APPLICATION OF KERNEL ESTIMATION IN ANALYSIS OF LABOUR PRODUCTIVITY OF THE LARGEST POLISH FIRMS IN 2004-2008

Barbara Batóg, Ph.D.

Jacek Batóg, Ph.D.

*Department of Econometrics and Statistics
Faculty of Economics and Management
University of Szczecin
Mickiewicza 64, 71-101 Szczecin
e-mail: batog@wneiz.pl*

Magdalena Mojsiewicz, Ph.D.

*Department of Insurance and Capital Markets
Faculty of Economics and Management
University of Szczecin
Mickiewicza 64, 71-101 Szczecin
e-mail: mojsiv@wneiz.pl*

Received 15 September 2010, Accepted 21 December 2010

Abstract

The parametric methods of statistical and econometric analysis are not always useful in examination of labour productivity of economic entities. In previous works the Authors found that the labour productivity is characterized by the lack of stable regularities in the range of structure and interdependency. In that case it is possible to apply non-parametric methods. In the paper the Authors tried to model the distributions of the labour productivity in time by means of kernel estimation using classical approaches (Epanechnikov, Rosenblatt) and the new proposition called kernel B. It seems that proposed approach is a useful merger of the statistical modeling theory and economic practice which allows to analyze the changes in the labour productivity – the essential factor for long-term economic growth and the welfare of society. The empirical results show that the labour productivity in the largest Polish companies had increased in 2004-2008 but the growths had not the same dynamics in different economic sectors.

Keywords: labour productivity, distributions, kernel estimation.

JEL classification: D24, C14, C16.

Introduction

As the theory of economics and the results of numerous empirical research show¹ labour productivity plays a significant role in the process of generating gross domestic product. Furthermore, the level of labour productivity is a factor which determines the differences in the growth of individual countries (regions)². Many authors claim that the average labour productivity level in a given economy largely depends on its changes in sectors³ and it is the largest companies that play the crucial role. Although they are not numerous, which is in accord with theoretical and practical principles of economy⁴, their revenue and the number of employees are of great importance in the economy of every country. In order to identify the regularities of the above mentioned phenomenon the study was carried out of labour productivity in the largest Polish companies from the 'List 2000' that has been published annually by the Rzeczpospolita daily for the last few years. It contains 2000 Polish companies who reached the highest sales revenue over a given year. In this case the labour productivity is defined as the ratio of sales revenue and employment.

In the theory of economics we use various methods of labour productivity analysis. It often happens that different methods lead to similar results. Such conclusion was drawn by J. Van Biesebroeck who analysed data characterising a panel of manufacturing plants in Columbia. All the five of methods he had applied (labour productivity indices, data envelopment analysis, instrumental variables, stochastic frontiers, semiparametric estimation) generated results that were practically identical, which confirms the fact that first of all labour productivity is strongly correlated with export status and investment outlays⁵.

The analysis of labour productivity in the largest Polish companies was carried out with the use of distribution models estimated by means of kernel estimation. This particular method helps to identify the distribution shape of an examined variable without assuming its type *a priori*, while kernel estimation applied in consequent periods allows to observe the changes undergoing in a given process. For example, R.Nicolini, when examining with the use of kernel estimation the 1977 and 2002 labour productivity of Spanish companies by regions, observed clear double-peak (bimodal) distribution⁶. Similar results of this method were obtained by D. Fiaschi and A.M. Lavezzi, who analysed labour productivity in individual regions of Europe, and D. Margaritis, R. Färe and S. Grosskopf who identified trends in labour productivity in the OECD countries in the period of 1979-2002⁷. In both of the works the authors carried out detailed analyses of labour productivity in basic economic sectors.

It should also be noted that beside the mainstream of studies on labour productivity in regions and sectors the kernel estimation method can be used in the individual labour productivity analysis⁸.

The principal aim of this study is to estimate the distribution of labour productivity in the largest Polish companies by means of the kernel estimation. Obtained results allow the authors to find out whether the distributions of this variable feature right-sided asymmetry and whether the shapes of these distributions change in time. Basing on the identification of sector membership of individual companies the types of economic activity prevailing in both the lowest and the highest values of labour productivity were determined. In the kernel estimation procedure the authors used two kernels best known in literature as Epanechnikov and Rosenblatt ones and a relatively new kernel B⁹.

1. Analysis methods and data characteristics

When we examine economic variables, their distributions are not usually known. In a classical approach we assume that the distribution of a given variable belongs to a certain class of theoretical distributions. In this case parameters of the location and scale of a given distribution type are estimated. Alternative to such approach is a non-parametric procedure (e.g. kernel estimation) that allows to construct a smooth histogram whose shape does not depend on the adopted bandwidth. When estimating the value of a density function for a given point in this procedure we attribute more weight to observations that are closer to this point¹⁰. There are also other solutions that can be used in the distribution analysis, such as D. Quah's stochastic kernel estimation¹¹.

Kernel density estimator is defined as:

$$\hat{f}_h(x) = \frac{1}{n \cdot h} \sum_{i=1}^n K\left(\frac{X - x_i}{h}\right), \quad (1)$$

where:

$h > 0$ – smoothing parameter called a bandwidth,

$K(x)$ – kernel function.

The key issue in kernel estimation is to determine the smoothing parameter h and the kernel function form. The value of constant h can be determined automatically or by means of the 2nd generation estimators¹². But the most frequently used kernel functions are the

Gaussian, Epanechnikov's and Rosenblatt's. In this study the two latter ones were accompanied by the kernel function B suggested by M. Mojsiewicz¹³:

$$K(x) = \begin{cases} -\frac{3}{(b-a)^3} \cdot (|x|-a) \cdot (|x|-b) & a \leq |x| \leq b \\ 0 & \text{for others} \end{cases} \quad (2)$$

Graphic presentation of kernel B for parameters $a = 0.2$ and $b = 0.95$ is presented in Figure 1.

Fig. 1. Graphic presentation of kernel B
Source: Mojsiewicz (2002).

M. Mojsiewicz¹⁴ demonstrates that kernel B more clearly determines the values of density functions in the right distribution tail and the calculated relative error becomes random.

Data used in the study came from 'Lists 2000' published by the Rzeczpospolita daily presenting 2000 Polish companies with the highest sales revenue in 2004-2008. In order to make the obtained results comparable the authors selected those companies that had been listed throughout the whole analysed period. The final group consisted of 560 companies. Additionally, the values of sales revenues were corrected with the price fluctuation indices that varied depending on the type of business activity of a given firm.

For the sake of sector analyses the selected companies were divided into four economic sectors according to NACE classification (Rev. 2):

1. Manufacturing (**C+D+E**).
2. Construction (**F**).
3. Non-financial Market Services (**G+H+I**).

4. Financial Market Services (J+K) and other services (L+M+N+O+P+Q).

The number of companies in individual sectors is presented in Table 1.

Table 1. The number of companies in sectors

Sector	1	2	3	4
Number of companies	272	32	224	32

Source: own calculations.

The analysed sectors differed greatly as far as their size is concerned. The largest were: sector 1 (Manufacturing) and sector 3 (Non-financial Market Services).

2. Empirical results

At the first stage of the analyses the authors identified the basic descriptive parameters characterising labour productivity distributions (see Table 2). When calculating labour productivity they used mean employment in a given year.

Table 2. Labour productivity mean and median for all the companies and labour productivity of highest concentration determined on the basis of theoretical distributions

Year	Mean	Median	Productivity of the highest concentration		
			Epanechnikov	Rosenblatt	B
2004	2 138.7	658.1	413.7	512.9	839.1
2005	2 037.2	659.0	393.3	502.4	784.6
2006	2 223.7	668.1	418.9	523.1	834.0
2007	2 488.5	709.2	455.9	564.9	940.1
2008	2 582.8	710.3	457.9	579.4	947.0

Source: own calculations.

Observation of both the mean value and the median leads to the conclusion that the mean labour productivity level in the largest Polish companies has been growing. However, the mean value increased faster (by 20.8% in 2008 in comparison to 2004) than the median (by 7.9%). That tendency was disturbed slightly in 2005. The productivity of the highest concentration followed the identical pattern. It should be noted, however, that the level of this type of productivity was changing along the change of kernel adopting the highest value for the kernel B.

On the basis of the calculated median values in individual sectors we can see that in companies of the sector 1 (Manufacturing) and sector 2 (Construction) the mean employment

was twice as high as in other sectors (see Table 3).

Table 3. Median of employment in sectors

Year	Sector			
	1	2	3	4
2004	694.5	600.0	250.0	252.0
2005	723.0	595.5	256.0	294.0
2006	749.5	624.5	289.5	318.0
2007	762.5	729.5	319.5	362.5
2008	749.0	644.0	335.0	385.0

Source: own calculations.

The comparison of the values of employment and labour productivity medians in individual sectors indicates a significantly higher level of labour productivity in those sectors where mean employment is lower. For example, in 2008 the median of labour productivity in the sectors 1 and 2 was PLN 562.5 and 355.5 thousand respectively, while in the sectors 3 and 4 – PLN 1,167.9 and 1,014.6 thousand.

At the next stage of the study the authors modelled the labour productivity distributions by means of the kernel estimation. The window widths were determined in line with the optimal window width rule depending on the number of companies and the mean labour productivity according to the following formulas:

$$- \text{for the Epanechnikov kernel } h = \sqrt[10]{20} \left(\frac{429}{2000} N^{-\frac{1}{5}} \right) \cdot \frac{\sum_i x_i}{N}, \quad (3)$$

$$- \text{for the Rosenblatt kernel } h = \sqrt[5]{54} \left(\frac{1}{\sqrt[5]{32\sqrt{\pi}}} N^{-\frac{1}{5}} \right) \cdot \frac{\sum_i x_i}{N}, \quad (4)$$

$$- \text{for the kernel B } h = 0,00001116 \left(N^{-\frac{1}{5}} \right) \cdot \frac{\sum_i x_i}{N}. \quad (5)$$

Figures 2–4 present distributions for the three types of kernels over all the analysed years. To make the diagrams more legible the upper limits of changeability intervals of the labour productivity distributions were set at PLN 6,000 thousand. It turned out that in all the years and kernel types the distributions demonstrated strong right-sided asymmetry¹⁵ and were characterised by a global maximum¹⁶. Another regularity was observed as well – the maximum was moving towards higher values of labour productivity and it was getting more and more flat in the subsequent years. In case of kernel B estimation of distributions a clear

minimum appeared of the low values of the analysed trait, which had been absent in case of the remaining two other kernels, the Epanechnikov and the Rosenblatt ones (see Figure 4).

Fig. 2. Labour productivity densities achieved by means of the Epanechnikov kernel
Source: own calculations.

Fig. 3. Labour productivity densities achieved by means of the Rosenblatt kernel
Source: own calculations.

Fig. 4. Labour productivity densities achieved by means of the kernel B

Source: own calculations.

Figures 5–7 present those parts of distributions which in 2008 included the values of labour productivity that were above the interval containing the productivity of the highest concentration. The lower limit of the analysed variable was set at PLN 1,500 thousand. Thanks to this limit we can observe vibrations on the graph of the density function for the labour productivity highest values. In addition, the introduction of different symbols for each of the sectors makes it possible to identify the sector structure of the presented distributions.

Fig. 5. Right tail of the labour productivity distribution estimated by means of the Epanechnikov kernel in individual sectors in 2008

Source: own calculations.

Fig. 6. Right tail of the labour productivity distribution estimated by means of the Rosenblatt kernel in individual sectors in 2008

Source: own calculations.

Fig. 7. Right tail of the labour productivity distribution estimated by means of the kernel B in individual sectors in 2008

Source: own calculations.

It can be seen that in the group of companies with labour productivity over PLN 1,500 thousand the majority of firms dealt with non-financial market services (sector 3).

The next stage of the study was the analysis of regularities in the right and left tails of distributions in the companies with the highest and the lowest levels of labour productivity. The criterion of selecting the range of the right tail was the phenomenon of flattening vibrations on the density function graph, which allowed setting the left limit of the right tail at PLN 10,000 thousand. The right limit of the left tail was set at the minimum value of the

density function obtained in the course of estimating the distribution by means of the kernel B.

Table 4. The number of companies in the right tail by years and sectors

Year	Sector 1		Sector 3	
	Number of firms	Percentage of firms*	Number of firms	Percentage of firms*
2004	2	0.7%	17	7.6%
2005	2	0.7%	18	8.0%
2006	2	0.7%	21	9.4%
2007	3	1.1%	21	9.4%
2008	5	1.8%	22	9.8%

* – the percentage in relation to all the other firms in the sector

Source: own calculations.

The group of companies of the highest productivity included only the ones dealing with manufacturing (sector 1) and the non-financial market services (sector 3), the latter ones being the majority (see Table 4). It should be noted, however, that mean employment in the service sector firms with the highest labour productivity was significantly lower than mean employment calculated for all the companies in the sector of non-financial services. The employment median in the former case was placed in the interval between 43–67 people, while in the whole sector it reached a much higher level (see Table 5).

Table 5 shows the sector structure of companies that found themselves in left distribution tails in subsequent years.

Table 5. The number of companies in the left tail by years and sectors

Year	Number of firms in sector				Total
	1	2	3	4	
2004	49	18	23	2	92
2005	45	16	20	1	82
2006	68	19	33	5	125
2007	57	17	29	4	107
2008	72	17	36	4	129
	Percentage of firms in sector				Total
	1	2	3	4	
2004	18.0%	56.3%	10.3%	6.3%	16.4%
2005	16.5%	50.0%	8.9%	3.1%	14.6%
2006	25.0%	59.4%	14.7%	15.6%	22.3%
2007	21.0%	53.1%	12.9%	12.5%	19.1%
2008	26.5%	53.1%	16.1%	12.5%	23.0%

Source: own calculations.

Although in view of the number of companies it is the manufacturing sector that predominates, when we take into account the relative share of representatives of individual sectors, we can see that over half of all the construction industry firms belong to the interval of the lowest labour productivity.

Table 6. The median of employment in the left tail

Years	Sector			
	1	2	3	4
2004	802.0	626.5	1 186.0	919.5
2005	870.0	668.5	1 670.0	989.0
2006	885.5	656.0	1 518.0	986.0
2007	887.0	756.0	2 635.0	1 125.5
2008	942.0	693.0	1 877.5	1 731.5

Source: own calculations.

Contrary to the right-tail companies the left tail ones demonstrated higher level of mean labour productivity in comparison to all the companies in a given sector (see Table 6). This regularity was particularly visible in case of service companies of both sectors 3 and 4.

Conclusions

In all the examined years the estimated distributions of labour productivity were characterised by very strong right-sided symmetry regardless of the type of kernel used for their estimation. The obtained results indicate the stable growth of labour productivity in the largest Polish companies. It should be noted, however, that sectors are strongly differentiated in relation to the examined variable. Among the four analysed economic sectors the best labour productivity was observed in the companies dealing with non-financial market services. They constituted a significant part of right distribution tails in all the studied years.

As far as the applied methodology is concerned, it is important to see the difference in character of distributions estimated by means of the kernel B in comparison to the two classical Epanechnikov and Rosenblatt kernels. The distributions obtained by means of the kernel B demonstrated local minimum of the density function for low values of labour productivity, which could not be seen in case of the other two kernels.

It seems that the choice of the research sample which consists of a set of companies identical in subsequent years and is corrected by eliminating the impact of goods and services price fluctuations has enabled the authors to draw reliable and valuable conclusions.

In further research the labour productivity analysis should cover a larger group of sectors, which would make it possible to study their nature more thoroughly. Also, it seems worthwhile to conduct a similar study on the international basis in order to identify possible differences and similarities concerning labour productivity in leading companies.

Finally, the issues presented in this study can be supplemented with research aiming at finding factors responsible for the regularities mentioned above.

Notes

¹ See e.g. the study on the EU presented by Dall'erba, Kamarianakis, Le Gallo, Plotnikova, 2005 and on the Polish economy: Polska 2005. *Raport o ...*, 2005.

² Benito, Ezcurra (2003).

³ E.g. see the works Scarpetta, Tressel (2002), Doyle, O'Leary, (1999), Tokarski, Gajewski (2002), Hultberg, Nadiri, Sickles (2004).

⁴ E.g. see Hozer (1991).

⁵ Van Biesebroeck (2003).

⁶ Nicolini (2007). See also Fingleton, López-Bazo (2003), Krüger (2004).

⁷ Margaritis, Färe, Grosskopf (2007).

⁸ E.g. see Batóg, Gazińska, Mojsiewicz (2002).

⁹ Mojsiewicz (2002).

¹⁰ Silverman (1986).

¹¹ Quah (1997).

¹² See Sheather, Jones (1991).

¹³ Mojsiewicz (2002).

¹⁴ Mojsiewicz (2002).

¹⁵ See Batóg (2001).

¹⁶ Unimodality of distributions can be verified by, for example, the bootstrap procedure proposed in the work of Efron and Tibshirani (1993), p.227.

References

Batóg, J., Gazińska, M. & Mojsiewicz, M. (2002). Ekonometryczne normowanie indywidualnej wydajności pracy. *Przegląd Statystyczny*, 49(1), 79-89.

Batóg, J. (2001). Dominujący typ rozkładu zmiennych w mikroskali. *Wiadomości Statystyczne*, 3, 18-24.

- Benito, J. M. & Ezcurra, R. (2003). *Spatial Disparities in Productivity and Industry Mix. The Case of the European Regions*. Department of Economics, Universidad Pública de Navarra.
- Dall'erba, S., Kamarianakis, Y., Le Gallo, J. & Plotnikova, M. (2005). Regional Productivity Differentials in Three New Member Countries What Can We Learn from the 1986 Enlargement to the South? *The Review of Regional Studies*, 35(1), 97-116.
- Doyle, E. & O'Leary, E. (1999). The Role of Structural Change in Labour Productivity Convergence among European Union Countries: 1970-1990. *Journal of Economic Studies*, 26(2), 106-120.
- Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman and Hall.
- Fiaschi, D. & Lavezzi, A.M. (2006). *Productivity Polarization and Sectoral Dynamics in European Regions*. Dipartimento di Scienze Economiche, Università di Pisa, and Dipartimento di Studi su Politica, Diritto e Società, Università di Palermo, October.
- Fingleton, B. & López-Bazo, E. (2003). *Explaining the Distribution of Manufacturing Productivity in the EU Regions*. ERSA conference papers ersa 03p197, European Regional Science Association.
- Hozer, J. (1991). Statystyka Bosego-Einsteina a skrajnie prawostronnie asymetryczne rozkłady demograficznych i ekonomicznych zmiennych losowych. In *Wyzwania dla nauk ekonomicznych w społecznej gospodarce rynkowej*. Szczecin: Fundacja na Rzecz Uniwersytetu Szczecińskiego, Polskie Towarzystwo Ekonomiczne.
- Hultberg, P.T., Nadiri, M.I. & Sickles, R.C. (2004). Cross-country Catch-up in the Manufacturing Sector Impacts of Heterogeneity on Convergence and Technology Adoption, *Empirical Economics*, 29, 753–768. DOI 10.1007/s00181-004-0209-5.
- Krüger, J.J. (2004). *Using the Manufacturing Productivity Distribution to Evaluate Growth Theories*. Jenaer Schriften zur Wirtschaftswissenschaft Arbeits- und Diskussionspapiere der Wirtschaftswissenschaftlichen Fakultät der Friedrich-Schiller-Universität Jena, 16.
- Margaritis, D., Färe, R. & Grosskopf, S. (2007). Productivity, Convergence and Policy: A Study of OECD Countries and Industries. In E. Thanassoulis & V.V. Podinovski (Guest Eds.), *Special issue on Applications and Application-Motivated Developments in Productivity Analysis*, *Journal of Productivity Analysis*, 28(1-2), 87-105. DOI 10.1007/s11123-007-0044-8.

- Mojsiewicz, M. (2002). *Badanie wpływu asymetrii informacji na ryzyko polskich instytucji ubezpieczeniowych. Analiza statystyczna*. Rozprawy i Studia T. (DV) 431, Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
- Nicolini, R. (2007). *Labor Productivity in Spain 1977-2002*. Institut d'Anàlis Econòmica, CSIC, November.
- Polska 2005. Raport o stanie gospodarki*. (2005). Warszawa: Ministerstwo Gospodarki i Pracy, Departament Analiz i Prognoz Ekonomicznych.
- Quah, D. (1997). *Regional Convergence from Local Isolated Actions Historical Outcomes*. CEP Discussion Papers 0378, Center for Economic Performance, LSE.
- Scarpetta, S. & Tressel, T. (2002). Productivity and Convergence in a Panel of OECD Industries: Do Regulations and Institutions Matter? *Economics Department Working Papers, OECD Publishing*, 342.
- Sheather, S.J. & Jones, M.C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society, Series B*, 53, 683-690.
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Tokarski, T. & Gajewski, P. (2002). Zależność między wydajnością pracy i zatrudnieniem w krajach OECD. *Wiadomości Statystyczne*, 8, 56-71.
- Van Biesebroeck, J. (2003). Revisiting Some Productivity Debates. *NBER Working Paper 10065*. Retrieved from <http://www.nber.org/papers/w10065>, NBER, Cambridge, October.