ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 4/2010

Stochastic Frontiers using a Fixed-effect Vector Decomposition Approach with an Application to ICT and Regional Productivity in Spain

William Greene, Luis Orea, Alan Wall

Departamento de Economía

Universidad de Oviedo

Available online at: www.unioviedo.es/economia/EDP.htm

Stochastic frontiers using a fixed-effect vector decomposition approach with an application to ICT and regional productivity in Spain

William Greene
Dept. of Economics, Stern School of Business, New York University

Luis Orea
Oviedo Efficiency Group
Department of Economics, University of Oviedo

Alan Wall
Oviedo Efficiency Group
Department of Economics, University of Oviedo

December 10, 2010

Abstract

Fixed effects estimators will generally produce imprecise estimates when the data contains variables with relatively low within variance. In a production setting, this can lead to unreliable estimates of key parameters. These problems carry over to production frontiers where technical inefficiency is estimated on the basis of the unit effects (Schmidt and Sickles, 1984) or through the use of true FE stochastic frontier (Greene, 2004; 2005). Plümper and Troeger (2007) propose what they label a Fixed Effect Vector Decomposition (FEVD) estimator which may permit greater precision in estimating variables with low within variance and they provide conditions under which their estimator performs better than the fixed effects estimator. In this paper we extend the FEVD estimator to a frontier setting. This allows the possible advantages of the FEVD estimator to be incorporated into Greene's true FE frontier model. In an empirical application of the estimator we analyze the effects of ICT capital on regional productivity in Spain using a balanced panel dataset at provincial level over the period 1986-2006. This data contains several variables with relatively low within variance, and the FEVD frontier provides much more plausible estimates of key parameters than the FE estimator or the true FE frontier estimator.

JEL: C23, O47, R10.

Keywords: Stochastic frontier, fixed effects vector decomposition, ICT, regional productivity.

1. Introduction

One of the main advantages of using panel data is that researchers can control for unobserved heterogeneity across units. Among panel data techniques, the fixed effects estimator has proven particularly popular as it permits arbitrary correlation between the (time-invariant) unobserved heterogeneity and the time-varying regressors. In the efficiency literature, fixed effects estimation of the stochastic frontier has a long tradition going back to Schmidt and Sickles (1984). In their model, the "fixed effect" was interpreted as a unit-specific inefficiency term. One of the criticisms of this interpretation is that the unit-specific term includes not only inefficiency but also any unobserved time-invariant cross-unit heterogeneity. To separate unobserved heterogeneity from inefficiency per se, Greene (2004) proposed the "true" fixed effects model which basically incorporated unit dummy variables into the standard normal-half-normal stochastic frontier model.

While these fixed effects stochastic frontier models make weak assumptions on the unit-specific effects, in that they can be arbitrarily correlated with the regressors, and have the virtue of being relatively easy to implement, they share the same drawbacks as any fixed effects model. For example, it is well-known that the effects of time-invariant variables cannot be estimated in these models as only within variance is used. Less well-recognized, as noted by Plümper and Troeger (2007), is the lack of precision of the FE estimator in estimating the effect of variables that have little within variance ("rarely changing" variables, in their terminology), which can lead to unreliable point estimates of the (deterministic) production function. This in turn will lead to unreliable estimates of relative technical inefficiency as this is measured by the distance to an estimated production frontier. Moreover, the presence of variables with low within variance is a problem likely to be found in many empirical applications using both macro and micro data sets in production economics. For instance, macro data sets contain aggregates which often move slowly over time. Firm-level data also often contain rarely changing variables, such as labour inputs farm data or capital stock in transport or electricity firms.

To address this issue, Plümper and Troeger (2007) proposed, in a non-frontier framework, a fixed effects vector decomposition (FEVD) estimator which may provide more precise estimates in a root mean-squared error sense than the FE model when estimating the effects of variables with low within variability. While some recent applications of this technique have appeared in fields such as sociology, energy and economic psychology, as far as we aware there have been no applications as yet to production economics.[1] One of the contributions of this paper will be to provide an empirical application that allows us to check whether the FEVD estimator works in practice in traditional production economics analysis.

In addition, and following the steps outlined in Plümper and Troeger (2007), in this paper we extend their model to a frontier framework and introduce an FEVD stochastic frontier estimator that can be viewed as a three-stage version of the "true" fixed effects frontier estimator proposed by Greene (2004). In particular, our estimator takes advantage of the insights of the FEVD model to permit an extension of the true FE frontier where estimation of rarely-changing variables may be more precise. The model has three stages. In the first stage, the true FE frontier is estimated. In the second stage we use the FEVD procedure to decompose the estimated unit-specific effects into observable and unobservable components. This is achieved by regressing the estimated unit effects from the first stage on variables with little or no within variation, thereby taking advantage of between-unit information. The estimated residuals from this regression are taken to represent truly unobserved unit heterogeneity. Finally, in the third stage the full frontier model without the unit effects is estimated by pooled OLS where the regressors include all time-varying and time-invariant regressors and the residuals from the second stage.

To illustrate the potential advantages of the FEVD frontier over previous FE models we provide an empirical application of our proposed model using production data from Spanish provinces. The data set comprises annual observations on the 50 Spanish provinces covering the period 1986-2006 and includes information on different types of capital (public, private and information and communications technology) and labour (skilled and unskilled). This data is particularly appropriate for the objectives of our study in that it contains several variables which move relatively slowly over time in the sense that have relatively low within-to-between variance ratios. This will render FE estimates imprecise and makes a case for an FEVD approach. We estimate an FEVD stochastic production frontier in order to check the influence of information and communications technology (ICT) capital on productivity and efficiency. Our results show that the FE estimator and the true FE frontier yield implausible estimates of key parameters of the

[1] See Worrall (2008), Carley (2009) and Boyce (2009) for recent applications of this technique.

regional production frontier. When observable heterogeneity is corrected for, the estimated parameters using the FEVD frontier are much more in line with expectations.

The contributions of this paper are therefore methodological and empirical. First, we extend the FEVD technique to a stochastic frontier framework, introducing an estimator which we believe can provide a useful extension to existing FE frontiers when the data contain slowly changing variables. Second, at an empirical level there are very few applications of the FEVD technique in economics and none to production functions or regional productivity.

2. The fixed effect vector decomposition model

The attraction of FE models is that they allow for arbitrary correlation between the unobserved individual heterogeneity and the observable inputs. The downside is that they only use within-unit variation in their estimation, grouping all between-unit variation (i.e., unobserved time-invariant heterogeneity across units) into the unit fixed effect, with the consequence that time-invariant variables cannot be estimated (Hsiao, 2003). This may or may not be a problem depending on the objectives of the study. However, a second problem with FE models is that they are inefficient in estimating variables with little within variance. Plümper and Troeger (2007) refer to these as "rarely changing" variables, providing several examples where within variance would be expected to much smaller than between variance. FE models can estimate coefficients for such variables but the standard errors will be large as their explanatory power will be captured largely by the estimated fixed effects. However, the problem is not just low significance of estimated parameters, as "point estimates are also unreliable since the influence of the error on the estimated coefficients becomes larger as the inefficiency of the estimator increases" (Plümper and Troeger, 2007).

As an alternative to FE, Plümper and Troeger (2007) propose what they label the fixed effect vector decomposition (FEVD) estimator, which is based on the FE estimator, and they provide conditions under which the FEVD estimator will be preferable to FE. To illustrate their procedure, assume we wish to estimate the following provincial production function:

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \varepsilon_{it} \quad (1)$$

where the x variables are time-varying and include observable inputs such as capital and labour, the z variables are time-invariant and would include characteristics such as the region to which the province belongs, and ε_{it} is the error term. α_i represents individual (province) heterogeneity and captures the effect of unobserved time-invariant provincial characteristics.

Using the within transformation, the FE estimator removes the individual effects α_i and the time-invariant variables z.:

$$y_{it} - \overline{y}_i = (\alpha_i - \alpha_i) + \beta_k \sum_{k=1}^{K}(x_{kit} - \overline{x}_{ki}) + \gamma_m \sum_{m=1}^{M}(z_{mi} - \overline{z}_{mi}) + (\varepsilon_{it} - \overline{\varepsilon}_i) \quad (2)$$

where

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}\,, \quad \overline{x}_{ki} = \frac{1}{T}\sum_{t=1}^{T} x_{kit}, \quad \overline{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$$

Plümper and Troeger (2007) note that the "estimated unit effects" $\hat{\alpha}_i$ do not equal the unit effects α_i in the DGP. Instead, denoting the fixed effects estimate of (2) as b_k^{FE} the estimated unit effects are:

$$\hat{\alpha}_i = \overline{y}_i - \sum_{k=1}^{K} b_k^{FE}\, \overline{x}_{ki} \quad (3)$$

Thus, the $\hat{\alpha}_i$ include not only the true unobserved unit-specific effects, α_i, but also the effect of the observed time-invariant variables, z, and the unit-means of the time-varying inputs, $\overline{x}_{ki}$.

The FEVD procedure proposed by Plümper and Troeger (2007) involves three stages. In the first, the FE model (1) is run to obtain estimates of the unit effects, $\hat{\alpha}_i$. In the second stage, these unit effects are regressed on the observed time-invariant and rarely changing variables, z. In doing so, the unit effects are decomposed into an unexplained part, θ_i, and a part explained by the available between-unit information contained in z:

$$\hat{\alpha}_i = \sum_{m=1}^{M} \gamma_m z_{mi} + \theta_i \quad (4)$$

In the third stage, the full model is run using pooled OLS without the unit effects but including the true unobservable component of α_i represented by the residual from the second-stage regression, θ_i:

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \delta\theta_i + \varepsilon_{it} \quad (5)$$

While the z variables may have been correlated with α_i, they are uncorrelated with θ_i by assumption.

A key issue here is what variables to include in the z vector in the second stage. Strictly time-invariant characteristics will obviously be included. Variables with sufficiently low within-variance should also be included but identifying such variables may not be so obvious. Plümper and Troeger (2007) carry out Monte Carlo simulations to provide the conditions under which a variable should be included in the second stage. Using the root mean squared error as their criterion, they find that the decision to treat a rarely changing variable as time-varying (x) or time-invariant (z) depends on the correlation between the variable and the unobserved heterogeneity and the ratio of the between to within variance. In general, for a given correlation, the greater the between-to-within ratio, the better the relative performance of the FEVD estimator. For a correlation of 0.3 between the variable and the unit heterogeneity, a between-to-within ratio of approximately 1.7 is sufficient for the FEVD estimator to be superior to FE. When the correlation rises to 0.5, the between-to-within ratio rises to about 2.8. While the correlation between the variable and the unit heterogeneity is unobservable, the inclusion of additional variables in z will reduce the potential for correlation, and Plümper and Troeger (2007) suggest that a between-to-within ratio of 2.8 is sufficient to justify the inclusion of the variable in question in the second stage. Moreover, even the within variance is large, as long as the between variance is much larger the FEVD was found to perform better on average than FE (though the absolute advantage in reliability will be smaller).

Recent research has raised questions about the validity of the FEVD model when only time-invariant variables are included in the second stage. In particular, Breusch *et al*. (2010) and Greene (2011) have shown that in this case the FEVD estimator simply reproduces the fixed effects estimates for the time-varying variables and produces spurious efficiency gains. On this basis, the FEVD estimator becomes a useful tool only when slowly changing variables are included in the second stage, which will be the case in our empirical application. Indeed, it is the very presence of slowly changing variables and the fact that their estimates in a FE setting can be imprecise in a mean-squared error (MSE) sense that may justify the use of the FEVD as a potentially interesting alternative estimator. At issue here is a trade-off between bias and efficiency. The FE estimator is consistent, as rarely changing variables are still time-varying, but will have high variance. The FEVD estimator, on the other hand, is biased but has low variance and under the conditions discussed above it will be more reliable than the consistent but high-variance FE estimator.[2]

On the other hand, Greene *et al*. (2010) show that when the unit means of all the time-varying variables are included in the second stage the FEVD reproduces the OLS estimates of the time-varying and time-invariant parameters. This provides an alternative interpretation of the FEVD estimator, i.e. as we incorporate the unit means of the rarely-changing variables we move away from FE towards OLS. The importance of the between-to-within variance as a criterion for the inclusion of time-varying variables in the second stage becomes clear under this interpretation as the aim is to maximize the use of between variation for those variables with relatively low within variation.

Yet another alternative is to use the Hausman and Taylor (1981) instrumental variable approach. Breusch *et al*. (2010) provide an IV interpretation of the FEVD in a time-invariant variable context (i.e., without rarely-changing variables) and argue that the Hausman-Taylor estimator is the most relevant direct competitor with the FEVD estimator. On the basis of a Monte-Carlo study they conclude that neither estimator completely dominates the other in a MSE sense, with FEVD performing better when endogeneity is not too severe. Greene (2011), on the other hand, sees no benefit to the FEVD approach in the time-invariant case but points to the comparison with Hausman-Taylor as being of possible relevance in the presence of rarely-changing variables. In any case, the performance of the Hausman-Taylor estimator will depend on the availability of enough time-varying exogenous variables in the model and the quality of the instruments available. There is also the difficulty of deciding which variables to treat as endogenous.[3]

3. Efficiency measurement in an FEVD framework

Fixed effect estimation of the frontier model was introduced by Schmidt and Sickles (1984). In a production function setting, the model can be written as:

[2] In Monte Carlo studies, Plümper and Troeger (2007) found that the FEVD estimator consistently outperformed both RE and OLS, so we will not consider these estimators.

[3] Note that when none of the time-invariant variables are endogenous and all the time-varying variables are permitted to be endogenous, the FEVD and Hausman-Taylor estimators are equivalent – see Breusch *et al*. (2010).

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} - u_i$$
$$= \alpha_i + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} \tag{6}$$

where v_{it} is a symmetric random error term, $u_i \geq 0$ represents technical inefficiency, $\alpha_i = \alpha - u_i$ and the technical inefficiency is estimated as $\hat{u}_i = \max_i(\hat{\alpha}_i) - \hat{\alpha}_i$.

In this model the u_i are treated as unit-specific constants and the model can be estimated by the usual fixed effects (FE) estimator or within transformation. In an FEVD framework this normalization can be carried out using either the original fixed effects obtained in the first stage or using the adjusted fixed effects, i.e. the residuals from the second stage. As the first-stage fixed effects capture time-invariant regressors and the unit-mean effects of time-varying inputs, they cannot be interpreted as "pure" efficiency measures. Hence, the second-stage fixed effect should be used. The lower the within variance of inputs, the higher will be the differences between the estimated first-stage and second-stage technical inefficiency scores.

While this model has the advantage that there is no need to assume that the unit-specific effects (including technical inefficiency) are uncorrelated with the explanatory variables, this model has some well-known restrictive features (Greene, 2004). First, inefficiency is time-invariant, which may not be plausible for "long" data sets. Second, and as seen above, the u_i terms will include not only inefficiency but also any time-invariant heterogeneity.

To relax these restrictions, Greene (2004, 2005) proposed the introduction of unit-specific constant terms in the stochastic frontier model. This "true" fixed effects model can be written:

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} - u_{it} \tag{7}$$

where it is commonly assumed that the symmetric random noise term, v_{it}, follows a normal distribution and the asymmetric inefficiency term follows a half-normal distribution, i.e., $v_{it} \sim N(0, \sigma_v^2)$ and $u_{it} \sim N^+(0, \sigma_u^2)$. In this model, inefficiency can now be seen to vary over time and is separated from unobserved heterogeneity.

The discussion in the previous section suggests that the FE frontier and true FE frontier models may yield imprecise parameter estimates in the presence of slowly changing variables and possibly misleading inference. Moreover, these frontier models do not allow us to estimate the effects of potentially interesting time-invariant variables. An alternative strategy would be to use a random effects (RE) specification. Heterogeneity bias can be controlled, for example, by the Mundlak (1978) transformation and a true random effects stochastic frontier using this transformation has been proposed by Farsi *et al*. (2005a, 2005b). However, for linear models the Mundlak transformation yields identical estimates to FE[4] so that it will not solve the problems that can be expected to arise in a FE setting with rarely changing variables.

Adapting the FEVD model to a stochastic frontier framework provides a potential solution to the problem of rarely changing variables and allows the incorporation of time-invariant variables, without the need for deciding on appropriate instrumental variables. To incorporate technical inefficiency when slowly changing variables are present, it seems natural therefore to extend the FEVD model into a stochastic frontier framework.

To estimate this FEVD stochastic frontier, in the first stage we estimate the true FE frontier in order to separate unobserved heterogeneity from inefficiency per se. In the second stage the estimated fixed effects are regressed on the time-invariant and rarely changing variables to decompose heterogeneity into its observable and truly unobserved components, the latter represented by the estimated error from this second stage regression. Finally, in the third stage we estimate the full stochastic frontier model without individual effects, including all time-varying and time invariant variables and the error term from the second stage. We now turn to our empirical illustration.

4. ICT and regional productivity: data and empirical specifications

The importance of ICT capital in productivity growth has been highlighted by several papers in the literature and ICT capital has been identified as a key source of the differences between US and European productivity (Jorgenson and Stiroh, 2000; Oliner and Sichel, 2000; Stiroh, 2002; Timmer and Van Ark, 2005). From a regional perspective, ICT capital can be expected to influence productivity and economic growth in different ways depending on the region considered. Barrios *et al*. (2007) point out that technical progress takes place fundamentally in high-tech sectors so that the higher the share of ICT-producing industries in regional production, the higher the impact of ICT on productivity and thereby on economic growth. Moreover, in its role as a capital input, the level of ICT capital determines the productive capacity of most sectors of the economy and

[4] See Hsiao (2003) for a proof.

affects the production processes of manufacturing, services and the primary sector. This implies that the share of ICT capital in total capital will affect productivity (Mas and Quesada, 2005; Barrios *et al*., 2007; Quesada, 2008).

One of the most common approaches to analysing the effect of ICT on productivity has been to estimate a Cobb-Douglas production function and a large body of literature now exists, particularly at firm and firm and industry level though far less at a macro level.[5] Very few papers at any level of disaggregation have used a stochastic frontier approach, though there have been some recent studies at country-level (Lin, 2009; Chen and Lin, 2009). At a regional level, the only studies that we are aware of are those by de la Fuente (2009) and Corrales (2009), both for Spain. De la Fuente (2009) estimates Cobb-Douglas and translog production functions to measure the impact of ICT on productivity for the 17 Spanish regions over the period 1977-2003. Corrales (2009) estimates translog stochastic frontiers using ICT capital as a determinant of technical efficiency for the 50 Spanish provinces, using many of the same variables we use in our study

The dataset consists of a balanced panel of annual observations on the fifty Spanish provinces covering the period 1986-2006. The output and input variables are as follows. Output (*Y*), is measured by real gross value added (thousands of euro, year 2000), provided by the Spanish National Statistics Institute (INE). The inputs used are public capital in infrastructures, labour, human capital, and non-residential private capital. Human capital is included in the production function as there is wide evidence of the complementary of skills and technology (see, e.g., Machin and van Reenan, 1998; also, Bresnahan *et al*., 2002, report a strong correlation and productive complementarities between IT and human capital). Moreover, in previous studies human capital has been found to have a significant role in regional productivity in Spain (see, e.g., Gumbau-Albert and Maudos, 2006). Human Capital (*HC*) is measured by the percentage of workers with post-secondary level studies as reported by the INE. Labour (*L*) is measured in number of workers and the data comes from the Labour Force Survey carried out by the INE. The data on infrastructures (*INF*) and private capital come from Mas *et al*. (2009). Non-residential private capital is disaggregated into non-ICT capital (*K*) and ICT capital (*IT*). The ICT capital series was elaborated from INE data and corresponds to capital expenditure on software, hardware and telecommunications. Some descriptive statistics are provided in Table 1 below.

[5] See Draca *et al*. (2007) for a comprehensive survey of studies at firm, industry and macro levels.

INSERT TABLE 1 HERE

This dataset has several advantages for our purposes. First, detailed data on capital disaggregated into public and private, and into ICT and non-ICT capital, are available at regional level and sub-regional (provincial) level for Spain covering a relatively long time period. Second, Spain is a large country and quite heterogeneous in geographical, cultural and economic terms; moreover, its Autonomous Communities (regions) have assumed an increasing role in policy-making over recent decades, complementing and at times substituting the central government in key areas such as health, education and innovation. Regional heterogeneity and the different policies carried out at this level by the governments of the Autonomous Communities may therefore be expected to affect the productivity of a province.[6] This would make it desirable to include this time-invariant variable (i.e., the Autonomous Community to which a province belongs) in a provincial production function, something which cannot be done in a fixed effects frontier but which can be done with an FEVD frontier. Moreover, this variable would be an obvious candidate to explain at least part of the provincial heterogeneity in the second stage of the FEVD procedure. Finally, despite the level of disaggregation of the data, they remain "macro" data, and it can be expected that some of the variables will change relatively slowly over time. This would argue in favour of an FEVD treatment as the FE provincial production frontier would be likely, in light of the discussion in the last section, to yield imprecise point estimates, with the fixed effects soaking up part of the explanatory of these variables.

Our full model is a fixed effects Cobb-Douglas provincial production frontier, which after taking logs can be expressed as:

$$\ln y_{it} = \alpha_i + \beta_G \ln INF_{it} + \beta_L \ln L_{it} + \beta_{HC} \ln HC_{it} + \beta_K \ln K_{it} + \beta_{IT} \ln IT_{it} + \gamma_m R_{mi} + \varepsilon_{it} \quad (8)$$

where the α_i are the individual province fixed effects captured by N dummy variables, the β's are parameters to be estimated and the R_{mi} are regional dummy variables which takes the value 1 if the province i is in Autonomous Community m (m = 2, ... , 17), and 0 otherwise. As these are time-invariant, they cannot be estimated with the FE estimator but can be estimated with an FEVD frontier. As usual, the error term ε_{it} is decomposed into

[6] Quesada (2008) lists some examples of regional-level features that would have an impact on the effectiveness of regional technology and innovation policies, including the strong industrial tradition of Autonomous Communities such as the Basque Country and Catalonia in capital goods and the Valencian Community in consumer goods, the existence of networks of technological or research institutes, the strength of regional employer associations, the availability of UE structural funds or the system of financing in an Autonomous Community.

two components, namely a symmetric random noise term, v_{it}, for which we assume $v_{it} \sim N(0, \sigma_v^2)$, and an asymmetric term, $u_{it} \geq 0$, representing technical inefficiency.

We will estimate two versions of the stochastic frontier model. The first assumes that technical inefficiency follows a half-normal distribution, i.e. $u_{it} \sim N^+(0, \sigma_u^2)$, which is the "true" fixed effects frontier proposed by Greene (2004, 2005). We then estimate the true fixed effects frontier by modelling the inefficiency term as a function of a series of variables, i.e. we assume that $u_{it} \sim N^+\left(0, \sigma_u^2 \cdot g(w_{it}; \delta)\right)$ where w is a vector of variables that explains inefficiency and δ is a set of parameters to be estimated (Caudill *et al.*, 1995). Both versions of the true fixed effects frontier are then compared to their equivalent FEVD frontiers using the procedure outlined in the last section.

5. Results and discussion

The estimates of the various versions of the provincial production frontier are presented in Table 2 below. A time trend (t) and its square have been included in the specifications to capture disembodied technical progress. The first two columns show the FE production function and its corresponding FEVD production function. The estimates of the FE model highlight the problems associated with this estimator when there is relatively little within variation. In particular, the coefficient on private capital (K) is not significant and takes a negative value. The labour coefficient is also quite low and the human capital coefficient is only significant at the 10% level, and the overall scale elasticity is a mere 0.46. These results clearly illustrate the imprecision of the FE estimator in the presence of slowly-changing variables.[7]

INSERT TABLE 2 HERE

The FE estimator serves as the first stage of the FEVD production function. In the second stage (equation 4), the estimated provincial effects are regressed on the time-invariant and rarely changing variables, z. For the components of z we use the regional (Autonomous Community) dummies and the inputs with between-to-within variance ratios of at least 1.7 as reported in Table 1, namely public infrastructure (*INF*), private

[7] We also carried estimations using the Hausman-Taylor estimator, allowing different combinations of time-varying variables to be endogeneous. None of the estimations performed well, with the private capital again generally coming out with a negative coefficient. We do not report these estimates here but they can be made available on request.

capital (K) and labour (L). The results from the second stage regressions for the FE estimator are reported in the first two columns of Table 3, and some comments are in order. The coefficients on the infrastructure, capital and labour variables are all highly significant, indicating a correlation between these variables and the unit effects which would render a random effects estimator inconsistent and justify the use of a fixed effects approach.[8] Also, the majority of the regional dummies are highly significant and therefore explain part of the observable heterogeneity in the estimated fixed effects. It turns out that the rarely changing variables (i.e., the provincial means of the infrastructure, capital and labour variables) account for the greatest share of the observable heterogeneity. The results from this second stage regression imply that the unit effects in the first stage FE regressions will have indeed soaked up a considerable amount of the explanatory power of these three variables. The extent of this will become apparent in the results from the third stage regression but before turning to these we first focus on the implications of slowly-changing variables on efficiency measurement using the estimated unit effects as proposed by Schmidt and Sickles (1984).

Figure 1 shows the estimated unit effects from the FE estimator and the unit effects having controlled for observable heterogeneity in the second stage regression of the FEVD procedure. Two series of second stage FEVD unit effects are presented. The first are partially-adjusted effects in the sense that they are the unit effects only the rarely changing variables but not the regional dummies are included in the regression. The second are the FEVD fully-adjusted unit effects, where both the rarely changing variable and the regional dummies are included (i.e., the first regression reported in Table 3). The provinces are ordered according to the size of the FE unit effect. The corresponding efficiency indices calculated on the basis of these estimated unit effects following Schmidt and Sickles (1984) are presented in Figure 2.

Figure 1 shows that the unit effects estimated by the FE model (Within) vary over a much wider range than those of the second stage FEVD regressions. More specifically, the estimated FE (within) unit effects range from 14.57 to 17.07 (standard deviation = 0.52), whereas the partially-adjusted FEVD effects range from 15.51 to 15.79 (standard deviation = 0.13) and the fully-adjusted FEVD effects from 15.57 to 15.68 (standard deviation = 0.04). This is consistent with equation (3), which showed that the estimated unit effects in the FE model include observable heterogeneity in the form of time invariant

[8] This was indeed confirmed by a Hausman test, which yielded a chi-squared statistic of 310.6, soundly rejecting the random effects model.

variables and the unit means of the time-variant inputs. Eliminating these in the FEVD second stage greatly reduces the variance of the estimated unit effects. Note, moreover, that the partially-adjusted and fully-adjusted series of unit effects are quite similar, highlighting the observation made above that the rarely changing variables accounted for most of the observable heterogeneity. When we decompose the full change in the estimated unit effects into the part due to the rarely changing variables and that due to the effect of the time-invariant regional dummies, our calculations showed that the rarely changing variables accounted for an average over provinces of 75% of the overall change in the estimated unit effects.

As expected, this has a profound impact on the efficiency indices calculated from these unit effects.[9] Figure 2 shows that the FE efficiency indices are implausibly low for the majority of provinces. Indeed, the minimum value was a mere 0.08 and the mean was only 0.24 (standard deviation = 0.17). Correcting the unit effects for the rarely changing variables substantially increases the efficiency indices: the partially-adjusted FEVD inefficiency indices had a minimum value of 0.52 and a mean of 0.76, more than three times higher than the mean of the FE indices (standard deviation = 0.09). Including the regional dummies to further control for observable heterogeneity leads to even higher indices: the fully-adjusted FEVD indices had a mean efficiency of 0.90 and a minimum of 0.83 (standard deviation = 0.04). Overall, the FEVD efficiency indices are more in line with what we would expect for the Spanish provinces. While Spain is a large and heterogeneous country and we might expect to see differences across provinces' efficiency levels, it is a developed country with a relatively homogeneous education system and good communications. There should be a relatively low dispersion in efficiency levels as a consequence.

The results from the third stage regressions of the full FEVD estimator (equations 5 and 8) where the unit effects have been replaced by the estimated residual from the second stage (θ) are shown in the second column of Table 2. The parameter estimates show a substantial improvement on those from the FE estimators. In particular, the coefficient on private capital is now positive and of plausible magnitude and the coefficient on labour is much more representative of labour's share in national income. The human capital variable also gains substantially in size and significance. Indeed, all the coefficients are now significant with p-values of well below 0.01. The effect of the Autonomous

[9] As the production function is estimated in logs, the technical efficiency indices, $\hat{u}_i$, are estimated as $\hat{u}_i = exp\{-(max(\hat{\alpha}_i) - \hat{\alpha}_i)\}$

Community can also be included and with the exception of one, all of these regional dummies were highly significant.

While the estimated parameters in the FEVD model are all highly significant, a word of caution is warranted at this point. In the case where only time-invariant variables are included in the second stage, Greene (2010) and Breusch *et al*. (2010) have shown that the FEVD standard errors are lower than those of the fixed-effects model to which it collapses and are underestimated. It is not clear yet whether and how any adjustment should be made to the standard errors in the rarely-changing variable case and this will doubtless be a subject of debate in the future. In our empirical model, it can be seen that the standard errors are almost identical to those from the fixed effects model and that gains in precision have arisen from more plausible parameter estimates, not from greatly reduced standard errors.

We now turn to the stochastic frontier specifications. The true FE frontier estimates for the standard normal-half-normal model where the variance of inefficiency is homoskedastic (Model 1) are shown in the third column of Table 2 and it can be seen that they quite similar to the estimates of the FE estimator. Thus, while the human capital coefficient gains significance, the private capital coefficient remains insignificant and negative and the labour coefficient is even slightly lower than the FE estimate. The true FE frontier serves as the first stage of the FEVD frontier estimator. The second stage regression estimates, reported in Table 3, are virtually identical to those of the FE model. The third stage estimates are shown in the fourth column of Table 2 and the parameter estimates are again very similar to those of the FEVD estimator (column two), being much more plausible than those of the true FE frontier.

Focusing on the contribution of ICT capital, we see that the coefficient is positive and highly statistically significant in all models, implying that greater intensity in ICT expenditure shifts the provincial production frontier upwards and raises productivity. The output elasticities of ICT capital are all close to 0.11 except for the true FEVD frontier (Model 1), for which the figure is 0.061. These elasticities are in line with those found in the previous literature. For example, de la Fuente (2009) estimated a series of production functions for the Spanish regions and reports output elasticities of ICT capital ranging from 0.062 to 0.122. At a country level, Dewan and Kraemer (2000), also using Cobb-Douglas specifications, report yearly output elasticities for developed countries for the period 1985-1993 with a range from 0.080 to 0.117 (see their Table 9). In a meta-study, Stiroh (2004) found a mean across studies of 0.05. However, he also estimated different Cobb-Douglas production functions for a data set of 58 U.S. industries. When using value added as output (as we do) he found an output elasticity of 0.114, very similar to ours.[10]

The estimated parameters of the input variables are not the only differences between the true FE frontier and the FEVD frontier. When we estimated the true FE frontier we found evidence of the existence of technical inefficiency, as the hypothesis that $\sigma_u^2 = 0$ was rejected. The average of the efficiency scores was 0.975. However, after adjusting the unit effects for observable heterogeneity in the FEVD frontier model the hypothesis of no technical inefficiency could not be rejected. Thus, taking account of observable heterogeneity can lead to different conclusions about the existence of technical inefficiency

The final two models reported in Table 2 are the true FE and FEVD stochastic frontiers where explanatory variables are included for the variance of the efficiency term following Caudill *et al*. (1995). We consider two explanatory variables. The first of these is a synthetic index of human capital (*IKH*) reported by IVIE-*Fundación Bancaja* based on the education level and experience of the provincial population and measures how many equivalent workers with no human capital would be necessary to achieve the productive capacity of the actual occupied provincial workforce (see Serrano and Soler, 2008). Note that this is more general indicator of human capital than the variable we used in the deterministic part of the production frontier. The second variable is the ratio of ICT capital to total capital (*ITIC*). A priori, we would expect both of these variables to improve efficiency.

As can be seen from the last two columns of Table 2, the input coefficients in these heteroskedastic frontier models are quite similar to their homoskedastic (standard normal-half normal) equivalents, with the FEVD frontier again yielding more reliable estimates. These models produce some interesting results. Note first that technical inefficiency was found to be present in both heteroskedastic frontiers. We are using the ICT share and the human capital index to model the variance of technical inefficiency, so that a positive (negative) coefficient implies that the variable increases (reduces) technical inefficiency. In both versions of the model (true FE frontier and FEVD frontier), human

[10] Interestingly, his other output elasticities were also very similar to ours for the value added output. His elasticity of labour was 0.606 and that of non-ICT capital was 0.295, very close to revenue share and to our FEVD frontier estimates. See Stiroh (2004), Table 5.

capital is found to reduce inefficiency, although in the FEVD frontier this effect loses significance. The share of ICT capital, on the other hand, is found to increase technical inefficiency in both models, though this effect is only significant in the FEVD frontier. Together with the fact that the sign of ICT in the deterministic part of the frontier is positive, this would imply that increased ICT capital enables provinces to attain a higher level of production but that they are not capable of taking full advantage of this. That is, greater ICT capital shifts the frontier upwards but as the share of ICT capital increases, the provinces produce at a greater distance from the frontier.

With regard to the technical efficiency estimates, the average efficiency scores for the heteroskedastic true FE and FEVD frontiers were 0.956 and 0.974 with minimum values of 0.814 and 0.799. It should be noted that although the average efficiency is quite similar in both models, the true FE and FEVD models yield very different efficiency scores. This is illustrated in Figure 3 where the average efficiency indices for the provinces are depicted. The difference between the models in terms of the provincial ranking according to the efficiency estimates which can be seen in the graph are underlined by a Spearman rank correlation coefficient of 0.109, a quite small value which moreover is not significantly different from zero (p-value = 0.22). This outcome is quite relevant as it casts doubts about the plausibility of the individual efficiency scores obtained in the first stage of our procedure (i.e., the true FE frontier). Indeed, if we conclude that some of the estimated parameters of the frontier using the true FE are not reasonable from an economic point of view (for instance, our true FE production frontier is decreasing in private capital), we should also conclude that the individual efficiency scores obtained from this frontier are implausible. The second and third stages of our procedure will "correct" these efficiency scores as they are computed using more reasonable estimates (from an economic point of view) of the production frontier. The small rank correlation indicates that the implausible estimates of the frontier using the true FE model generate very different efficiency rankings.

6. Conclusions

In this paper we have proposed an FEVD stochastic frontier based on the work of Plümper and Troeger (2007) and which can be interpreted as an extension of the true FE frontier which potentially permits more precise parameter estimates in the presence of slowly-changing variables. In an empirical application to Spanish regional data, we find that the presence of slowly changing variables, defined as those with a high ratio of between-to-within variance, yields implausible estimates of the parameters of a production estimated using the FE estimator. The same occurs when estimating stochastic production frontiers when the true FE estimator is used. The estimates of the FEVD estimator and the FEVD frontiers, on the other hand, are found to be much more reasonable from an economic point of view. Although the average efficiency scores for the true FE and FEVD frontiers were quite similar, they yield very different efficiency rankings. The second and third stages of our procedure will "correct" these efficiency scores as they are computed using more reasonable estimates of the production frontier.

Many data sets, especially those using macro data, can be expected to contain rarely changing variables and FE estimators will therefore be likely to yield poor estimates. This may lead researchers to opt for random effects estimators, possibly correcting for endogeneity by using techniques such as the Hausman-Taylor estimator or the Mundlak transformation. However, these estimators will not deal with the problems associated with slowly changing variables and can be expected to provide unreliable estimates of production frontiers. We believe that the FEVD frontier has the potential to provide a valuable alternative to true random effects and true fixed effects frontiers for researchers whose data includes variables with much greater between than within variance. Moreover, the conditions under which the FEVD estimator can be expected to perform better than FE, based on the ratio of between-to-within variance, are easily checked.

We have illustrated the advantages of the FEVD frontier model with an empirical application using production data from Spanish provinces where we investigate the effect of ICT capital on regional productivity. We find that increased use of ICT capital shifts the frontier upwards, but as the share of ICT capital in total capital increases, the provinces produce at a greater distance from the frontier i.e., we find that they are not capable of taking full advantage of the possibilities offered by ICT capital. To confirm these results, future research should be carried out using other data sets and/or carrying out Monte Carlo experiments to compare different estimators in non-linear models such us traditional stochastic frontiers.

References

Barrios, S., Mas, M., Navajas, E. and J. Quesada (2007). "Mapping the ICT in EU regions: Location, employment, factors of attractiveness and economic impact." *JRC Scientific and Technical Reports*, European Commission.

Boyce, C.J. (2009). "Understanding fixed effects in human well-being." *Journal of Economic Psychology*, doi: 10.1016/j.joep.2009.08.009.

Bresnahan, T.F., Brynjolfsson, E. and L. Hitt (2002). "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence." *Quarterly Journal of Economics* 117, 339-376.

Breusch, T., Ward, M.B., Nguyen, H. and T. Kompas (2010). "On the fixed-effects vector decomposition." *MPRA Paper No. 21452*. Online at http://mpra.ub.uni-muenchen.de/21452/

Carley, S. (2009). "State renewable energy electricity policies: An empirical evaluation of effectiveness." *Energy Policy* 37, 3071-3081.

Caudill, S.B., Ford, J.M. and D.M. Gropper (1995). "Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity." *Journal of Business Economics and Statistics* 13, 105-111.

Chen, Y.H. and W.T. Lin (2009). "Analyzing the relationships between information technology, inputs substitution and national characteristics based on CES stochastic frontier production models." *International Journal of Production Economics* 120, 552-569.

Dewan, S. and K.L. Kraemer (2000). "Information technology and productivity: Evidence from country-level data." *Management Science* 46, 548-562.

Draca, M., Sadun, R. and J. van Reenen (2007). "Productivity and ICTs: A review of the evidence." In Mansell, R., Avgerou, C., Quah, D. and R. Silverstone (eds): *The Oxford Handbook of Information and Communication Technologies*. Oxford University Press.

Farsi, M., Filipini, M. and W. Greene (2005a). "Efficiency measurement in network industries: Application to the Swiss railway companies." *Journal of Regulatory Economics* 28, 69-90.

Farsi, M., Filipini, M. and M. Kuenzle (2005b). "Unobserved heterogeneity in stochastic cost frontier models: An application to Swiss nursing homes." *Applied Economics* 37, 2127-2141.

Greene, W. (2004). "Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization's panel data on National Health Care Systems." *Health Economics* 13, 959-980.

Greene, W. (2005). "Reconsidering heterogeneity in panel data estimators of the stochastic frontier models." *Journal of Econometrics* 126, 269-303.

Greene, W. (2011). "Fixed effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models?" *Political Analysis* (in press). http://pages.stern.nyu.edu/~wgreene.

Greene, W., Orea, L. and A. Wall (2011). "A one-stage random effect counterpart of the fixed-effect vector decomposition model with an application to UK electricity distribution utilities." *Efficiency Series Paper 01/2011*, Department of Economics, University of Oviedo, Spain. http://www.unioviedo.es/economia/EDP/ESP0111.pdf.

Gumbau-Albert, M. and J. Maudos (2006). "Technological activity and productivity in the Spanish regions." *Annals of Regional Science* 40, 55-80.

Hausman, J.A. and W.E. Taylor (1981). "Panel data and unobservable individual effects." *Econometrica* 49, 1377-1398.

Hsiao, C. (2003). *Analysis of Panel Data*, 2nd ed. Cambridge University Press.

Jorgenson, D.W. and K. Stiroh (2000). "Raising the Speed Limit: U.S. Economic Growth in the Information Age." *Brookings Papers on Economic Activity* 1, 125-211.

Lin, W.T. (2009). "The business value of information technology as measured by technical efficiency: Evidence from country-level data." *Decision Support Systems* 46, 865-874.

Machin, S. and J. Van Reenan (1998). "Technology and changes in skill structure: Evidence from seven OECD countries." *Quarterly Journal of Economics* 113, 1215-1244.

Mas, M., Pérez, F. and E. Uriel (2009), *El stock y los servicios de capital en España y su distribución territorial (1964-2007)*. Fundación BBVA.

Mas, M. and J. Quesada (2005). *Las Nuevas Tecnologías y el Crecimiento Económico en España*. Fundación BBVA, Bilbao.

Mundlak, Y. (1978). "On the pooling of time-series and cross-section data." *Econometrica* 64, 69-85.

Oliner, S.D. and D.E. Sichel (2000). "The Resurgence of Growth in the Late 1990's: Is Information Technology the Story?" *Journal of Economic Perspectives* 14, 3-22.

Plümper, T. and V.E. Troeger (2007), "Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects," *Political Analysis* 15, 124-139.

Quesada, J. (2008). "Política regional de innovación." *Investigaciones Regionales* 12, 181-210.

Serrano, L. and A. Soler (2008). *Metodología para la estimación de las series de capital humano. 1964-2007*. Fundación Bancaja-IVIE.

Stiroh, K. (2002). "Information Technology and U.S. Productivity Revival: What Do the Industry Data Say?" *American Economic Review* 92, 1559-1576.

Timmer, M. and B. van Ark (2005). "Does Information and Communication Technology Drive EU-US Productivity Growth Differentials?" *Oxford Economic Papers* 57, 693-716.

Worrall, J.L. (2008). "Racial composition, unemployment, and crime: Dealing with inconsistencies in panel designs." *Social Science Research* 37, 787-800.

Table 1. Summary statistics of variables

Variable	Mean	Min	Max	Standard deviations			
				Overall	*Between*	*Within*	*Between-to-Within*
ln Y	15.65	13.76	18.63	0.89			
ln INF	15.03	13.43	17.53	0.67	0.61	0.29	2.10
ln K	16.20	14.32	19.16	0.83	0.80	0.26	3.05
ln L	5.24	3.39	8.00	0.84	0.83	0.16	5.15
ln HC	-1.96	-3.14	-1.14	0.34	0.19	0.27	0.71
ln IT	13.13	10.64	16.71	1.02	0.88	0.54	1.63
IKH	2.57	1.99	3.24	0.27			
ITIC	0.04	0.01	0.09	0.01			

Table 2. FE, FEVD, True FE frontier and FEVD frontier estimations

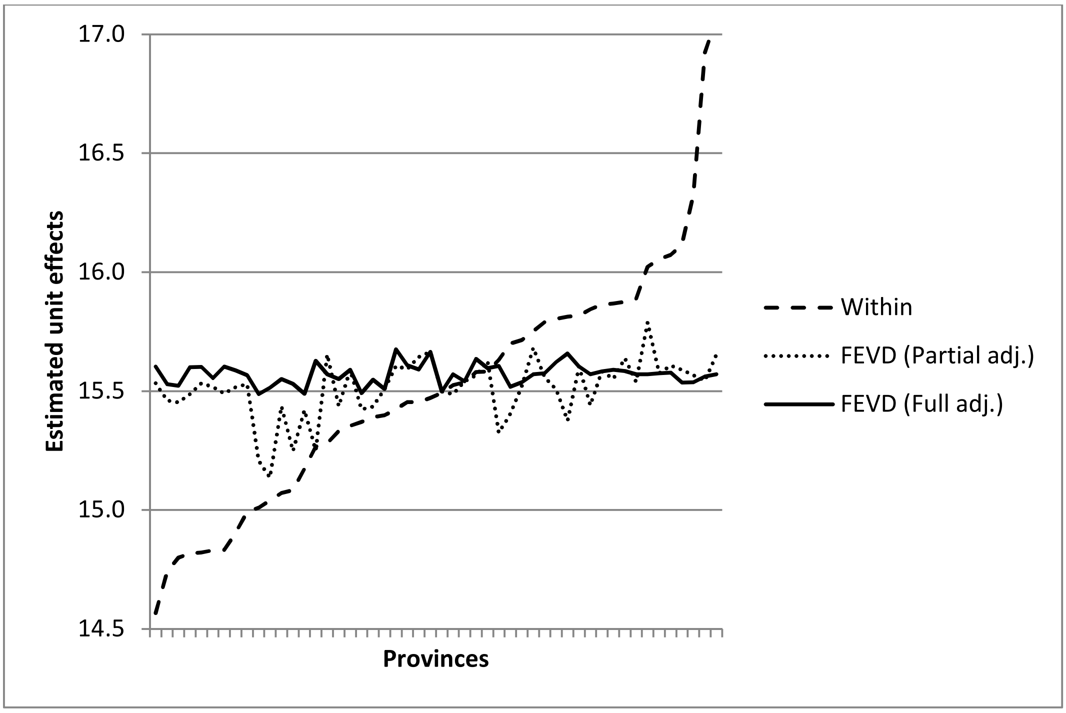
Variable	FE	FEVD	True FE Frontier Model 1	FEVD Frontier Model 1	True FE Frontier Model 2	FEVD Frontier Model 2
Dependent variable: ln Gross Value Added (Y)						
ln INF	0.0491***	0.0465***	0.0488***	0.0461***	0.0407***	0.0325***
	(*0.0146*)	(*0.0113*)	(*0.0131*)	(*0.0102*)	(*0.0127*)	(*0.0093*)
ln K	-0.0274	0.2969***	-0.0217	0.3039***	-0.0175	0.2628***
	(*0.0250*)	(*0.0232*)	(*0.0232*)	(*0.0226*)	(*0.0228*)	(*0.0217*)
ln L	0.3078***	0.6187***	0.2971***	0.6203***	0.3016***	0.6280***
	(*0.0135*)	(*0.0126*)	(*0.0132*)	(*0.0124*)	(*0.0136*)	(*0.0113*)
ln HC	0.0176*	0.0663***	0.0192**	0.0669***	0.0004	0.0498***
	(*0.0097*)	(*0.0132*)	(*0.0093*)	(*0.0119*)	(*0.0095*)	(*0.0110*)
ln IT	0.1165***	0.0690***	0.1148***	0.0612***	0.1087***	0.1096***
	(*0.0165*)	(*0.0234*)	(*0.0144*)	(*0.0215*)	(*0.0162*)	(*0.0211*)
t	0.0174***	0.0061***	0.0159***	0.0062***	0.0148***	0.0022
	(*0.0015*)	(*0.0018*)	(*0.0014*)	(*0.0015*)	(*0.0014*)	(*0.0014*)
t^2	-0.0007***	-0.0013***	-0.0005***	-0.0012***	-0.0004***	-0.0007***
	(*0.0001*)	(*0.0002*)	(*0.0001*)	(*0.0001*)	(*0.0001*)	(*0.0001*)
θ		1.0196***		1.0086***		0.9707***
		(*0.0442*)		(*0.0410*)		(*0.0368*)
Constant	15.511***	15.747***	15.484***	15.746***	15.483***	15.772***
	(*0.0141*)	(0.0143)	(*0.0194*)	(*0.0303*)	(*0.0189*)	(*0.0125*)
Region Dummies	No	Yes	No	Yes	No	Yes
No Technical Inefficiency (H_0: $\sigma_u^2 = 0$)			Rejected	Not rejected	Rejected	Rejected
ITIC					0.1350	1.7539***
					(*0.0838*)	(*0.1819*)
IKH					-3.5295***	-0.9543
					(*0.6450*)	(*1.0414*)
Constant					2.5852*	-6.9666**
					(*1.5271*)	(*2.8014*)
Number of observations: 1050						

Note: *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table 3. Second stage estimations of FE model and True FEVD frontiers

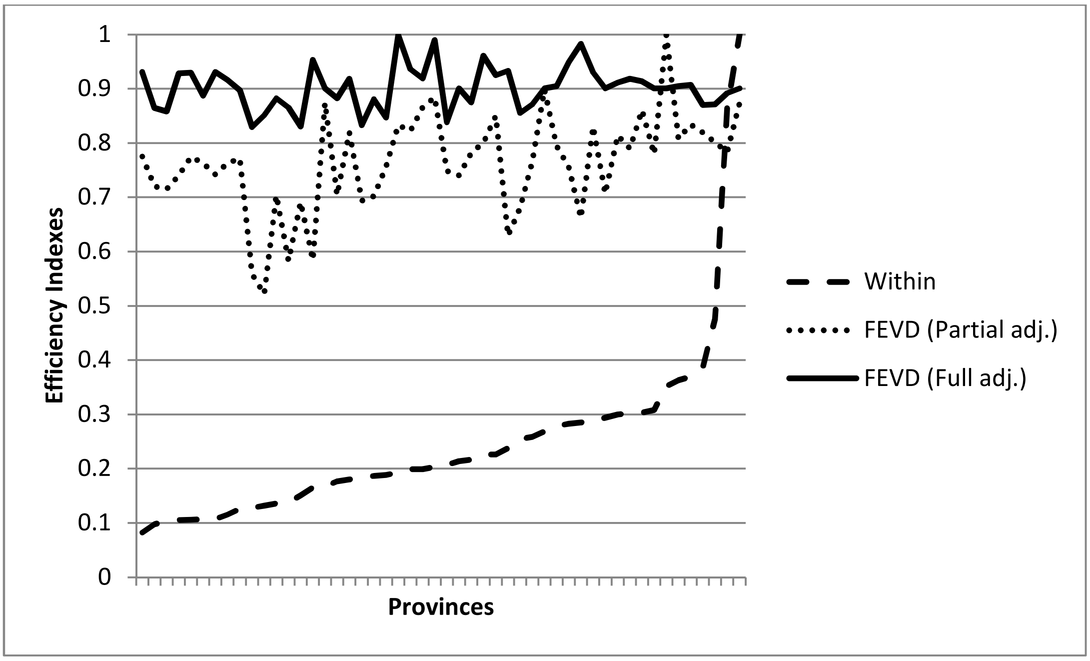
	FE		Homoskedastic True FEVD Frontier (Model 1)		Heteroskedastic True FEVD Frontier (Model 2)	
Variable	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat
ln INF	-0.091	-2.095	-0.090	-2.061	-0.082	-1.913
ln K	0.275	5.517	0.255	4.973	0.231	4.204
ln L	0.386	8.137	0.408	8.361	0.427	8.208
Region 2	-0.038	-1.585	-0.037	-1.578	-0.034	-1.372
Region 3	-0.130	-4.423	-0.127	-4.336	-0.125	-4.188
Region 4	0.218	8.552	0.221	8.465	0.213	7.654
Region 5	-0.115	-2.539	-0.109	-2.414	-0.096	-2.091
Region 6	-0.083	-3.453	-0.082	-3.437	-0.078	-3.174
Region 7	-0.070	-2.243	-0.070	-2.282	-0.072	-2.293
Region 8	-0.114	-5.382	-0.112	-5.239	-0.108	-4.675
Region 9	-0.017	-0.503	-0.009	-0.264	0.007	0.217
Region 10	0.029	1.321	0.028	1.298	0.025	1.057
Region 11	-0.377	-7.932	-0.362	-7.501	-0.341	-6.700
Region 12	-0.282	-7.304	-0.283	-7.129	-0.282	-6.788
Region 13	0.083	1.918	0.091	2.107	0.107	2.441
Region 14	-0.032	-1.738	-0.034	-1.814	-0.039	-1.907
Region 15	0.111	5.843	0.111	5.882	0.117	5.942
Region 16	0.053	1.664	0.057	1.819	0.065	2.026
Region 17	0.083	4.448	0.079	4.245	0.081	4.083
Constant	15.571	858.48	15.606	867.03	15.607	823.53
R^2	0.993		0.993		0.992	
No. observations: 50						

Figure 1. Estimated unit effects: within and FEVD (second stage)*



* Partial-adj. refers to FEVD second stage excluding regional dummies. Full-adj. refers to FEVD second stage including regional dummies.

Figure 2. Efficiency based on unit effects: within and FEVD (second stage)*



* Partial-adj. refers to FEVD second stage excluding regional dummies. Full-adj. refers to FEVD second stage including regional dummies.

Figure 3. Average provincial efficiency indices from heteroskedastic frontier models