A note on joint estimation of scale economies and productivity growth parameters

Efthymios G. Tsionas\textsuperscript{a,*}, John Loizides\textsuperscript{b}

\textsuperscript{a}Council of Economic Advisers, Ministry of National Economy, 5 Nikis Street, Constitution Square, 10180 Athens, Greece
\textsuperscript{b}Department of International and European Economic Studies, Athens University of Economics and Business, Athens, Greece

Received 4 August 1998; accepted 23 March 2000

Abstract

In applied research it is common to estimate complete production models, obtain measures of scale economies and total factor productivity growth and then regress such measures on exogenous variables. Such procedures result in inconsistent estimates of technological as well as regression parameters: estimation of total factor productivity and scale economies in the first step, does not take into account that these measures depend on exogenous variables in the second step. Therefore, their dependence on exogenous variables is not properly taken into account in the first step. The study proposes to estimate jointly the cost function, the share equations as well as total factor productivity and scale economies measures, using full system estimation to account for all the restrictions implied by their endogeneity. The approach is illustrated using data from British, French, and German railways. \copyright~2001 Elsevier Science B.V. All rights reserved.

Keywords: Scale economies; Productivity growth; Complete production models; Duality; Econometric estimation of technology characteristics; European railway systems

1. Introduction

In recent years there has been an explosion of interest in estimating technical characteristics of production and cost structure, using complete production models. The purpose of formulating and estimating complete econometric models of production is to assess the extent of substitutability of certain factors of production and evaluate the magnitude of scale economies and productivity growth.\textsuperscript{1} However, in many situations of interest researchers believe that scale economies and productivity growth have been affected by certain exogenous variables, and then they seek to test empirically whether or not this is the case. To that end, the applied researcher starts with a cost function

\[ C(p_t; y_t, t) = C(p_{1t}, p_{2t}, \ldots, p_{mt}, y_t, t), \quad t = 1, \ldots, T, \quad (1) \]

where \( p_{it} \) denotes the price of input \( i \) at date \( t \), \( y_t \) denotes output and \( t \) denotes technological progress. Using standard duality theory [9,10], one can derive input demand functions of the form

\[ x_{it} = \frac{\partial C(p_t, y_t, t)}{\partial p_{it}}, \quad i = 1, \ldots, n. \quad (2) \]
Scale economies (SE) and total factor productivity growth (TFP) are obtained as
\[
SE_t = \left( \frac{\partial C(p_t, y_t, t)}{\partial y_t} \frac{y_t}{C(p_t, y_t, t)} \right)^{-1} = \frac{\partial \ln C(p_t, y_t, t)}{\partial \ln (y_t)}
\]
and
\[
TFP_t = \frac{\partial C_y}{\partial t} Y_t = \frac{\partial \ln y_t}{\partial t}.
\]
It is common to estimate the complete system composed of (1) and (2) using FIML or iterative Zellner’s technique (IZEF), obtain estimates of SE\(_t\) and TFP\(_t\) at the final estimates and then regress them on certain exogenous variables \(z_t\):
\[
\begin{align*}
\hat{SE}_t &= z_t' \alpha + v_{1t}, \\
\hat{TFP}_t &= z_t' \beta + v_{2t}.
\end{align*}
\] (5)
In (5), \(\alpha\) and \(\beta\) denote coefficient estimates, \(v_{1t}\), \(v_{2t}\) denote error terms and a hat denotes evaluation at the final IZEF estimates. A popular way to estimate the practical importance of certain quantitative and qualitative variables on the level of technical efficiency is to estimate first a technical efficiency index from the residuals of a production function, which includes as exogenous variables some basic inputs of production\(^2\) (capital, labor, energy, land, fertilizer, etc., see for example [11–13] among others) and second to regress technical efficiency or productivity growth on such variables as education, farm assets, credit, etc. The argument has been advanced (e.g. [14]) that although it would be more meaningful to include such variables as formal inputs in the production function, this would be undesirable because it raises some conceptual problems regarding substitutability: For example, it makes no sense to argue that there is substitution between fertilizer and education – in an agricultural economics context, or between average trip length and labor – in a railway economics context. For that reason, there is considerable incentive for defending the “regression approach”. However, there is little reason to support the two-step approach that is commonly used in the literature. In fact, the two-step approach suffers from several problems:

(i) Regressions like (5) ignore parameter uncertainty and simply compute the dependent variable at the final estimates, without taking proper account of their variability. As a matter of fact, these regressions ignore the fact that TFP or SE have been computed in the first round, under the assumption that they do not depend on exogenous variables, i.e. that they are strictly exogenous variables. In the second round, however, this assumption is (silently) violated.

(ii) Regressions like (5) are problematic because they introduce additional variables \(z_t\) into the problem in an ad hoc manner. Formally, variables in \(z_t\) should enter (1) as inputs, however, in that case one would have substitution between \(z_t\) and the existing inputs (unless variables in \(z_t\) enter as fixed factors in the short run). This may not always be a reasonable approach to the problem. Also, (5) is problematic because it does not have a direct interpretation.

(iii) Regressions (5) in conjunction with (3) and (4) suggest that prices or output are endogenous and should depend on \(z_t\). If that is the case, joint estimation of (1) and (2) by IZEF will yield inconsistent results, and the same is true for OLS applied to (5).

A way out of these problems is to estimate (1) and (2) jointly with (3) and (4). In general, (3) and (4) are functions of prices, output, time and the structural parameters of the cost function, so joint estimation introduces additional restrictions (on top of homogeneity, symmetry, etc.) on structural parameters imposed by the endogeneity of TFP and SE.

2. Application to the translog technology

To fix ideas, assume that the cost function can be described by the generalized translog technology, in which case it follows that
\[
\ln C(p; y) = \alpha_0 + \alpha_y \ln y + \frac{1}{2} \gamma_{yy}(\ln y)^2 \\
+ \sum_{i=1}^{n} \beta_i \ln p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln p_i \ln p_j
\]

\[ C(p, y) \]
\[ \sum_{i=1}^{n} \beta_t = 1, \]
\[ \sum_{j=1}^{n} \gamma_{ij} = \sum_{i=1}^{n} \delta_{yi} = \sum_{i=1}^{n} \theta_{iT} = 0, \quad \gamma_{ij} = \gamma_{ji}. \]

Scale economies are measured using the concept of inverse output cost elasticity:

\[ SE_t = \left( \alpha_y + \alpha_{y} \ln y_t + \sum_{i=1}^{n} \delta_{yi} \ln p_{it} + \alpha_{yt} t \right)^{-1}. \]

The rate of growth in total factor productivity is given by

\[ TFP_t = -\frac{\theta_T + \theta_{TT} t + \sum_{i=1}^{n} \theta_{i} \ln p_{it} + \alpha_{yt} \ln y_t}{\alpha_y + \alpha_{y} \ln y_t + \sum_{i=1}^{n} \delta_{yi} \ln p_{it} + \alpha_{yt} t}. \]

Append ing normal disturbances to (5) – as they apply to (8) and (9), produces stochastic equations that can be estimated jointly with (1) and (2). The appropriate estimator in this context is the nonlinear three-stage least-squares estimator (NL3S).

3. Empirical application

In this section, the new technique will be applied to productivity growth and scale economies in three European Union railway systems, namely French, German and British railways. Analysis of productivity growth and the cost structure of railway systems has been given particular emphasis in the European Union in recent years due to mounting subsidies and alleged inefficiencies in the sector, see [8, 15–18, 20, 22]. The translog technology with three inputs (K, L and E) is used. The nature of the data and their construction are detailed in Appendix A. There has been active interest in the literature regarding the issue of dependence of scale economies and productivity growth on railway technical characteristics, mainly passenger or freight density and the degree of electrification of the system. The present study will not deviate from this practice, and to that effect Eq. (5) are further specified as

\[ \hat{SE}_t = b_0 + b_1 EL_t + b_2 DEN_t + \nu_{1t}, \]
\[ \hat{TFP}_t = c_0 + c_1 EL_t + c_2 DEN_t + \nu_{2t}, \]

where \( EL_t \) denotes percentage of electrification of the railway network and \( DEN_t \) denotes train density–essentially a capacity measure. The joint estimation approach appends the following two equations to the system of cost function and share equations:

\[ \left( \alpha_y + \alpha_{y} \ln y_t + \sum_{i=1}^{n} \delta_{yi} \ln p_{it} + \alpha_{yt} t \right)^{-1} = B_0 + B_1 EL_t + B_2 DEN_t + V_{1t}, \]
\[ -\frac{\theta_T + \theta_{TT} t + \sum_{i=1}^{n} \theta_{i} \ln p_{it} + \alpha_{yt} \ln y_t}{\alpha_y + \alpha_{y} \ln y_t + \sum_{i=1}^{n} \delta_{yi} \ln p_{it} + \alpha_{yt} t} = C_0 + C_1 EL_t + C_2 DEN_t + V_{2t}, \]

where \( V_{1t} \) and \( V_{2t} \) are error terms and \( B_0, B_1, B_2, C_0, C_1 \) and \( C_2 \) are coefficients to be estimated. The disturbances of the cost function, share equations and (11) and (12) are assumed to be jointly normally distributed with zero mean and positive-definite covariance matrix. This approach allows the researcher to capture the cross-equation dependence in error terms. The regression approach, on the contrary, is forced to assume that \( V_{1t} \) and \( V_{2t} \) have to be statistically independent of cost function and share equation disturbances.

Before proceeding with the estimation, it must be noted that the price of energy is estimated across electricity, diesel and lubricants following standard practice in the literature of railway economics. This makes it dependent on the amount of electrification, since the underlying energy prices are different. This

---

3 This is based on passenger traffic density (defined as passenger km to length of lines worked) and freight traffic density (defined as freight train km to length of lines worked).
drawback, however, is corrected by joint estimation, because (3) and (4) make output and factor prices, functions of the available exogenous variables. Therefore, a simultaneous equation bias does not arise because energy prices are automatically instrumented out by the nature of joint estimation. For this to happen, of course, electrification must be included in the exogenous variable list in $z_t$.

Table 1
Parameter estimates for the traditional and joint estimation approaches.*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Germany</th>
<th>France</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Traditional approach</td>
<td>Joint estimation approach</td>
<td>Traditional approach</td>
</tr>
<tr>
<td>$\beta_K$</td>
<td>0.604 (2.25)</td>
<td>7.036 (1.16)</td>
<td>−0.222 (0.19)</td>
</tr>
<tr>
<td>$\beta_L$</td>
<td>0.443 (1.536)</td>
<td>24.62 (1.44)</td>
<td>1.026 (0.89)</td>
</tr>
<tr>
<td>$\gamma_{KK}$</td>
<td>0.026 (4.27)</td>
<td>0.019 (1.21)</td>
<td>0.092 (10.5)</td>
</tr>
<tr>
<td>$\gamma_{KL}$</td>
<td>−0.197 (3.18)</td>
<td>−0.068 (1.81)</td>
<td>−0.087 (9.94)</td>
</tr>
<tr>
<td>$\gamma_{LL}$</td>
<td>0.062 (8.32)</td>
<td>−0.038 (0.34)</td>
<td>0.113 (12.36)</td>
</tr>
<tr>
<td>$\delta_{Ks}$</td>
<td>−0.042 (1.84)</td>
<td>−0.563 (1.13)</td>
<td>0.022 (0.21)</td>
</tr>
<tr>
<td>$\delta_{Ls}$</td>
<td>0.0068 (0.276)</td>
<td>−1.98 (1.42)</td>
<td>−0.027 (0.28)</td>
</tr>
<tr>
<td>$\theta_{KT}$</td>
<td>0.0020 (9.23)</td>
<td>0.0018 (2.02)</td>
<td>0.014 (12.8)</td>
</tr>
<tr>
<td>$\theta_{LT}$</td>
<td>−0.04 (16.14)</td>
<td>−0.0066 (2.65)</td>
<td>−0.015 (13.3)</td>
</tr>
<tr>
<td>$\alpha_t$</td>
<td>−48.01 (1.14)</td>
<td>−37.01 (1.40)</td>
<td>75.27 (1.85)</td>
</tr>
<tr>
<td>$\alpha_{yt}$</td>
<td>4.16 (1.15)</td>
<td>3.20 (1.40)</td>
<td>−6.43 (1.85)</td>
</tr>
<tr>
<td>$\theta_T$</td>
<td>0.330 (1.67)</td>
<td>0.128 (1.05)</td>
<td>0.269 (1.25)</td>
</tr>
<tr>
<td>$\theta_{TT}$</td>
<td>−0.0011 (3.38)</td>
<td>−0.111 (1.06)</td>
<td>0.0015 (3.5)</td>
</tr>
<tr>
<td>$\alpha_{yt}$</td>
<td>−0.028 (1.66)</td>
<td>−10.83 (1.05)</td>
<td>−0.024 (1.29)</td>
</tr>
</tbody>
</table>

*Absolute t-values appear in parentheses beneath parameter estimates. “Traditional approach” refers to estimation of cost function and share equations accompanied with separate simple regressions of TFP and SE on technical characteristics. Results for this approach have been obtained using Full Information Maximum Likelihood. “Joint estimation approach” involves joint estimation of cost function, share equations as well as SE and TFP equations. Parameters estimates for this approach have been obtained using iterative nonlinear three-stage least squares. Results were obtained using TSP 4.3 (subroutines FIML and LSQ). Standard errors were computed using the quadratic form of analytic first derivatives. Equation $R^2$’s were in excess of 0.9 in all cases.
Notice that (11) and (12) are implicit nonlinear-in-the-parameters equations. Statistical estimates for the traditional and joint estimation approaches (using FIML and NL3S respectively) are reported in Table 1. It is clear that some parameter estimates are greatly sensitive to inclusion of TFP and SE equations in the system of share and cost functions. These parameters are mainly output- and time-related parameters ($\alpha_y$, $\alpha_{yy}$, $\theta_T$ and $\theta_{TT}$). This sensitivity implies that specifying a complete system that includes TFP and SE equations makes an important difference.

Results for (10) are reported in Table 2. The most striking result from Table 2 is that while, according to the traditional approach, electrification is highly statistically significant in explaining the time-series behavior of productivity growth and scale economies, this is simply an artifact since estimates of $b_1$ and $b_2$ from the joint estimation approach are not statistically significant at conventional significance levels. Therefore, using the straightforward traditional approach could prove dangerous from the policy viewpoint, as this approach would seem to suggest further electrification in Germany and de-electrification of railway lines in France and the UK in order to increase productivity growth. These results, however, are based on the strong statistical significance of $b_1$ and $b_2$ which—according to joint estimation approach—is not supported by the data. It must also be noted that joint estimation seems to lower the overall significance of the model. This fact may be due to the fact that too many parameters have to be estimated jointly based on a finite sample of data. This, of course, implies that joint estimation of factor demands, scale economies and productivity growth is not a panacea, but its results very much depend on the available data and the model.

Table 2
Dependence of scale economies and productivity growth on technical characteristics.a

<table>
<thead>
<tr>
<th>Country</th>
<th>Explanatory variable</th>
<th>Constant</th>
<th>Electrification</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Traditional</td>
<td>Joint estimation</td>
<td>Traditional</td>
</tr>
<tr>
<td>I. Scale economies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td>-0.549</td>
<td>0.025</td>
<td>-0.0066</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.8)</td>
<td>(0.18)</td>
<td>(4.53)</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td>32.62</td>
<td>0.074</td>
<td>-0.161</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.75)</td>
<td>(0.016)</td>
<td>(1.45)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td>-0.302</td>
<td>0.435</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.69)</td>
<td>(0.15)</td>
<td>(6.26)</td>
</tr>
<tr>
<td>II. Productivity growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td>-0.12</td>
<td>-1.38</td>
<td>0.00038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.57)</td>
<td>(1.66)</td>
<td>(6.1)</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td>0.184</td>
<td>0.092</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.02)</td>
<td>(0.028)</td>
<td>(2.73)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td>0.021</td>
<td>-0.029</td>
<td>-0.0022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.90)</td>
<td>(0.18)</td>
<td>(6.91)</td>
</tr>
</tbody>
</table>

*aAbsolute t-values are reported in parentheses beneath parameter estimates. The “traditional approach” involves computation of SE and TFP at FIML parameter estimates (FIML applied to the cost function and share equations) and simple regressions on electrification and density. “Joint estimation” involves equations for SE and TFP which are estimated jointly with the cost function and share equations.*
4. Conclusions

The paper provided a practical way to incorporate total factor productivity growth (TFP) and scale economies (SE) equations in a complete cost function–share equation system and examine their dependence on exogenous variables. Such systems can be estimated using NL3S. This approach is a consistent alternative to the traditional practice of estimating cost functions and share equations, computing TFP and SE at FIML or IZEF parameter estimates and regressing them on exogenous variables using OLS. The latter approach suffers from several econometric problems, including inconsistency of parameter estimates due to simultaneity biases. In an empirical application that involved productivity measurement and SE in three European railways (French, German and British) the study documented significant differences in the way exogenous variables have affected TFP and SE. The traditional approach implied statistically significant parameters of TFP and SE regressions on electrification. On the contrary, the approach developed in the present paper showed that this is an artifact. The proposed approach can be useful in several fields, including agricultural applied economics where the problem of determinants of productivity growth and scale economies arises on a routine basis.

Acknowledgements

The authors wish to thank Dimitris Christophoulou for useful comments.

Appendix A. Data construction and sources

Annual data for three countries of the European Union have been used, namely France, Germany, and the United Kingdom, over the period 1969–1992. The cost of capital is user's cost defined as the sum of interest and depreciation costs. Capital prices were obtained by dividing user costs by the capital stock. Capital stock includes land and fixed installations, transport stock and other equipment. The source of data is the International Union of Railways (UIC, 1969–1993 [19]). For France, and the UK, we have used the price deflator of transport equipment (obtained from the National Accounts of OECD [20]) as the price of capital, because frequent re-evaluations of the capital stock distorted capital prices.

The quantity of labor is the number of employees. Labor costs have been obtained as the sum of total wages and salaries paid, including benefits and pensions. Labor costs divided by the number of employees gives the price of labor. Electricity, diesel oil and lubricants expressed in equivalent thermal units consumed, give the energy input. The price of energy is defined as the energy cost divided by equivalent thermal units. Since the International Union of Railways Statistics does not report energy values for 1990 and 1992, these were taken from OECD [20] adjusted for conformity to those used for the period 1969–1990. The definition of output is dictated, to a large extent, by the availability of data. We use as output the sum of kilometric passengers and kilometric freight as reported by the International Union of Railways (UIC, [19]). See also Nash [21]. The revenue-weighted sum was used as well, but the results were robust to the choice.

References


