calculated from equation (2) for high skilled labor are 106.5 thousand DM p.a., the median is 109.7 thousand DM p.a. For medium skilled labor, the related figures are 69.7/70.6 thousand DM, for low skilled labor they are 52.6/52.5 thousand DM p.a. Interestingly, the standard deviation of labor cost decreases with decreasing skills. This reflects the fact that wages for unskilled workers and for workers with completed vocational training are negotiated between employer's associations and trade unions in Germany. The more qualified employees are, the more often wages are bargained over bilaterally and are not bound to collective wage agreements (Fitzenberger et al., 1999).

How do these estimated labor costs compare to actual labor cost as recorded in official statistics? Pfeiffer (1999) uses data from the a detailed statistic on wages of different sectors and skill groups.[8] He calculates skill–specific labor cost for German manufacturing industries from these statistic. His figures relate to 1995 and differentiate between East

[6] The standard error was calculated using the 'δ method', see Greene (1997, ch. 6.7.5) for details.

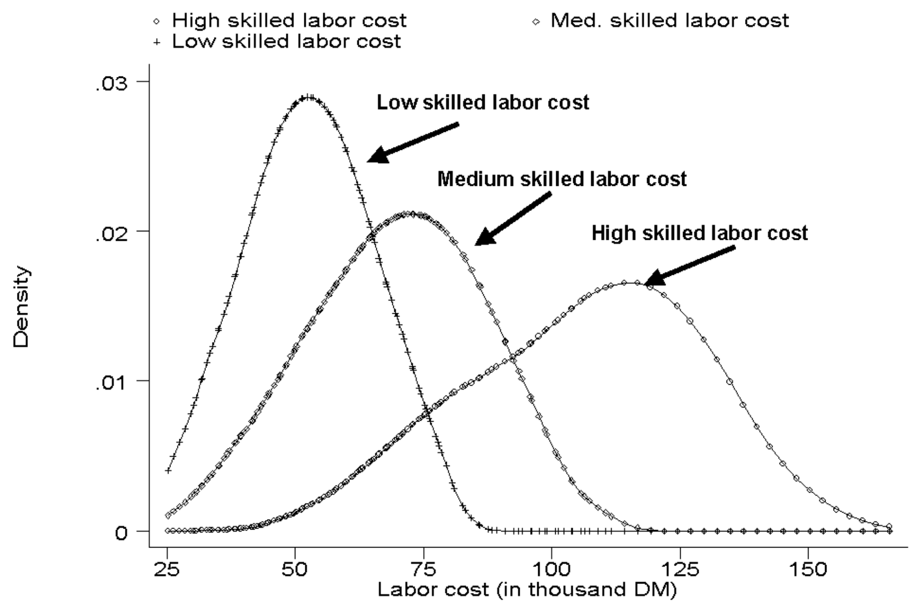
[7] I used STATA6.0's `kdensity` option do estimate the densities, with the Epanechnikov kernel function and halfwidth 10.

[8] His data sources are the 'Fachserie 16' of 1995 provided by the Federal Statistical office and the German Socio–Economic Panel, an annual household panel survey (`http://www.diw.de/soep/soepe.htm`).

and West Germany. A first difference between Pfeiffer's (1999) and my calculation is that his figures are related to manufacturing industries while mine correspond to services.[9] A second and minor distinction concerns skill definition. His 'engineers/scientist' correspond to my 'high skilled' workers group. Pfeiffer's 'technicians/foremen' and his 'skilled' workers are grouped in my 'medium skilled' category.

FIGURE 1

Kernel density estimates of high, medium and low skill labor cost per employee (in thousands of DM)



In Table 3, I compare Pfeiffer's (1999) figures with my estimated labor cost.[10] Estimated and actual labor cost are very close to one another both with respect to both means and minima and maxima. This highlights that the method proposed in this paper leads to reliable results.

4 Conclusion

In this paper I suggest a method to calculate firm–and skill–specific labor cost from information on total labor cost and firms' skill mix only. Since firm–level data usually lack information on such skill– and firm–specific labor cost, virtually all studies on the demand for heterogeneous labor do not capture labor cost although factor cost are definitely important determinants of factor demand.

Comparisons of estimated and actual labor cost show that the approach proposed here leads to reliable results.

[9] German official statistics do not provide data labor cost for the service sector.

[10] Note that I included dummy variables for East German firms in my specification which enable me to also differentiate between East and West German labor cost.

TABLE 3

Comparison of means, minima and maxima of the estimated labor cost and of Pfeiffer's (1999) figures (both in thousands of DM)

	West Germany			East Germany		
	Mean	Min.	Max.	Mean	Min.	Max.
Pfeiffer (1999)						
engineers/scientists	104.2	80.7	136.3	62.7	50.2	75.3
technicians/foremen	81.4	60.8	99.6	52.7	39.9	65.8
skilled workers	64.5	48.4	83.7	41.6	30.6	59.5
unskilled	54.7	42.1	69.9	36.3	25.9	51.4
Labor cost decomposition						
High skilled	121.3	88.2	155.8	87.5	49.4	132.0
Medium skilled	80.4	66.9	100.6	55.9	40.2	81.1
Low skilled	58.9	49.4	68.5	44.5	35.2	53.5

References

Adams, J. D. (1999), 'The Structure of Firm R&D, the Factor Intensity of Production, and Skill Bias', *The Review of Economics and Statistics* 81(3), 499–510.

Bartel, A.P., and F.R. Lichtenberg (1987) 'The Comparative Advantage of Educated Workers in Implementing New Technology', *The Review of Economics and Statistics* 69(1), 1–11.

Berman, E, J. Bound, and Z. Griliches (1994), 'Changes in the Demand for Skilled Labor within US Manufacturing Industries', *Quarterly Journal of Economics* 109(2), 367–397.

Berndt, E., C. Morrison, and L. Rosenblum (1992), 'High–Tech Capital Formation and Labor Composition in US Manufacturing Industries: An Exploratory Analysis', *NBER working paper 4010.*

Blanchflower, D.G., and S.M. Burgess (1998), 'New Technology and Jobs: Comparative Evidence from A Two Country Study', *Economics of Innovation and New Technology* 5, 109–138.

Chennels, L., and J. van Reenen (1999), 'Has Technology Hurt Less Skilled Workers? An Econometric Survey of the Effects of Technical Change on the Structure of Pay and Jobs', *The Institute for Fiscal Studies Working paper Series No. W99/27.*

Doms, M., T. Dunne, and K.R. Troske (1997), 'Workers, Wages, and Technology', *Quarterly Journal of Economics*, 112 (1), 253–290.

Fitzenberger, B., I. Haggeney and M. Ernst (1998), 'Wer ist noch Mitglied in Gewerkschaften?', *Zeitschrift für Wirtschafts– und Sozialwissenschaften* 119(2), 223–264.

Greene, W.H. (1997) *Econometric analysis*, Prentice–Hall, Upper Saddle River.

Haskel, J., and Y. Heden (1999), 'Computers and the Demand for Skilled Labour: Industry– and Establishment–Level Panel Evidence for the UK', *The Economic Journal* 109, C68–C79.

Janz, N., and G. Licht (1999) , *Innovationsaktivitäten der deutschen Wirtschaft*, Nomos–Verlag, Baden–Baden.

Katz, L.F., and D.H. Autor (1999), 'Changes in the Wage Structure and Earnings Inequality', in: O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, North–Holland.

OECD (1997), *Guidelines for Collecting and Interpreting Innovation Survey Data — the OSLO Manual*, OECD, Paris.

Pfeiffer, F. (1999), 'Human Capital and Innovation in Eastern and Western Germany', in: Fritsch, M. and H. Brezinski (eds.), *Innovation and Technological Change in Eastern Europe*, Edward Elgar, Northampton.