Bank i Kredyt 51(3), 2020, 293-316

BANK I KREDYT
BANK & CREDIT

A control function approach to measuring the total factor productivity of enterprises in Poland

Mariusz Górajski*, Mirosław Błażej#

Submitted: 3 June 2019. Accepted: 23 December 2019.

Abstract

Investigation of external and internal determinants of total factor productivity (TFP) is one of the main issues in the economics of growth. This paper aims to measure TFP and to identify the determinants of productivity for enterprises in Poland in the period of 2005–2016. Moreover, we examine sector heterogeneity of productivity and identify the sectors of the Polish economy in which enterprises achieve significantly higher TFP indicators. We estimate the production function by applying the econometric method of control functions. Under weak assumptions, this method allows for a consistent estimation of labour and capital elasticities of gross value added. We determine empirical distributions of TFP for the whole sample and conditional to selected productivity determinants. By applying econometric panel data models for the individual firms, we confirm the dependence of TFP of the enterprise on the form of ownership, investment rate, firm-level export status and their size. Finally, we observe a sector differentiation of TFP distributions and their strong dependence on the market concentration index.

Keywords: total factor productivity, production function estimation, control function methods, Levinsohn-Petrin model, Olley-Pakes model, TFP determinants

JEL: C14, C23, D21, D24

* University of Lodz, Faculty of Economics and Sociology; e-mail: mariusz.gorajski@uni.lodz.pl; Statistics Poland, Department of Macroeconomic Studies and Finance.

Statistics Poland, Department of Macroeconomic Studies and Finance; e-mail: miroslaw.blazej@stat.gov.pl.

1 Introduction

The total factor productivity (TFP) of an enterprise is an unobservable endogenous variable that determines the effectiveness of outlays on all production factors on the production volume of the enterprise. Both the correct measurement of total factor productivity and the indication of the main determinants of the enterprise productivity are necessary to describe the production process and resource management correctly.

TFP measurement is available mainly based on determining the residual component from the production function equation. Therefore, econometric models of the production function are used to determine the individual productivity of enterprises (see van Beveren 2012). The individual total factor productivity of an enterprise can be used directly to analyse the impact of other variables on the performance of a given enterprise or, after aggregation, serve as an indicator of productivity at the level of a selected sector or the entire economy.

The results of many studies show that after the recent global financial crisis, a slowdown in the growth of the global technological frontier has been confirmed (see IMF 2016; OECD 2015). At the same time, there is retardation (or an impediment) in technology transfer from most developed economies to developing countries (cf. IMF 2016; OECD 2015). As observed for many economies, the lack of return on the TFP growth path from before the global financial crisis is the so-called "productivity puzzle".

In recent years, against the background of the digital revolution and its integration with global value chains, we could expect significant increases in total factor productivity. However, for many countries, the expected increases in productivity are not observed. Therefore, the search for external factors, appropriate economic policy, market regulations or institutional settings that will stimulate the growth of individual productivity of enterprises is currently one of the main challenges in the field of economic growth. If the increase in productivity has not slowed down in recent years, then the appearance of the productivity puzzle can be associated only with TFP measurement problems. Thus, the answer to the question: to what extent the different methods of measuring TFP can lead to similar conclusions about the determinants of productivity of enterprises is of great scientific and practical importance.

For instance, according to Melitz (2003), the productivity of companies is the most crucial element impacting on the decision to export. Only the most productive companies with low marginal costs achieve the possibility of entering foreign markets (see Hagemejer 2006). It follows that in each sector there is a minimum productivity value below which the company is not able to maintain a positive export status. Besides, variables such as company size, level of market concentration, the form of ownership, and investment rate are among the main determinants of enterprise productivity.

There are only a few empirical TFP studies for the Polish economy in the literature. These are mainly analyses performed for the aggregate Cobb-Douglas function and assuming constant returns to scale in subregions or by voivodships (cf. Dańska-Borsiak 2011; Dańska-Borsiak, Laskowska 2012; Ciołek, Brodzicki 2016; Gosińska, Ulrichs 2020 and references to literature therein). Sulimierska (2014), for three levels of data aggregation, presents a very comprehensive study of total factor productivity in the manufacturing sector of the Polish economy. Hagemejer (2006) determines individual TFP indicators for enterprises from the manufacturing sector in order to indicate the determinants of the decisions to enter foreign markets. Moreover, Hagemejer and Kolasa (2011) estimate the production function for panel micro-data and measure the productivity for enterprises in Poland in the period of 1996–2005. Recently, for companies in Poland, Gradzewicz and Mućk (2019) have analysed the dynamics of markups of prices over marginal costs in the years 2002–2016. They estimate a translog production function based on the Ackerberg, Caves and Frazer model (2015) and prove that the globalisation of markets and changes in the global value chains are the main factors lying behind the recent fall in markups in Poland.

This paper aims to indicate the main determinants and conditions of individual total factor productivity indices for a large sample of enterprises in Poland in the period of 2005–2016. In particular, we attempt to answer the following question: in which sectors of the Polish economy do enterprises achieve significantly higher total factor productivity levels? The specification of the above--mentioned research question is the hypothesis that there is a significant sector diversification of total factor productivity in the Polish economy. Our analyses allow to identify sectors of the Polish economy in which enterprises do achieve significantly higher total factor productivity. This study is, according to the authors' knowledge, one of only a few attempts to measure TFP based on micro-panel data (cf. Pavcnik 2002; Breunig, Wong 2005; Criscuolo, Martin 2009; van Beveren 2012; Ackerberg, Caves, Frazer 2015 and references to literature therein).

The construction of the enterprise database used in this empirical study is based on a census of Polish enterprises employing more than nine employees and required preliminary data pre-processing (see Appendix A). We apply a control function method to solve the problem of endogenous explanatory variables in the enterprise production function (cf. Wooldridge 2015; Ackerberg, Benkard, Pakes 2007). The control functions introduced into the econometric model represented by the so-called proxy variables are designed to approximate the unobserved individual productivity of companies in such a way that a consistent estimation of the production function elasticities is feasible. Our paper compares the TFP estimation results obtained based on the Olley-Pakes (OP) model (cf. Olley, Pakes 1996) and two versions of the Levinsohn-Petrin (LP) model (see Levinsohn, Petrin 2003). The application of the title approach to the production function estimation allows to control the selection bias in the sample resulting from the natural process of exiting of companies from the market. We calculate empirical distributions of individual TFP indices for the whole sample and conditional to selected productivity determinants. In the second part of the paper, we examine dynamic panel models to measure the impact of selected productivity determinants on TFP level, including export status, ownership form, investment rate, company size and degree of market concentration.

The conducted analyses confirmed the sector diversity of empirical TFP distributions. We observe that companies operating in the information and communication sector (J), supporting financial and insurance activities (K), and dealing with professional, scientific and technical activities (M) achieve significantly higher levels of log*TFP*.[1] We confirm equally high levels of total factor productivity among companies producing and supplying electricity, gas, water and air conditioning (D). Besides, enterprises with dominant foreign owners have significantly higher total factor productivity levels – on average by over 23% – than companies with predominantly public share capital (state-owned enterprises). Exporters hold an advantage in the levels of total factor productivity over enterprises operating only on the Polish market, for which TFP ratios are on average 26% lower. We obtain the highest indicators of TFP for companies operating on highly concentrated markets, where they are on average twice as

[1] Hereinafter J, K, M, D are the names of sections according to the Polish Classification of Activities (PKD 2007) and/or the European NACE classification system Revision 2. PKD 2007 is coherent and comparable with the classification NACE Revision 2.

high as those for enterprises operating on markets with a structure close to the conditions of perfect competition.

The rest of this paper is organised as follows. The next section briefly presents the methodology of TFP measuring. In section 3, we determine the total factor productivity of Polish companies using classical panel-data models and the control function approach to estimating neoclassical Cobb-Douglas production function. Section 4 presents empirical TFP distributions. In section 5, using dynamic panel-data models, we investigate the determinants and conditions of individual total factor productivities. The last section provides some conclusions and discusses the implications of our findings.

2 TFP measurement

The enterprise production process is most often described by production functions that meet the assumptions of the neoclassical growth theory. In the second half of the 20th century, representatives of the Cambridge school proposed the concept of the neoclassical production function with constant elasticity of substitution (CES). CES functions are still very often used in theoretical and empirical analyses of the production process (see, e.g. Sztaudynger 2003; Klump, McAdam, Willman 2007; Growiec 2012 and references therein). Establishing uniform parameterisation for the CES family of functions is still discussed in the literature (cf. e.g. Klump, Preissler 2000). Hence the interpretation and estimation of CES function parameters can be cumbersome and confusing. A little later, a transcendental logarithmic function (translog) was used in theoretical considerations and empirical analyses of the production process. Kmenta (1967) showed that with some restrictions regarding the parameters, the translog function is a linear approximation of the CES function around the value of the elasticity-of-substitution parameter, which corresponds to the Cobb-Douglas function.

The question about the existence of the aggregate production function remains a crucial theoretical problem. Conformable with the microfoundations, the constructions of aggregated Cobb-Douglas and CES production functions were carried out by Jones (2005) and Growiec (2008a, 2008b). In particular, the authors, based on assumptions about the optimal behaviour of enterprises and the probabilistic definition of the technology frontier, in which it was assumed that unit-factor productivities come from independent Pareto distributions, have derived the aggregate Cobb-Douglas production function.

Estimates of the aggregate elasticity of substitution between production inputs known in the literature are inconclusive. Tables 1 and 2 in the work of Klump, McAdam and Willman (2004) collect estimates of various production function models for the US economy and other world economies. It is worth noting that a large part of the estimates of aggregate elasticity of substitution oscillates around unity – which corresponds to the Cobb-Douglas production function. Therefore, and due to the available microeconometric tools, in our considerations, we assume that the production process is determined by a function with constant elasticity of substitution. At the same time, no restrictions were imposed on the returns to scale, that will be subject to statistical verification.

Further in the paper, we assume that the gross value added, Y_{it}, for enterprise i in period t is determined by the Cobb-Douglas function in the form of:

$$Y_{it} = A_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} \tag{1}$$

where L_{it}, K_{it}, respectively, are the quantities of labour and capital, and A_{it} is the coefficient of the natural Hicks efficiency of the production process. L_{it} is the number of employees at the end of period *t*. The variables Y_{it}, K_{it} express production and capital values, respectively, and are not fully observable, but it is possible to measure them by setting actual real levels of gross value added and physical capital in the enterprise (values at constant prices 2010 = 100, see Appendix B).

The level of individual technology used in the production process is an unobservable variable, which is decomposed into a product of the average productivity of companies in the economy e^{β_0}, the coefficient of individual productivity V_{it} and the independent white-noise idiosyncratic component $U_{it} = e^{u_{it}}$. As a result, we obtain:

$$A_{it} = TFP_{it} U_{it} = e^{\beta_0} V_{it} U_{it} \tag{2}$$

From now on let $y_{it}, l_{it}, k_{it}, v_{it}, u_{it}$ denote the logarithms of the variables $Y_{it}, L_{it}, K_{it}, V_{it}, U_{it}$, respectively. Let the sum:

$$\omega_{it} = \beta_0 + v_{it} \tag{3}$$

define the total factor productivity coefficient at the level of *i*-th company. Then the production equation (1) can be presented in the log-linear form:

$$y_{it} = \omega_{it} + \beta_k k_{it} + \beta_l l_{it} + u_{it} \tag{4}$$

The productivity coefficient ω_{it} is often interpreted as a state variable in the company decision problem consisting in the selection of production factors, while the random component u_{it} is associated with all errors in measuring the variables and represents the so-called unpredictable productivity shock.

In order to determine the enterprise individual total factor productivity, equation (4) is estimated. As a consequence, we get the following estimation of the company productivity coefficient:

$$\hat{\omega}_{it} = \hat{\beta}_0 + \hat{v}_{it} = y_{it} - \hat{\beta}_k k_{it} + \hat{\beta}_l l_{it} \tag{5}$$

Hence the company productivity is calculated according to the formula:

$$\widehat{TFP}_{it} = e^{\hat{\omega}_{it}} \tag{6}$$

The values of $\widehat{TFP}_{it}$ for $i \in S$ can be used direcly to analyze the impact of other variables on company performance or, once aggregated, serve as an indicator of productivity at the economic sector level, *S*.

When estimating the production function, the emerging problems of endogeneity of explanatory variables, endogeneity of attrition and omitted individual prices should be addressed (see Ackerberg, Benkard, Pakes 2007 or van Beveren 2012). The problem of simultaneity or endogeneity of explanatory variables lies in the company decisions on the workloads of labour and capital outlays. These inputs are not independently chosen but rather linked to the current level of productivity. Thus in the case of estimation of the production function using the classical least-squares method, we often get an upward bias in the input coefficients for labour and materials, and an underestimation of gross value added elasticity relative to capital (see Ackerberg, Caves, Pakes 2007, p. 4206). The problem of selection bias occurs when balanced panel data are used in the estimation, from which incomplete statistical units have been removed, i.e. companies that have ceased operations or entered the market during the sample period. Exit or entry decisions are endogenous for companies and strongly correlated with the level of productivity. Hence the estimates of the production function coefficients derived from the preselected samples (i.e. samples without the enterprises who have left the market) become biased and inconsistent.

The problem of omitted price bias is related to the lack of unit prices of products and production factors. This problem cannot be fully solved by applying an appropriate estimation method, but the use of maximally disaggregated price deflators is a partial solution.[2] We note that due to the lack of unit prices, there is a potential impact of the monopoly rent on the total factor productivity level (cf. Criscuolo, Martin 2009). The TFP measurement procedure used in our paper may overestimate the value of this indicator for enterprises with high monopolistic charges. When measuring TFP, average prices determined at the level of divisions of the Polish Classification of Activities (PKD 2007) are used in place of individual prices to determine the real gross value added Y_{it}. As a result, we overestimate the values of Y_{it} for companies with high monopoly rents. As a consequence, the TFP indices may also be disturbed by the effect of high monopolistic margins. In section 5, we compensate for the omitted--price effect by adding the Herfindahl-Hirschman market concentration index[3] at the PKD-division level to the set of explanatory variables in dynamic panel-data models for TFP.

The OP model and the LP model, which belong to the class of control function methods, are robust against the first two problems mentioned above, due to the use of a variable approximating productivity shocks (the so-called proxy variables), as well as due to the probit model estimates of the company probability of survival on the market. The OP model uses investments as a proxy variable, while the LP model assumes that expenditures on materials and energy control for unobserved TFP indices. The estimation of LP and OP models is a three-stage procedure. In the first stage, we estimate the labour elasticity by building a non-linear regression model, where we approximate the unobservable productivity indices by a higher-order polynomial of the capital and proxy variable. In the second step, we estimate the conditional probabilities of the company survival on the market. In the last stage, substituting the results from the first two estimation steps, we obtain a non-linear regression equation for the gross value added of those enterprises that survived on the market. In this step, thanks to the non-linear least squares method, we obtain an estimate of the elasticity of the capital. Standard errors of parameter estimates are determined using bootstrap methods.

In summary, the measurement of total factor productivity in control function models is possible by indicating proxy variables for an unobservable factor of productivity. The approximation of productivity by investments proposed by Olley-Pakes raises several doubts because it is based on the assumption of a monotonic, positive relationship between the productivity TFP and investment outlays. Moreover, only entities for which positive investment expenditures have been recorded can take part in the estimation, which significantly reduces the sample size. Levinsohn and Petrin initiated the search for other control variables for productivity, among which were considered, inter alia, intermediate consumption and its components (cf. e.g. Levinsohn, Petrin 2003; Gradzewicz, Mućk 2019) or corporate profits (Criscuolo, Martin 2009).

[2] The authors hereby thank Dariusz Kotlewski for providing investment, capital and gross-value added deflators at 4 digits PKD sectors.

[3] The high levels of the Herfindahl-Hirschman index indicate companies operating on the markets with a monopolistic structure.

In section 3, we used the following methods to estimate the production function:

– a linear regression model with the ordinary least-squares estimator (pooled OLS model which is referred as a benchmark model),

– panel data models with individual effects, including an estimator of fixed effects (FE model) and a model with random individual effects (RE model),

– methods of control functions, including OP model, and two specifications of the Levinsohn--Petrin model.

3 Estimation of production function parameters

The data used in our study originate from annual reports for years 2005–2016 on the business activity of all Polish enterprises employing at least ten employees. All data are reported in the SP survey of Statistics Poland (Annual Enterprise Survey, hereinafter SP sample, see also Appendix A). The results of the gross value added production function estimation using six alternative econometric models for the sample of all enterprises from SP survey are collected in Table 1.[4] The standard errors of estimators in control function models were determined using the bootstrap procedure. In all the analysed models, the results of Student's t-tests indicate a statistically significant positive impact of labour and capital on the gross value added of companies. We note significant differences between the estimators of production function elasticities determined from classic panel-data models (pooled OLS, RE, FE models) and models based on control function methods (LP and OP models). For the first group of models (pooled OLS, RE, FE) and in the OP model, the capital coefficient β_k is underestimated.[5] The impact of labour input on production volume is much stronger in the pooled OLS, RE, FE models. These results are directly related to the problem of endogeneity of input factors in the production function described in the previous part of the paper.

LP models significantly increase the role of capital in the production process, while the output elasticity of labour is significantly lower than in the other proposed models. The OP model indicates a shrink output elasticity of capital, which may be due to the relatively small sample size, as we are limited to enterprises with positive investment outlays. In order to verify the occurrence of constant returns to scale, we perform Wald's tests, in which the null hypothesis assumes that $H_0: \beta_k + \beta_l = 1$. The value of Wald's statistics indicates that at the significance level of 0.001 the hypothesis about constant returns to scale should be rejected for the production function estimated using control function methods and panel-data FE and RE econometric models. Only the pooled OLS estimator indicates the occurrence of constant returns to scale. The mechanism correcting for enterprise exits from the market in the LP model allows for a slight increase in the estimate of the output elasticity of the capital coefficient.

Considering the properties of control function models, in further analyses, we ultimately choose the estimates of production function parameters obtained from the Levinsohn-Petrin model with the market exit rule (indicated by LPe later on).

[4] All calculations and estimations were made in R and STATA programs, in particular the following packages were used: prodest (Rovigatti 2017), estprod, dplyr, plm.

[5] Using aggregated data for sectors of the Polish economy, it is shown that the share of gross operating surplus reaches 45% of gross-value added (see e.g. Kotlewski, Błażej 2020). This value may constitute a good approximation of the capital share in the gross-value added c^S_{share} in a given sector *S*. Hence, using the following equation $\beta^I_K = \mu \cdot c^S_{share}$ combining the gross value added elasticity of capital β^I_K with the monopolistic markup μ and capital share c^S_{share}, we should expect that the aggregate elasticity of capital β^I_K will oscillate around 50%.

4 Empirical TFP distributions

In this section, we present and analysel og*TFP* distributions based on the SP sample and in subsamples of companies defined by selected enterprise and market characteristics (the so-called explanatory variables or TFP determinants, see Table 2) and by PKD sections[6] (see Table 3). The set of explanatory variables contains six qualitative variables:

1) *ownership* is a variable with three categories: state-owned enterprises (*ownership* = *SOE*) are enterprises with a majority public ownership; private domestic enterprises (*ownership* = *PDE*) with a majority private domestic ownership, and in the third category, we include foreign-owned enterprises (*ownership* = *FOE*);

2) *export intensity* is determined by the *ES* ratio of the exports sale revenues to total sale revenues and has been divided into three categories: non-exporters (*export intensity* = *no export*) if $ES = 0\%$, exporters (*export intensity* = *moderate*) for $ES \in (0\%, 50\%]$ and companies with dominant shares of exports in revenues constitute the last category (*export intensity* = *high*) if $ES > 50\%$;

3) *investment intensity* is a qualitative variable with four determined categories based on the investment rate (*IR*), where *IR* is the quotient between changes in fixed assets and gross value added in a given year; in the first category, there are companies with no investments in a given year, i.e. $IR \leq 0\%$ (*investment intensity* = *disinvestment*); companies with low investment rates $IR \in (0.10\%]$ are included in the next group (*investment intensity* = *low*); in the third category there are companies with average investment rates, $IR \in (10\%, 25\%]$ (*investment intensity* = *moderate*); in the last group we distinguish companies with high investment rates IR > 25% (*investment intensity* = *high*);

4) *market concentration* is determined using the Herfindahl-Hirschman index (*HH_index*) determined on the basis of PKD divisions; for this variable, three categories are distinguished for each year: in the first group, companies are operating on the market with low concentration where market conditions similar to perfect competition do prevail (*market concentration* = *low* if $HH\_index < 0.01$); in the second category we group companies with average concentration indices $HH\_index \in (0.01; 0.2]$ (*market concentration* = *moderate*); the last group includes companies operating in PKD 4 digits sectors for which the Herfindahl-Hirschman concentration indices are above 0.2 (*market concentration* = *high*);

5) *enterprise size* is determined by the average number of employees L (in full-time equivalents) employed in a given year; there are three categories: medium-sized companies (*enterprise size* = *medium* if $L < 50$), large companies (*enterprise size* = *large* if $L \in (50, 250]$) and the last category includes very large companies (*enterprise size* = *very large* if $L > 250$);

6) *sector* indicates the enterprise PKD section, thus we distinguish 19 categories of the sector variable: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S.

We use the LPe model to estimate the production function parameters (see Table 1) and calculate the productivity coefficients $\hat{\omega}_{it} = \widehat{\log TFP_{it}}$ according to (5). The empirical distributions of $\hat{\omega}_{it}$ for all companies in the SP sample and most of the subsamples selected by the categories of TFP determinants are symmetric with the features of the Gaussian distribution (see Figures 1–5 and Tables 2–3). In few cases, e.g. for enterprises operating on markets with very high concentrations (*market concentration* = *high*) or for companies from the public administration sector (section O in PKD), we observe asymmetrical and bimodal distributions (see Figure 2 right panel and Figure 5). Moreover, we report heterogeneity of the empirical densities of $\hat{\omega}_{it}$ across PKD sections (see Figure 5).

[6] In these analyses there are no differences between the Polish PKD and the European NACE classification systems.

In the SP sample, the average level of productivity coefficients $\hat{\omega}_{it} = \widehat{\log TFP_{it}}$ is 3.93 (median of $\hat{\omega}_{it} = 3.90$) and the standard deviation is 0.76. The sectors with the highest average level of total factor productivity include (cf. Figure 5 and Table 3): enterprises supporting financial and insurance services (section K with the median of $\hat{\omega}_{it} = 4.71$), companies from the information and communication section (section J with the median of $\hat{\omega}_{it} = 4.61$) and enterprises engaged in professional, technical and scientific activities (section M with the median $\hat{\omega}_{it} = 4.46$). Comparing the measures of central tendency for $\hat{\omega}_{it}$ distributions in cross-sections determined by categories of explanatory variables (see Table 2), it is easy to notice that the groups of companies with the highest productivity coefficient values include: enterprises operating in highly concentrated markets (median of $\hat{\omega}_{it} = 4.44$, see right panel of Figure 2), companies with dominant foreign-owned capital (median of $\hat{\omega}_{it} = 4.40$, see left panel of Figure 1) and a group of very large enterprises (median $\hat{\omega}_{it} = 4.15$, see Figure 2 left panel).

The high share of exports in revenues (see Figure 1 – right panel), and significant value of foreign-owned capital (see Figure 1 – left panel) shifts to the right empirical density functions of productivity coefficients $\hat{\omega}_{it} = \widehat{\log TFP_{it}}$ in relation to non-exporting companies and state-owned enterprises, respectively.

In PKD sections with high levels of the Herfindahl-Hirschman concentration index (see Figure 2 right panel), we observe a shift in the distribution of the total factor productivity indicator $\hat{\omega}_{it}$ towards higher values. At the same time for enterprises with high concentration indices, we distinguish two groups of companies focused around two modal values of $\hat{\omega}_{it}$ distribution: the first consists of companies that have levels of total factor productivity similar to the average for the SP sample oscillating around 4.2, the second group includes companies with high levels of productivity indicator exceeding 4.5.

The empirical distributions of the total factor productivity indicator for the sample of companies with different levels of investment rates are similar to each other (see Figure 3). Companies with very high investment rates have similar $\hat{\omega}_{it}$ distributions to non-investing companies (medians of $\hat{\omega}_{it}$ for both these categories are around 3.8), while companies with low investment rates show higher levels of total factor productivity (median of $\hat{\omega}_{it}$ = 4.12).

Figure 4 shows the dynamics of $\hat{\omega}_{it}$ distributions in the period 2005–2016. The empirical distributions of the logs of total factor productivity index determined in subsequent years have similar shapes, and as the years go by, we observe their shift towards higher values (median of $\hat{\omega}_{it}$ in 2005 = 3.70; median of $\hat{\omega}_{it}$ in 2016 = 4.04).

The distribution of logs of total factor productivities among manufacturing companies (section C) is similar to the empirical density function of productivity coefficients in the SP sample. At the same time, we observe a significant increase in the economic efficiency of enterprises in the following sections: K (activities supporting financial and insurance services), J (Information and communication), M (Professional, scientific and technical activity) and B (Mining and quarrying) (see Figure 5). We note that the log*TFP* distribution for companies supporting financial and insurance activities (section K) is very diffuse and right-skewed; hence this sector of the economy includes companies with the highest values of total factor productivity.

5 Determinants of individual total factor productivity

In this section, we look for the determinants of the total factor productivity of Polish enterprises. We assume that the firm-level total factor productivity $\hat{\omega}_{it} = \widehat{\log TFP_{it}}$, which is derived from the LPe model according to (5), is a response variable. The set of explanatory variables is specified in section 4 and includes categorical variables describing: the form of ownership (*ownership*), export status (*export intensity)*, investment intensity (*investment intensity*), company size (*size*), market concentration (*concentration*), firm economic sector (λ_{is} = 1 if *sector* = *s*). We also add a full set of time dummies λ_t = 1 for t = 2006, …, 2016. The basic model for the total factor productivity of enterprises has the form of:

$$\begin{aligned}\hat{\omega}_{it} &= \lambda_t + \lambda_{is} + \alpha_i \\ &+ \sum_{k=1}^{2} \left(\alpha_{1,k} \cdot ownership_{it,k} + \alpha_{2,k} \cdot export\_intensity_{it,k} + \alpha_{3,k} \cdot size_{it,k} + \alpha_{4,k} \cdot concentration_{it,k} \right) \\ &+ \sum_{k=1}^{3} \left(\alpha_{5,k} \cdot investment\_intensity_{it,k} + \alpha_{6,k} \cdot investment\_intensity_{it-\varepsilon_{it}} \right) + \varepsilon_{it}\end{aligned} \tag{7}$$

where α_i corresponds to individual effect for enterprise i, and $ownership_{k,it}$, $export\_intensity_{it,k}$, $size_{it,k}$, $concetration_{it,k}$, $investment\_intensity_{it,k}$, $investment\_intensity_{it\text{-}1,k}$ are binary variables with the value 1 for the *k*-th category of the explanatory variable, and ε_{it} is the white noise error term.

The dynamic specification of the model with the autoregressive component is also considered:

$$\begin{aligned}\hat{\omega}_{it} &= \alpha_0 \hat{\omega}_{it-1} + \lambda_t + \lambda_{is} + \alpha_i + \\ &+ \sum_{k=1}^{2} \left(\alpha_{1,k} \cdot ownership_{it,k} + \alpha_{2,k} \cdot export\_intensity_{it,k} + \alpha_{3,k} \cdot size_{it,k} + \alpha_{4,k} \cdot concentration_{it,k} \right) \\ &+ \sum_{k=1}^{3} \left(\alpha_{5,k} \cdot investment\_intensity_{it,k} + \alpha_{6,k} \cdot investment\_intensity_{it-1,k} \right) + \varepsilon_{it}\end{aligned} \tag{8}$$

The estimation of equation (7) is based on two-way panel-data models with individual and time effects. In addition to the classic pooled OLS and between estimator (BE) estimators, we also apply the fixed effect (FE) and the random effect (RE) estimators adding time and sector effects to each model specification. In the search for the best model, continuous explanatory variables were also considered and lags in investment expenditures were taken into account.

Statistical tests indicate the occurrence of both individual random effects (LM Breusch-Pagan test) and individual fixed effects (Wald's chi-squared test). Wald's tests confirmed the statistical significance of time and sector effects. Due to yearly effects in the model (7), the Hausman test cannot be used to choose between FE and RE models. For this purpose, the Mundlak test was used, which rejects RE models in favour of the FE specification. This result should be treated with caution due to the relatively low value of the correlation coefficient between random individual effects and the response variable. Moreover, the Wooldridge autocorrelation test for residuals from the model (7) indicates a strong autocorrelation of residuals and, as a consequence, an incorrect specification of the model.[7] Hence, the model (7) was extended to equation (8), which additionally contains a lagged response variable.

[7] In the class of classic panel-data models, the FE estimator, preferred by the Mundlak test, leads to estimates of parameters that are not in line with economic intuition. In particular, the FE model shows that foreign-owned enterprises have lower total factor productivity indices than companies with public capital, and very large companies operate less efficiently than medium-sized companies. These results are a consequence of unfulfilled assumptions of the FE model regarding the lack of autocorrelation of the random effect component.

Model (8) was estimated using pooled OLS, FE, RE, BE estimators (see Table 4). Due to the lagged endogenous variable on the right-hand side of equation (8), estimates from these models may be inconsistent and biased. In order to solve the problem of regressor endogeneity in (8), an estimation of the system generalized method of moments was used for estimation (hereinafter briefly sGMM, cf. Blundell, Bond 1999), where instrumental variables are constructed using orthogonal increments of the response variable. The last two columns of Table 3 contain the results of model estimation (8) obtained using sGMM estimators for two different sets of instrumental variables (sGMM1, sGMM2). Furthermore, only the sGMM2 model includes sector effects. For both specifications, Arellano-Bond tests confirm the lack of autocorrelation of the error term. In order to verify the set of instrumental variables, the results of the Sargan and Hansen test of overidentifying moment restrictions confirm that the selection of instrumental variables for the sGMM1 model is correct.

The verification of the impact of selected explanatory variables on the individual productivity coefficient, $\hat{\omega}_{it} = \widehat{\log TFP_{it}}$, is carried out based on Student t and Wald's tests (details are presented in Tables 4 and 5). The results indicate a significant combined effect of all variables included in the models sGMM1, sGMM2 on the levels of enterprise total factor productivity.

Considering the above statistical-econometric analysis, ultimately panel autoregressive models: sGMM1, sGMM2 were chosen to formulate economic conclusions. Firstly, we observe a quite strong persistence in the dynamics of logarithms of total factor productivity (the autoregression coefficient is over 0.6). Secondly, foreign-owned enterprises have on average at least 23% higher productivity rates than companies with public-owned capital ($\widehat{\alpha}_{1,2} = 0.239$ in sGMM1 model, $\widehat{\alpha}_{1,2} = 0.232$ in sGMM2 model). An increase in the share of exports in sale revenues stimulates the total factor productivity of enterprises. Companies with a high share of exports increase their productivity on average by about 26% in relation to units operating on the domestic market ($\widehat{\alpha}_{2,2} = 0.261$ in sGMM1 model, $\widehat{\alpha}_{2,2} = 0.076$ in sGMM2 model). Large enterprises with the same inputs of production factors can achieve slightly higher gross value added, on average by up to 3% than medium-sized companies ($\widehat{\alpha}_{3,2} = 0.031$ in sGMM2 model).

For companies operating on markets with very high concentrations, we observe on average a more than twofold increase in the total factor productivity compared to markets with a structure similar to perfect competition ($\widehat{\alpha}_{4,2} = 2.072$ in sGMM1 model). It is worth noting here that the significant increase in TFP indicators for companies operating in highly concentrated markets may be mainly due to the problem of the lack of firm-individual prices. As a result, the estimated TFP indices can be further strengthened by the high monopoly markups of these enterprises. Hence, adding the market concentration variable to the set of explanatory variables allows for excluding the impact of monopolistic power on TFP indices and leads to the correct interpretation of model parameters.

The investment activity of enterprises is approximated in the model (8) using the current and lagged investment rates. Estimates of parameters standing at the categories of current investment intensity are consistent with the conclusions obtained based on the conditional log*TFP* distributions from section 4 (see Figure 3 and Table 2). An increase in current investments with a rate not exceeding 10% results in an increase in TFP indices by an average of over 8% ($\widehat{\alpha}_{5,1} = 0.092$ in sGMM1 model, $\widehat{\alpha}_{5,1} = 0{,}084$ in sGMM2 model), but the companies with high current investments (where the investment rates exceed 25%) report a decrease in the individual total factor productivity by about 9%. The impact of investments from the previous year on the current productivity of enterprises is weak or even statistically insignificant. Only for enterprises with investment rates between 10%

and 25% do we observe the negative impact of the lagged investment expenditures on the TFP levels (decrease in TFP by about 2% compared to companies without investments in the last year). However, the obtained results should be interpreted with some caution due to difficulties in econometric modelling of investment processes in the enterprise. The difficulties are related to finding the transmission of current and lagged investments in the future value of physical capital, but also organizational changes that occur during the investment process are difficult to quantify.

The estimation results of sectoral effects from the sGMM2 model (8) are summarized in Table 5. The Student's t-tests confirm the hypothesis about significant sectorial differentiation of the total factor productivity in the Polish economy. Sectorial effects were expressed in relation to the level of log*TFP* indices in the manufacturing sector (section C). For most PKD sections (apart from sections A, O, S) we note a significantly higher average total factor productivity than for companies from section C. The highest average log TFP ratios are among enterprises operating in the information and communication sector (J), supporting financial and insurance activities (K) and those involved in professional, scientific and technical activities (M), where companies achieve higher total factor productivity by 28.1%, 25.8%, 21.9%, respectively, above the average TFP level in section C. Equally high productivity measures can be confirmed among the companies producing and supplying electricity, gas, water and air conditioning (D), where logs of TFP are on average 23.7% higher.

6 Summary

Accurate measurement of the firm-level total factor productivity is feasible thanks to appropriate methods of estimating the production function equation. In this paper, we estimate the neoclassical production function based on micro-data of enterprises from the Polish economy. Estimation of the production equation is related to, among others, problems of input endogeneity, sample selection bias or omitted individual prices. Consequently, OLS estimators can lead to inconsistent estimates. In order to solve these problems, the control function method was used, which allows, with fairly general assumptions, for consistent estimation of output elasticities.

On the one hand, LP and OP econometric models partly solve the problem of underestimation of capital elasticity reported in the literature. However, the obtained values of output elasticity of capital are still lower than expected. Besides, the observed decreasing returns to scale lead, among others, to the issue on how to measure capital correctly. On the other hand, the tools used to measure TFP indices do not take into account the lack of individual prices, which, despite decreasing markups over the companies' production costs in the last decade, could be the reason for underestimating the capital coefficient in the production function and the occurrence of decreasing returns to scale.

In the second part of the paper, the primary productivity determinants were indicated using dynamic panel-data models for individual TFP indicators. Firstly, we confirm the positive relationship between the export intensity and the value of the total factor productivity of the enterprise (TFP increases up to 26% on average for companies with positive export status). Secondly, very large companies report a higher TFP of 3%. Moreover, companies with foreign-owned capital have a TFP productivity rate, on average, at least 23% higher than state-owned companies. Thirdly, sector diversification of enterprise productivity distributions and their strong dependence on the market concentration indices have been observed. Enterprises from the information and communication sector (J), supporting financial and insurance activities (K) and dealing with professional, scientific and technical activities (M) achieve higher total factor productivity levels, 28.1%, 25.8% and 21.9%, respectively, above the average TFP level in the manufacturing sector (C).

References

Ackerberg D., Benkard B.S., Pakes A. (2007), Econometric tools for analyzing market outcomes, in: J. Heckman, E. Leamers (eds.), *Handbook of Econometrics*, vol. 6a, Elsevier.

Ackerberg D., Caves K., Frazer G. (2015), Identification properties of recent production function estimators, *Econometrica*, 83(6), 2411–2451.

van Beveren I. (2012), Total factor productivity estimation: a practical review, *Journal of Economic Surveys*, 26, 98–128.

Blundell R., Bond S. (2000), GMM estimation with persistent panel data: an application to production functions, *Econometric Reviews*, 19(3), 321–340.

Breunig R., Wong M.-H. (2005), Estimation of total factor productivity, in: *Productivity Commission 2005, Quantitative Tools for Microeconomic Policy Analysis, Conference Proceedings,* 17–18 November 2004, Canberra.

Ciołek D., Brodzicki T. (2016), Determinants of productivity of Polish LADs, *Bank i Kredyt*, 47(5), 463–494.

Criscuolo C., Martin R. (2009), Multinationals and U.S. productivity leadership: evidence from Great Britain, *The Review of Economics and Statistics*, 91, 263–281.

Dańska-Borsiak B. (2011), Zróżnicowanie łącznej produktywności czynników produkcji według województw, *Wiadomości Statystyczne,* 12(607), 13–26.

DańskaBorsiak B., Laskowska I. (2012), The determinants of total factor productivity in Polish subregions. Panel data analysis, *Comparative Economic Research*, 15(4), 17–29.

Gosińska E., Ulrichs M. (2020), Sektorowe funkcje produkcji – wnioski z modeli panelowych dla Polski, *National Economy*, forthcoming.

Growiec J. (2008a), Production functions and distributions of unit factor productivities: uncovering the ink, *Economocs Letters*, 101, 87–90.

Growiec J. (2008b), A new class of production functions and an argument against purely labor-augmenting technical change, *International Journal of Economic Theory*, 4, 483–502.

Growiec J. (2012), *Zagregowana funkcja produkcji w ekonomii wzrostu gospodarczego i konwergencji*, Oficyna Wydawnicza SGH.

Gradzewicz M., Mućk J. (2019), *Globalization and the fall of markups*, NBP Working Paper, 304, Narodowy Bank Polski.

Hagemejer J. (2006), Factors driving the firms decision to export. Firm-level evidence from Poland, *Bank i Kredyt*, 37(7), 30–43.

Hagemejer J., Kolasa M. (2011), Internationalisation and economic performance of enterprises: evidence from Polish firm-level data, *The World Economy*, 34(1), 74–100.

IMF (2016), *Central, Eastern and Southeastern Europe: How to get back on the fast track?,* International Monetary Fund.

IMF (2019), *Country Report,* 19/38, International Monetary Fund.

Jones C.I. (2005), The shape of production functions and the direction of technical change, *Quarterly Journal of Economics*, 120, 517–549.

Klump R., McAdam P., Willman A. (2004), *Factor substitution and factor-augmenting technical progress in the United States: a normalized supply-side system approach*, EBC Working Paper Series, 367.

Klump R., McAdam P., Willman A. (2007), Factor substitution and factor-augmenting technical progress in the United States: a normalized supply-side system approach, *The Review of Economics and Statistics*, 89, 183–192.

Klump R., Preissler H. (2000), CES production functions and economic growth, *The Scandinavian Journal of Economics*, 102, 41–56.

Kmenta J. (1967), On estimation of the CES production function, *International Economic Review*, 8, 180–189.

Kotlewski D., Błażej M. (2020), Sustainability of the convergence between Polish and EU developed economies in the light of KLEMS growth accounting, *Bank i Kredyt*, 51(2), 121–142.

Levinsohn J., Petrin A. (2003), Estimating production functions using inputs to control for unobservables, *Review of Economic Studies*, 70, 317–341.

Melitz M.J. (2003), The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica*, 71, 1695–1725.

OECD (2015), *The Future of Productivity*.

Olley G.S., Pakes A. (1996) The dynamics of productivity in the telecommunications equipment industry, *Econometrica*, 64(6), 1263–1282.

Pavcnik N. (2002), Trade liberalization, exit, and productivity improvement: evidence from Chilean plants, *Review of Economic Studies*, 69(1), 245–276.

Rovigatti G. (2017), Production function estimation in R: the prodest package, *Journal of Open Source Software*, 2(371), 1–19.

Sulimierska M. (2014), *Total factor productivity estimation for Polish manufacturing industry – a comparison of alternative methods*, Working Paper Series, 67, University of Sussex.

Sztaudynger J.J. (2003), *Modyfikacje funkcji produkcji i wydajności pracy z zastosowaniami*, Wydawnictwo Uniwersytetu Łódzkiego.

Wooldridge J.M. (2015), Control function methods in applied econometrics, *Journal of Human Resources*, 50(2), 420–45.

Acknowledgements

The results presented in this paper are an extension of the research carried out under the cooperation of Statistics Poland with the International Monetary Fund in 2018 (see IMF 2019).

The authors would like to thank for all valuable comments and remarks provided during this work elaboration. We want to thank the reviewers, P. Baranowski, D. Kotlewski, K. Krogulski, R. Sierhej, J.J. Sztaudynger and S. Wójcik, as well as colleagues from the Department of Econometrics of the University of Lodz, and the participants of the conferences: *Produktywność gospodarki: uwarunkowania, determinanty, perspektywy* (Wrocław, April 2019) and VIII Scientific Conference *Modelowanie i prognozowanie gospodarki narodowej* (Sopot, May 2019).

The views and opinions presented in this article are those of the authors and have not been endorsed by Statistics Poland.

Appendix

Table 1
Production function estimators

	LP	LPe	OPe	FE	RE	OLS
β_l	0.7607***	**0.7607*****	0.7355***	0.7907***	0.8156***	0.8443***
β_k	0.1560***	**0.1666*****	0.0578***	0.0809***	0.0969***	0.1260***
$\beta_l + \beta_k = 1$	0.9167***	**0.9273*****	0.7933***	0.8717***	0.9124***	0.9703
β_0	4.0074	**3.9323**	4.7895	4.3589	4.1802	3.9162
Volatility of log(*TFP*)	0.7582	**0.7590**	0.7917	0.7817	–	–
Sample size	584 893	**584 893**	511 369	584 893	584 893	584 893

Notes:
LP – the Levinsohn-Petrin model, LPe – a Levinsohn-Petrin model with exit rule, OPe – the Olley-Pakes model with exit rule, FE – a panel-data fixed-effect model, RE – a panel-data random-effects model, OLS – the pooled OLS model.
* p-value = 0.05, ** p-value = 0.01, *** p-value = 0.001 in Student's t-tests for $H_0 : \beta_l = 0$, $H_0 : \beta_k = 0$ and the Wald test for $H_0 \colon \beta_l + \beta_k = 1$.

Table 2

Descriptive statistics of log*TFP* distributions in cross-sections defined by the categories of explanatory variables

		N	min	Q1	me	m	Q3	max	sd	IQR
SP sample		584893	-5.74	3.48	3.90	3.93	4.36	10.31	0.76	0.87
Ownership	FOE	75283	-3.16	3.90	**4.40**	4.41	4.94	9.60	0.86	1.04
	PDE	320684	-5.74	3.53	3.93	3.94	4.34	9.87	0.72	0.81
	SOE	38597	-3.39	3.53	3.85	3.88	4.21	9.45	0.70	0.67
Export intensity	no export	392997	-4.17	3.42	3.83	3.85	4.25	10.31	0.74	0.83
	moderate	135655	-5.74	3.64	4.06	4.10	4.54	9.84	0.77	0.90
	high	54625	-3.38	3.64	**4.09**	4.12	4.59	10.27	0.77	0.94
Investment intensity	high	74880	-4.17	3.38	3.81	3.79	4.24	9.45	0.80	0.85
	moderate	52612	-0.37	3.57	3.95	3.97	4.34	8.38	0.64	0.77
	low	142389	-2.89	3.70	**4.12**	4.18	4.59	10.27	0.73	0.89
	disinvestment	313396	-5.74	3.41	3.82	3.85	4.26	10.31	0.76	0.86
Enterprise size	very large	35356	-5.74	3.75	**4.15**	4.18	4.59	7.65	0.71	0.83
	large	160772	-4.73	3.51	3.89	3.92	4.30	10.27	0.69	0.79
	medium	387149	-4.17	3.45	3.88	3.92	4.36	10.31	0.79	0.90
Market concentration	high	2549	-1.51	3.82	**4.44**	4.52	5.09	10.31	1.13	1.28
	medium	299515	-4.73	3.44	3.87	3.90	4.33	9.87	0.77	0.89
	low	256344	-4.17	3.56	3.96	3.99	4.39	10.15	0.72	0.82

Notes:
FOE – foreign-owned enterprises, PDE – private domestic enterprises, SOE – state-owned enterprises, N – subsample size, min – minimum value of log*TFP* in a sample, Q1 – first quartile, me – median, m – mean, Q3 – third quartile, max – maximum value of log*TFP* in a sample, sd – standard deviation, IQR – interquartile range.

Source: calculations based on the LPe model.

Table 3

Descriptive statistics of log*TFP* distributions in cross-sections defined by PKD sectors

PKD section	**N**	**min**	**Q1**	**me**	**m**	**Q3**	**max**	**sd**	**IQ**
A	1462	-1.87	3.13	3.51	3.53	3.93	7.18	0.83	0.80
B	3098	-1.42	3.96	**4.39**	4.34	4.77	7.15	0.71	0.81
C	157113	-4.73	3.41	3.81	3.81	4.21	8.35	0.68	0.80
D	5475	-3.21	3.77	4.05	4.21	4.48	9.37	0.76	0.70
E	13438	-1.13	3.59	3.85	3.87	4.12	6.55	0.51	0.53
F	59236	-3.66	3.68	4.06	4.07	4.45	10.15	0.71	0.77
G	182323	-4.17	3.43	3.84	3.90	4.30	9.45	0.75	0.87
H	28480	-4.16	3.75	4.12	4.10	4.46	10.31	0.67	0.71
I	10558	-1.78	3.31	3.65	3.64	4.00	6.27	0.63	0.69
J	14918	-2.53	4.14	**4.61**	4.60	5.06	8.83	0.77	0.92
K	848	0.33	4.13	**4.71**	4.84	5.47	9.84	1.06	1.34
L	22582	-2.92	3.30	3.66	3.74	4.08	8.56	0.74	0.78
M	22202	-3.38	3.99	**4.46**	4.47	4.96	9.26	0.80	0.97
N	16041	-1.14	3.50	4.04	4.07	4.57	9.75	0.85	1.07
O	25	2.80	3.76	4.01	4.12	4.48	5.42	0.62	0.72
P	2360	-3.25	3.29	4.07	3.84	4.69	6.97	1.26	1.40
Q	14403	-1.08	3.74	4.04	4.07	4.36	8.39	0.53	0.62
R	1910	-1.54	3.36	4.00	4.06	4.89	8.48	1.19	1.53
S	1936	-1.34	3.35	3.81	3.79	4.19	6.68	0.75	0.84

Notes:
N – sample size, min – minimum value of log*TFP* in a sample, Q1 – first quartile, me – median, m – mean, Q3 – third quartile, max – maximum value of log*TFP* in a sample, sd – standard deviation, IQR – interquartile range;
A – Agriculture, forestry and fishing, B – Mining and quarrying, C – Manufacturing, D – Electricity, gas, steam and air conditioning supply, E – water supply; sewerage, waste management and remediation activities, F – Construction, G – Wholesale and retail trade; repair of motor vehicles and motorcycles, H – Transportation and storage, I – Accommodation and food service activities, J – Information and communication, K – Financial and insurance activities, L – Real estate activities, M – Professional, scientific and technical activities, N – Administrative and support service activities, O – Public administration and defence; compulsory social security, P – Education, Q – Human health and social work activities, R – Arts, entertainment and recreation, S – Other service activities.

Source: calculations based on the LPe model.

Table 4

Estimation results of the total factor productivity model (8)

	Pooled OLS	**FE (8)**	**RE (8)**	**BE (8)**	**sGMM1 (8)**	**sGMM2 (8)**
	log*TFP*$_{it}$	**log*TFP*$_{it}$**	**log*TFP*$_{it}$**	**log*TFP*$_{it}$**	**log*TFP*$_{it}$**	**log*TFP*$_{it}$**
α_0	0.798***	0.287***	0.587***	0.894***	**0.601*****	**0.631*****
Ownership (base: SOE)						
$\alpha_{1,1}$, *PDE*	0.035***	-0.038***	0.058***	0.014***	**0.124***	**0.068*****
$\alpha_{1,1}$, *FOE*	0.135***	-0.016*	0.227***	0.068***	**0.239*****	**0.232*****
Export intensity (base: low)						
$\alpha_{2,1}$, *moderate*	0.053***	0.014***	0.060***	0.033***	**0.199***	**0.078*****
$\alpha_{2,2}$, *high*	0.052***	0.066***	0.079***	0.041***	**0.261***	**0.076*****
Size (base: medium)						
$\alpha_{3,1}$, *large*	-0.006***	-0.038***	-0.017***	0.017***	0.000	**-0.013*****
$\alpha_{3,2}$, *very large*	0.025***	-0.066***	0.014***	0.043***	0.086	**0.030*****
Concentration (base: low)						
$\alpha_{4,1}$, *moderate*	-0.020***	0.000	-0.017***	-0.007**	0.008	**-0.026*****
$\alpha_{4,2}$, *large*	0.060***	0.080***	0.105***	0.021	**2.073***	**0.138*****
Investment intensity (base: disinvestment)						
$\alpha_{5,1}$, *low*	0.071***	0.073***	0.074***	0.138***	**0.092*****	**0.084*****
$\alpha_{5,2}$, *moderate*	0.007***	0.024***	0.013***	0.040***	0.063	**0.019*****
$\alpha_{5,3}$, *high*	-0.097***	-0.079***	-0.095***	-0.131***	-0.048	**-0.087*****
Investment intensity (*t* – 1) (base: disinvestment)						
$\alpha_{6,1}$, *low*	-0.025***	-0.005***	-0.012***	-0.125***	-0.005	-0.006
$\alpha_{6,2}$, *moderate*	-0.023***	-0.020***	-0.022***	-0.0699***	0.019	**-0.018*****
$\alpha_{6,3}$, *high*	0.023***	-0.039***	-0.011***	0.037***	0.031	0.001
N	458052	458052	458052	458052	458052	458052

Notes: * p-value = 0.05, ** p-value = 0.01, *** p-value = 0.001 in Student's t-tests for $H_0 : \alpha_{i,k} = 0$.

Source: calculations based on the LPe model.

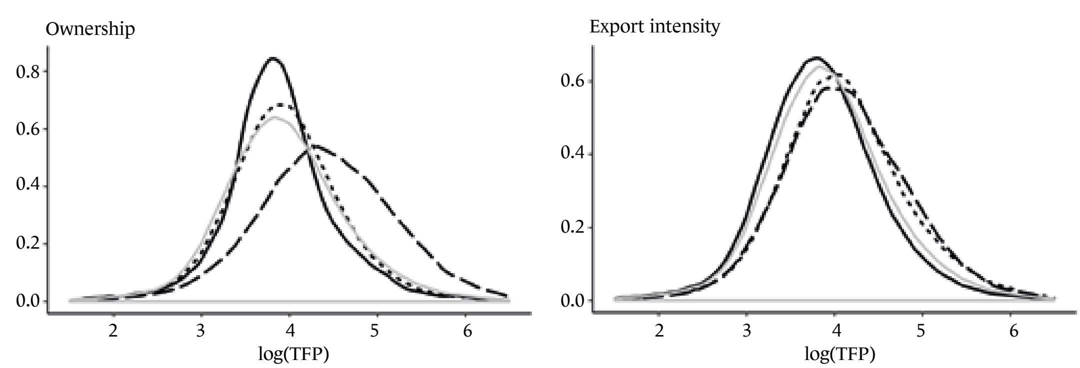
Table 5
Sectoral diversity of TFP indicators

	PKD sectors base: section C								
	A	**B**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
sGMM2, λ_{is}	-0.115***	0.185***	**0.237*** **	0.102***	0.111***	0.043***	0.104***	0.024***	**0.281*** **
	PKD sectors base: section C								
	K	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**
sGMM2, λ_{is}	**0.258*** **	0.069***	**0.219*** **	0.119***	-0.150**	0.084***	0.125***	0.151***	0.022

Notes:
NACE sections are described under Table 3.
* p-value = 0.05, ** p-value = 0.01, *** p-value = 0.001 in Student's t-tests for H_0: $\lambda_{is} = 0$.

Source: estimation of sector effects in the sGMM2 model (8).

Figure 1
Conditional distributions of log*TFP* by categories of ownership and export intensity

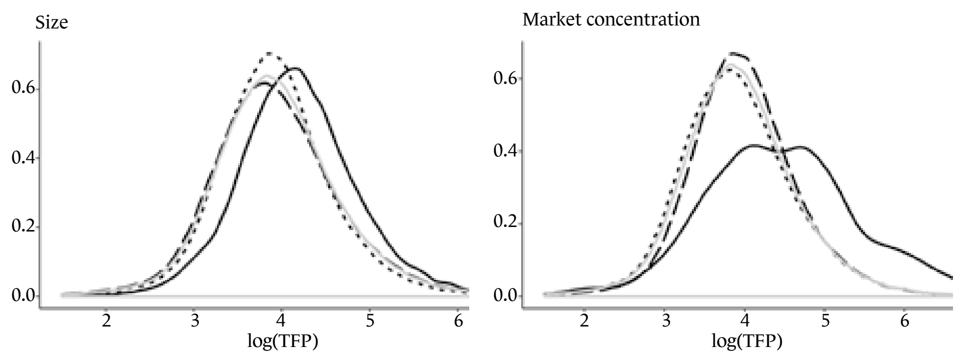


Notes:
Estimation of the empirical density function of log*TFP* using the Gaussian kernel density estimator; grey solid lines – empirical density function for the SP sample.
Left panel: conditional empirical density functions by ownership categories: dashed line – foreign-owned enterprises (FOE), dotted line – private domestic enterprises (PDE), solid black line – state-owned enterprises (SOE).
Right panel: conditional empirical density functions by export intensity categories: dashed line – high export intensity, dotted line – moderate export intensity, solid black line – non-exporters.

Source: calculations based on the LPe model.

Figure 2

Conditional distributions of log*TFP* by categories of enterprise size and market concentration

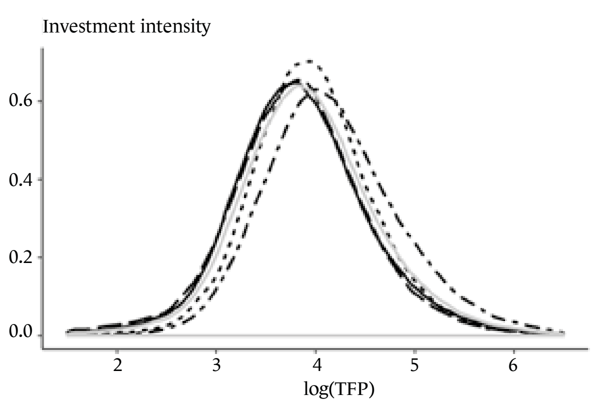


Notes:
Estimation of the empirical density function of log*TFP* using the Gaussian kernel density estimator, grey solid lines – empirical density function for the SP sample.
Left panel: conditional empirical density functions by enterprise size categories: dashed line – medium enterprises, dotted line – large enterprises, solid black line – very large enterprises.
Right panel: conditional empirical density functions by market concentration categories, dashed line – low market concentration, dotted line – moderate market concentration, solid black line – high market concentration.

Source: calculations based on the LPe model.

Figure 3

Conditional distribution of logTFP by categories of investment intensity

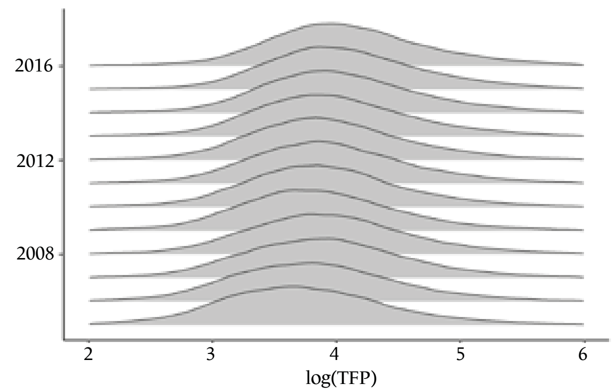


Notes:
Estimation of the empirical density function of log*TFP* using the Gaussian kernel density estimator, grey solid lines – empirical density function for the SP sample; conditional empirical density functions by investment intensity categories: dashed line – high investment intensity, dotted line – moderate investment intensity, dash-dotted line – low investment intensity, solid black line – disinvestment.

Source: calculations based on the LPe model.

Figure 4

Conditional distributions of log*TFP* by years 2005–2016

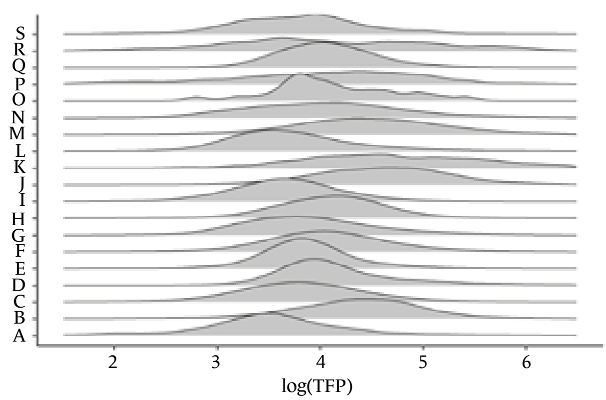


Note: estimation of the empirical density function of log*TFP* using the Gaussian kernel density estimator.

Source: calculations based on the LPe model.

Figure 5

Conditional distributions of log*TFP* by PKD sections



Note: estimation of the empirical density function of log*TFP* using the Gaussian kernel density estimator.

Source: calculations based on the LPe model.

Appendix A. Databases and data pre-processing

The data used in the study originate from the annual reports on the business activity of enterprises for years 2005–2016 reported in the Annual Enterprise Survey of Statistics Poland (SP sample for short). In order to build the final database, it is necessary to:

– combine micro-data from many tables based on the REGON identification number,

– clear data sets from outlier observations and expert imputation of missing data,

– solve the problem of transition in 2008 from the classification of activities PKD-2004 to PKD-2007.

Based on the collected SP sample and attached external variables (e.g. investment, capital and gross-value added deflators) the values of endogenous and explanatory variables are calculated according to formulas in Appendix B.

Imputation and data editing

Expert data cleaning is based on the following rules:

R1: removing from the database firms without positive net revenue from sales of products, goods and materials, as well as firms where altogether material and energy consumption, costs of external services and travel expenses were zero.

R2: companies meeting at least one of the following conditions are removed:

– the value of fixed assets at the beginning and end of the year is zero,

– the average number of full-time equivalent employees is zero,

– the labour cost is zero.

The data was edited manually to replace missing values with zeros for the following variables: business travel expenses, intangible assets, costs of production for own use, value of goods and materials sold, excise tax, value of semi-finished products and production in progress at the beginning and end of the year, stock of finished products at the beginning and end of the year.

As a result, we obtain a consistent firm-level database, cleaned from outliers, containing auxiliary, output and control variables. The dataset used in the study includes over 585 thousand statistical units, constituting 67% of all observations registered in the SP reports for the years 2005–2016. On average, over 48 thousand companies are found in the database annually, in total about 100 thousand companies. The data include 4.6 million employees and annual sales at the level of PLN 1.4 trillion. The SP sample constitutes the majority of the non-financial enterprise sector in Poland; however, it is worth emphasizing that the sample was not selected using the representative method.

Appendix B. Measurement of endogenous and explanatory variables

The company global output is defined as

$$GO = R_P + C_{own} + R_{C\&M} - V_{C\&M} - T + \left(UGO_{12} - UGO_0\right) + \left(FGO_{12} - FGO_0\right) \tag{9}$$

where:

R_p – net revenues from the sale of products (goods and services),
C_{own} – cost of production for own use,
$R_{C\&M}$ – net revenues from sales of commodities and raw materials,
$V_{C\&M}$ – value of sold commodities and raw materials,
T – excise tax,
UGO_0, UGO_{12} – semi-finished products and production in progress – beginning of the year and end of the year,
FGO_0, FGO_{12} – finished goods – beginning and end of the year.

The following formulas give the labour costs (*CL*) and intermediate consumption (*IC*):

$$CL = CL_W + CL_{SS} + BT \tag{10}$$

$$IC = IC_{ME} + IC_S + IC_{other} - BT \tag{11}$$

where:

CL_W – remunerations,
CL_{SS} – social security contributions,
BT – business travel costs,
IC_{ME} – use of raw materials and energy,
IC_S – outside services,
IC_{other} – other costs.

Company gross value added (*Y*) is the difference between its global output (*GO*) and intermediate consumption (*IC*):

$$Y = GO - IC \tag{12}$$

The company physical capital is defined as the average annual level of fixed assets $K = \left(FK_{12} + FK_0\right)/2$. The final measurement of variables Y and K is determined by calculating the real gross value added and real physical capital of the enterprise at constant average prices from 2010. For this purpose, capital and gross-value added deflators at 4 digits PKD sectors are used. The individual productivity coefficients obtained from the LPe model are subjected to a winsorization procedure, where the top and bottom 1% of observations are removed from the sample.

Zastosowanie metody funkcji kontrolnych do pomiaru łącznej produktywności czynników produkcji przedsiębiorstw w Polsce

Streszczenie

Badanie uwarunkowań zewnętrznych i wewnętrznych łącznej produktywności czynników produkcji (TFP) stanowi jedno z głównych zagadnień w ekonomii wzrostu gospodarczego. Celem niniejszej pracy jest pomiar TFP oraz wskazanie głównych determinant łącznej produktywności czynników produkcji dla przedsiębiorstw w Polsce w latach 2005–2016. Ponadto zbadano sektorowe zróżnicowanie produktywności przedsiębiorstw oraz wskazano sektory gospodarki polskiej, w których przedsiębiorstwa osiągają najwyższe wskaźniki łącznej produktywności czynników produkcji. Do estymacji funkcji produkcji zastosowano ekonometryczną metodę funkcji kontrolnych, pozwalającą na zgodną estymację elastyczności wartości dodanej brutto względem nakładów kapitałowych i pracy. Wyznaczono rozkłady wskaźników TFP dla całej badanej próby oraz rozkłady warunkowe TFP względem wybranych uwarunkowań produktywności. Za pomocą dynamicznych modeli panelowych opisano wpływ wybranych determinant produktywności na poziom wskaźnika TFP. Po pierwsze, potwierdzono zależność łącznej produktywności czynników produkcji w przedsiębiorstwie od formy własności, stopy inwestycji, statusu eksportowego oraz wielkości firmy. Po drugie, zauważono sektorowe zróżnicowanie rozkładów TFP oraz ich silną zależność od indeksu koncentracji rynku.

Słowa kluczowe: łączna produktywność czynników produkcji, estymacja funkcji produkcji, metody funkcji kontrolnych, model Levinsohna-Petrina, determinanty TFP