

Firm Level Productivity Estimates¹:

Methodological Note

Date of last update: February 16, 2021

Please see summary of updates at the end of this document

Introduction

Firm productivity — the ability to generate greater outputs with less inputs — is one of the key elements of economic growth. Considerable scholarly analysis has been devoted to measuring productivity, especially since Robert Solow’s groundbreaking work (1957). The recent increased availability of detailed, firm-level datasets from developed economies has further intensified the interest into the subject, including investigations of how productivity varies by firm and economy characteristics (e.g. Syverson 2011). Indeed, Solow’s seminal, neoclassical growth model anticipates that a sizeable portion of comparable income inequality can largely be explained by underlying differences in productivity. To the extent that data can be disaggregated, researchers can delve into within economy differences; as long as data are comparable across economies and across time, cross-economy differences can be explored (see, for example, Acemoglu and Dell 2010).

To estimate the productivity of firms in less developed economies, researchers have turned to analysis using survey-based data, often in the absence of comparable census data. The Enterprise Surveys (ES), a detailed firm-level data collected by the World Bank’s (WB) Enterprise Analysis unit, is well suited for such inquiry. To ease the study of productivity by researchers and policymakers, the Enterprise Analysis unit is providing two types of firm level productivity estimates based on the ES. First, we estimate total factor productivity (TFP) — the portion of output that is not explained by the amount of inputs utilized — as outlined in Section II. Because this estimation may be troublesome for multiple reasons — e.g. selection, simultaneity, issues with using monetary (as opposed to physical) measures of output and inputs (Olley and Pakes 1996, Levinsohn and Petrin 2003, Foster et al. 2008, Hsieh and Klenow 2009, Restuccia and Rogerson 2013) — we also provide a much simpler but more robust estimates of firms’ productivity in the form of factor ratios. Unlike TFP estimates, some of these are also available for non-manufacturing firms.

The rest of this note is organized as follows. Section I briefly summarizes the ES data. Section II discusses estimation of revenue-based TFP, so-called TFPR. The factor ratio measures are discussed in Section III. The final section explains certain practicalities such as how to link the data with other ES datasets. The appendix contains the exact wording of questions associated with

¹ We thank Jan De Loecker, Chad Syverson, and other participants of the “Productivity Boot Camp” organized by the World Bank Group’s Equitable Growth, Finance and Institutions Chief Economist Office and Trade & Competitiveness Global Practice ¹. Corresponding authors are: David C. Francis - dfrancis@worldbank.org, and Nona Karalashvili – nkaralashvili@worldbank.org both from the Enterprise Analysis Unit. Jorge Rodriguez Meza, Program Manager, Enterprise Analysis Unit.

the relevant variables. Summary of updates given at the end of the note provides brief information about all updates to the data and methodology.

Section I: Data

The Enterprise Analysis unit has been conducting surveys using a methodology that allows cross economy analysis since 2006. To date, over 146,000 face-to-face interviews with top managers and business owners in 143 economies have taken place under this, “Global Methodology”. The surveys not using this methodology are excluded, so are the surveys conducted earlier than 2006. The data from Zimbabwe 2011 are excluded from analysis due to the complications with the data that reflect the consequences faced by this economy while recovering from a period of hyperinflation. Additionally, the survey in Cambodia in 2013 was not conducted under the supervision of the unit, which is why it is also excluded. A further 24 surveys were dropped because at least one of the key variables used in the analysis were not collected there.² This leaves 255 surveys in 134 economies and more than 154,000 interviews with top managers and business owners of firms spanning more than 40 different industries by two-digit ISIC code.³ Of these interviews, more than 84,000 are with top managers and business owners of manufacturing firms, for which we provide TFPR estimates. Factor ratios using labor costs and revenues are provided for all firms in the sample.

Section II: Firm Level Estimates of Total Factor Productivity

We begin with a Cobb-Douglas production function in the following form:

$$VA_i = A_i K_i^{\alpha} L_i^{\beta} \quad (1)$$

where firm-level value-added VA_i is a function of inputs of capital (K_i), and labor (L_i). Firms’ efficiency of production is measured by the term A_i which is the portion of output that cannot be directly attributed to the utilized inputs. We refer to the above model, or rather its version with natural logarithm applied on both sides, as VAKL. To allow for a more flexible functional form, we check if the second-order Taylor polynomial around the Cobb-Douglas production function fits the data better, and if so, use this translog specification. In addition to VAKL, we provide estimates based on YKLM specification where VA_i in (1) is replaced with Y_i , output, and the righthand-side has an additional input variable, materials (M_i).⁴

² These are: Bangladesh 2007, Benin 2009, Bhutan 2009, Cape Verde 2009, Central African Republic 2011, Chad 2009, Congo 2009, Eritrea 2009, Fiji 2009, Gabon 2009, Lesotho 2009, Liberia 2009, Malawi 2009, Micronesia 2009, Niger 2009, Pakistan 2007, Rwanda 2011, Samoa 2009, Sierra Leone 2009, Timor-Leste 2009, Togo 2009, Tonga 2009, Vanuatu 2009, and Venezuela 2006.

³ Note that the ES uses ISIC Rev. 3.1.

⁴ Note that other versions of (1) are also possible, e.g. YKL, or YKELM with E for Electricity. We do not analyze these here.

While analytically straightforward, estimation of A_i can be troublesome. Importantly, a selection and simultaneity problem exists as firms' choices regarding continuation of operations and quantities of inputs are guided by their productivity (e.g. Olley and Pakes 1996, Levinsohn and Petrin 2003). Since the past values of investment or intermediates may be used to remedy this problem, the ES has recently started collecting the corresponding information. This will be used in the future as the data becomes available. In addition to selection and simultaneity problems, there are issues associated with the fact that often only monetary (as opposed to physical) output and inputs are observed. Such revenue-based TFP is often referred to as TFPR, where R stands for revenue. Estimates of TFPR are based on an implicit assumption that markets are perfectly competitive. Consequently, market dynamics through clearing prices are incorporated in the resulting productivity estimates thus conflating productivity and market power. Since the ES collect information on revenues and firm-level line item costs (rather than physical inputs and outputs), TFPR is the measure that can be estimated using the ES data. For a recent discussion of these and other issues in estimation see for instance Foster et al. (2008), Hsieh and Klenow (2009), Restuccia and Rogerson (2013).

Section II.1: Production Functions and Regression Equations

TFPR is estimated using two models: VAKL, and YKLM. Each of these models, as given in the form of equation (1), assumes a common production technology (i.e. α_k and α_l are the same within the sample). To avoid this rather restrictive assumption, equation (1) is enriched as follows: TFPR is estimated separately for each industry — grouped by two-digit ISIC codes, denoted with s (for sector) — over pooled economies. Furthermore, wherever possible, elasticities of output with respect to capital, labor, and materials (i.e. α_k and α_l) are allowed to vary by the income-level grouping of the corresponding economy. The income levels are grouped according to the WB classification as of the year in which each survey was conducted and are denoted with w . Empirical investigation of stability of our estimates revealed that this income grouping is appropriate if the number of observations per industry and income group is at least 500. For industries with fewer than 500 observations per income group the elasticities are estimated across all economies. Finally, to control for an average economy-level and time effects, dummy variables for each economy c and year y are included (e.g. Halvorsen and Palmquist 1980).

The regression functions are thus as follows:

$$\begin{aligned} \ln(VAct_i) = & c_{VAKL} + \alpha_1 \ln(Kct_i) + \alpha_2 \ln(Lct_i) + \alpha_3 \ln(Kct_i) \cdot I_c + \alpha_4 \ln(Lct_i) \cdot I_c \quad (VAKL) \\ & + \alpha_5 [\ln(Kct_i)]^2 \\ & + \alpha_6 [\ln(Lct_i)]^2 + \alpha_7 \ln(Kct_i) \cdot \ln(Lct_i) + FE_I + FE_c + FE_t \\ & + u_{ctiVAKL} \end{aligned}$$

$$\begin{aligned} \ln(Yct_i) = & c_{YKLM} + \beta_1 \ln(Kct_i) + \beta_2 \ln(Lct_i) + \beta_3 \ln(Mct_i) + \beta_4 \ln(Kct_i) \cdot I_c + \beta_5 \ln(Lct_i) \cdot I_c \quad (YKLM) \\ & + \beta_6 \ln(Mct_i) \cdot I_c + \beta_7 [\ln(Kct_i)]^2 + \beta_8 [\ln(Lct_i)]^2 + \beta_9 [\ln(Mct_i)]^2 \\ & + \beta_{10} \ln(Kct_i) \cdot \ln(Lct_i) + \beta_{11} \ln(Kct_i) \cdot \ln(Mct_i) + \beta_{12} \ln(Lct_i) \cdot \ln(Mct_i) \\ & + \beta_{13} \ln(Kct_i) \cdot \ln(Lct_i) \cdot \ln(Mct_i) + FE_I + FE_c + FE_t + u_{ctiYKLM} \end{aligned}$$

The regression is estimated separately for each industry, s .

where

- c denotes country
- t denotes year
- i denotes firm
- I_c denotes a dummy variable for income group of the economy (equals 1 for high income based on the WB classification as of the year in which each survey was conducted). Note that these interaction terms are included only if the number of observations per industry and income group is at least 500.⁵
- c_s^{VAKL} , and c_s^{YKLM} , are constants
- FE_l , FE_c , and FE_t are income level, economy, and year fixed effects

To reiterate, the above regressions are run for each industry separately. Note that the underlying assumption of equations (VAKL) and (YKLM) is that elasticities of output with respect to inputs are the same across economies within the same income groups. Also note that the above regressions are weighted using sampling weights.

The TFPR is estimated as:

$$TFPR_{ctif} = \hat{u}_{ctif} + \hat{c}_f + FE_{lf} + FE_{cf} + FE_{tf}$$

where $f \in \{VAKL, YKLM\}$.

The data contains the firm-level estimates, i.e. \hat{TFPR}_{sci}^{VAKL} and \hat{TFPR}_{sci}^{YKLM} , alongside the variables used as described below. The regression tables are provided in the annex attached to this document in the Microsoft Excel format.

All estimates take into consideration the survey design for the ES by incorporating both stratification and probability (survey) weight information.

Section II.2: Description of Variables

VA, K, and L are proxied using the questions available in the ES. More precisely, VA is proxied by the difference between the total annual sales of the establishment (variable d2 in the data) and total annual cost of inputs (variable n2e), K is proxied by the replacement value of machinery, vehicles, and equipment (variable n7a); L is proxied by the total annual cost of labor (variable n2a). For YKLM specification, Y is proxied by total annual sales of the establishment (variable d2); K and L are the same as in VAKL specification, and M is proxied by the total annual cost of inputs (variable n2e). The exact wording of questions are given in the appendix.

⁵ More precisely, following this principle, the input elasticities vary across countries for ISIC group of 15 and 16, also for ISIC 17, 18, group of 23 and 24, ISICs 25, 26, 28, 29, group of 30, 31, 32, and 33, and ISIC 36. The input elasticities are the same across countries for ISICs 19, 20, 21, 22, 27, and group of 34 and 35.

Note that before estimating equations (*VAKL*) and (*YKLM*), outliers in *d2* (capturing *Y*), *n7a* (capturing *K*), *n2a* (capturing *L*), *n2e* (capturing *M*), and *VA* (*d2* minus *n2e*), as well as outliers in ratios *n7a/VA* and *n2a/VA* were turned into missing. To find outliers in levels, we first transform variables as $\ln(x + 1)$, and group observations by economy and broadly defined sector (more precisely, manufacturing and services). Next, we calculate (unweighted) means and standard deviations of these transformed variables within each group. Observations that are more than three standard deviations away from the mean are then marked as outliers and turned into missing. To find outliers in ratios, we first transform variables as $\ln(x)$, and group observations by industry. The three-standard-deviation rule is then applied (unweighted) and the corresponding observations are turned into missing. Additionally, note that in cases of negative value added, logarithms cannot be defined, and thus these observations are not included.

Note that all the above variables are collected in local currency units (LCUs) which is specific to the survey and year. Consequently, the data span different fiscal years. For the estimation of crosseconomy regressions all data must be transformed to a common currency-year. To do this, all variables are first exchanged into U.S. Dollars (USD) using the official exchange rate (period average) from the World Development Indicators (WDI).⁶ The data are then deflated to 2009 using the GDP deflator for the United States from the relevant reference fiscal year.⁷ Note that information on the closing month of the firms' fiscal year is used to adjust exchange rates and deflators for each firm.⁸

Section II.3: Cutoffs for Sample Size

For the purposes of precision of estimation, the regressions (*VAKL*) and (*YKLM*) were run only if the total number of observations in a sector is at least 120. To make the most use of the available data, some industries were grouped together giving rise to the following set of industries:

ISICs 15 and 16:	Manufacturing of food products and beverages, and manufacturing of tobacco products
ISIC 17	Manufacture of textiles
ISIC 18	Manufacture of wearing apparel; dressing and dyeing of fur
ISIC 19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
ISIC 20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
ISIC 21	Manufacture of paper and paper products
ISIC 22	Publishing, printing and reproduction of recorded media

⁶ WDI indicator code: PA.NUS.FCRF

⁷ WDI indicator code: NY.GDP.DEFL.ZS

⁸ The fiscal year and its closing month information are given in variables “*d2_l1_year_perf_indicators*” and “*d2_n3_last_month_fy_perf_ind*” respectively.

ISICs 23 and 24:	Manufacturing of coke, refined petroleum products and nuclear fuel, and manufacturing of chemicals and chemical products
ISIC 25	Manufacture of rubber and plastics products
ISIC 26	Manufacture of other non-metallic mineral products
ISIC 27	Manufacture of basic metals
ISIC 28	Manufacture of fabricated metal products, except machinery and equipment
ISIC 29	Manufacture of machinery and equipment not elsewhere classified (n.e.c.)
ISICs 30, 31, 32, and 33:	Manufacturing of office, accounting and computing machinery; manufacturing of electrical machinery and apparatus n.e.c., manufacturing of radio, television and communication equipment and apparatus, and manufacturing of medical, precision and optical instruments, watches and clocks
ISICs 34 and 35:	Manufacturing of motor vehicles, trailers and semi-trailers, and manufacturing of other transport equipment
ISIC 36	Manufacture of furniture; manufacturing n.e.c.

The minimum of 120 observations per industry alongside with the difficulty of grouping the recycling sector (ISIC 37) with other industries results in the dropping of ISIC 37. Note that each row in the above table represents separate regressions for each, VAKL and YKLM, specification.

Section III: Factor Share Based Estimates of Productivity

The following much simpler but very informative measures of productivity are also provided: (i) labor cost per USD of sales, (ii) total sales per worker (in USD 2009, as in the rest of the measures), (iii) cost of inputs per USD of sales, available for manufacturing firms only; and (iv) cost of finished goods per USD of sales, available for services firms only. These are simple ratios of the corresponding variables.

Section IV: Practical Aspects of Using the Data

This section covers several practicalities. First, the calculation of average productivity estimates is discussed. Next, brief comments on the treatment of item non-response (and lack thereof) are provided. This is followed by a description of how to link the data with other ES datasets. Specific sets of commands in Stata are suggested for analysis of the data. And finally, we provide a summary of data changes, as this dataset is periodically updated.

The calculation of averages or central tendencies (e.g. means, medians) of firm-level productivity at the economy level is possible using these estimates. While different approaches to this calculation are possible, the most straightforward and available method is by using survey weights,

which are available in the data. All the ES included in the sample were implemented following a stratified simple randomized sampling methodology, with strata defined by sector, firm size (by employees), and geographic location. The use of sampling weights in the process of calculating averages thus helps account for the industry composition within economies. All necessary variables for these estimations (including in regression analysis) are provided in the dataset.

While item non-response may be consequential for most analysis, we do not attempt to address it in the data. We do not employ any of the available imputation or re-weighting methods that assume that data “missingness” is not ignorable and is related to underlying firm characteristics (Little and Rubin 1987). Additionally, the survey (probability) weights included in the data are agnostic to item non-response and the missingness of productivity estimates.

To link the data with other ES datasets that are available on the ES portal, we recommend using the variable *idstd*.⁹ This is a unique identifier of each interview in the entire set of ES interviews. It corresponds to the same variable that is included in all relevant firm-level ES datasets, including the dataset that combines all surveys conducted since 2006.

We suggest taking advantage of the survey commands available in statistical software when using the dataset (e.g. *svy* in Stata). These commands appropriately leverage the survey design expressed in strata (contained in variable *strata*) and the survey (probability) weights (contained in variable *wt*).

The dataset is current as of the date noted in the heading of this document. It will be updated upon the publication of additional global ES datasets. Please note that changes should also be expected due to updates in WDI variables. All changes will be documented in detail.

We ask users of the data to cite the dataset as follows:

World Bank Group, Enterprise Analysis Unit. 2017. “Firm Level Productivity Estimates”.
www.enterprisesurveys.org/

References

- Acemoglu, D. and M. Dell 2010. “Productivity Differences Between and Within Countries”, *American Economic Journal: Macroeconomics*, 2(1), pp. 169–88
- Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. “Reallocation, firm turnover and efficiency: Selection on productivity or profitability?” *American Economic Review*, 98(1): 394-425
- Halvorsen, R., and Palmquist, R. 1980. The interpretation of dummy variables in semilogarithmic equations. *American Economic Review* 70(3): 474-475.
- Hsieh, Chang-Tai and Peter J. Klenow. 2009. “Misallocation and Manufacturing TFP in China and India.” *Quarterly Journal of Economics* 124 (4): 1403–48.

⁹ The data portal is available at <http://www.enterprisesurveys.org/>

Levinsohn J, Petrin A. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*. 2003 Apr 1;70(2):317-41.

Little, Roderick J and DB Rubin. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Olley, G.S. and Pakes, A., 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6), pp.1263-1297.

Restuccia, Diego and Richard Rogerson. 2008. "Policy distortions and aggregate productivity with heterogeneous establishments." *Review of Economic Dynamics* 11(4), 707–20.

Solow, Robert M. "Technical change and the aggregate production function." *The review of Economics and Statistics* (1957): 312-320.

Syverson, Chad. 2011. "What determines productivity?" *Journal of Economic Literature* 49(2): 326-65.

Appendix

Total annual sales of establishment is measured by variable d2, which records responses to the following question:

“In [last complete] fiscal year, what were this establishment’s total annual sales for all products and services?”

Total annual cost of labor is measured by variable n2a, with the corresponding question as follows:

“From this establishment’s Income Statement for fiscal year please provide total annual cost of labor including wages, salaries, bonuses, social security payments”

Total annual cost of inputs is measured by variable n2e, with the corresponding question asked only to the manufacturing firms as follows:

“From this establishment’s Income Statement for fiscal year please provide total annual cost of raw materials and intermediate goods used in production”

Total annual cost of finished goods is measured by variable n2i, with the corresponding question asked only to the services firms as follows:

“From this establishment’s Income Statement for fiscal year please provide total annual cost of finished goods and materials purchased to resell”

Total number of workers is measured by variable l1, with the corresponding question as follows:

“At the end of [the last complete] fiscal year, how many permanent, full-time individual worked in this establishment? Please include all employees and managers (Permanent, fulltime employees are defined as all paid employees that are contracted for a term of one or more fiscal years and/or have a guaranteed renewal of their employment contract and that work a full shift)

Price of machinery, vehicles, and equipment is measured by variable n7a, with the corresponding question as follows:

“Hypothetically, if this establishment were to purchase [machinery, vehicles, and equipment] it uses now, in their current condition and regardless of whether the establishment owns them or not, how much would they cost, independently of whether they are owned, rented or leased?

Summary of Updates

As of February 16, 2021

Three recently completed Enterprise Surveys, namely Belgium 2020, Luxembourg 2020, and South Africa 2020, were added since the last release on September 10, 2020.

As of September 10, 2020

Four recently completed Enterprise Surveys, namely Armenia 2020, Azerbaijan 2019, Romania 2019, and Tunisia 2020, were added since the last release on July 07, 2020.

As of July 07, 2020

Two recently completed Enterprise Surveys, namely Egypt 2020, and Hungary 2019, were added since the last release on May 11, 2020.

As of May 11, 2020

Four recently completed Enterprise Surveys, namely Bulgaria 2019, Lebanon 2019, Slovakia 2019, and Zambia 2019, were added since the last release on April 15, 2020.

As of April 15, 2020

Two recently completed Enterprise Surveys, namely Czech Republic 2019 and Rwanda 2019, were added since the last release on March 2, 2020.

As of March 2, 2020

Three main changes were made since the last release on April 9, 2018. First, recently completed Enterprise Surveys, namely Albania 2019, Belarus 2018, Bosnia and Herzegovina 2019, Chad 2018, Colombia 2017, Croatia 2019, Cyprus 2019, Estonia 2019, Gambia 2018, Georgia 2019, Greece 2018, Guatemala 2017, Italy 2019, Jordan 2019, Kazakhstan 2019, Kenya 2018, Kosovo 2019, Kyrgyz Republic 2019, Lao PDR 2018, Latvia 2019, Lithuania 2019, Malta 2019, Moldova 2019, Mongolia 2019, Montenegro 2019, Morocco 2019, Mozambique 2018, North Macedonia 2019, Poland 2019, Portugal 2019, Russia Federation 2019, Serbia 2019, Slovenia 2019, Suriname 2018, Tajikistan 2019, Turkey 2019, Ukraine 2019, Uzbekistan 2019, and West Bank and Gaza 2019. Second, we included the interaction between $n7a$, $n2a$ and $n2e$ as additional regressor in the YKLM specification. Third, in the previous release, we the translog terms for sectors 17, 19, 21, 29, 30313233, and 3435 were dropped in the VAKL specification because they were not found to be jointly significant at 5% level. For these sectors, $tfprVAKL$ was estimated using a Cobb-Douglas specification. In this version, $tfprYKLM$ and $tfprVAK$ are estimated using translog production function for all sectors.

As of April 9, 2018

Five main changes were made since the last release on September 11, 2017. First, as it has become usual, recently completed Enterprise Surveys, namely Bolivia 2017, Ecuador 2017, Honduras 2016, Liberia 2017, Paraguay 2017, Sierra Leone 2017, and Uruguay 2017 were added.

Second, we replaced the YKL specification with VAKL. Consequently, we added VA to the list of variables for which we remove outliers (as described in the main text), and modified ratios

accordingly (e.g. instead of removing outliers on ratio $n2a/d2$, we use $n2a/VA$). Note that unlike previously, we remove outliers in ratios without applying sampling weights.

Third, we added translog terms to the regressions as outlined in the main text. We drop these terms in each regression where these terms are not found to be jointly statistically significant (at 5% level). Furthermore, we also added a dummy for income level (term FEI).

Fourth, one of our diligent users noticed unusual patterns in the Brazil 2009 data. The ES team considered the issue in much detail, going through communications that occurred at the time of the fieldwork of this survey. As a result, the team was able to update few observations on sales and costs. All respective data on the ES portal was updated, so was the data included in the estimation.

Finally, the exchange rates and deflators have been updated following the updates in the WDI databases.

As of September 11, 2017

Three changes were made since the last release on May 1, 2017. First, recently completed Enterprise Surveys, namely Dominican Republic 2016, Egypt 2016, Myanmar 2016, Nicaragua 2016, and Niger 2017 were added. Note that Turkey 2015 data is also included in the current database. The Turkey data was used in estimation previously but, as the data had not been released publicly, these observations were omitted from the database.

Second, one of our diligent users noticed that we have been using incorrect exchange rates for El Salvador 2006, Kosovo 2009, and Slovenia 2009. We have fixed these errors, and are grateful to the user.

Third, exchange rates and deflators are adjusted for the structure of each firm's fiscal year. In particular, we use the information about the closing month of the firms' fiscal year, and weight the annual figures from WDI accordingly for each firm.

As of May 1, 2017

Four changes were made since the last release on February 21, 2017. First, recently completed Enterprise Surveys, namely Côte d'Ivoire 2016 and Zimbabwe 2016 were added to the sample.

Second, the proxies of Y , K , L and M were transformed into the common currency-year as described on page 4, instead of being used in local currency units (LCUs).

Third, redenomination of currencies for the relevant surveys (namely Belarus 2008, Belarus 2013, Ghana 2007, Zambia 2007, and Zambia 2013) were taken into account. This change is reflected in the variable `exrate_d2`, which together with variable `deflator_usd_adjust_d2`, is used in transforming Y , K , L and M into the common currency-year.

And finally, exchange rates and GDP deflators were updated following the most recent update of the WDI (on April 17, 2017).

As of February 21, 2017

The sample has changed significantly since the original release on August 1, 2016. We decided to include all available surveys in each country instead of only the latest one. We also decided to use all surveys conducted in 2006 or later rather than dropping surveys that were conducted earlier than 2009. The surveys that were completed between August 1, 2016 and February 21, 2017 were also added to the sample. Furthermore, the survey conducted in Turkey in 2015 as part of the Regional Investment Climate Assessment project that was undertaken in collaboration with the Ministry of Development of Turkey was also added to the analysis, excluding micro firms covered in that survey. Since we are not at liberty to release the Turkey 2015 data, it is dropped from the database, the TFPR estimates however are based on these observations. Moreover, in addition to excluding outliers in $d2$, $n7a$, $n2a$, and $n2e$, we also exclude outliers in ratios $n2a/d2$, $n7a/d2$, and $n2e/d2$. Additionally, instead of grouping countries by income in three categories (high, medium, low), we decided to group them in two (high, low) if the number of observations per industry and income group is at least 500. The income grouping is not applied if the number of observations per industry and income group is below 500 and the input elasticities in these industries are assumed to be constant across countries. This happens for ISICs 19, 20, 21, 22, 27, and group of 34 and 35. Finally, the restriction on the minimum number of observations per country and ISIC code has been lifted (previously, if the number of observations per country and ISIC code was below 15, then these observations were excluded from the regressions).

Due to these changes, the following 86 surveys comprising 46,216 observations have been added to the original sample: Afghanistan 2008, Albania 2007, Angola 2006, Argentina 2006, Armenia 2009, Azerbaijan 2009, Belarus 2008, Benin 2016, Bosnia and Herzegovina 2009, Bolivia 2006, Botswana 2006, Bulgaria 2007, Bulgaria 2009, Burundi 2006, Cameroon 2016, Chile 2006, Colombia 2006, Croatia 2007, Czech 2009, DRC 2006, DRC 2010, Ecuador 2006, El Salvador 2006, El Salvador 2016, Estonia 2009, Ethiopia 2011, Macedonia 2009, Gambia 2006, Georgia 2008, Ghana 2007, Guatemala 2006, Guinea 2006, Guinea 2016, Guinea-Bissau 2006, Honduras 2006, Hungary 2009, Indonesia 2009, Kazakhstan 2009, Kenya 2007, Kosovo 2009, Kyrgyzstan 2009, Lao PDR 2009, Lao PDR 2012, Latvia 2009, Lesotho 2016, Lithuania 2009, Madagascar 2009, Mali 2007, Mali 2016, Mauritania 2006, Mexico 2006, Moldova 2009, Mongolia 2009, Montenegro 2009, Mozambique 2007, Namibia 2006, Nepal 2009, Nicaragua 2006, Nigeria 2007, Panama 2006, Paraguay 2006, Peru 2006, Philippines 2009, Poland 2009, Romania 2009, Russia 2009, Rwanda 2006, Senegal 2007, Serbia 2009, Slovakia 2009, Slovenia 2009, South Africa 2007, Swaziland 2006, Swaziland 2016, Tajikistan 2008, Tanzania 2006, Togo 2016, Turkey 2008, Turkey 2015, Uganda 2006, Ukraine 2008, Uruguay 2006, Uzbekistan 2008, Vietnam 2009, Yemen 2010, Zambia 2007.

As the sample includes all available surveys in each country, we deemed it appropriate to include country and year fixed effects in the regressions as depicted in equations (YKL) and (YKLM) (previously only the country fixed effect was included).