NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Andrei Akhremenko, Alexander Petrov,
Egor Yureskul

TFP ESTIMATES AND FRONTIER-GENERATED PRODUCTION FUNCTION: A NEW DATASET FOR POLITICAL SCIENCE

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: POLITICAL SCIENCE
WP BRP 54/PS/2017

Andrei Akhremenko[1], *Alexander Petrov*[2],

Egor Yureskul[3]

TFP ESTIMATES AND FRONTIER-GENERATED PRODUCTION FUNCTION: A NEW DATASET FOR POLITICAL SCIENCE[4]

There have been numerous studies concerning productivity, representing two general approaches toward measuring the concept. Parametric approaches specify the actual form of the production function based on theoretical assumptions, while non-parametric approaches use empirical best-practice cases as a benchmark for productivity measures. We propose a new approach to obtain cross-country total factor productivity estimates and a method to derive the production function, which builds on empirical data and does not require a priori assumptions about its functional specification. The approach is based on radial model of data envelopment analysis. The obtained TFP estimates are validated through comparing to PWT and UNIDO datasets. Some preliminary analysis is also provided concerning application of the TFP estimates to cross-country divergence/convergence in productivity. We also demonstrate the potential utility of the estimates for political science research.

Keywords: productivity, production function, production possibility frontier, data envelopment analysis

JEL Classification: O47, C14

---

[1] National Research University Higher School of Economics. Laboratory for Qualitative and Quantitative Analysis of Political Regimes, project leader; email: aakhremenko@hse.ru

[2] National Research University Higher School of Economics. Laboratory for Qualitative and Quantitative Analysis of Political Regimes, senior researcher. Keldysh Institute of Applied Mathematics (Russian Academy of Sciences), leading researcher; email: petrov.alexander.p@yandex.ru

[3] National Research University Higher School of Economics. Laboratory for Qualitative and Quantitative Analysis of Political Regimes, project manager; email: eyureskul@hse.ru

[4] This Paper is an output of a research project implemented by Laboratory for Qualitative and Quantitative Analysis of Political Regimes within the Annual Thematic Plan for basic and applied research at National Research University – Higher School of Economics. Project title: "Economic efficiency and the evolution of political regimes: a theoretical model of interdependence, cross-country and dynamical analysis".

1. Introduction

Studies involving total factor productivity (TFP) date back as far as the 1930s when first growth accounting methods were developed [Copeland 1937]. Although researchers agree on the concept of TFP not being "deeply theoretical" [Hulten 2001], it still provides a basic conceptual framework for measuring economic development which adds a lot to common growth indicators such as current prices GDP or GDP per capita. Initial attempts at measuring TFP were undertaken in the form of the output per unit input index [Copeland and Martin 1938] which introduced the scaling factor into the fundamental GDP accounting formula, allowing researchers to estimate the role that capital and labor productivity changes played in economic growth. The seminal work by Robert Solow further expanded the concept of TFP through the analysis of an aggregate production function with a Hicks-neutral shift parameter and constant returns to scale. The so-called "Solow residual" produced by this method measures the residual growth rate of output not explained by input increases, i.e. total factor productivity growth [Solow 1956]. By design, Solow's measure of productivity change is conceptually equivalent to a "measure of ignorance" [Abramowitz 1956] in that it captures all the factors contributing to total factor productivity as well as possible measurement error.

Building on Solow's work, productivity research has expanded in two general directions. The first branch of studies is usually called the parametric approach: within this direction, researchers make assumptions about the production function's parametric form and estimate all of its parameters using econometric methods. The inherent value of parametric productivity analysis lies in the ability to break down the abstract notion of TFP into its key components: the increase in input efficiency, technological innovation and measurement error [Prucha, Nadiri 1981]. Modern applications of the parametric approach include methods such as Stochastic Frontier Analysis that uses a pre-defined functional form to estimate a production, cost or profit frontier that serves as a best-practice model for measuring inefficiencies [Battese, Coelli 1995; Battese, Prasada Rao 2002]. Despite its advantages (such as being able to incorporate random error into generating the best-practice frontier), the parametric approach forces the researcher to impose unnecessary restrictions on function parameters and brings about several unwanted sources of bias [Hulten 2001].

The non-parametric approach originates from successful attempts by researchers to adapt Solow's conceptual framework to discrete-time data using distance functions in order to calculate productivity indices. The latter technique includes Tornqvist indices as approximations for continuous translog production functions [Diewert 1976] or less data-demanding Malmquist indices to build best-practice production frontiers [Caves, Christensen, Diewert 1982; Fare et al 1994]. The aforementioned frontier approach has promoted an entire class of non-parametric methods commonly referred to as "envelopment methods" which include, among others, the popular Free Disposal Hull and Data Envelopment Analysis [Coelli et al. 2005].

While economists use distance function indices to produce macro-level TFP estimates [PWT 2013; Issaksson 2007], the frontier approach has been widely used for productivity analysis at all levels from firms to nations, both in private and public sector [Milliken et al. 2011; De Witte, Geys 2011; Mahmood 2012].

The underlying assumption within the frontier approach is as follows: the most productive firms [called "Decision-Making Units" or DMUs to stress the universal nature of the approach] in the analyzed dataset – that is, the ones with the best output-to-input ratios – form the best-practice production possibility frontier for other firms. Linear programming techniques are then used to calculate distance functions between less efficient DMUs and the frontier [Banker Charnes Cooper 1984]. From this point on, we'll focus on Data Envelopment Analysis (DEA) when talking about the frontier approach since it is one of the most commonly used techniques in non-parametric productivity analysis.

From both the conceptual and the practical standpoint, frontier methods add a lot to productivity research. First and foremost, they do not require as many assumptions on the researcher's part as the parametric approach: most methods have to at least assume the actual parametric form of the production technology (like a Cobb-Douglas or Leontief production function). Still, there is no sound theoretical basis for picking one functional form over the other: the conventional choice of the Cobb-Douglas production function seems to be based rather on the utility of its mathematical form than on a solid theoretical ground.

In contrast, frontier methods operate with a conceptually transparent idea of best-practice empirical input-output sets which, in turn, allow the researcher to further break down the TFP measures into key components [Battese et al. 2002, Arora 2013]. Second, the only data DEA and similar techniques require are quantities of inputs and outputs: price data is not necessary for obtaining efficiency estimates (although the basic DEA model may be expanded to include relative input prices, see [Coelli et al. 2005]). In addition, the basic DEA model may be expanded to obtain statistical inference from productivity estimates [Ray 2002; Simar, Wilson 2007], to incorporate a stochastic component in the form of an error measure [Kenneth et al. 1994] or to account for frontier shifts over time [Ray, Desli 1997].

It is worth noting that the non-parametric frontier approach is data-sensitive: since empirical data points shape the best-practice frontier, adding or removing observations may completely alter both the frontier's shape and the resulting TFP estimates, thus reducing the utility of DEA for measuring TFP change. However, clear research design allows the researcher to formulate proper DEA specifications that avoid such bias.

From a theoretical standpoint, non-parametric frontier approaches bolster the conceptual scope of TFP due to the lack of production function specification. This may be viewed as a problem by economic theorists, but for political science it sometimes is the only way to estimate productivity measures that are inherently related to popular research subjects such as institutions, regimes and social capital.

While not a major strand in contemporary political science literature, studies connecting TFP to institutional development do exist. It is probably considered common knowledge by now that property rights protection is the basic determinant of market efficiency. Hall and Jones [Hall and Jones 1999] point out that a number of political and social institutions may lead to increased productive activity through fair allocation of market resources to efficient firms (i.e. economic policy), as well as by suppressing private and state diversion (i.e. reducing corruption and theft). Government effectiveness has also been viewed as a determinant of economic productivity in many studies, particularly by economic organizations such as OECD.

Our results show, however, that productivity can be something more than a dependent variable. TFP appears to be an important prerequisite of a successful institutional change; the rise in productivity may enforce cooperation and social trust [Akhremenko, Petrov, Yureskul 2017]. Mathematical simulations have shown that TFP may strongly affect policy decisions available to governments under different regime types [Akhremenko, Petrov 2014] and be a factor of sustainable economic development in the case of economic retrospective voting [Akhremenko et al. 2015].

All in all, a reliable estimate of TFP is needed on the first place.

In this paper, we propose an approach to frontier methods in measuring total factor productivity different from most popular estimation techniques based on index numbers. The basis of the methodology is an early DEA model [Farrell,1957] that uses input/output ratios instead of pure input and output quantities. While imposing constant returns to scale restriction on the production possibility frontier, the specification we use reduces the dimensionality of the model and allows for illustrative graphical visualization in the case of two inputs and one output.

However, the main methodological contribution of this paper is as follows: by obtaining the vertices of the best-practice frontier through DEA, we are able to generate a production function specification. The proposed method approaches productivity research from a different angle: instead of specifying the production function beforehand or foregoing the specification altogether, we specify the production function's form after estimating the best-practice frontier, thus making it frontier-generated. This is in contrast to most studies that exogenously specify the production function's form (in practice, using the Cobb-Douglas function in most cases) and use empirical data only to estimate its parameters.

The frontier-generated production function (hereafter referred to as FgPF) on the other hand, is unambiguously specified based on empirical data. This broadens the scope of research questions that can be answered by analyzing TFP changes: instead of simply measuring TFP [Rao 1996] or its connection with exogenous factors [Veeramani, Goldar 2004], we can measure the rate of technical substitution, as well as classify DMUs according to production factor deficiencies. In addition, we can estimate the direction of technical progress for a given dataset by comparing FgPF isoquants.

In a manner of speaking, production functions widely used is economic theory are smooth and meet Inada conditions, and our FgPF can be viewed as a piecewise linear approximation to such function .

The productivity dataset we provide here possesses a number of distinct advantages for political science research. First and foremost, our estimation technique allows us to create a cross-section-time-series dataset, something that the index number approach (such as Tornqvist quantity indices) cannot accomplish. Second, in this paper we show that our TFP estimates demonstrate strong statistical relationships with key social, political and institutional indicators which would allow researchers to use our measures of TFP as explanatory and control variables in further studies. Moreover, there are strong indications that TFP estimates may serve as integral indicators of overall institutional and social effectiveness (quality) of "state-society" systems. In contrast to survey-based measures and expert estimates of many "political" variables available to researchers, we obtain "hard data" from macroeconomic indicators.

The rest of the paper is organized as follows.

In Section 2 we introduce the method for calculating TFP estimates with an overview of the data on inputs and outputs. In Section 3, the frontier-generated production function is presented. Section 4 provides basic descriptive statistics for the TFP estimates produced. In Section 5, we compare our TFP measures with existing TFP datasets. In Section 6, we employ clustering methods to break down the country sample into two groups according to productivity measures. In Section 7, we look at TFP estimates in the context of institutional and social capital variables. Section 8 serves to conclude. The TFP estimates for the selected country sample are provided in Appendix A.

2. Method

2.1 Basic productivity model

A country's output Y is supposed to be produced by employing two factors, namely capital K and labor L. The specific form of the production function $Y = F\left(K, L\right)$ is generated by DEA and will be constructed in Section 5. We also assume constant returns to scale, so $F\left(\lambda K, \lambda L\right) = \lambda F\left(K, L\right)$.

To obtain cross-country TFP estimates, we use DEA, which takes the particularly simple form under these assumptions [Farrell,1957].

Taking the data for a certain year, we represent each country as a point in the $\left(L/Y, K/Y\right)$ plane. The enveloping broken line (i.e. the empirically constructed best-practice frontier) indicates the position of the countries that are most productive for a given proportion of factors. These countries are marked C, D, E in Fig.1, which serves to illustrate. The TFP value for these countries is taken as unit: $A_C = A_D = A_E = 1$. For countries behind the frontier, we have TFP $A < 1$. For example, the TFP value of country B is calculated as $A_B = \left|OB_f\right| / \left|OB\right|$.

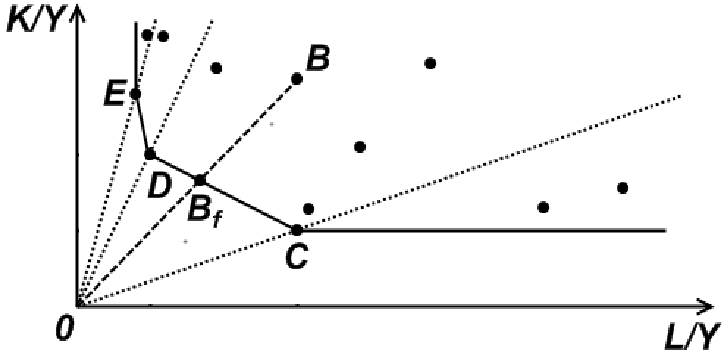

Figure 1. The calculation of TFP

When dealing with successive years, we use two variations of DEA. The first of them is LMDEA (Long-memory DEA), developed by Forstner and Isaksson, 2002. The LMDEA model presumes global TFP change to be non-negative and therefore allows the frontier to move only in west and south directions. Technically, this means that for every year (except for the first year in the dataset, i.e. 1990), the set of current-year country data is supplemented with the previous-year data for countries that shaped the frontier in the previous year.

As a result, for a given year, some countries may be represented by several points on the diagram. The extreme case is given by Equatorial Guinea which grew rapidly and almost proportionally in capital and output, thus shaping the south part of the frontier – Fig.2.

The second variation we used considers each year by itself. In our TFP database this variation is referred to as PLDEA (PlainDEA). For any fixed year and country, the PlainDEA score is not greater than LMDEA score. Both PlainDEA and LMDEA scores increase for those countries which productivity increases faster than that of the frontier countries with similar capital-labor ratio.

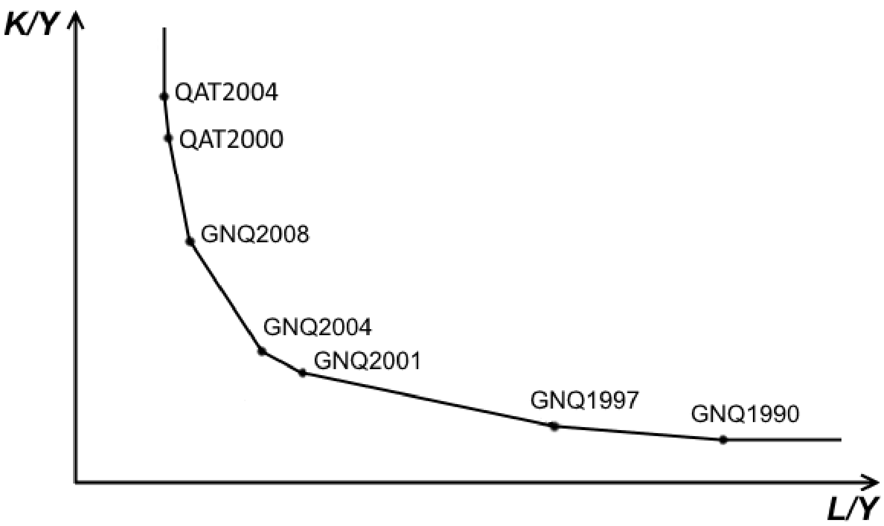

Figure 2. The frontier in 2013, LMDEA (this figure is not drawn to scale). Marks like QAT2004 and GNQ1990 refer to Qatar,2004 and Equatorial Guinea,1990

2.2 Data sources and design of cross-country comparisons

We obtain Y and K estimates from IMF data[5]:

- *Y: "*General government capital stock, PPP, Constant 2005 Dollars";

[5] http://www.imf.org/en/Data

- K_g (public capital): estimated on the basis of the "General government capital stock, percent of GDP" variable: $100K_g / Y$;

- K_p (private capital): estimated on the basis of the "Private capital stock, Percent of GDP" variable: $100K_p / Y$.

Total capital is assumed to be a sum of private and public capital: $K = K_g + K_p$.

To estimate L, we use the World Bank's data for "Labor force, total".

The goals for the empirical study in this paper are as follows:

- A comprehensive validity check for the proposed TFP estimates, including both internal consistency and correlations with existing TFP measures (Penn World Tables, UNIDO World Productivity Database).
- Analysis of structural and dynamical patterns that may be captured by the proposed TFP estimates, but not any other existing measurement technique;
- A clustering of world's economies according to TFP levels and TFP growth.

The country sample varies from 119 countries for 1990 to 127 countries for 2013, due to the availability of data. The list of countries for each year in the sample, together with corresponding TFP estimates, is provided in Appendix A.

The empirical study provided two types of TFP estimates: long-memory DEA estimates accounting for best-practice frontier shifts in the previous years (hereafter dubbed LM estimates); and PLDEA estimates without accounting for previous best-practice countries (hereafter dubbed PL estimates).

3 Frontier-generated Production Function

In this section we proceed to the derivation of FgPF. Figure 3 serves to illustrate. The straight lines drawn through point O and each of the points C,D,E (i.e. countries that shape the frontier) separate the quadrant $K > 0, L > 0$ into several areas. We number them counterclockwise starting from 0.

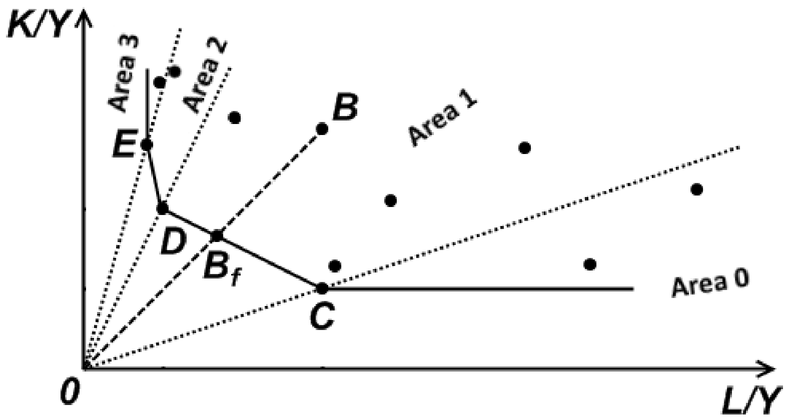

Figure 3. The derivation of Frontier-generated Production Function

The equation of the frontier in area DOC is $a_1 L / Y + b_1 K / Y = 1$ where values of a_1 and b_1 can be easily calculated from the coordinates of points C, D (the index "1" refers to the area number). Hence, for any country on this part of the frontier (namely, for C, D and countries between them, if any) we have $Y = a_1 L + b_1 K$. This is the FgPF form for these countries.

In order to derive FgPF form for country B, which falls behind the frontier, consider the "virtual country" B_f that has the same amounts of resources L_B, K_B as country B, and whose

TFP is $A_{Bf} = 1$. This means that $Y_B = A_B Y_{Bf}$. Obviously, the output of country B_f is given by $Y_{Bf} = a_1 L_B + b_1 K_B$. Therefore, $Y_B = A_B\left(a_1 L_B + b_1 K_B\right)$.

This consideration is applicable to any area of the diagram. It should be noted that $a_0 = 0$ and $b_m = 0$ (here m+1 is the number of areas).

So the FgPF of country i which falls within area s takes the form

$$Y_i\left(L_i, K_i\right) = A_i\left(a_s L_i + b_s K_i\right) \tag{1}$$

Here A_i is country-specific TFP, and a_s, b_s are area-specific coefficients.

As for the countries that fall on the lines separating the areas, two alternative formulae may be written. For definiteness we assume that each area contains its south-east border.

So the formal definition of Frontier-generated Production Function is as follows. Let the frontier be shaped by countries $C_1\left(L_1, K_1\right), C_2\left(L_2, K_2\right), \ldots, C_m\left(L_m, K_m\right)$, numbered from south-east to north-west. Then the equation of the frontier is given by

$$\begin{array}{ll}
b_0 \dfrac{K}{Y} = 1, & if\ 0 \le \dfrac{K}{L} < \dfrac{K_1}{L_1}, \\[2ex]
a_1 \dfrac{L}{Y} + b_1 \dfrac{K}{Y} = 1, & if\ \dfrac{K_1}{L_1} \le \dfrac{K}{L} < \dfrac{K_2}{L_2}, \\[2ex]
\multicolumn{2}{c}{\ldots\ldots\ldots\ldots\ldots\ldots} \\[2ex]
a_{m-1} \dfrac{L}{Y} + b_{m-1} \dfrac{K}{Y} = 1, & if\ \dfrac{K_{m-1}}{L_{m-1}} \le \dfrac{K}{L} < \dfrac{K_m}{L_m} \\[2ex]
a_m \dfrac{L}{Y} = 1, & if\ \dfrac{K_m}{L_m} \le \dfrac{K}{L}
\end{array}$$

and FgPF is given by

$$Y\left(L, K\right) = \begin{cases}
A b_0 K, & if\ 0 \le \dfrac{K}{L} < \dfrac{K_1}{L_1}, \\[2ex]
A\left(a_1 L + b_1 K\right), & if\ \dfrac{K_1}{L_1} \le \dfrac{K}{L} < \dfrac{K_2}{L_2}, \\[2ex]
\ldots\ldots\ldots\ldots\ldots\ldots & \\[2ex]
A\left(a_{m-1} L + b_{m-1} K\right), & if\ \dfrac{K_{m-1}}{L_{m-1}} \le \dfrac{K}{L} < \dfrac{K_m}{L_m} \\[2ex]
A a_m L, & if\ \dfrac{K}{L} > \dfrac{K_m}{L_m}\ \ (or\ L = 0)
\end{cases}$$

Here A is a country-specific TFP.

The properties of FgPF are as follows:

1. It follows from the continuity of the frontier that $Y\left(L, K\right)$ is a continuous function of two variables

2. $Y\left(L, K\right)$ is piecewise linear and both first-order partial derivatives are non-negative and piecewise constant functions

3. $Y\left(0, K\right) = Y\left(L, 0\right) = 0$

4. $Y(\lambda L, \lambda K) = Y(L, K)$ for any $\lambda > 0$

5. $Y(L, K)$ is concave (but not strictly concave)

6. If $K = K_0 = const > 0$, then $Y(L, K_0)$ is non-strictly increasing, concave (but not strictly concave), and $\lim_{L \to \infty} Y(L, K_0) = Ab_0 K_0$ (Fig.4). Similarly, if $L = L_0 = const > 0$, then $Y(L_0, K)$ is non-strictly increasing, concave (but not strictly concave), and $\lim_{K \to \infty} Y(L_0, K) = Aa_m L_0$.

7. Isoquants $Y(L, K) = const$ are broken lines like those shown in Fig.5. Those isoquants are geometrically similar to the frontier but they are drawn in the (L, K) plane instead of the $(L/Y, K/Y)$ plane.

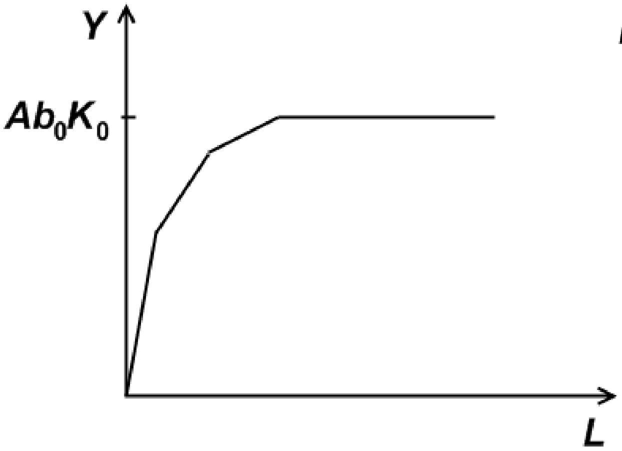

Figure 4.FgPF at $K = K_0 = const > 0$

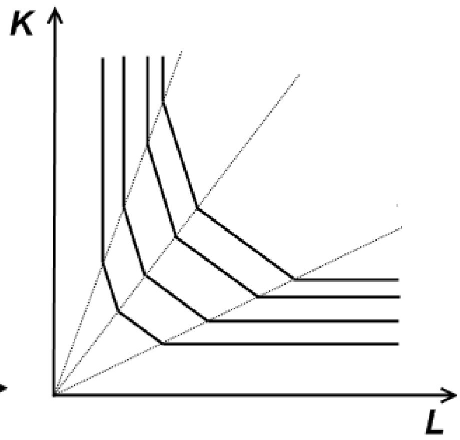

Figure 5. Isoquants of FgPF

It is clear from Fig.5 that FgPF is similar to the Leontief production function in the border areas and to the linear production function in the inner areas.

In case of factor imbalance, the deficient factor becomes more significant for the production process. If factor imbalance is so prominent that the country falls within one of the two areas bordering the axes, then the output is determined solely by the deficient factor.

To summarize, the properties of FgPF are similar to Inada conditions [Inada 1963; Uzawa 1963], but do not satisfy them to the full. However, the FgPF functional form is defined not through theoretical assumptions, but as a result of empirical estimates of the best-practice frontier. This allows to avoid methodological problems associated with parametric approaches (such as negative elasticity in the Cobb-Douglas function).

4 Overview of cross-country estimates

Estimates obtained with LMDEA and PLDEA demonstrate high correlations: Pearson coefficients for each cross-section in the data are 0.94 or higher and highly significant (p< 0,001). Therefore, relative TFP values for most countries are identical regardless of the DEA model used. However, for a number of countries the LMDEA model provides lower TFP estimates than the PLDEA model. This is due to the fact that the LM frontier either lies further from most data points than the PL frontier or is in line with it.

The basic structure of TFP distribution remains the same for both LM and PL estimates: it is bimodal, suggesting a mixture of two normal distributions (see Fig. 6a,b). Therefore, we can clearly define two groups of countries according to their TFP levels: “forerunners” and "catching-up". This is one of the key patterns recognizable through our measurement method; this pattern will be looked into in more detail in Section 5.

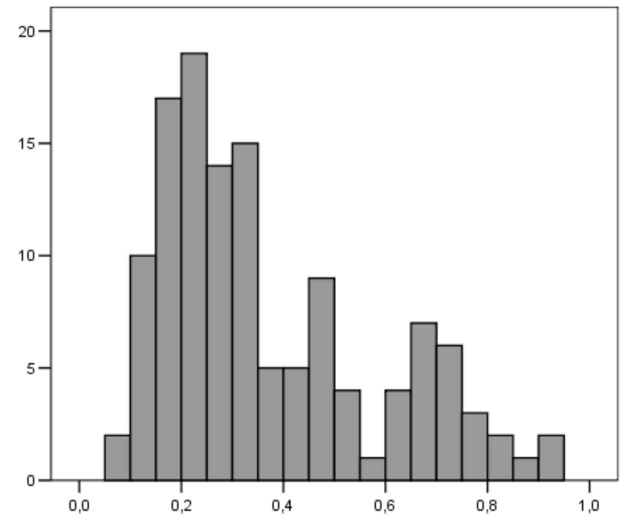

Figure 6a. LM-estimates distribution in 2009

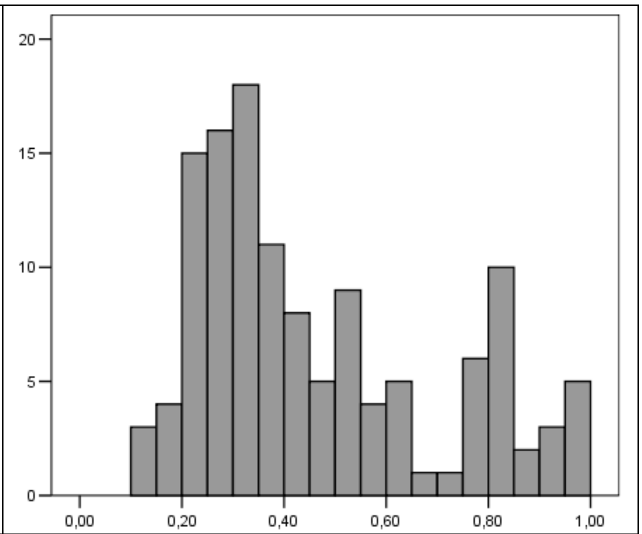

Figure 6b. PL-estimates distribution in 2009

Basic theory behind TFP suggests that it should change at a slow rate, without drastic "leaps". We expect TFP frequency distribution to change slowly from year to year, and therefore to demonstrate high positive correlations between years. Indeed, correlation coefficients are 0.98 or higher and highly significant (p<0,001) for both LM and PL estimates. TFP consistency over time is further backed by autocorrelation analysis for average LM estimates over each year: the autocorrelation function demonstrates a single significant peak at lag number +1. Thus, TFP change is a good fit for a first-order autoregression function $TFP(t) = \theta TFP(t-1) + \varepsilon(t)$.

5 Validity check

In this section we compare our TFP estimates with existing estimates produced by different methods. From the few existing datasets we have chosen those that estimate TFP for similar country and year samples, were produced by respectable organizations and demonstrate drastically different approaches to measuring TFP: Penn World Tables 8.1 estimates (PWT[6]) and World Productivity Database estimates (WPD[7]).

Our statistics coincide with the PWT dataset in 88 countries and 22 years (1990-2011), and with the WPD dataset in 88 countries and 11 years (1990-2000), which should be enough for a full comparison. The PWT dataset uses a parametric production function specification; the WPD dataset uses DEA, although a different set of production factors and a different model for estimating the best-practice frontier.

The following variables have been chosen for cross-country analysis (Table 1):

PWT: CTFP – TFP levels at current PPP prices; TFP level for United States is treated as 1.

WPD: TFP_K06_US – TFP levels (with no unit of measurement);TFP level for United States is treated as 1.

We compare these variables to both LM and PL estimates which we transform to satisfy the condition of TFP(US) = 1.

[6]Robert C. Feenstra, Robert Inklaar, Marcel P. Timmer The Next Generation of the Penn World Table VOL. 105, NO. 10, OCTOBER 2015 (pp. 3150-82). Available at www.ggdc.net/pwtиwww.internationaldata.org

[7] Isaksson A. World Productivity Database: a Technical Description. UNIDO, Working paper 10/2007. Vienna, 2007. https://www.unido.org/fileadmin/user_media/Publications/documents/world_productivity_database_technical_description.pdf. Data available at https://www.unido.org/data1/wpd/Index.cfm

For dynamic comparison we use only PWT estimates, namely the RTFPNA variable (TFP at constant prices, with 2005 TFP levels treated as 1). We compare these data to our LM estimates, also transformed to TFP(2005) = 1. We don't use WPD estimates for comparison, since 11 coinciding time points are not enough for most statistical methods.

Table 1 TFP estimate comparisons

Project	Variable	Description	Years	Coinciding country cases	Estimates for comparison	Comparison type
PWT	*CTFP*	TFP level at current PPPs (USA=1)	1990-2011	88	LM, PL	Cross-country
	RTFPNA	TFP at constant national prices (2005=1)	1990-2011	88	LM	Dynamical
WPD	*TFP_K06_US*	TFP level (USA = 1)	1990-2000	88	LM, PL	Cross-country

For cross-country comparisons our hypothesis was as follows: our TFP estimates will demonstrate high correlations with existing estimates (due to measuring the same concept), yet the relationship will be far from functional, which will let us differentiate our estimates from existing ones. The results confirm this hypothesis: Spearman correlations between our TFP estimates and both PWT and WPD estimates are highly significant and range from 0.7 to 0.9 over the years (Fig. 7).

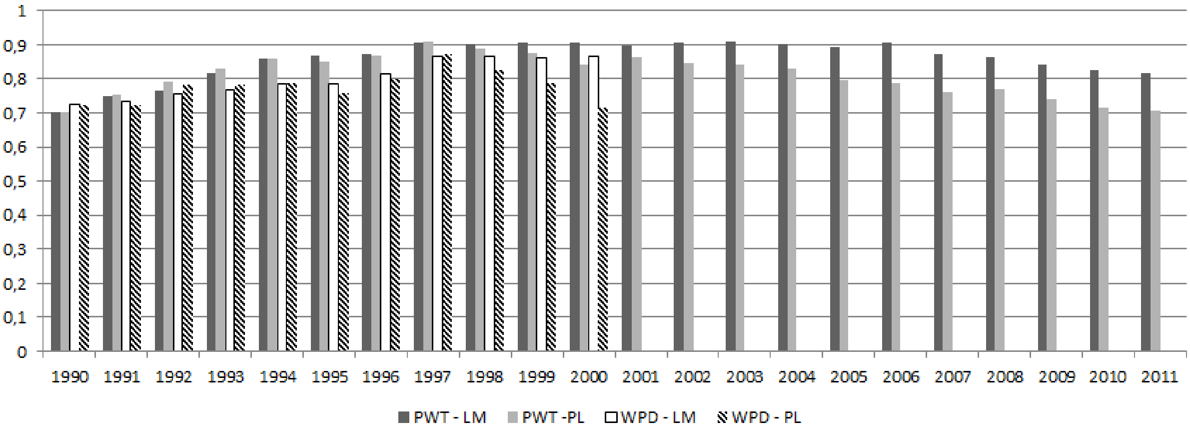

Figure 7.Correlations between existing TFP estimates and proposed estimates. All correlations are significant at p < 0,001

Descriptive statistics show the differences between TFP estimates produced by our method and other projects. It is worth noting that average and median TFP estimates generally lie between PWT and WPD values, the former being higher on average and the latter being lower on average (Fig. 8a,b). In addition, average estimates are closer to PWT, while median estimates are closer to WPD.

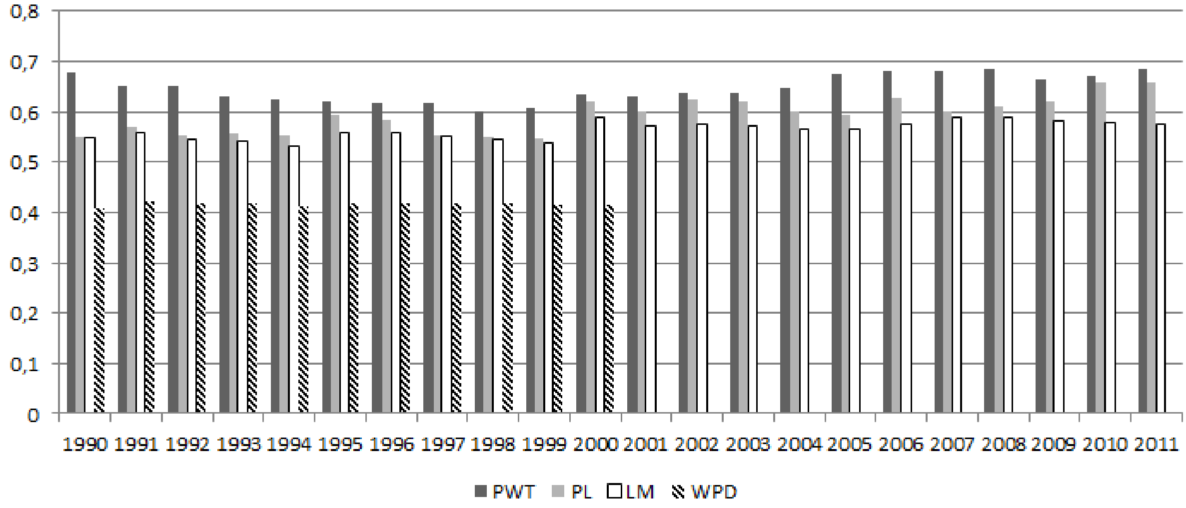

Fig. 8a. Average TFP estimates

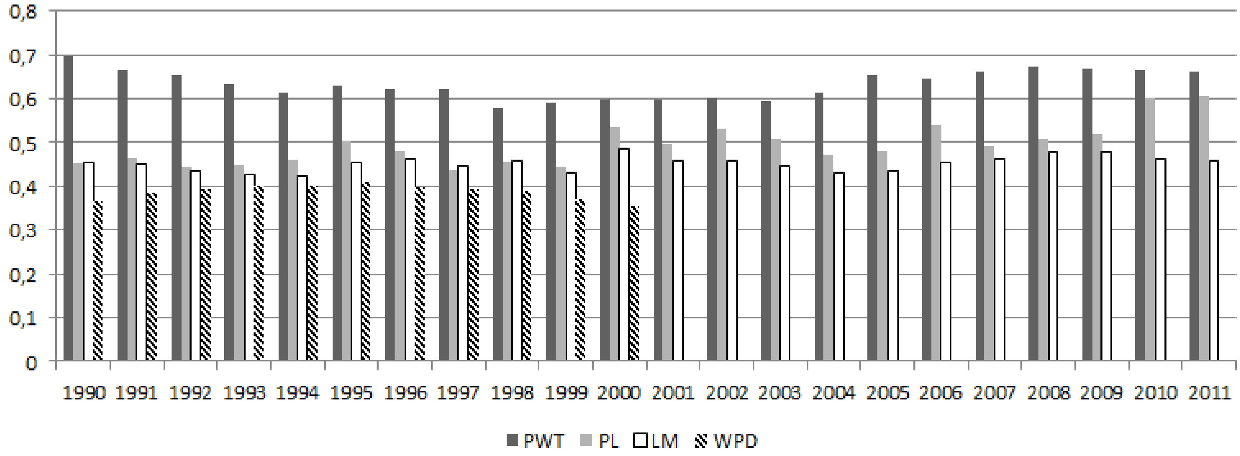

Fig. 8b. Median TFP estimates

The observed differences between TFP estimates may partly be explained by their frequency distributions: PWT estimates are close to being normally distributed (Fig 9a), which is due to the method employed by PWT; WPD estimated demonstrate strong asymmetry and are close to being log-normally distributed (Fig 9b). Our estimates demonstrate bimodal distributions with a more massive right component (Fig. 9c)

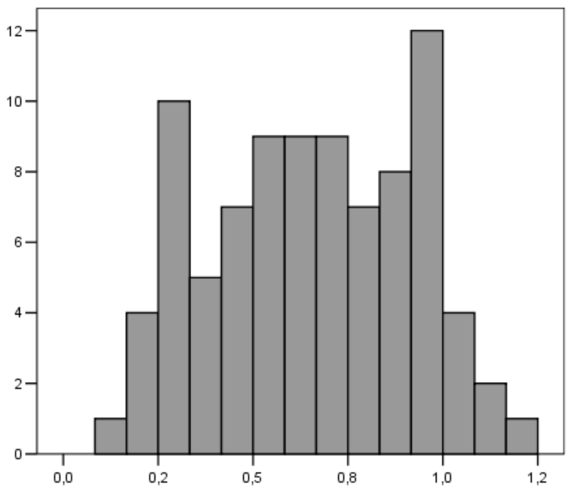

Figure 9a Typical PWT estimate distribution

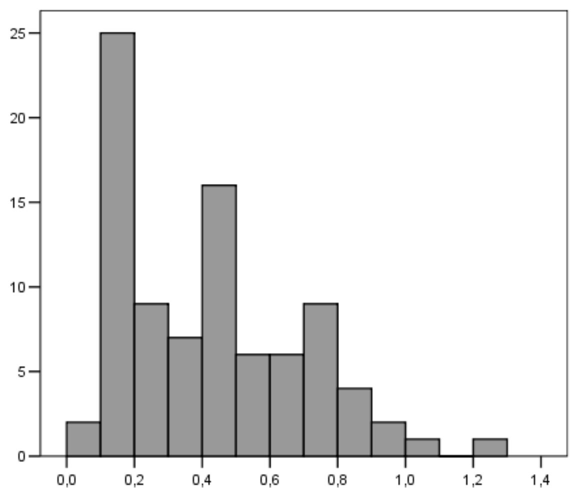

Figure 9b Typical WPD estimate distribution

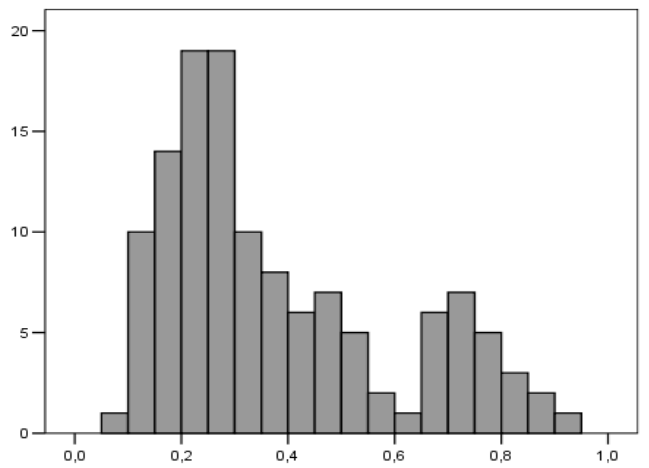

Figure 9c Typical LM/PL estimate distribution

The differences in estimate distributions are quite significant and point to certain assumptions behind different approaches to measuring TFP:

PWT: the country sample is homogenously and normally distributed, which means most countries have TFP close to sample average, with high and low TFP cases being rare.

WPD: the country sample is homogenously distributed, with most countries demonstrating low TFP, high TFP cases are rare.

LM/PL: the country sample is heterogeneous, with two distinct groups of productivity forerunners and catching-up.

Distribution heterogeneity in LM/PL estimates will be further analyzed below.

Preliminary comparative analysis of TFP dynamics between PWT estimates (the RTFNA variable) and LM estimates was done by looking at average estimate dynamics. On the one hand, average TFP variation over time is similar for both estimates. The cross-correlation function (CCF[8]) for the two time-series demonstrates a single significant peak at time lag number 0 (Table 2, Fig 12).

Table 2 LM/PWT cross-correlation

Lag	Cross Correlation	Std.Error
-7	-0,102	0,267
-6	-0,186	0,258
-5	-0,115	0,250
-4	-0,063	0,242
-3	-0,292	0,235
-2	-0,014	0,229
-1	0,248	0,223
0	**0,553**	**0,218**
1	0,149	0,223
2	0,018	0,229
3	0,162	0,235
4	0,087	0,242
5	0,045	0,250
6	0,300	0,258
7	0,178	0,267

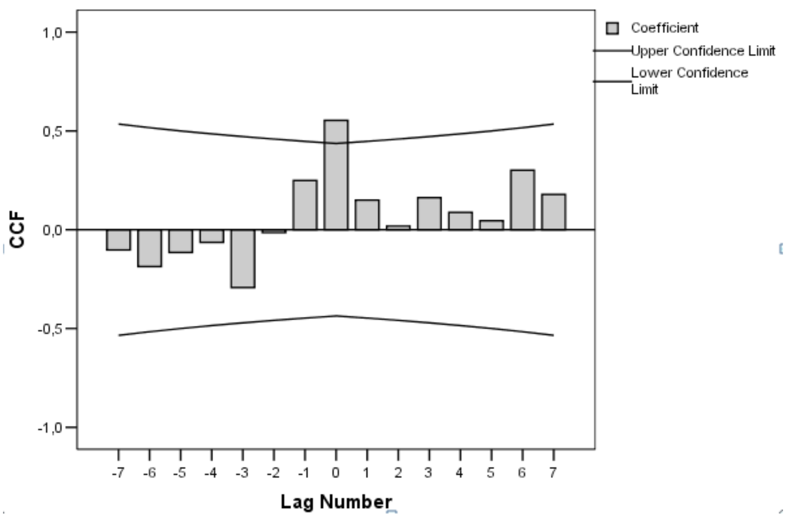

Fig. 10 LM/PWT cross-correlation

8 The CCF was calculated using first differences to remove the trend.

On the other hand, dynamic trends are somewhat dissimilar: the differences are clear during the time period between 1990 and 2001. PWT estimates demonstrate a non-monotonous dynamic overall (decline in the 1990s with consequent growth), while LM estimates show an almost linear growth (Fig 11).

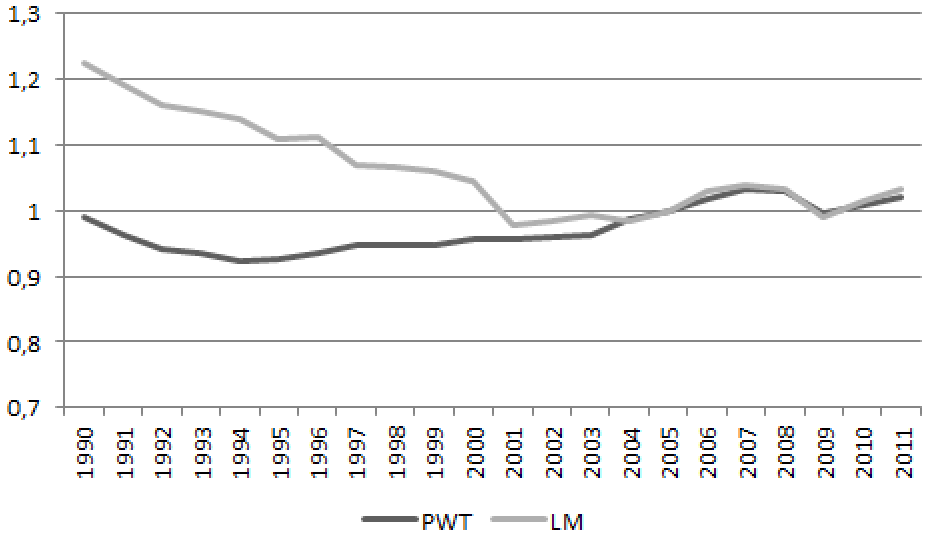

Figure 11: Average PWT and LM estimate dynamics

All in all, the TFP estimates proposed in this paper correlate well with existing TFP measures. The differences in both frequency distributions and average dynamics are due to differences in the estimation method.

6. TFP distribution across countries

As mentioned earlier, one of the most salient features of our TFP estimates is the heterogeneity of the distribution. That is, there are two distinct groups (“forerunners” and “catching-up”) among countries of the world, which in turn raises multiple questions worth investigating. Are the shares of the two groups constant over time? Do countries move from one group to another and if they do, how often?

We employ cluster analysis to answer these questions. We apply K-means clustering and statistical mixture models to both both PL and LM estimates for each year in the dataset. For the K-means model we set the number of clusters to two (K = 2). For the mixture model the number of mixture components was determined during the computation process. The mixture algorithm has successfully identified two groups within the LM estimates for each of the 24 years in the dataset (1990-2013) and for 22 years out of 24 in the case of PL estimates. This serves as strong evidence in favor of the two-group structure in country productivity statistics.

Since mixture models assign group membership probabilities for every observation and we only have two groups, we only look at cases that have high ($p > 0.5$) probability to belong to the “forerunner” group.

As a result, we produce two types of cluster membership statistics (K-means and mixture) for two types of TFP estimates (LM and PL).

Fig. 12 shows the percentage of countries assigned to the group of “forerunners” for each type of clustering statistics and each year in the dataset.

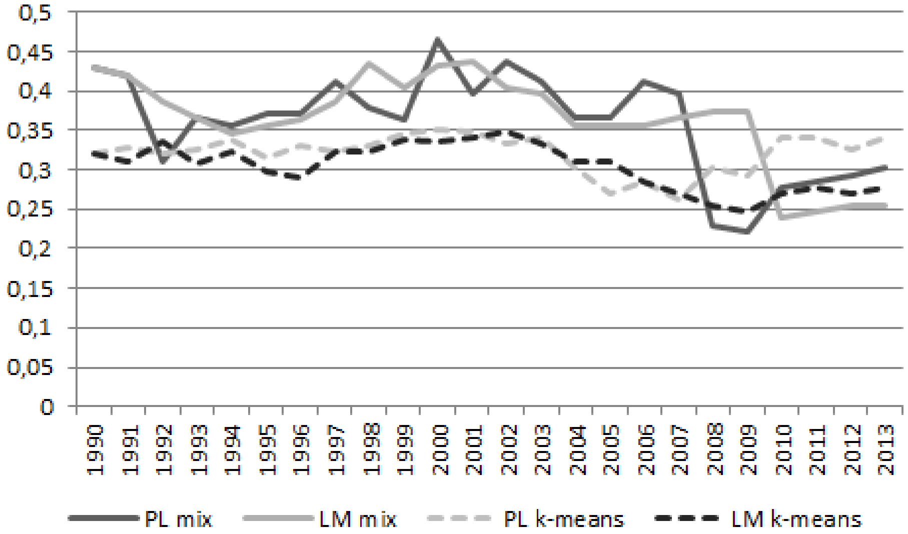

Figure 12. Share of the countries in the “forerunners” cluster for both LM and PL estimates according to k-means and mixture clustering

Three of the four clustering procedures capture the steady reduction in the percentage of “leaders” in the period between 1990 and 2013. For LM estimates and mixture models, the share of “leaders” is 1.7 times lower in 2013 than in 1990; for PL estimates and mixture the share is 1,4 times lower; for K-means clustering and LM estimates it is 1.2 times lower.

This tendency for “forerunners” share reduction is easily observable by comparing group membership probabilities for 1990 and 2013 (Fig. 13a,b). The reduction in the number of countries clearly defined as “forerunners” (membership probability close to 1) is accompanied by an increase in the density of the “catching-up” cluster. The same process may be observed by comparing LM estimate distributions for 1990 and 2013 (Fig. 13c,d): the heterogeneity in the distribution becomes more pronounced over time.

Therefore, the TFP data suggest that the divergence in productivity between groups of countries is growing over time, mostly due to “transitional” countries moving toward the “catching-up” group.

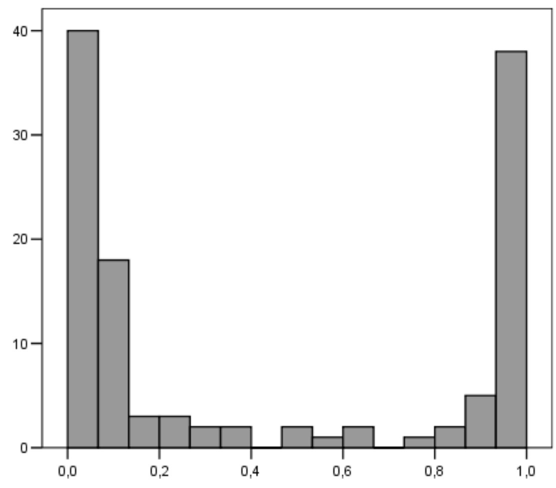

Figure13a. Probability distribution for “forerunners” membership in 1990

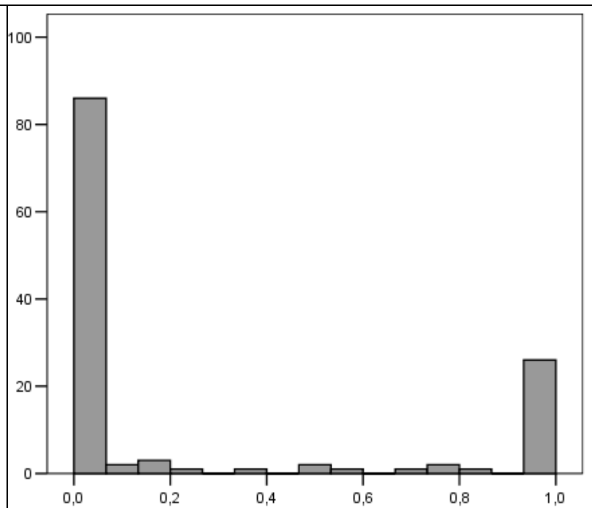

Figure13b. Probability distribution for “forerunners” membership in 2013

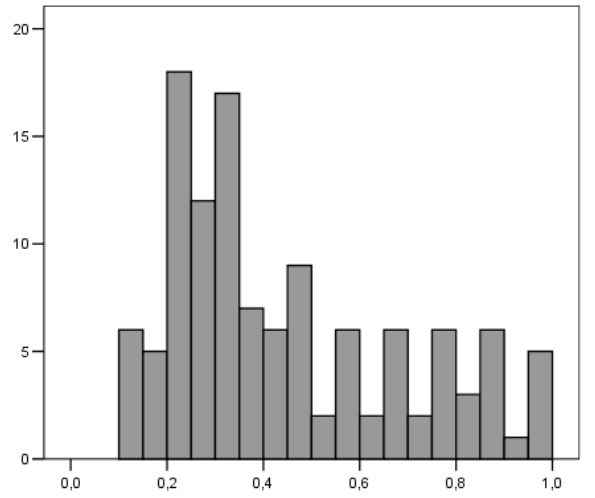	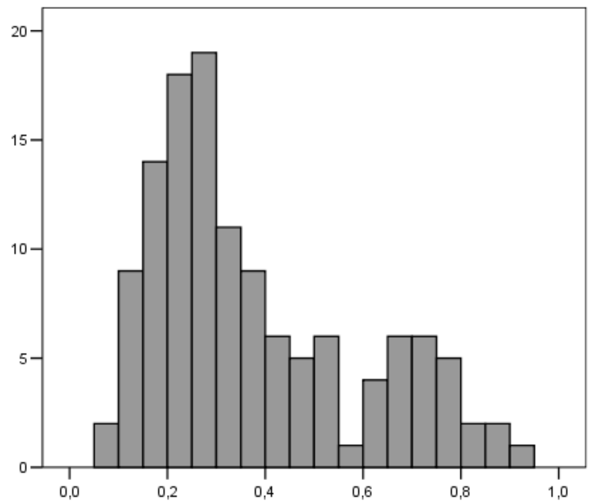
Figure13c. LM estimate frequency distribution in 1990	Figure13d. LM estimate frequency distribution in 2013

This observation is further backed by analyzing TFP dynamics for individual countries. We've classified country dynamics based on four clustering results (two clustering methods for each of the two types of estimates) into three basic groups:

- "Core" countries that never change their group membership. There are 75 countries in this group (60% of the total), 22 of which form the core of the "leader" group.
- "Transitional" countries that change their cluster membership for each method employed. There are 22 of these countries located on the "border" between two groups. This effect is probably due to minor fluctuations in TFP estimates.
- Countries with indeterminate dynamics (about 23% of the total number) that change their cluster membership for some methods and don't change it for others. In most cases, these countries firmly belong to the "catching-up" group, while their membership shifts are most likely due to clustering artifacts.

The "forerunners in 2009 and 2013" group is formed mostly by developed Western countries, as well as developed Eastern countries represented by Hong Kong and oil producers represented by Qatar and Kuwait. The presence of Equatorial Guinea and Barbados in this group is due to the following: TFP is a measure of productivity given the existing values of labour and capital, not a measure of "productivity in general". In other words, a country without large capital stock and incapable of effective investment can still have a high TFP measure if it is efficient in its use of existing capital.

The countries that change their group membership mostly demonstrate nonlinear TFP dynamics over the years; cluster membership for these countries changed several times during the period. Only 4 cases (18%, namely, Lithuania, Poland, Nigeria, Trinidad and Tobago) demonstrate generally positive changes, i.e. toward the "leader" cluster. The remaining 18 countries mostly tend to shift toward the "catching-up" group (Table 3). This serves as additional proof for the hypothesis that the "forerunners" group "shrinks" over the years.

Table 3. Shifts in cluster membership between 2009 and 2013

		2013	
		Catching-up	Forerunners
1990	Catching-up	53 countries: Albania, Armenia, Bangladesh, Benin, Bhutan, Botswana, Brazil, Burkina Faso, Cameroon, Cape Verde, Central African Republic, China, Colombia, Comoros, Congo Dem.Rep., Congo Rep., Cote d'Ivoire, Djibouti, Ecuador, El Salvador, Ethiopia, Ghana, Guinea-Bissau, India, Jordan, Kazakhstan, Kenya, Lesotho, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mongolia, Morocco, Mozambique, Namibia, Nepal, Niger, Pakistan, Paraguay, Peru, Philippines, Romania, Russia, Tanzania, Thailand, Togo, Tunisia, Ukraine, Venezuela, Zambia, Zimbabwe	4 countries: Lithuania, Nigeria, Poland, Trinidad and Tobago,
	Leaders	18 countries: Azerbaijan, Bahrain, Belize, Chad, Chile, Egypt, Gambia, Georgia, Honduras, Mexico, Oman, Portugal, St. Lucia, St. Vincent and Grenadines, Tajikistan, Uzbekistan, Vietnam, Yemen	22 countries: Australia, Austria, Bahamas, Barbados, Belgium, Canada, Denmark, Equatorial Guinea, Finland, France, Germany, Hong Kong, Iceland, Ireland, Italy, Kuwait, Luxembourg, Netherlands, Norway, Qatar, United Kingdom, United States

Finally, an important observation can be made for countries gradually moving from one group to another: there are very few of such cases, especially cases with increasing productivity. Among such countries one can name Poland, Lithuania, Trinidad and Tobago and possibly Estonia. Steady decrease in productivity is observed for Belize, Chad, Oman, Portugal, St. Lucia, St. Vincent/Grenadines. This observation may be seen as evidence in favor of the path dependence problem [David 2000].

7. TFP as a variable in political science research

As an additional external validity check, we look at statistical relationships between our TFP estimates and key institutional and political variables available to researchers.

In order to achieve this goal, we choose three key dimensions important to political science research: social capital, institutional quality and state stability.

For measures of social capital we use World Value Survey data [World Value Survey] and C. Welzel’s survey dataset [Welzel 2013]. The WVS variable is the share of people answering positively to the question “Do you trust unknown people?”; the other variable is a dummy-coded response to a similar questions. The variables are respectively called “Unknown” and “wel_trstdSCALE” in Table 4.

While a multitude of expert-survey based measures of institutional quality exist, most of them are heavily correlated. To give an overview of the relationship between our TFP estimates and institutional quality, we employ principal component analysis and construct a complex variable combining several institutional measures available to researchers. These three key groups of variables:

a) transparency and absence of corruption (*Anticorruption & Transparency*);

b) property rights protection;

c) government effectiveness.

For transparency and absence of corruption, we use the following data: Bayesian Corruption Index, Andrew Williams Transparency Index, Index of Political Corruption from the “Varieties of Democracy” project, TI’s Corruption perceptions index as well as WGI’s Control of Corruption Index. We conduct a principal component analysis on the five variables and use the first component as a proxy ("Institutional"), since it explains 87.1% of the variation in data. However, all the results are valid for individual variables as well. For the sake of simplicity we present most results for the whole country sample, while all the calculations have been checked for individual countries.

Finally, we use Polity IV's State Fragility Index as a proxy for state stability.

Table 4. Spearman correlations between TFP estimates and key political and social variables

	PL_TFP	LM_TFP
Unknown	0,438794	0,341563
p-value	0,000001	0,000001
N	441	441
wel_trstdSCALE	0,342485	0,374216
p-value	0,000001	0,000001
N	230	230
State fragility index	-0,54762	-0,64422
p-value	0,000001	0,000001
N	2580	2581
Institutional	0,621443	0,685267
p-value	0,000001	0,000001
N	1678	1678

The correlations between these variables and our estimates for TFP are presented in Table 4. Both plain (PL_TFP) and long-memory (LM_TFP) estimates demonstrate strong significant relationships with state stability and institutions, while the relationship between TFP and social capital remains inconclusive. This may be due to possible non-linear nature of the relationship or to the existence of an intermittent component. In any case, this particular relationship warrants further research into the problem.

8. Conclusion

In this paper we propose a new method of measuring total factor productivity. We expand upon existing non-parametric techniques used to estimate TFP. Our method employs a Data Envelopment Analysis variant with input to output ratios serving as dimensions for the production possibility space. We use a two-factor production technology with labor and capital as inputs and GDP at constant prices as an output.

The method allows us, after constructing the non-parametric best-practice production frontier, to specify the actual form of the production function, which we dub the Frontier-generated Production Function (FgPF). The FgPF form generated by our model is a piecewise linear function, the properties of which are close to Inada conditions.

By employing our variant of the non-parametric TFP estimation technique, we were able to produce TFP estimate for a large sample of countries for the years between 1990 and 2013. Statistical analysis shows that the estimates produced in this study differ from existing TFP estimates (namely, published within Penn World Tables and UNIDO's World Productivity Database), yet demonstrate strong significant correlations with the latter.

Our preliminary analysis of the TFP estimates has yielded notable results: we were able to successfully determine two distinct clusters within the country sample, corresponding to high-productivity ("forerunners") and low-productivity ("catching-up") economies. The data suggest that the clusters become more distinct over time, and therefore the differences in productivity between the two types of countries are actually increasing.

Since the TFP estimation method allows us to specify the FgPF, this opens up the possibility to answer several interesting research questions in the future. By comparing frontiers for different years, one can deduce the direction in which actual technical progress is moving (i.e. labor-efficiency or capital-efficiency). Moreover, since non-parametric approaches do not require price data, we can enhance the time-span of the estimate dataset for more in-depth analysis of dynamics (such as testing whether the perceived productivity divergence is actually significant). By estimating TFP for a large country sample, we can look into the interactions between institutional data (such as regime characteristics or quality of government) and economic productivity. This would in turn allow for decomposition of the TFP estimates into both scale and factor efficiency and the role of political institutions in productivity.

We've also shown that our TFP estimates demonstrate significant relationships with key variables common for most political science studies, namely institutional quality, state stability and social capital. This is a preliminary external validity check that shows our dataset to be a feasible alternative to other economic variables such as GDP per capita.

References

- Abramovitz, Moses. 1956. Resource and output trends in the United States since1870.American Economic Review46 (2): 5–23.
- Akhremenko A., Petrov A., Yureskul Ye. 2017. Barriers to Institutional Development in Europe // Social Sciences., vol. 48-4. P. 91-110 (в печати)
- Akhremenko A., Petrov A. 2014 Efficiency, Policy Selection and Growth in Democracy and Autocracy: A Formal Dynamical Model / NRU Higher School of Economics. Series PS "Political Science". No. WP BRP 16/PS/2014.
- Akhremenko A. S., Petrov A., Yureskul E. 2015 Cyclically Balanced Growth Paths in a Model of Economic Growth with Endogenous Policy Switching / NRU Higher School of Economics. Series WP BRP "Economics/EC". No. 109/EC/2015
- Arora, Nitin, 2013. "Does Catching-up or Innovations Drive Total Factor Productivity Growth in Indian Sugar Industry? A Non-Parametric Analysis," Indian Economic Review, Department of Economics, Delhi School of Economics, vol. 48(2), pages 351-380.
- Banker R.D., Charnes A., Cooper W.W. 1984 Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science, 30, pp. 1078-1092.
- Battese G.E., Prasada Rao D.S. 2002. Technology Gap, Efficiency, and a Stochastic Metafrontier Function. International Journal of Business and Economics, Vol.1, No.2, 87-93.
- Battese G.E., Rao D., O'Donnell C.J. 2002. A Metafrontier Production Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating Under Different Technologies. - Journal of Productivity Analysis 21(1):91–103.
- Battese, G. E. &Coelli, T. J. 1995. 'A model for technical inefficiency effects in a stochastic frontier production function for panel data', Empirical Economics 20(1), 325–332.
- Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert. 1982. The economic theory of index numbers and the measurement ofinput, output, and productivity. Econometrica50 (6): 1393–1414.
- Coelli, T.J.; Rao, D.S.P.; O'Donnell, C.J.; Battese, G.E. 2005. An Introduction to Efficiency and Productivity Analysis, 2nd Edition.

- Cooper, W.W., Seiford, L.M., Tone, K. 2007. Data envelopment analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Springer.
- Copeland, Morris A. 1937. Concepts of national income. Studies in Income and Wealth.Vol. 1, 3–63. New York: National Bureau of Economic Research.
- Copeland, Morris A., and E. M. Martin. 1938. The correction of wealth and income estimates for price changes. Studies in Income and Wealth.Vol. 2, 85–135. New York: National Bureau of Economic Research.
- David, Paul A. 2000. Path Dependence, its Critics and the Quest for “Historical Economics”. - P. Garrouste and S. Ioannides (eds), Evolution and Path Dependence in Economic Ideas: Past and Present, Edward Elgar Publishing, Cheltenham, England.
- De Witte K., Geys B. 2011. Evaluating Efficient Public Good Provision: Theory and Evidence from a Generalized Conditional Efficiency Model for Public Libraries. Journal of Urban Economics, 69, pp. 319-327
- Diewert, W. Erwin. 1976. Exact and superlative index numbers. Journal of Econometrics,4:115–45.
- Fare, Rolf, Shawna Grosskopf, Mary Norris, and Zhongyang Zhang. 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. American Economic Review84 (1): 66–83.
- Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A (General), 120(3), 253-290.
- Feenstra R C, Inklaar R, Timmer M P. 2015. The Next Generation of the Penn World Table vol. 105, no. 10, october 2015 (pp. 3150-82). www.ggdc.net/pwt
- Forstner, H., & Isaksson, A. 2002. Productivity, Technology, and Efficiency: An Analysis of the World Technology Frontier; when Memory is Infinite. Statistics and Information Networks Branch of UNIDO.
- Hall, Robert E., and Charles I. Jones. 1999. “Why Do Some Countries Produce So Much More Output Per Worker Than Others?” The Quarterly Journal of Economics, vol. 114, no. 1, pp. 83–116.
- Hulten, C.R. 2001. Total Factor Productivity: A Short Biography. Chapter in NBER book New Developments in Productivity Analysis, Charles R. Hulten, Edwin R. Dean and Michael J. Harper, editors (p. 1 - 54)
- IMF Data. 2016. http://data.imf.org/?sk=1CE8A55F-CFA7-4BC0-BCE2-256EE65AC0E4. Retrieved 25.04.2016
- Inada, Ken-Ichi. 1963. "On a Two-Sector Model of Economic Growth: Comments and a Generalization". The Review of Economic Studies. 30 (2): 119–127
- Issaksson, A. 2007. World Productivity Database: a technical description UNIDO Research And Statistics Branch Staff Working Paper 10/2007
- Kenneth, C., Land, C., Lovell, K., &Thore, S. 1994. Productive efficiency under capitalism and state socialism: An empirical inquiry using chance-constrained data envelopment analysis. Technological Forecasting and Social Change, 46, 139-152
- Mahmood, T. 2012. Effects of Input Composition on Technical Efficiencies of Textile Industries in Pakistan. The Pakistan Development Review, Vol. 51, No. 2 (Summer 2012), pp. 117-130
- Milliken O, Devlin R A, Barham V, Hogg W and Dahrouge S. 2011. Comparative Efficiency Assessment of Primary Care Service Delivery Models Using Data Envelopment Analysis, Canadian Public Policy, 37, issue 1, p. 85-109.
- Prucha, I R, and M. Ishaq Nadiri. 1981. Endogenous capital utilizationand productivity measurement in dynamic factor demand models: Theory and an application to the U.S. electrical machinery industry. Journal of Econometrics15:367–96.Rao 1996

- Ray, S. 2002. Did India's Economic Reforms improve Efficiency and Productivity? A Nonparametric Analysis of the Initial Evidence from Manufacturing. Indian Economic Review, New Series, Vol. 37, No. 1 (January-June 2002), pp. 23-57
- Ray SC and E Desli. 1997. "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment." American Economic Review 87(5): 1033-1039.
- Rao, J.M. 1996. Manufacturing Productivity Growth: Method and Measurement. Economic and Political Weekly, Vol. 31, 2927-2936.
- Simar L., Wilson P. 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136, pp. 31–64
- Solow R M. 1956. "A Contribution to the Theory of Economic Growth," The Quarterly Journal of Economics, Oxford University Press, vol. 70(1), pages 65-94.
- Uzawa, H. 1963. "On a Two-Sector Model of Economic Growth II". The Review of Economic Studies. 30 (2): 105–118.
- Veeramani, C., Goldar, B. 2004. Investment Climate And Total Factor Productivity In Manufacturing: Analysis Of Indian States. Indian Council For Research On International Economic Relations. Working Paper No. 127.
- World Bank. 2016a. GDP at market prices (constant 2005 US$).

http://data.worldbank.org/indicator/NY.GDP.MKTP.KD. Retrieved 16.05.2016

- World Bank. 2016b. Labor force, total.

http://data.worldbank.org/indicator/SL.TLF.TOTL.IN. Retrieved 19.05.2016

Andrei Akhremenko
National Research University Higher School of Economics (Moscow, Russia). Laboratory for Qualitative and Quantitative Analysis of Political Regimes. Project leader.
Email: aakhremenko@hse.ru

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.