# ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 04/2005

# Estimating the Productivity of Public Infrastructure using Maximum Entropy Econometrics

Jorge Rodriguez-Vález, Antonio Álvarez, Carlos Arias y Esteban Fernández

Departamento de Economía

Universidad de Oviedo

# UNIVERSIDAD DE OVIEDO

# DEPARTAMENTO DE ECONOMÍA

## PERMANENT SEMINAR ON EFFICIENCY AND PRODUCTIVITY

# ESTIMATING THE PRODUCTIVITY OF PUBLIC INFRAESTRUCTURE USING MAXIMUM ENTROPY ECONOMETRICS

**Jorge Rodriguez-Vález♦, Antonio Alvarez♣, Carlos Arias♠ and Esteban Fernandez♥**

**Efficiency Series Paper 04/2005**

**Abstract:** The empirical analysis of public infrastructure productivity has been limited by several econometric problems. These problems are probably the cause of the great variability in results reported in the literature. One of these econometric problems is the existence of multicollinearity when infrastructure is included as an input in a production function. In this paper, we propose to deal with multicollinearity by using prior information on parameter values. For this purpose, we explore the use of maximum entropy econometrics to estimate the parameters of a regional production function in Spain. Our results show that infrastructure plays an important role on production. This result is not affected by quite large changes on prior information on parameter values.

**Key words:** Infraestructures, maximum entropy, multicollineairity, productivity

♦ Corresponding author. Departamento de Economía, Universidad de León, Spain.
e-mail: Jorge.rodriguez@unileon.es

♣ Departamento de Economía, Universidad de Oviedo, Spain.

♠ Departamento de Economía, Universidad de León, Spain.

♥ Departmento de Economía Aplicada, Universidad de Oviedo, Spain.

## 1. Introduction

Empirical analysis in economics is very often hampered by the characteristics of data. For example, multicollinearity caused by a large correlation among explanatory variables in a model increases the variance of the least square estimator reducing the accuracy of the estimator. As a result, the estimates are highly sensitive to changes in the specification of the model, the levels of significance are low and unreasonable estimates in sign or magnitude appear [Greene (2000), p. 256]. Under these circumstances, the estimates are of little if any value.

Goldberger (1991, p. 246) discusses the similarities between multicollinearity in multivariate analysis and having a small sample in a univariate population. In other words, he sees multicollinearity as a simple problem of lack of data to estimate the parameters of a model. Both problems (small datasets and multicollinearity) are pervasive in economics, but researchers trained in classical econometrics do not feel comfortable dealing with them. Facing this situation, some authors argue for the feasibility of using small or multicollinear datasets in applied economics because, in a sense, there is some information in all datasets, even with a single observation, and the amount of information does not decrease with the number of observations [Paris and Howitt (1998)]. Therefore, the relevant issue is how to use rigorously the information contained in small or collinear datasets. A solution lies in the use of prior information in the form of possible values of the parameter. Maximum entropy (ME) econometrics provides a rigorous but operationally simple method to deal with prior information (Marsh and Mittelhammer, 2004).

The present paper adds to a voluminous literature that deals with the productivity of public capital or infrastructure (See surveys by Gramlich, 1994 and de la Fuente, 2002). It is fair to say that policy makers expect a lot from public infrastructure in terms of increasing productivity, economic growth and decreasing regional disparities. However, finding empirical evidence of the effects of public infrastructure has proven to be a difficult task. One of the reasons is the high correlation between private inputs and public capital in the estimation of (regional) production functions which include this type of capital as input.[1] Ai and Cassou (1997) and Vijverberg, Vijvergerg and Gamble

---

[1] For example, the condition numbers and variance decomposition for each eigenvalue of the matrix X'X with the sample information of explanatory variables used in this paper (data for

(1997) point to multicollinearity (together with other econometric problems that sometimes have not been fully addressed in empirical work) as the cause of the disparity of estimates obtained in the literature. In fact, a non negligible number of papers in the literature do not find evidence of a positive contribution of infrastructure to growth [Holtz-Eakin (1994), Baltagi and Pinnoi (1995), Garcia-Milà, McGuire and Porter (1996)] while others find clear evidence of such contribution [Aschauer (1989), Munnell (1990) and Garcia-Milá and McGuire (1992)].

In this paper, we estimate an aggregate production function by ME to obtain evidence on the contribution of public capital to output growth in the Spanish regions. The available panel dataset suffers from quite severe multicollinearity. Since the estimation by maximum entropy uses prior information on parameter values, we check the influence of such information in the estimates. In our empirical application, the output elasticity of public capital is always larger than 0.13 in spite of quite large changes in prior information. We also compare the results obtained with the estimates provided by classical econometric methods. The sensitivity of estimates to changes in prior information is often seen as a weakness of the method. However, we use such peculiarity to provide new evidence against a negative or null value of productivity of infrastructure. Precisely, we look at ME estimates of the output elasticity of public capital when increasingly negative values of this parameter are used as prior information. The results of our empirical exercise point out to ME as a method that provides a venue for empirical analysis when multicollinearity is an issue.

The paper is organized as follows. In section 2, we discuss the main features of maximum entropy econometrics. In section 3, we specify a production function estimable by ME. In section 4, we describe the data and present the results of the estimation by ME. The paper ends with some conclusions.

## 2. Maximum entropy econometrics: an overview

We want to estimate the parameters of a linear model in which a variable $y$ depends on $H$ explanatory variables $x_h$:

---

Spanish Regions) show clearly the presence of potential harmful problems of multicollinearity in all inputs. The correlation among the variables exceeds 0.9.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e} \tag{1}$$

where **y** is a $(T \times 1)$ vector of observations, **X** is the $(T \times H)$ matrix of explanatory variables for each observation, $\boldsymbol{\theta}$ is the $(H \times 1)$ vector of parameters to be estimated, and **e** is a $(T \times 1)$ error vector. We estimate this model using the maximum entropy estimator proposed by Golan, Judge and Miller (1996).[2]

The starting point is the specification of each parameter $\theta_h$ as a discrete random variable that can take ( $M \geq 2$ ) different values which are grouped in the so called support vector: $\mathbf{b}_h = (b_{h1}, b_{h2}, \ldots, b_{hM})$. The support vector is chosen using prior information on the value of the parameter. For example, if prior information determines a range of possible values for the parameter the support vector could contains *M* values in this interval. Each value of the support vector has a probability of being the 'real' parameter. Hence, the support vector has an associated probability vector $\mathbf{p}'_h = (p_{h1}, p_{h2}, ..., p_{hM})$ such that $\sum_{m=1}^{M} p_{hm} = 1, \forall h$. Finally, each parameter in the model is written as a linear combination of the elements of the support vector (chosen) weighted by their probability (unknown):

$$\theta_h = \mathbf{b}_h \mathbf{p}_h = \sum_{m=1}^{M} b_{hm} p_{hm} \tag{2}$$

The vector of *H* unknown parameters in the model can be written as:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_H \end{bmatrix} = \mathbf{Bp} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{b}_H \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_H \end{bmatrix} \tag{3}$$

A similar procedure is used to specify the error vector of the model (**e**). For each element $e_t$ of the vector, we assume the existence of a support vector

[2] There are several applied papers that have used this estimator. See, for example, Paris and Howitt (1998), Fraser (2000), Golan, Perloff and Shen (2001) and Gardebroek and Oude Lansink (2004).

$\mathbf{v}=(v_1, v_2,..., v_J)$[3] with probabilities $\mathbf{w'_t}=(w_{t1}, w_{t2},..., w_{tJ})$ where $J \geq 2$. The error vector **e** can be written as:

$$\mathbf{e}=\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix}=\mathbf{Vw}=\begin{bmatrix} \mathbf{v} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{v} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{v} \end{bmatrix}\begin{bmatrix} \mathbf{w_1} \\ \mathbf{w_2} \\ \vdots \\ \mathbf{w_T} \end{bmatrix} \tag{4}$$

And the value of the error, for an observation *t* is:

$$e_t = \mathbf{vw_t} = \sum_{j=1}^{J} v_j w_{tj} \tag{5}$$

Finally, the model in equation (1) can be written as:

$$\mathbf{y}=\mathbf{XBp}+\mathbf{Vw} \tag{6}$$

In this specification, the problem of estimating the unknown vector of parameters $\boldsymbol{\theta}$ is transformed into the estimation of *M*×*H* probabilities of the values of the support vectors of the parameters and *J*×*T* probabilities of the values of the support vector of the error. Using the estimated probabilities, an estimate of each parameter can be recovered as:

$$\hat{\theta}_h = \sum_{m=1}^{M} \hat{p}_{hm} b_{hm} \; ; \quad \forall h=1,...,H \tag{7}$$

ME estimates the probabilities in (6) by maximizing an entropy function. The entropy function is a measure of uncertainty or ignorance about the outcomes of an event represented by a random variable. The entropy function defined by Shannon (1948) is:

$$EF(\mathbf{p}) = -\sum_{m=1}^{M} p_m \ln p_m \tag{8}$$

where *EF* is the value of the entropy function and $\mathbf{p}=[p_1, p_2,..., p_M]$ are the probabilities of *M* possible outcomes $x_1, x_2,...,x_M$ of a discrete random variable *x,* such that $\sum_{m=1}^{M} p_m = 1$.

[3] The common practice is to choose a set of values centered around zero.

The interpretation of the entropy function as a measure of ignorance is quite intuitive. The function *EF*(**p**) goes to zero when the probability of one of the possible outcomes of the random variable goes to one. In other words, the function goes to zero as uncertainty vanishes. This is the minimum value of the entropy function since it can not take negative values. In the other hand, the entropy function achieves its unrestricted maximum for the uniform distribution: $\left( p_m = \frac{1}{M}, \forall m = 1,...,M \right)$. This is a quite intuitive result since the uncertainty about the outcome of an event is maximum when all outcomes have the same probability.

In empirical analysis, the availability of a dataset put us somewhere between these two polar cases (no uncertainty and complete uncertainty). In this case, the focal point is the level of ignorance that can be claimed when there are some observations of the outcome of a random variable. The existence of data consisting of several observations of the outcomes of the random variable reduces the uncertainty and, therefore, the entropy. In fact, it is possible to reject probability distributions that could not have generated the observed data. For example, two different realizations of a random variable mean that the sample could not come from a probability distribution that attributes a probability of one to a realization and zero to the others. The maximum level of ignorance compatible with a given dataset is measured by the maximum entropy conditional on the constraints on probability defined by the dataset. The maximum entropy principle consists of choosing as estimates the probabilities associated with the maximum entropy conditioned on the dataset.

The estimation of the model in (1) requires the estimation of the probabilities of the elements of the support vectors. The probabilities can be calculated using the following optimization program:

$$\underset{\mathbf{p},\mathbf{w}}{Max}\, EF(\mathbf{p},\mathbf{w}) = -\sum_{h=1}^{H}\sum_{m=1}^{M} p_{hm} \ln p_{hm} - \sum_{t=1}^{T}\sum_{j=1}^{J} w_{tj} \ln w_{tj} \tag{9a}$$

Subject to:

$$\sum_{h=1}^{H}\sum_{m=1}^{M} b_{hm} p_{hm} x_{ht} + \sum_{j=1}^{J} v_j w_{tj} = y_t,\ \forall t = 1,...,T \tag{9b}$$

$$\sum_{m=1}^{M} p_{hm} = 1, \ \forall h = 1,...,H \tag{9c}$$

$$\sum_{j=1}^{J} w_{tj} = 1, \ \forall t = 1,...,T \tag{9d}$$

The equation in (9a) is the entropy function of Shannon adapted to the estimation of $M \times H + J \times T$ probabilities. The equation in (9b) contains sample information in terms of the model in (6) ensuring that the estimated probabilities are compatible with the $T$ available observations. The equations in (9c) and (9d) ensures that probabilities add-up to one. The Lagrangian associated to the constrained maximization can be written as:

$$\begin{aligned} L = & -\sum_{h=1}^{H}\sum_{m=1}^{M} p_{hm} \ln p_{hm} - \sum_{t=1}^{T}\sum_{j=1}^{J} w_{tj} \ln w_{tj} + \sum_{t=1}^{T} \lambda_t \left[ y_t - \sum_{h=1}^{H}\sum_{m=1}^{M} x_{ht} p_{hm} b_m - \sum_{j=1}^{J} w_{tj} v_j \right] \\ & + \sum_{h=1}^{H} \mu_h \left[ 1 - \sum_{m=1}^{M} p_{hm} \right] + \sum_{t=1}^{T} \gamma_t \left[ 1 - \sum_{j=1}^{J} w_{tj} \right] \end{aligned} \tag{10}$$

The solution of the optimization program in (9) gives the estimates of the probabilities of the elements of the support vectors. The parameters $\theta$ of the model can be recovered using expression (7).[4]

### 3. Empirical model

The role of infrastructure in economic growth has usually been studied by including private capital along with private inputs as explanatory variables in a production function. The basic model can be written as:

$$Y = f(A, K, L, G) \tag{11}$$

where $Y$ is the production, $K$ is the stock of private capital, $L$ is Labor, $G$ denotes the stock of public infrastructure and $A$ is a measure of total factor productivity.

4 The large sample properties of the ME estimators are analyzed in Golan, Judge and Miller (1996; pp. 96-123 and 131-133). ME estimators are shown to be consistent and asymptotically normal. These authors analyze also the small sample properties using Monte Carlo simulation.

In log form, a technology represented by a Cobb-Douglas production function can be written as:

$$\ln Y_{it} = \theta_i + \theta_\tau t + \theta_{\tau\tau} t^2 + \theta_K \ln K_{it} + \theta_L \ln L_{it} + \theta_G \ln G_{it} + e_{it} \qquad (12)$$

where subscript *i* denotes regions, subscript *t* denotes time[5] and *e* is an error term. Exogenous technical change is introduced in the model as a quadratic function of time. The productivity of infrastructures depends on the value of the parameter $\theta_G$. A value of the parameter greater than zero can be interpreted as evidence of a positive contribution of infrastructures to private production.

The model in (12) is estimated using data on 17 Spanish regions observed for 21 years (1980-2000). The estimation by ME of the 22 parameters in the model (17 regional effects, 3 output elasticities and two parameters that model technical change) requires to choose 22 support vectors. Each support vector contains 3 values. Under constant returns to scale and perfect competition, the parameters of the private inputs in the Cobb-Douglas production function are the output shares of these inputs. Therefore the value of these shares in national accounts can be used as prior information on the output elasticities of private inputs. In Spain, the output share of capital is around 0.4 and the output share of labor is around 0.6. Therefore, these values are chosen as the central values of the support vectors for these output elasticities: (0.3, 0.4, 0.5) for private capital and (0.5, 0.6, 0.7) for labor. The output elasticity of public capital ($\theta_G$) can not be interpreted as an output share since this input is supplied free of charge by the government. In this case, we choose a support vector that contains two widely accepted assumptions: the non negativity of the output elasticity of public capital and that the output elasticity of public capital is smaller than the output elasticity of private capital. Therefore, the lower value of the support vector is 0 and the upper value is 0.3.
The support vector of the regional effects and the trend are chosen using as a guide the estimates of model (12) obtained using the within estimator for panel data. Finally, for the support vector of the elements of the error vector we follow the current practice in the literature. Golan, Judge and Miller (1996: p. 88) suggest to set the lower and upper values of the support vector of the error as $\pm 3$ times the standard deviation of

[5] In the previous section, *t* was used to denote observations. Here, we use the conventional notation in the panel data literature.

the dependent variable.[6] Following this rule, we choose a vector with three values centered in zero. Therefore the support vectors chosen are:

$$\begin{aligned}
&\boldsymbol{b}_i = (-12.5,\ 0.0,\ 12.5)\ \forall i = 1,\ 2, \ldots,\ 17 \\
&\boldsymbol{b}_\tau = (0.0,\ 0.05,\ 0.1) \\
&\boldsymbol{b}_{tt} = (-0.05,\ 0.0,\ 0.05) \\
&\boldsymbol{b}_K = (0.3,\ 0.4,\ 0.5) \\
&\boldsymbol{b}_L = (0.5,\ 0.6,\ 0.7) \\
&\boldsymbol{b}_G = (0.0,\ 0.15,\ 0.3) \\
&\boldsymbol{v} = (-2.6,\ 0.0,\ 2.6)
\end{aligned} \tag{13}$$

The model in (12) written in terms of support vector and probabilities requires to estimate 66 probabilities for the values of the support vector of the parameters (3 probabilities for each of the 22 parameters). Additionally, the sample has 357 observations (17 regions observed from 1980 to 2000), which implies that it is necessary to estimate 1071 probabilities for the values of the support vector of the errors (3 probabilities for each error). It is important to point out that this large number of probabilities is ancillary results for the estimation of the 22 parameters and the errors. The estimation consists of the maximization of the following entropy function:

$$\begin{aligned}
EF(\mathbf{p},\mathbf{w}) = &-\sum_{i=1}^{17}\sum_{m=1}^{3} p_{im}\ln p_{im} - \sum_{m=1}^{3} p_{\tau m}\ln p_{\tau m} - \sum_{m=1}^{3} p_{\tau\tau m}\ln p_{\tau\tau m} - \\
&-\sum_{m=1}^{3} p_{Km}\ln p_{Km} - \sum_{m=1}^{3} p_{Lm}\ln p_{Lm} - \sum_{m=1}^{3} p_{Gm}\ln p_{Gm} - \sum_{i=1}^{17}\sum_{t=1}^{21}\sum_{m=1}^{3} w_{itm}\ln w_{itm}
\end{aligned} \tag{14a}$$

Subject to:

- 357 constraints stemming from the sample using expression (12):

$$\begin{aligned}
\ln Y_{it} = &\sum_{m=1}^{3} b_{im} p_{im} + \sum_{m=1}^{3} b_{\tau m} p_{\tau m} t + \sum_{m=1}^{3} b_{\tau\tau m} p_{\tau\tau m} t^2 + \sum_{m=1}^{3} b_{Km} p_{Km}\ln K_{it} + \\
&+ \sum_{m=1}^{3} b_{Lm} p_{Lm}\ln L_{it} + \sum_{m=1}^{3} b_{Gm} p_{Gm}\ln G_{it} + \sum_{m=1}^{3} v_m w_{itm}
\end{aligned} \tag{14b}$$

[6] For example, in the normal distribution this interval covers 99% of the range of the dependent variable. Therefore, the range of the error likely includes even the limiting case in which the independent variables do not have any explanatory power. For a more formal discussion of this issue see Pukelsheim (1994).

- 22 adding-up constraints for the probabilities of the support vectors of the parameters:

$$\sum_{m=1}^{3} p_{hm} = 1 \qquad h = 1,\ldots,22 \tag{14c}$$

- 357 adding-up constraints for the probabilities of the support vector of the error terms.

$$\sum_{m=1}^{3} w_{itm} = 1 \qquad i = 1,\ldots,17;\ t = 1,\ldots,21 \tag{14d}$$

- An adding-up constraint of the errors.

$$\sum_{i=1}^{17}\sum_{t=1}^{21}\sum_{m=1}^{3} v_m w_{itm} = 0 \tag{14e}$$

## 4. Data and estimation

We have used the following variables for the estimation of equation (12). The private production (*Y*), measured by the Gross Value Added, is taken from the official Regional Accounting of Spain [INE (2003)], linked with data on previous periods published in Cordero and Gayoso (1997). We subtract the part of the production corresponding to services not for sale to obtain a measure of private production. The data on public[7] and private capital have been collected by Mas, Pérez y Uriel (2005). Labor is the number of employees in the private sector as reported by Mas *et al.* (2002).

In the third column of Table 1 we show the results obtained using the within estimator.[8] The elasticity of private capital is not significantly different from zero and the public capital affects negatively production. These results are probably the consequence of multicollinearity. The estimates by ME of the parameters of equation (12) using the

[7] Social capital (education and health care infrastructure) is not included as part of public capital. Several papers have shown the negligible effect of these expenditures on production. Baltagi y Pinnoi (1995) claim that social public capital stock may not be the best index of education and health services.

[8] We estimate a 'fixed effects' model. The null hypothesis of no correlation between the specific effects and the explanatory variables can be rejected at the 0.01 significance level using the Hausman test.

support vector defined in (13) are shown in the fourth column of Table 1.[9] In this case, the output elasticity of public capital is 0.17 what indicates that public capital affects positively and substantially private production.

**Table 1: Estimates of the parameters of the production function**

| Variable | Parameter | Within | ME [a] | ME [b] | ME [c] |
|---|---|---|---|---|---|
| Infrastructures | $\theta_G$ | -0.049 | 0.174 | 0.188 | 0.190 |
| | | (-2.23) | | | |
| Labor | $\theta_L$ | 0.480 | 0.605 | 0.608 | 0.609 |
| | | (15.28) | | | |
| Private capital | $\theta_K$ | 0.009 | 0.413 | 0.423 | 0.430 |
| | | (0.25) | | | |
| $t^2$ | $\theta_{\tau\tau}$ | -0.0002 | -0.002 | -0.002 | -0.002 |
| | | (-3.90) | | | |
| $t$ | $\theta_\tau$ | 0.029 | 0.042 | 0.042 | 0.042 |
| | | (13.96) | | | |

t-ratios in parenthesis

*a*: Estimates obtained using the support vector in (13)

*b*: Estimates obtained using the support vector in (15a)

*c*: Estimates obtained using the support vector in (15b)

The dependence of ME estimates on prior information on the parameter values is an interesting feature of the method but also a potential weakness. Therefore, it is reasonable to be concerned about the relative contribution of support vectors and sample data to ME estimates. In this paper, we explore this issue by estimating the model in (12) using two additional sets of support vectors:

[9] The estimates were obtained using the CONOPT2 algorithm for nonlinear optimization in GAMS.

$$
\begin{aligned}
&\boldsymbol{b}_i = (-12.5,\ 0.0,\ 12.5)\ \forall i = 1,\ 2,\ ...,\ 17 \\
&\boldsymbol{b}_\tau = (0.0,\ 0.05,\ 0.1) \\
&\boldsymbol{b}_{\tau\tau} = (-0.05,\ 0.0,\ 0.05) \\
&\boldsymbol{b}_K = (0.2,\ 0.4,\ 0.6) \\
&\boldsymbol{b}_L = (0.4,\ 0.6,\ 0.8) \\
&\boldsymbol{b}_G = (-0.15,\ 0.15,\ 0.45) \\
&\boldsymbol{v} = (-2.6,\ 0.0,\ 2.6)
\end{aligned} \tag{15a}
$$

$$
\begin{aligned}
&\boldsymbol{b}_i = (-12.5,\ 0.0,\ 12.5)\ \forall i = 1,\ 2,\ ...,\ 17 \\
&\boldsymbol{b}_\tau = (0.0,\ 0.05,\ 0.1) \\
&\boldsymbol{b}_{\tau\tau} = (-0.05,\ 0.0,\ 0.05) \\
&\boldsymbol{b}_K = (0.1,\ 0.4,\ 0.7) \\
&\boldsymbol{b}_L = (0.3,\ 0.6,\ 0.9) \\
&\boldsymbol{b}_G = (-0.30,\ 0.15,\ 0.60) \\
&\boldsymbol{v} = (-2.62,\ 0.0,\ 2.62)
\end{aligned} \tag{15b}
$$

The supports vectors in (15) are chosen keeping the central value but increasing the range of the support vectors of output elasticities twofold in (15a) and three times in (15b) with respect to the range of the support vectors in (13). The results of the estimation are shown in the last two columns of Table 1. The estimates are not very sensitive to changes in the values of the support vector. Therefore, the prior information included through the support vectors do affect the estimation by ME but the empirical evidence obtained is by no means completely determined by these vectors.

An important bottom line is that the results in Table 1 imply positive estimates of the productivity of public capital for different support vectors. In the first estimation, we construct the support vector using seemingly relevant information on parameter values while in the second and third case the support vectors are quite wide and seem less relevant in guiding the estimation. However, all the estimates support the idea of a positive productivity of public capital.

The previous exercise shows how the point estimates could be affected by the choice of the support vector. This result is usually seen as a problem of ME. However, in this paper we propose to use the differences in the estimates caused by changes in the support vector to check the strength of the empirical evidence obtained by ME. In particular, we explore the chances of obtaining a negative or null estimate of the output elasticity of public capital by changing the support vector of the parameters. A null or negative productivity of public capital is a counterintuitive result that has been found sometimes in empirical research. Using different support vectors for the output elasticity of public capital, we are able to show some evidence against this result. For this purpose, we estimate the model in (12) using the support vectors *a*, *b*, *c* and *d* in Table 2. These vectors are constructed fixing the higher value of the support vector in 0.3 and choosing increasingly negative values for the lower value. For the other parameters in the model, we use the support vectors in (13). The results in Table 2 show that the output elasticity of public capital has remained above 0.13 in all instances. In other words, ME does not produce a null or negative estimate of the output elasticity of public capital when the support vector clearly contemplates that outcome. This result can be interpreted as additional evidence favorable to the existence of a positive effect of public capital in production.

**Table 2: Estimation by ME with different support vectors**

| Variable | Parameter | Support vector *a* | Support vector *b* | Support vector *c* | Support vector *d* |
|---|---|---|---|---|---|
| Infrastructures | $\theta_G$ | 0.155 | 0.145 | 0.140 | 0.138 |
| Labor | $\theta_L$ | 0.606 | 0.607 | 0.607 | 0.607 |
| Private capital | $\theta_K$ | 0.416 | 0.418 | 0.419 | 0.420 |
| $t^2$ | $\theta_{\tau\tau}$ | -0.002 | -0.002 | -0.002 | -0.002 |
| $t$ | $\theta_{\tau}$ | 0.043 | 0.043 | 0.043 | 0.043 |

Support vector *a* : $\mathbf{b}_G$ = (-0.1,,0.1,,0.3)
Support vector *b*: $\mathbf{b}_G$ = (-0.2, 0.05, 0.3)
Support vector *c*: $\mathbf{b}_G$ = (-0.3, 0.0, 0.3)
Support vector *d*: $\mathbf{b}_G$ = (-0.4, -0.05, 0.3)

The ME and Within estimates have been used above to highlight the differences between classical econometrics and ME estimation. The basic difference is that ME uses prior information to deal with samples with multicollinearity. The comparison can be strengthened by assuming the existence of constant returns to scale (CRS) in the production function. In this case, the econometric estimation is done under the assumption that the output elasticities of private inputs add up to one ($\theta_K + \theta_L = 1$). In ME, the same parametric restriction can be used to estimate the parameters without a support vector for these two parameters because they can be estimated directly as probabilities since they add-up to one. Therefore, under CRS, both methods use similar prior information on the output elasticities of private capital and labor. In a model with CRS, the entropy function becomes:

$$\underset{p,w}{Max}\, EF(p,w) = -\sum_{i=1}^{17}\sum_{m=1}^{3} p_{im} \ln p_{im} - \sum_{m=1}^{3} p_{\tau m} \ln p_{\tau m} - \sum_{m=1}^{3} p_{\tau\tau m} \ln p_{\tau\tau m} - \\ -\theta_K \ln\theta_K - \theta_L \ln\theta_L - \sum_{m=1}^{3} p_{Gm} \ln p_{Gm} - \sum_{i=1}^{17}\sum_{t=1}^{21}\sum_{m=1}^{3} w_{itm} \ln w_{itm} \tag{16}$$

where, $\theta_K + \theta_L = 1$ is now one of the constraints of the optimization program.

The results of estimating equation (12) under CRS by both methods are shown in Table 3. The support vector for the regional effects have been constructed using the Within estimates under CRS (-5.0, 0.0, 5.0) following the same rule adopted for previous estimates. We use a wide support vector for the output elasticity of public capital that, apparently, does not contain relevant information for the estimation by ME (0.0, 0.5, 1.0). The support vectors of the trend and the error terms are not changed for this estimation.

**Table 3: Production function under CRS**

| Variable | Parameter | ME | Within |
|---|---|---|---|
| Infrastructure | $\theta_G$ | 0.285 | -0.030 |
| | | | (-0.97) |
| Labor | $\theta_L$ | 0.651 | 0.646 |
| | | | (8.46) |
| Private capital | $\theta_K$ | 0.349 | 0.354 |
| | | | (8.46) |
| $t^2$ | $\theta_{\tau\tau}$ | -0.002 | -0.0005 |
| | | | (-6.51) |
| $t$ | $\theta_\tau$ | 0.040 | 0.024 |
| | | | (8.22) |

t-ratios in parenthesis

The results in Table 3 show that under CRS the estimates of the output elasticities of private capital and labor are very similar in ME and Within. However, both methods give different estimates of the output elasticity of public infrastructure. The within estimator gives a negative estimate(although not significantly different from zero) while by ME gives a positive and substantive estimate for the output elasticity of public capital.

As a summary, the results in this section illustrate the opportunities of empirical analysis raised by ME in models with multicollinearity and provide fresh empirical evidence about the productivity of public capital.

## 5. Conclusions

In this paper, we have explored the use of maximum entropy estimation for empirical analysis when multicollinearity is an issue. This problem emerges in the estimation of a Cobb-Douglas production function for the Spanish regions in which public capital is included as an input. The key element of the estimation by maximum entropy is the introduction of prior information on parameter values. As a result, it is important to analyze the changes in estimates with changes on this information. For that purpose, we first estimate the production function with three quite different sets of support vectors. In all cases, the output elasticity of public capital is positive and quite substantive. Using a third set of support vectors, we have shown that the data provide quite strong evidence against a negative value of the output elasticity of public capital.

Although the results obtained with maximum entropy estimation need a detailed analysis of sensibility, the methodology appears as a useful procedure to gather empirical evidence when the use of more common econometric methods is impeded or precluded by multicollinearity. This conclusion stems mainly from the thorough comparison of maximum entropy estimation with the within estimator. Due to multicollinearity, the within estimator fails to provide evidence of a positive productivity of public capital while that evidence is quite strong estimating the model by maximum entropy.

## References


Ai, C. and Cassou, S., (1997): "On public capital analysis with state data", *Economics Letters*, Vol. 57, 209-212.

Aschauer, D.A., (1989), "Is public expenditure productive?", *Journal of Monetary Economics*, 23(2), 177-200.

Baltagi, B.H. and N. Pinnoi, (1995), "Public Capital Stock and State Productivity Growth: Further Evidence from an Error Components Model", *Empirical Economics*, Vol. 20, 351-359.

Cordero, G. and A. Gayoso, (1997), *Evolución de las economías regionales en los primeros 90*, Dirección de Análisis y Programación Presupuestaria, Ministerio de Economía y Hacienda, November, Madrid.

de la Fuente, Á., (2002), "Infraestructuras y productividad: un panorama de la literatura", in de la Fuente, A., *Fondos Estructurales, inversión en infraestructuras y crecimiento regional*, Documentos de Economía, Nº 18, Centro de Investigación Económica y Financiera, Fundación Caixa Galicia, 81-142.

Fraser, I., (2000), "An application of maximum entropy estimation: the demand for meat in the United Kingdom", *Applied Economics*, Vol. 32(1), 45-59.

Garcia-Milà, T. and T.J. McGuire, (1992), "The contribution of publicly provided inputs to states' economies", *Regional Science and Urban Economics*, 22, 229-241.

Garcia-Milà, T., McGuire, T., and H. Porter, (1996), "The effect of public capital in state-level production functions reconsidered", *The Review of Economics and Statistics*, Vol. 78(1), 177-180.

Gardebroek, C. and A. Oude Lansink, (2004), "Farm-specific Adjustment Costs in Dutch Pig Farming", *Journal of Agricultural Economics*, Vol. 55(1), 3-24.

Golan, A., Judge, G. and D. Miller, (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York, John Wiley & Sons.

Golan, A., Perloff, J.M. and E.Z. Shen, (2001), "Estimating a Demand System with Nonnegativity Constraints: Mexican Meat Demand", *The Review of Economics and Statistics*, Vol. 83(3), 541-550.

Goldberger, A.S., (1991), *A Course in Econometrics*, Cambridge-USA, Harvard University Press.

Gramlich, E. M., (1994), "Infrastructure Investment: A Review Essay", *Journal of Economic Literature*, Vol. XXXII, sep, 1176-1196.

Greene, W., (2000), *Econometric analysis*, New Jersey, Prentice Hall.

Holtz-Eakin, D., (1994), "Public-sector capital and the productivity puzzle", *The Review of Economics and Statistics*, Vol. 76 (1), feb., 12-21.

INE, (2003), *Contabilidad Regional de España. Base 1995*, Madrid.

Marsh, T.L. and C. Mittelhammer, (2004), *Generalized Maximum Entropy Estimation of a First Order Spatial Autoregressive Model*, Advances in Econometrics, Vol.18, 199-234, Elsevier.

Mas, M., Pérez, F. and E. Uriel, (2005), *El stock de capital en España y su distribución territorial (1964-2002)*, Bilbao, Fundación BBVA.

Mas, M., Pérez, F., Serrano, L, Soler, A. and E. Uriel, (2002), *Capital humano, series históricas 1964-2001*, Valencia, BanCaja.

Munnell, A., (1990), "How does public infrastructure affect regional economic performance?", *New England Economic Review*, Federal Reserve Bank of Boston, set.-oct, 11-32.

Paris, Q. and R.E. Howitt, (1998), "An Analysis of Ill-Posed Production Problems Using Maximum-Entropy", *American Journal of Agricultural Economics*, Vol. 80, 124-138.

Pukelsheim, F., (1994), "The three sigma rule", *The American Statiscian*, Vol. 48(4), 88-91.

Shannon, C.E., (1948), "A Mathematical Theory of Communication", *Bell System Technical Journal*, Vol. 27, 379-423.

Vijverberg, W.P.M., Vijverberg, C.P.C., and J.L. Gamble, (1997), "Public capital and private productivity", *The Review of Economics and Statistics*, Vol.79(2), 267-278.