Productivity premia for many modes of internationalization. A replication study of Békés and Muraközy (*Economics Letters*, 2016)

Joachim Wagner[*]

Abstract
This study attempts to replicate estimation results from Gábor Békés and Balázs Muraközy, Measuring productivity premia with many modes of internationalization, published in *Economics Letters* (2016). In this paper the authors use comparable firm-level data for seven European countries based on the EFIGE dataset to estimate the productivity premia of firms with various modes of internationalization by several empirical methods to demonstrate how results differ due to the method applied. While the EFIGE data are available free of charge from the web one core variable used by Békés and Muraközy is not, because total factor productivity (tfp) as computed by the authors is based on data from a commercial data base and, therefore, is available for users with a license to this database only. The freely available EFIGE data, however, come with another tfp-variable that can be used instead. In this replication study I use the EFIGE data with this publicly available tfp-variable to replicate (parts of) the estimations of Békés and Muraközy (2016) to see whether their results hold with the widely used public use version of the EFIGE data, too. It turns out that the big picture that emerges from using both productivity measures tends to be very similar. The use

of the public use version of the EFIGE data for empirical investigations that deal with productivity, therefore, seems to be feasible.

1 Introduction

Performance premia of internationally active firms over their counterparts that are active on the national market only – defined as the difference in a performance measure like productivity, growth or profitability between firms from both groups – are at the core of the huge empirical literature on international firm activities and firm performance. While exporter productivity premia were among the most important topics at the beginning (see the survey by Wagner (2007)) other forms of international firm activity besides exports and other dimensions of firm performance besides productivity have been investigated, too (see Wagner 2012).

Recently, Békés and Muraközy (2016) made an important contribution to this literature by looking at the consequences of the use of various modes of internationalization by firms for the estimation of productivity premia for these modes. They argue that the presence of many single or combined modes necessitates classification across modes and demonstrate that the way researchers proceed here can influence the conclusions drawn from an empirical investigation. The authors use comparable firm-level data for seven European countries based on the EFIGE dataset described in detail in Altomonte and Aquilante (2012). While the EFIGE data are available free of charge from the web one core variable used by Békés and Muraközy is not, because total factor productivity (tfp) as computed by the authors is based on data from a commercial data base and, therefore, is available for users with a license to this database only. The publicly available EFIGE data, however, come with another tfp-variable that can be used instead. Békés and Muraközy (2016, p. 62) point out that they measure tfp more appropriately by using data for firms from the whole respective economy, while the tfp measure that comes with the EFIGE data is based on information for firms included in the EFIGE sample only.

In this replication study I use the EFIGE data with the public use tfp-variable to replicate (parts of) the estimations of Békés and Muraközy (2016) to see whether their results hold with this widely used version of the EFIGE data, too.[1] In doing so I intend to document, on the one hand, the degree to which the results of Békés and Muraközy (2016) do depend on the exact way total factor productivity is measured. According to Syverson (2011, p. 332) one might expect that this should not be the case to a large degree: "The inherent variation in establishment- or firm-level microdata is typically so large as to swamp any small measurement-induced differences in productivity metrics. Simply put, high-productivity producers will tend to look efficient regardless of the specific way that their productivity is measured." If, on the other hand, results differ considerably when the two different tfp measures are used, and when the (not publicly available) tfp measure applied in Békés and Muraközy (2016) is more appropriate than the publicly available tfp measure, this should be kept in mind when using the public use version of the EFIGE data for empirical investigations that deal with productivity.

2 Replication study

Békés and Muraközy (2016) look at five modes of internationalization of firms: (1) indirect exports (selling goods or services on a foreign market through an intermediary based in the home country);

---

[1]See the large number of entries listed in Google Scholar citing Altomonte and Aquilante (2012) for an (inclomplete) list of papers that use the EFIGE data.

(2) direct exports; (3) outsourced manufacturing in a foreign market (running at least part of the firms production activity in another country via contracts and arms length agreement with local firms); (4) foreign direct investment (FDI) in services (firms have any foreign affiliates but have no manufacturing FDI); (5) FDI in manufacturing (firms that have foreign affiliates where products are produced, and that may have services FDI, too). Note that any one firm may have between zero and five of these modes of internationalization in a year. The performance indicator looked at by Békés and Muraközy (2016) is total factor productivity (TFP), demeaned at industry and country level. Note that Békés and Muraközy (2016) do not use the TFP-variable that comes with the public use EFIGE data (labelled `tfp_va`) but compute their own variant estimated by fixed effects panel regression using the whole economy data available in the commercial data base Amadeus for the seven countries in EFIGE (and not for the firms in the EFIGE sample with the necessary information only). Given that access to the Amadeus dataset is restricted to users with an (expensive) license the newly computed TFP-measure (which uses more information and which can be considered as a better measure of firm-level productivity than `tfp_va` that comes with the EFIGE data) cannot be distributed freely.

In this replication study I use the EFIGE data with the tfp-va variable to replicate (parts of) the estimations of Békés and Muraközy (2016) to see whether their results hold with the widely used public use version of the EFIGE data, too. In these exercises I estimate OLS regressions $y_i = a + \beta X_i + e_i$ where y is demeaned `tfp_va` and X is a (set of) dummy-variable(s) representing different modes of internationalization; i is an index of the firm, and e is an error term.

As a first step productivity premia of internationalization modes are estimated by including one dummy variable for each mode at a time. This is labelled Approach [1] by Békés and Muraközy (2016). Table 1 reports the original results (based on Table 1 in Békés and Muraközy (2016)) and the results from the replication study. In line with results reported in the wider literature on international firm activities and firm performance (surveyed in Wagner (2007, 2012)) all estimated regression coefficients of the dummy-variables that indicate the use of one mode of internationalization at a time are positive and highly statistically significantly different from zero. No matter which mode of international firm activity is considered, and which measure for TFP is used, internationally active firms are more productive than their counterparts from the same industry and the same country that are active in their home country only.[2]

Firms may use different modes of international activity simultaneously. If they do so the reported estimates for productivity premia of one mode estimated by Approach [1] include the premia for other modes used. In an alternative approach that is labeled Approach [2] by Békés and Muraközy (2016) an additional structure is added. It is assumed that the premia of the combined modes is the sum of the single modes. In the empirical model dummy variables for each of the single modes are included. Table 2 reports results for this approach that are taken from either the original study by Békés and Muraközy (2016, Table 1, column 6) or from the replication study. Irrespective of the TFP-measure used the big picture is identical: The premia for direct export, service FDI and manufacturing FDI are positive and significant, while this no longer holds for the premia for indirect export and outsourced manufacturing when direct export and FDI are controlled for.

---

[2]Note that the point estimates of the premia differ by order of magnitude between the original study and the replication study. Unfortunately Békés and Muraközy (2016) do not report any descriptive statistics for their TFP-measure so it is not possible to investigate these differences further. Note, however, that Altomonte et al. (2012, p. 33, Table 12, column 1) report an exporter premium of 0.0999 based on an empirical model that uses the public use EFIGE data set with tfp_va, and that is identical to the one used in the replication study here, where the estimated premium for direct exporters is 0.106 (which is very close to 0.0999).

Table 1: Estimated productivity premia of various modes of internationalization - Approach [1]:
Dummy-variables for single modes

Mode		Indirect export	Direct export	Outsourced manufacturing	Service FDI	Manufacturing FDI
Method: OLS						
TFP - Békés/Muraközy	β	0.377	0.551	0.691	1.232	1.391
	p	<0.01	<0.01	<0.01	<0.01	<0.01
Number of observations		4,199	8,780	3,825	3,892	3.917
R-squared		0.026	0.062	0.051	0.158	0.185
TFP - EFIGE data set	β	0.077	0.106	0.115	0.270	0.220
	p	0.001	0.000	0.001	0.000	0.000
Number of observations		3,323	7,011	3,043	3,094	3,130
R^2		0,004	0.012	0.006	0.038	0.025

Note: β is the estimated regression coefficient of a dummy variable indicating whether a firm used the internationalization mode or not; p is the prob-value for the estimated coefficient (based on heteroscedasticity-robust standard errors). For a definition of internationalization modes see text. TFP is total factor productivity, demeaned by country and industry. Results for TFP - Békés/Muraközy are taken from Békés and Muraközy (2016), Table 1 (where p-values are only reported to be smaller than 0.01). Results for TFP EFIGE data set are own computations using the variable tfp_va that is available in the EFIGE data set. For details, see text. Note that all empirical models include a constant term; results for the estimated coefficients are not reported here to economize on space.

Table 2: Estimated productivity premia of various modes of internationalization - Approach [2]: Dummy-variables for each mode

Mode		Indirect export	Direct export	Outsourced manufacturing	Service FDI	Manufacturing FDI	Number of observations	R^2
Method: OLS								
TFP - Békés/Muraközy	β	-0.00284	0.391	0.103	0.884	1.051	9,342	0.124
	p	>0.01	<0.01	>0.01	<0.01	<0.01		
TFP - EFIGE data set	β	0.0056	0.076	-0.0035	0.203	0.154	7,432	0.021
	p	0.801	0.000	0.914	0.000	0.000		

Note: β is the estimated regression coefficient of a dummy variable indicating whether a firm used the internationalization mode or not; p is the prob-value for the estimated coefficient (based on heteroscedasticity-robust standard errors). For a definition of interna-tionalization modes see text. TFP is total factor productivity, demeaned by country and industry. Results for TFP - Békés/Muraközy are taken fromBékés and Muraközy (2016), Table 1 (where p-values are only reported to be smaller than 0.01 or not). Results for TFP EFIGE data set are own computations using the variable tfp_va that is available in the EFIGE data set. For details, see text. Note that all empirical models include a constant term; results for the estimated coefficients are not reported here to economize on space.

Table 3: Estimated productivity premia of various modes of internationalization - Approach [3]: topcoding (version 1)

	Mode	Indirect export	Direct export	Outsourced manufacturing	Service FDI	Manufacturing FDI	Number of observations	R^2
	Rank	1	2	3	4	5		
Method: OLS								
TFP - Békés/	β	-0.072	0.340	0.0353	1.160	1.391	9,341	0.123
Muraközy	p	>0.01	<0.01	>0.01	<0.01	<0.01		
TFP - EFIGE	β	0.031	0.084	-0.010	0.271	0.221	7,433	0.022
data set	p	0.350	0.000	0.782	0.000	0.000		

Note: β is the estimated regression coefficient of a dummy variable indicating whether a firm used the internationalization mode or not; p is the prob-value for the estimated coefficient (based on heteroscedasticity-robust standard errors). For a definition of internationalization modes see text. TFP is total factor productivity, demeaned by country and industry. Results for TFP - Békés/Muraközy are taken from Békés and Muraközy (2016), Table 2 (where p-values are only reported to be smaller than 0.01). Results for TFP EFIGE data set are own computations using the variable tfp_va that is available in the EFIGE data set. For details, see text. Note that all empirical models include a constant term; results for the estimated coefficients are not reported here to economize on space.

Békés and Muraközy (2016) argue that the additivity assumption used in Approach [2] is quite restrictive when firms supply multiple countries or sell multiple products. As an alternative they suggest Approach [3] that they call topcoding. As sorting theories predict a pecking order of modes, one may rank the single modes either based on theory or their unconditional premia and code each firm to the highest single mode it conducts. (Békés and Muraközy (2016), p. 63) They report results for two variants of topcoding that rank outsourced production as third or second mode, respectively. Results from the original study and from the replication study using these two alternative classifications are reported in Table 3 and Table 4. Here results between the original study and the replication study do differ in a point that is considered as important by Békés and Muraközy (2016, p. 63): "(O)utsourced manufacturing is only significant when it is ranked high to start with, i.e. classifying a mode high may generate spurious sorting." In contrast to this finding, the estimated coefficient of outsourced manufacturing is not statistically significant at any conventional level in the replication study, irrespective of the classification applied. Note, however, that the p-value of this coefficient drops sharply when outsourced manufacturing is ranked low instead of high, and that the sign switches from positive to negative. These results lead to a similar conclusion as the one based on the original study.

3   Concluding remarks

This study replicates estimation results from Békés and Muraközy (2016). While the EFIGE data used by the authors are available from the web one core variable used by them is not, because total factor productivity (tfp) as computed by the authors is based on data from a commercial data base and, therefore, available for users with a license to this database only. The freely available EFIGE data, however, come with another tfp-variable that can be used instead. In this replication study I use the EFIGE data with this publicly available tfp-variable to replicate (parts of) the estimations

Table 4: Estimated productivity premia of various modes of internationalization - Approach [3]: topcoding (version 2)

	Mode	Indirect export	Direct export	Outsourced manufacturing	Service FDI	Manufacturing FDI	Number of observations	R^2
	Rank	1	2	3	4	5		
Method: OLS								
TFP - Békés/ Muraközy	β	-0.072	0.340	0.0353	1.160	1.391	9,341	0.123
	p	>0.01	<0.01	>0.01	<0.01	<0.01		
TFP - EFIGE data set	β	0.031	0.084	-0.010	0.271	0.221	7,433	0.022
	p	0.350	0.000	0.782	0.000	0.000		

Note: β is the estimated regression coefficient of a dummy variable indicating whether a firm used the internationalization mode or not; p is the prob-value for the estimated coefficient (based on heteroscedasticity-robust standard errors). For a definition of internationalization modes see text. TFP is total factor productivity, demeaned by country and industry. Results for TFP - Békés/Muraközy are taken from Békés and Muraközy (2016), Table 2 (where p-values are only reported to be smaller than 0.01). Results for TFP EFIGE data set are own computations using the variable tfp_va that is available in the EFIGE data set. For details, see text. Note that all empirical models include a constant term; results for the estimated coefficients are not reported here to economize on space.

of Békés and Muraközy (2016) to see whether their results hold with the widely used public use version of the EFIGE data, too. It turns out that the big picture that emerges from using both productivity measures tends to be very similar. This is in line with Syverson (2011, p. 332) who argues that high-productivity producers will tend to look efficient regardless of the specific way that their productivity is measured. The use of the public use version of the EFIGE data for empirical investigations that deal with productivity, therefore, seems to be feasible.

4 References

Altomonte, Carlo and Tommaso Aquilante (2012). The EU-EFIGE/Bruegel-Unicredit dataset. *Bruegel Working Paper*, 2012/13.

Altomonte, Carlo, Tommaso Aquilante and Gianmarco I. P. Ottaviano (2012). The triggers of competitiveness: The EFIGE cross-country report. *Bruegel Blueprint* 17. Brussels: Bruegel.

Békés, Gábor and Balázs Muraközy (2016). Measuring productivity premia with many modes of internationalization. *Economics Letters* 139 (1), 61-64. *DOI:*10.1016/j.econlet.2015.12.016

Syverson, Chad (2011). What Determines Productivity? *Journal of Economic Literature* 49 (2), 326-365.

Wagner, Joachim (2007). Exports and Productivity: A survey of the evidence from firm level data. *The World Economy* 30 (1), 60-82.

Wagner, Joachim (2012). International Trade and Firm Performance: A Survey of Empirical Studies since 2006. *Review of World Economics* 148 (2), 235-267.