EXTENDED ABSTRACT

Title: Estimating regional production functions: The EU NUTS-2 case

Authors and e-mail of all: Ilias Kostarakos (ilias.kostarakos@ec.europa.eu), Javier Barbero Jimenez (Javier.BARBERO-JIMENEZ@ec.europa.eu), Andrea Conte (andrea.conte@ec.europa.eu)

Department: Unit B7 – Knowledge for Finance, Innovation and Growth

University: European Commission, Joint Research Centre

Subject area: Growth, convergence and development

Abstract: *(minimum1500 words)*

The analysis of productivity differences across countries and/or regions has been a central issue in macroeconomics for decades. Questions related to, among others, whether there are signs of convergence across regions and what are the main drivers of productivity and economic performance have been of interest to academics and policymakers alike (see, among others, Dettori et al. (2012), Ladu (2012), Marrocu et al. (2013), Beugelsdijk et al. (2018), Männasoo et al. (2018) and Siller et al. (2019)).

In the context of EU regional-level productivity analysis, (estimates of) production functions have been extensively used in order to obtain a measure of the level and/or the growth rate of Total Factor Productivity (henceforth, TFP).

In particular, using a standard Cobb-Douglas production function, two alternative approaches are used. Either the deterministic Solow-type sources of growth analysis is employed, usually assuming that the shares of labour and capital are 0.7 and 0.3, respectively or, the labour share is calculated based on available data and the capital share is obtained as a residual. The econometric specification is either based on the cross-section of countries (in the spirit of Mankiw et al. (1992)) or the panel as whole (see Schatzer et al. (2019) for a recent overview of the methods utilized in the

regional/spatial literature for obtaining an estimate of TFP)[1]. In this paper, we will focus on the panel approach as it will allow us to exploit the full information provided by both the time and space dimensions.

One common denominator of the panel techniques utilized in the literature (mainly, the different variants of the Fixed Effects estimator) is that they are characterized by a number of rather restrictive assumptions, namely:

(i) they impose parameter homogeneity, that is, the production function is assumed to be the same for all the regions in the sample,

(ii) they disregard the potential impact of cross-sectional dependence, that is, the cross-region dependencies stemming from factors other than geographical proximity, and

(iii) from a more technical point of view, they assume that all processes are stationary.

Each of these three assumptions could have an important bearing on the empirical results and influence the estimates of the technology parameters (i.e. the coefficients of the factors of production) and, as such, TFP, casting doubts on their reliability. As an example of the impact that each of these assumptions is bearing, consider the case where the variables are demonstrably non-stationary: this would imply that, unless a long-run (cointegrating) relationship exists, there is the potential that the estimates obtained are spurious and, thus, unreliable.

Given that among the various alternative approaches there is no consensus as to which one should be implemented, in this paper we propose to compare and contrast the main approaches utilized in the literature for estimating regional-level production functions and, hence, productivity levels. In that respect, the paper closest to ours in spirit is that of Schatzer et al. (2019), who estimate regional production functions using variants of the Fixed Effects estimator and compare the obtained estimates of TFP in order to assess whether the results vary significantly. Here, we propose an alternative route. In particular, rather than comparing the estimates of TFP across different estimators, we take a step back and focus on the estimated production functions. We examine the impact of the three-abovementioned modelling assumptions in an attempt to identify which method seems to fit the available data best. That is, we pay close attention to the time-series properties of the data and employ formal residual diagnostic tests in order to identify the model that is not misspecified.

[1] There is a distinct body of literature that employs frontier techniques (like the Data Envelopment Analysis or the Stochastic Frontier) in order to estimate TFP – see Filippetti and Peyrache (2015), Rogge (2019) and Burger et al. (2021). However, as these approaches fall outside the scope of the present paper, we refer the interested reader to the references cited before.

In particular, using data from the ARDECO database spanning the 1980-2019 period for 243 NUTS-2 regions of the EU27, we revisit the issue of estimating regional production functions through the lens of recently developed panel time series econometrics techniques – the so-called second-generation panel estimators. In particular, our analysis builds on the unobserved common factor framework and utilizes the Common Correlated Effects (henceforth, CCE) estimators introduced in the seminal paper of Pesaran (2006).

The main merit of this approach is that it allows for a very flexible modelling of time-varying unobserved heterogeneity, which in our case is reflected in the measure of Total Factor Productivity – the `measure of our ignorance'. To our knowledge, this is the first paper that applies this estimation approach to the issue of estimating regional production functions.

As already mentioned, one important issue that has not been thoroughly examined in the relevant empirical literature is that of interdependence across panel units (in this case, regions), otherwise known as cross-sectional dependence. In particular, the literature focuses only on local spillover effects that arise because of factors such as geographical proximity. These factors are usually modelled via the use of exogenously defined spatial weight matrices (see, for example, Dettori et al. (2012)).

However, especially in the case of the EU regions, these interdependencies and linkages are more likely to be highly pervasive. In particular, merely from their participation in the Union, these countries and regions are subject to increased interconnectedness stemming from a multitude of economic, political, institutional and cultural reasons. These include, the formation of the Eurozone and the establishment of the European Central Bank that essentially saw member states delegate monetary policy to a single authority and the implementation of fiscal policy rules in the context of the excessive deficit procedure, to name but a few. These features are particularly salient in our case, given that during the period under examination (namely, the post-1980s) the EU Member States were subject to large, global shocks that were common yet had quite distinct, country- and region-specific impacts. These shocks, such as the Global Financial Crisis of 2008, the subsequent debt crisis of the Southern Periphery of the EU etc. had a heterogeneous impact due to differences in the institutional setup of the regions and other local factors. Moreover, we can think of other secular processes such as the intensification of globalization (both in terms of trade and the ever-increasing importance of the financial sector) and the integration into global value chains that, if anything, have led to stronger dependencies across the regions of the EU. The potential effects of such cross-country correlations are quite important from an estimation point of view and, to our knowledge, have not been accounted for in the relevant literature. In particular, it has been shown (see Kapetanios et al. (2011)) that the presence of these correlations does not allow for the identification of the parameters of interest, leading to biased estimates and inconsistent inference.

In order to tackle the impact of cross-sectional dependence and of the other limiting assumptions already mentioned, we employ the unobserved common factor framework and estimate the specification using the CCE estimators of Pesaran (2006). This is an all-encompassing framework that essentially nests the approaches that have already been utilized in the literature, allowing us to directly be able to compare the results obtained by removing the various assumptions imposed by each approach.

The main empirical specification used in this paper is the (log-linearized) Cobb-Douglas production function:

$$y_{it} = \boldsymbol{\beta'}_i \boldsymbol{x}_{it} + u_{it}$$

where y denotes the (log of) Gross Value Added (GVA) per worker and **x** is the vector of the factors of production. The coefficients of the factors of production, i.e. the technology parameters **β** are region-specific but are assumed constant over time. The main novelty introduced by the unobserved common factor framework is the flexibility related to the modelling of TFP. In particular, we assume that TFP has a multifactor structure:

$$u_{it} = a_i + \boldsymbol{\lambda'}_i \boldsymbol{f}_t + \varepsilon_{it}$$

where α are the region fixed effects that capture the level of TFP, while **f** is a vector of time-varying unobserved common factors that have a region-specific impact, captured by the heterogeneous factor loadings λ. The **f** factors are allowed to be non-stationary, thus introducing stationarity in the panel. Additionally, they may be correlated with the factors of production, **x**. In particular, we assume:

$$x_{it} = \boldsymbol{\delta'}_i \boldsymbol{g}_t + \rho_{1i} f_{1t} + \ldots + \rho_{ni} f_{nt} + v_{it}$$

where **g** is a vector of unobserved common factors that exert an impact only on the inputs **x**, while a subset of the unobserved common factors **f** (so that $\boldsymbol{f}_{.t} \subset \boldsymbol{f}_t$) affects both y and **x**. This induces endogeneity in the panel since the regressors are now correlated with the error term in the production function. To be more precise, if the **f** factors are left unaccounted for, an omitted variables bias is introduced.

As such, the main, distinctive features of our empirical specification are: (i) the heterogeneity across regions of both the factors of production and the unobservables, (ii) the endogeneity of the observables caused by the unobserved common factors (omitted variables bias) and, (iii) the potential for non-stationary observables and unobservables.

The all-encompassing nature of our empirical specification is made evident by the fact that if we assume that the technology coefficients are identical across countries ($\boldsymbol{\beta_i}=\boldsymbol{\beta}$) and the unobserved common factors are proxied by time effects ($\boldsymbol{\lambda'_i f_t} = \tau_t$) then we have the two-way Fixed Effects estimator that has been extensively utilized in the relevant literature (see, among others, Dettori et al. (2012), Ladu (2012), Marrocu and Paci (2011), Schatzer et al. (2019) and Siller et al. (2021)).

In this paper we take an agnostic view in terms of the estimator that is ultimately to be employed and we let the data "speak". In particular, we compare and contrast the results of alternative estimation approaches and modelling assumptions related to TFP. That is, by using a wide array of estimators ranging from those that are very restrictive in the way they allow TFP to evolve over time to the ones that are very flexible and allow a nonlinear time path, we can compare and contrast the estimated factor coefficients and subject the error term to formal testing. In particular, by testing whether the residuals are `well-behaved' (i.e. stationary and weakly cross-sectionally dependent) we can ultimately decide which empirical model emerges as the appropriate specification to describe the panel dataset at hand.

Another focal point in the literature focusing on estimating production functions is the issue of endogeneity. In our case, endogeneity can arise due to the presence of unobserved common factors and due to feedback effects, i.e. reverse causality. While the first point can be readily tackled via the Common Correlated Effects estimators that were previously introduced, with regards to reverse causality an alternative approach is needed.

The main tool that has been extensively utilized in the literature is the established GMM approach, pioneered by Arellano and Bond (1991) and Blundell and Bond (1998). However, the GMM-type estimators –which rely on the use of own instrumentation in order to circumvent the issue of finding valid external instruments- cannot be utilized in panels which are characterized by non-stationarity and (strong) cross-sectional dependence (see Pesaran and Smith (1995) and Eberhardt and Teal (2019)). As such, we resort to Granger-type weak exogeneity tests in order to ensure that our estimated factor coefficients and, consequently, our productivity estimates are not driven by reverse causality.

Overall, our analysis indicates that there exist significant differences when moving away from pooled estimators to ones that allow for factor coefficient heterogeneity and between models that account for and those that disregard the potential effects of cross-sectional dependence. We draw the following set of conclusions from our empirical analysis:

firstly, technology heterogeneity matters. That is, the commonly adopted assumption of production factor coefficients being identical across regions is rejected by the data. The widely used two-way Fixed Effects estimator, usually with constant returns to scale imposed, is found to be misspecified on the basis of formal residual diagnostics giving rise to concerns regarding potentially spurious results. Moreover, the results obtained by estimators that allow for parameter heterogeneity are characterized by well-behaved residuals, indicating that we do not face concerns of model misspecification.

secondly, accounting for cross-section dependence in the form of unobserved common factors that exert a differential impact across regions has an important bearing on the empirical results. In particular, the heterogeneous capital coefficients are –on average- lower compared to the estimates obtained under the assumption of cross-section independence.

thirdly, our results indicate that the assumption of constant returns to scale is rejected in our panel dataset. As such, approaches that rely on constant returns (like the deterministic sources of growth analysis a la Solow) in order to obtain TFP estimates may result in misleading estimates.

finally, weak exogeneity tests indicate that our results can be reasonably assumed to represent production functions only in the case of the CCE estimators, again highlighting the need to take into account the time series properties of the data and cross-sectional dependence.

Overall, one overarching theme stemming from our results is the importance of heterogeneity. In particular, it is of high importance to account for region-specific characteristics when estimating the production structure and analyzing economic performance, both in terms of output and productivity, at the regional level.

References

Arellano, M and Bond, S. (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". *Review of Economic Studies*. 58(2): 277-297

Beugelsdijk, S., Klasing, M.I. and P. Milionis (2018). "Regional economic development in Europe: the role of total factor productivity." *Regional Studies*, Vol. 52(4): 461-476

Blundell, R. and Bond, S. (1998). "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1): 115–143

Burger, M. J., Kounetas, K., Napolitano, O. and S. Stavropoulos (2021) "Do innovation and human capital actually narrow the technology gap? Champions and laggards of European regional productive performance"." *Regional Studies, forthcoming*

Dettori, B., Marrocu, E. and R. Paci. (2012). "Total factor productivity, intangible assets and spatial dependence in the European Regions." *Regional Studies*, Vol. 46: 1401–1416.

Eberhardt, M. and Teal, F. (2020). "The magnitude of the task ahead: Macro implications of heterogeneous technology." *Review of Income and Wealth*, Vol. 66(2): 334-360

Filippetti, A and Peyrache, A. (2015) "Labour Productivity and Technology Gap in European Regions: A Conditional Frontier Approach." *Regional Studies*, Vol. 49(4): 532-554

Kapetanios, G., Pesaran, M. H. and T. Yamagata. (2011). "Panels with nonstationary multifactor error structures." *Journal of Econometrics*, Vol. 160(2): 326–348.

Ladu, M. G. (2012). "The relationship between total factor productivity growth and employment: some evidence from a sample of European Regions." *Empirica*, Vol. 39: 513–524.

Mankiw, N. G., Romer, D. and D. N. Weil, (1992). "A contribution to the empirics of economic growth." *The Quarterly Journal of Economics*, Vol. 107(2): 407-437.

Männasoo, K., Hein, H., and R. Ruubel (2018). "The contributions of human capital, R&D spending and convergence to total factor productivity growth." *Regional Studies*, Vol. 52(12): 1598–1611

Marrocu, E., and Paci, R. (2011). "They arrive with new information. Tourism flows and production efficiency in the European Regions." *Tourism Management*, Vol. 32:750–58

Marrocu, E., Paci, R. and S. Usai (2013). "Productivity growth in the old and new Europe: The role of agglomeration externalities." *Journal of Regional Science*, Vol. 53: 418–42.

Pesaran, M. H. (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, Vol. 74(4): 967–1012.

Pesaran, M. H. and Smith, R. P. (1995). "Estimating long-run relationships from dynamic heterogeneous panels." *Journal of Econometrics*, Vol. 68(1): 79–113.

Rogge, N. (2019). "Regional productivity growth in the EU since 2000: Something is better than nothing." *Empirical Economics*, Vol. 56 (2): 423-444

Schatzer, T., Siller, M., Walde, J., and G. Tappeiner (2019). "The impact of model choice on estimates of regional TFP." *International Regional Science Review*, Vol. 42(1): 98–116

Siller, M. Schatzer, T., Walde, J. and Tappeiner, G. (2021) "What drives total factor productivity growth? An examination of spillover effects." *Regional Studies*, Vol. 55(6): 1129-1139

Keywords: *regional economic development; total factor productivity; European regions; technology heterogeneity; common factor model; nonstationary panels*

JEL codes: C23, O47, O11