Departamento de Economía Aplicada
Universidad de Oviedo

(DPAE/09/03)

Discussion Papers on Applied Economics.

Department of Applied Economics. University of Oviedo.

# R&D AND PRODUCTIVITY GROWTH: ALLOWING FOR SPATIAL HETEROGENEITY BY ENTROPY ECONOMETRICS

Esteban Fernández-Vázquez (evazquez@uniovi.es)*

Fernando Rubiera-Morollón (frubiera@uniovi.es)

University of Oviedo, Department of Applied Economics, Faculty of Economics, Campus del Cristo, Oviedo, 33006, Spain.

* Corresponding author.

Abstract:

In this paper we adopt an entropy econometrics-based estimator to study regional variations in regression coefficients and apply it to analyze productivity growths generated by R&D activities at a regional level. Considering the possible effects of the region's own R&D stock as well as the spillovers produced in other regions, the paper proposes the use of an entropy-based technique to estimate these effects for a specific location. Depending on the degree of heterogeneity of the set of regions analyzed, it is possible that some of these regions present characteristics that enable them to more easily convert R&D efforts (generated in the region itself or obtained from other regions by R&D spillovers) into productivity gains, whereas in other regions the effect of (direct or spillover generated) R&D activities may be irrelevant. We illustrate this idea with an empirical application for Spanish regions.

Keywords: R&D spillovers; productivity growth, entropy econometrics, Spain.
JEL codes: O30, O40, R11.

# 1. Introduction.

Since the beginnings of Economics as a science, economists have attempted to understand the factors of economic growth. The first approaches focused on the role of physical capital; see the classical works by Harrod (1939) and Domar (1946). However, contributions by the American economists Denison, Kuznets and Solow produced a veritable revolution in the approach to economic growth analysis. In their research, these authors understand growth as a complex phenomenon in which human capital, social cohesion and the capacity to innovate are even more relevant factors than physical capital. The relevance of the innovation and capacity of territories to experiment techological change was heightened following the major contributions of Romer (1990), Grossman and Helpman (1991), Aghion and Howitt (1992), Coe and Helpman (1995) or Jones (1995), among others. All the aforementioned papers pointed out the remarkable effects of technology, in general, and R&D activity, in particular, on economic growth.

In keeping with this line of thought, a growing literature measuring the effect of technological activities on economic growth has appeared over the last decade. The majority of recent empirical studies have attempted to estimate how R&D spending and knowledge accumulation contribute to territorial (i.e., national or regional) productivity growth and, through increased productivity, how they lead to long-term economic growth. The goal of the present research is to measure how higher stocks of knowledge accumulate comparative advantages for achieving economic growth and how the effects derived from accumulating knowledge in one specific area can be transmitted to other locations; i.e., through technology spillovers.

Among other authors, the spatial effects of technological change has been studied by Brugger and Stuckey (1987), Todtling (1990) or, more recently, by Wakelin (2001), Bottazzi and Peri (2003) and Varga and Schalk (2004), detecting a positive correlation between technological activities and economic development. For the specific case of Spain, De la Fuente (2002) links R&D effects on regional productivity with regional convergence and growth patterns of Spanish provinces. Other recent studies by Gumbau-Albert and Maudos (2006) or Lopez-Bazo *et al.* (2006) investigated the link between regional TFP growth and R&D activities, observing an important effect of regional spillovers associated with technological activities[1].

In this paper we present a new methodological approach to measuring technological spillovers based on Entropy Econometrics. The paper is organized in six sections. The first section introduces some basic concepts regarding the general framework and econometrical approach. Section 2 presents the general basis of the entropy econometric that provides the basis for the estimation we shall apply. In Section 3 we present a particular entropy-based estimator that allows for incorporating several types of *a priori* information. The main advantage of this estimator is that, even in situations characterized by a lack of large data samples, it is capable of identifying extraneous variables at the same time as it estimates the

[1] See Breschi and Lissoni (2001) or Döring and Schnellenbach (2006) for an extensive review of research works that study the impact of regional knowledge spillovers.

relevant parameters of the specified model. This estimation technique can be very useful when we wish to estimate models at a specific-regional level. Section 4 shows an empirical application to Spanish regions and discusses the main results. Finally, Section 5 offers the conclusions drawn.

## 2. Basic Framework and Econometric Approach.

Similarly to many of the works previously cited, we will focus our interest on the TFP measure, based on Solow's residual. To calculate the value of TFP, let us consider an extended Cobb-Douglas regional production functions with constant returns to scale of the physical factors like the following:

$$y_{it} = \mu_t L_{it}^{\alpha} K_{it}^{(1-\alpha)} R_{it}^{\beta_R} IR_{it}^{\beta_{IR}} \qquad (1)$$

in which *y* stands for value added, *K* indicates the stock of physical capital, *L* denotes employment, *R* indicates the stock of R&D in a region, *IR* is the indirect stock of R&D obtained through spillovers from other regions and *t* and *i* are time and regional index respectively. The term $\mu$ is a parameter related to the initial level of productivity in the region. Taking logarithms, (1) is transformed into:

$$ln\, y_{it} = ln\, \mu_{it} + \beta_L \ln L_{it} + (1-\beta_L) \ln K_{it} + \beta_R \ln R_{it} + \beta_{IR} \ln IR_{it} \qquad (2)$$

and:

$$\begin{aligned} TFP_{it} &= ln\, y_{it} - \beta_L\, ln\, L_{it} - (1-\beta_L)\, ln\, K_{it} \\ &= ln\, \mu_{it} + \beta_R\, ln\, R_{it} + \beta_{IR}\, ln\, IR_{it} \end{aligned} \qquad (3)$$

Taking first differences on (3), TFP growth between two periods is given by the following equation:

$$T\dot{F}P_{it} = \dot{\mu}_{it} + \beta_R\, \dot{R}_{it} + \beta_{IR} I\dot{R}_{it} \qquad (4)$$

where the dots denote growth rates. Equation (4) shows the empirical model that is usually estimated for measuring the contribution of R&D activities to regional productivity growth. It is important to note that this equation is a global model with constant parameters for all the regions considered. The general problem is that in regression models where the cases are located geographically, sometimes regression coefficients do not remain fixed, but they might vary over space. In other words, imposing a common structure to all the set of regions studied can be unrealistic, given that the size of the effects of technology activities on productivity growth can vary across space depending on the extent of agglomeration economies in some regions. For example, in Varga (2000), Varga and Schalk (2004) or Ciriaci and Palma (2008) knowledge spillovers are allowed to vary across regions depending on economic concentration of economic activities.

Previous solutions have been proposed to capture this spatial drift from the global model: the geographically weighted regression (GWR, see Brundson et al., 1996) which can be seen as an extension of the parameter expansion method (Casetti, 1972). Basically, GWR consists on estimating regressions like:

$$T\dot{F}P_{it} = \dot{\mu}_{it}(n_i, e_i) + \beta_R(n_i, e_i)\,\dot{R}_{it} + \beta_{IR}(n_i, e_i)I\dot{R}_{it} \quad (5)$$

where the parameters are defined as functions of the geographical location $(n_i, e_i)$ of region *i*. If these functions are constant for all the $(n_i, e_i)$, then equations (4) and (5) are the same.

In this paper we propose a different approach. Similarly to GWR, we will assume that parameters $\beta_R$ and $\beta_{IR}$ can vary across the different territories. But we do not fix an *a priori* geographical function to explain parameter variation; instead we let the data speak for themselves and we will estimate equations like (4) for each and every region independently. In other words, if we have data for a group of G regions, we estimate the following set of G regression equations:

$$\begin{aligned} T\dot{F}P_{1t} &= \dot{\mu}_{1t} + \beta_{1R}\,\dot{R}_{1t} + \beta_{1IR}I\dot{R}_{1t} \\ T\dot{F}P_{2t} &= \dot{\mu}_{2t} + \beta_{2R}\,\dot{R}_{2t} + \beta_{2IR}I\dot{R}_{2t} \\ &(\ldots) \\ T\dot{F}P_{Gt} &= \dot{\mu}_{Gt} + \beta_{GR}\,\dot{R}_{Gt} + \beta_{GIR}I\dot{R}_{Gt} \end{aligned} \quad (5)$$

Estimating this type of equations can be problematic if we use traditional estimation techniques. Basically, the lack of large series of data at a regional level prevents of using Least Squares-Maximum Likelihood estimators, given the reduced numbers of degrees of freedom. Instead, we propose the application of Entropy Econometrics (EE) to estimate such equations, given that these techniques have interesting properties when dealing ill-conditioned estimation problems (small samples or data sets affected by large collinearity). In Golan et al. (1996) or Kapur and Kesavan (1992) extensive descriptions of the entropy estimation approach can be found.

## 3. A basis for Entropy Econometrics.

### *3.1. The Cross-Entropy technique*

Generally speaking, EE techniques are used to recover unknown probability distributions of discrete random variables that can take M different known values. The estimate $\tilde{\boldsymbol{p}}$ of the unknown probability distribution $\boldsymbol{p}$ must be as similar as possible to an appropriate *a priori* distribution $\boldsymbol{q}$, constrained by the observed data. Specifically, the Cross-Entropy (CE) procedure estimates $\tilde{\boldsymbol{p}}$ by minimizing the Kullback-Leibler divergence $D(\boldsymbol{p}\|\boldsymbol{q})$ (Kullback, 1959):

$$\underset{\boldsymbol{p}}{\text{Min}}\, D(\boldsymbol{p}\|\boldsymbol{q}) = \sum_{m=1}^{M} p_m \ln\left(\frac{p_m}{q_m}\right) \quad (7)$$

The divergence $D(\boldsymbol{p}\|\boldsymbol{q})$ measures the dissimilarity of the distributions $\boldsymbol{p}$ and $\boldsymbol{q}$. This measure reaches its minimum (zero) when $\boldsymbol{p}$ and $\boldsymbol{q}$ are identical and this minimum is reached when no constrains are imposed. If some information (for example, observations on the variable) is available, each piece of information will lead to a Bayesian update of the *a priori* distribution $\boldsymbol{q}$.

The underlying idea of the CE methodology can be applied for estimating the parameters of general linear models, which leads us to the so-called generalized Cross Entropy (GCE). Let us suppose a variable $y$ that depends on H explanatory variables $x_h$:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{8}$$

where $\boldsymbol{y}$ is a $(T \times 1)$ vector of observations for $y$, $\boldsymbol{X}$ is a $(T \times H)$ matrix of observations for the $x_h$ variables, $\boldsymbol{\beta}$ is the $(H \times 1)$ vector of unknown parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_H)$ to be estimated, and $\boldsymbol{\epsilon}$ is a $(T \times 1)$ vector with the random term of the linear model. Each $\beta_h$ is assumed to be a discrete random variable. We assume that there is some information about its $M \geq 2$ possible realizations. This information is included for the estimation by means of a support vector $\boldsymbol{b}' = (b_1, \dots, b_M)$ with corresponding probabilities $\boldsymbol{p'}_h = (p_{h1}, \dots, p_{hM})$. The vector $\boldsymbol{b}$ is based on the researcher's a priori belief about the likely values of the parameter. For the sake of convenient exposition, it will be assumed that the *M* values are the same for every parameter, although this assumption can easily be relaxed. Now, vector $\boldsymbol{\beta}$ can be written as:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_H \end{bmatrix} = \boldsymbol{BP} = \begin{bmatrix} \boldsymbol{b'} & 0 & \cdots & 0 \\ 0 & \boldsymbol{b'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{b'} \end{bmatrix} \begin{bmatrix} \boldsymbol{p_1} \\ \boldsymbol{p_2} \\ \vdots \\ \boldsymbol{p_H} \end{bmatrix} \tag{9}$$

where $\boldsymbol{B}$ and $\boldsymbol{P}$ have dimensions $(H \times HM)$ and $(HM \times 1)$ respectively. Now, the value of each parameter $\beta_h$ is given by the following expression:

$$\beta_h = \boldsymbol{b'}\boldsymbol{p_h} = \sum_{m=1}^{M} b_m \, p_{hm}; \ \forall h = 1, \dots, H \tag{10}$$

For the random term, a similar approach is followed. Oppositely to other estimation techniques, GCE does not require rigid assumptions about a specific probability distribution function of the stochastic component, but it still is necessary to make some assumptions. $\boldsymbol{\epsilon}$ is assumed to have mean $E[\boldsymbol{\epsilon}] = 0$ and a finite covariance matrix. Basically, we represent our uncertainty about the realizations of vector $\boldsymbol{\epsilon}$ treating each element $\epsilon_t$ as a discrete random variable with $J \geq 2$ possible outcomes contained in a convex set $\boldsymbol{v'} = \{v_1, \dots, v_J\}$, which for the sake of simplicity is assumed as common for all the $\epsilon_t$. We also assume that these possible realizations are symmetric around zero ($-v_1 = v_J$). The traditional way of fixing the upper and lower limits of this set is to apply the three-sigma rule (see Pukelsheim, 1994). Under these conditions, vector $\boldsymbol{\epsilon}$ can be defined as:

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix} = \boldsymbol{VW} = \begin{bmatrix} \boldsymbol{v'} & 0 & \cdots & 0 \\ 0 & \boldsymbol{v'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{v'} \end{bmatrix} \begin{bmatrix} \boldsymbol{w_1} \\ \boldsymbol{w_2} \\ \vdots \\ \boldsymbol{w_T} \end{bmatrix} \tag{11}$$

and the value of the random term for an observation *t* equals:

$$\epsilon_t = \boldsymbol{v'}\boldsymbol{w_t} = \sum_{j=1}^{J} v_j w_{tj}; \ \forall t = 1, \dots, T \tag{12}$$

and, consequently, model (8) can be transformed into:

$$\boldsymbol{y} = \boldsymbol{XBp} + \boldsymbol{Vw} \tag{13}$$

So we need also to estimate the elements of matrix $\boldsymbol{W}$ (denoted by $\widetilde{w}_{tj}$) and the estimation problem for the general linear model is transformed into the estimation of $H + T$ probability distributions. For this estimation, once specified the a priori probability distributions $\boldsymbol{Q}$ and $\boldsymbol{W^0}$ respectively for $\boldsymbol{P}$ and $\boldsymbol{W}$, the GCE problem is written in the following terms:

$$\underset{\boldsymbol{P},\boldsymbol{W}}{\text{Min}}\, D(\boldsymbol{P}, \boldsymbol{W} \| \boldsymbol{Q}, \boldsymbol{W^0}) = \sum_{h=1}^{H} \sum_{m=1}^{M} p_{hm} \, ln\left(\frac{p_{hm}}{q_{hm}}\right) + \sum_{t=1}^{T} \sum_{j=1}^{J} w_{tj} \, ln\left(\frac{w_{tj}}{w_{tj}^{0}}\right) \tag{14}$$

subject to:

$$y_t = \sum_{h=1}^{H} \sum_{m=1}^{M} b_m p_{hm} x_{ht} + \sum_{j=1}^{J} v_j w_{tj}; \ \forall t = 1, \dots, T \tag{15}$$

$$\sum_{m=1}^{M} p_{hm} = 1 ; \ \forall h = 1, \dots, H \tag{16}$$

$$\sum_{j=1}^{J} w_{tj} = 1 ; \ \forall t = 1, \dots, T \tag{17}$$

The restrictions in (15) ensure that the posterior probability distributions of the estimates and the errors are compatible with the observations. The equations in (16) and (17) are just normalization constraints. This GCE estimation procedure can be seen as an extension of the particular Generalized Maximum Entropy (GME) principle (or alternatively the GME can be considered as a particular case of the GCE procedure), given that the solutions of both approaches are the same when the *a priori* probability distribution contained in $\boldsymbol{Q}$ are all uniform. In other words, the ME solutions are obtained by minimizing the Kullback-Leibler divergence $D(\boldsymbol{P}\|\boldsymbol{Q})$ between the unknown $p_{hm}$ and the *a priori* $q_{hm} = \frac{1}{M}\ \forall m = 1,..,M$ . The same happens for the estimation of $w_{tj}$ when the *a priori* probabilities are fixed as uniform ($w_{tj}^0 = \frac{1}{J}\ \forall j = 1,..,J$ ), which is the natural (and most frequently applied) point of departure to reflect the uncertainty about $\boldsymbol{\epsilon}$.

### *3.2. A data-weighted prior estimator*

The sketched GCE procedure can be extended in order to develop a more flexible estimator that allows for simultaneous parameter estimation and variable selection in linear statistical models. Related to the Bayesian Method of Moments (BMOM, see Zellner, 1996, 1997), the technique has been proposed in Golan (2001) as data-based method of estimation that uses both sample and non-sample information in determining a basis for coefficient reduction and extraneous variable identification. In other words, this technique allows for shrinking the coefficient of the explanatory variables that can be classified as irrelevant in the linear model. A recent empirical application of this method can also be found in Bernadini (2008).

*Coefficient estimation.*

Our objective is to identify the extraneous variables included in the model and simultaneously produce an estimator with a good sampling performance over the whole range of the parameter space, which can be achieved by the GCE estimator if we combine uniform priors with spike priors. We start by specifying a discrete support space ***b*** for each $\beta_h$ (and the same for ***v***) symmetric around zero, with large lower and upper bounds for ***b*** and the three-sigma rule for ***v***, so that each $\beta_h$ and $\epsilon_t$ are contained in the chosen interval with high probability. Besides fixing an uniform probability distribution that will be used as ***q*** in the GCE estimation (i.e.; $q_m = \frac{1}{M}$ $\forall\, b_m$)., we also specify a “spike” prior for each $\beta_h$, with a very high probability at $b_m = 0$ (i.e.; $q_m \cong 1$ if $b_m = 0$ and $q_m \cong 0$ for the remaining values). Thus, a flexible, data-based prior is specified such that for each $\beta_h$ coordinate either a spike prior at the $b_m = 0$, a uniform prior over the discrete support space ***b***, or any convex combination of the two can result. If we denote with $\boldsymbol{q^u}$ and $\boldsymbol{q^s}$ the uniform and spike a priori distributions respectively, the objective proposed can be achieved by modifying the previous GCE program in the following way:

$$\underset{\boldsymbol{P},\boldsymbol{P}^{\gamma},\boldsymbol{W}}{\text{Min}}\, D(\boldsymbol{P},\boldsymbol{P}^{\gamma},\boldsymbol{W}\|\boldsymbol{Q},\boldsymbol{Q}^{\gamma},\boldsymbol{W}^{\mathbf{0}}) = \sum_{h=1}^{H}(1-\gamma_h)\sum_{m=1}^{M} p_{hm}\, ln\left(\frac{p_{hm}}{q_{hm}^{u}}\right) + \sum_{h=1}^{H}\gamma_h \sum_{m=1}^{M} p_{hm}\, ln\left(\frac{p_{hm}}{q_{hm}^{s}}\right) + \sum_{h=1}^{H}\sum_{n=1}^{N} p_{hn}^{\gamma}\, ln\left(\frac{p_{hn}^{\gamma}}{q_{hn}^{\gamma}}\right) + \sum_{t=1}^{T}\sum_{j=1}^{J} w_{tj}\, ln\left(\frac{w_{tj}}{w_{tj}^{0}}\right) \quad (18)$$

subject to:

$$y_t = \sum_{h=1}^{H}\sum_{m=1}^{M} b_m p_{hm} x_{ht} + \sum_{j=1}^{J} v_j\, w_{tj};\ \ \forall t = 1,\dots,T \quad (19)$$

$$\sum_{m=1}^{M} p_{hm} = 1\,;\ \forall h = 1,\dots,H \quad (20)$$

$$\sum_{j=1}^{J} w_{tj} = 1\,;\ \forall t = 1,\dots,T \quad (21)$$

$$\sum_{n=1}^{N} p_{hn}^{\gamma} = 1\,;\ \forall h = 1,\dots,H \quad (22)$$

The $\gamma_h$ parameters are estimated simultaneously with the rest of $\beta_h$ coefficients of the model. Each $\gamma_h$ measures the weight given to the spike prior $\boldsymbol{q}^{s}$ for each parameter $\beta_h$ and it is defined as $\tilde{\gamma}_h = \sum_{n=1}^{N} b_{hn}^{\gamma}\, \tilde{p}_{hn}^{\gamma}$, where $b_{h1}^{\gamma} = 0$ and $b_{hN}^{\gamma} = 1$ are respectively the lower and upper bound defined as the support of these parameters ($\boldsymbol{b'}_h = (0,\dots,1) \rightarrow 0 \leq \gamma_h \leq 1;\ \forall h = 1,..,H$). The *a priori* probability distributions fixed for them are always uniform ($q_h^{\gamma} = \frac{1}{N}\ \forall n = 1,..,N$) and the same is applied for the errors (again $w_{tj}^{0} = \frac{1}{J}\ \forall j = 1,..,J$).

To understand the logic of this data-weighted prior (DWP) estimator an explanation on the objective function of the previous minimization program is required. Note that equation (18) is divided in four terms. The last term is exactly the same as in the GCE program and it measures the Kullback divergence between the posterior and the prior probabilities for the noise component of the model. The first term quantifies the divergence between the recovered probabilities and the uniform priors for each $\beta_h$ coefficient, being this divergence weighted by $(1-\gamma_h)$. On the contrary, the second element of (18) measures the divergence with the spike prior and it s weighted by $\gamma_h$. The third element in (18) relates to the Kullback divergence of the weighting parameters $\gamma_h$.

The solutions of this minimization program are:

$$\tilde{p}_{hm} = \frac{q_{hm}^{\tilde{\gamma}_h/A_h} exp\left[\left(A_h{}^{-1}\right)\sum_{t=1}^{T}\tilde{\lambda}_t b_m x_{ht}\right]}{\sum_{m=1}^{M} q_{hm}^{\tilde{\gamma}_h/A_h} exp\left[\left(A_h{}^{-1}\right)\sum_{t=1}^{T}\tilde{\lambda}_t b_m x_{ht}\right]}; \forall h = 1, \dots H; m = 1, \dots, M \tag{23}$$

$$\widetilde{w}_{tj} = \frac{w_{tj}^{0} exp\left[\tilde{\lambda}_t v_j\right]}{\sum_{j=1}^{J} w_{tj}^{0} exp\left[\tilde{\lambda}_t v_j\right]}; \forall t = 1, \dots T; j = 1, \dots, J \tag{24}$$

were:

$$\tilde{\gamma}_h = \sum_{n=1}^{N} b_{hn}^{\gamma}\, \tilde{p}_{hn}^{\gamma} \tag{25}$$

$$A_h = [1 - \tilde{\gamma}_h] \Big/ \left[(\tilde{\gamma}_h - 1) lnM - \tilde{\gamma}_h \sum_{m=1}^{M} q_{hm}^{u} ln(q_{hm}^{u}) + \tilde{\gamma}_h\right] \tag{26}$$

and $\tilde{\lambda}_t$ are the Lagrangian multipliers associated with restrictions (19). From the recovered $\tilde{p}_{hm}$ probabilities, the estimated value of each parameter $\beta_h$ is obtained as:

$$\tilde{\beta}_h = \sum_{m=1}^{M} b_m\, \tilde{p}_{hm}; \ \forall h = 1, \dots, H \tag{27}$$

Under some mild assumptions (see Golan 2001, page 177) the consistency and asymptotic normality of the DWP estimates can be ensured. Additionally, these assumptions also guarantee that the approximate variances of the DWP estimator is lower than the approximate variance of the GCE estimator, which in turn is lower than the approximate variance of a ML-LS estimator (see Golan, 2001, page 179).

*Variable selection.*

Simultaneously to the estimation of the parameters of the model, the DWP estimator discriminates between relevant and extraneous explanatory variables. The proposed estimation strategy provides two indications for this objective. Firstly, estimates of the weighting parameters $\gamma_h$, obtained as:

$$\tilde{\gamma}_h = \sum_{n=1}^{N} b_{hn}^{\gamma}\, \tilde{p}_{hn}^{\gamma}; \ \forall h = 1, \dots, H \tag{28}$$

can be used as a tool for this purpose: as $\tilde{\gamma}_h \rightarrow 0$ the prior becomes more uniform and the estimates approach those of the GME estimator. On the contrary, large values of $\tilde{\gamma}_h$, the GCE estimator with spike prior on zero takes over.

Consequently, the irrelevant variables of the model will be characterized by large values of $\tilde{\gamma}_h$ (Golan considers sufficiently large values when $\tilde{\gamma}_h > 0.49$) together with estimates of $\beta_h$ close to zero.

Moreover, a $\chi^2$ statistic can be constructed in order to test if the estimate for $\beta_h$ is significantly different from zero (and, in consequence, variable $x_h$ is not irrelevant). The basic idea is to test if the recovered $\tilde{p}_{hm}$ are significantly different from the respective spike prior $q_{hm}^s$. The Kullback-Leibler divergence between our posterior and these a priori probabilities is:

$$D_h(\tilde{\boldsymbol{p}}_{\boldsymbol{h}} \| \boldsymbol{q}_{\boldsymbol{h}}^{\boldsymbol{s}}) = \sum_{m=1}^{M} \tilde{p}_{hm} \, ln \left( \frac{\tilde{p}_{hm}}{q_{hm}^s} \right) \qquad (29)$$

And the chi-squared divergence between both distributions is:

$$\chi_{M-1}^2 = M \sum_{m=1}^{M} \frac{(\tilde{p}_{hm} - q_{hm}^s)^2}{q_{hm}^s} \qquad (30)$$

A second-order approximation of $D_h(\tilde{\boldsymbol{p}}_{\boldsymbol{h}} \| \boldsymbol{q}_{\boldsymbol{h}}^{\boldsymbol{s}})$ is the entropy-ratio statistic for evaluating $\tilde{\boldsymbol{p}}_{\boldsymbol{h}}$ versus $\boldsymbol{q}_{\boldsymbol{h}}^{\boldsymbol{s}}$:

$$D_h(\tilde{\boldsymbol{p}}_{\boldsymbol{h}} \| \boldsymbol{q}_{\boldsymbol{h}}^{\boldsymbol{s}}) \cong \frac{1}{2} \sum_{m=1}^{M} \frac{(\tilde{p}_{hm} - q_{hm}^s)^2}{q_{hm}^s} \qquad (31)$$

Consequently,

$$2MD_h(\tilde{\boldsymbol{p}}_{\boldsymbol{h}} \| \boldsymbol{q}_{\boldsymbol{h}}^{\boldsymbol{s}}) \to \chi_{M-1}^2 \qquad (32)$$

Given this relationship, we can use the measure $2MD_h(\tilde{\boldsymbol{p}}_{\boldsymbol{h}} \| \boldsymbol{q}_{\boldsymbol{h}}^{\boldsymbol{s}})$ in order to test the hypothesis $H_0$: $\beta_h = 0$. If the null hypothesis is not rejected, an extraneous variable $x_h$ is identified[2].

## 4. An empirical application for the Spanish regions.

### *4.1. The specific formulation for the case of the Spanish regions.*

The above sketched estimation technique can be very useful when we want to estimate models like the group of equation regressions depicted on (6). Depending on the degree of heterogeneity of the set of regions analyzed, it is possible that some of them present characteristics to convert more easily R&D efforts (generated on the region itself or obtained from other regions by R&D spillovers) into productivity gains, whereas in other regions the effect of (direct or spillover generated) R&D activities

[2] To prevent computational problems that appear when computing log(0), in the empirical application on the next section the spike priors $\boldsymbol{q}_{\boldsymbol{h}}^{\boldsymbol{u}}$ have been specified with a point mass at zero equal to 0.999 and 0.0005 respectively for the other points of the support vectors.

could be irrelevant. By applying the DWP estimator we have estimated to the Spanish regions a set of equations as the following:

$$\dot{y}_{it} = \mu_i + \eta_i d_t + \beta_{iR}\, \dot{R}_{it} + \beta_{iIR}\, \dot{IR}_{it} \tag{33}$$

We have 15 equations to estimate, where *i* = 1, ..., 15 (the 15 Spanish inland regions)[3] and *t* = 1,...,20 (annual growth rates from 1980-1981 to 1999-2000). The dependent variable $\dot{y}_{it}$ is the annual growth rate of total factor productivity. As explained previously, the TFP growth in a region *i* in a time period *t* is assumed to depend basically on the own R&D stock growth rate ($\dot{R}_{it}$) and the growth rate of the stock of R&D obtained through spillovers ($\dot{IR}_{it}$). Additionally, equation (33) contains a constant $\mu_i$ and the effect of a time dummy variable $d_t$ that takes value 1 from period 1990-1991 to 1999-2000.

In order to obtain the indirect stock of R&D obtained through spillovers from other regions ($IR$), the specification of a matrix **S** of spatial weights is necessary, given that variable *IR* is defined as the weighted sum of the R&D stocks of the neighbor regions:

$$IR_{it} = \sum_{j \neq i}^{15} s_{ij} R_{jt} \tag{34}$$

Several approaches can be taken for defining the elements $s_{ij}$.[4] Specifically, we have considered two different possibilities:

$$\boldsymbol{S}_B = \begin{cases} s_{ij} = 1; \text{ if } j \text{ and } i \text{ have a common border} \\ s_{ij} = 0; \text{ anywhere else;} \end{cases}$$

$$\boldsymbol{S}_T = \begin{cases} s_{ij} = t_{ij} \Big/ \sum_{j \neq i} t_{ij} \\ s_{ij} = 0; \text{ anywhere else;} \end{cases}$$

being $t_{ij}$ the volume of trade (in Euros) between the regions *i* and *j* in 2001.

These two matrices correspond to different approaches of representing channels for regional R&D spillovers. When the binary matrix $\boldsymbol{S}_B$ is considered, we are considering a geographical dimension, assuming that R&D spillovers are generated only between regions that share a common border. Additionally, matrix $\boldsymbol{S}_T$ introduces the possibility that the R&D spillovers may be generated not by a spatial dimension but through interregional trade. For the estimations we have collected data of R&D stocks per region from the BDMores database elaborated by the Spanish Ministry of Economy from 1980 until 2000 at constant prices of 1980, as well as data of regional labor, private physical capital and gross value added (also at constant prices of 1980). This database also provides information of the share of remuneration of workers in the

[3] We exclude of the analysis the cases of the Balearic and Canary Islands, given their location relatively far from the remaining set of Spanish regions.

[4] Following the traditional practice in spatial econometrics, the $s_{ij}$ elements are row-standardized in all the cases.

gross value added (that is used of a proxy of parameter $\alpha$ to obtain the regional TFP levels). The data of interregional trade necessary to construct matrix $\boldsymbol{S}_T$ are obtained from the C-Interreg database elaborated by the Klein Institute at the Autonomous University of Madrid.

With these data, we have estimated 15 different equations like (33) for the 15 Spanish regions using the DWP estimator. Consequently, specifying some support for the set of parameters to estimate and the errors is required. We have fixed the same vector **b** for parameters $\beta_{iR}$ and $\beta_{iIR}$. In order to use the DWP estimator, these supports should be centered on zero and with values large enough to cover all the parameter space. Specifically, we have considered M=3 with vectors $\boldsymbol{b}' = (-1,0,1)$. In the recent works by Beneito (2001), López-Bazo et al. (2006), Gumbau-Albert and Maudós (2006), Balmaseda and Melguizo (2007) or Escribá and Murgui (2007) the estimates obtained for the own region R&D stocks elasticity ranging from almost zero to around 0.45, which seems to suggest that an upper bound of 1 for these parameters $\beta_{iR}$ is sufficiently high. The variability of previous empirical results about the contribution of other regions' R&D on regional TFP growth is even larger. Gumbau-Albert and Maudos (2006) found it quite large, between 0.18 and 0.41 (although they obtain negative estimates for the own R&D), whereas López-Bazo et al. (2006) obtained very small positive values (and even negative estimates in some specifications of their model). In any case, the vector ***b*** considered for parameters $\beta_{iIR}$ seems adequate as well. Given the high uncertainty about the sign and magnitude of the constant and the time dummy, we opted for specifying wider support vectors like $\boldsymbol{b}' = (-5,0,5)$ for parameters $\mu_i$ and $\eta_i$. For the weighting parameters $\gamma_h$ we fixed supporting vectors composed only by N=2 values $\boldsymbol{b}' = (0,1)$. Finally, the usual three-sigma rule (with the standard deviations of the dependent variables) has been applied for specifying the supports of the error terms.

### *4.2. Brief discussion of some results.*

The estimation results following the two different proposed approaches are reported in Table 1 (Matrix $S_B$ and Matrix $S_T$). For the sake of simplicity, estimates of the intercept and time dummy parameter are not reported. The symbol (*) means that the estimate is significantly different from zero at the 5% level according to $\chi^2_{M-1}$ statistics. In the remaining cases, the variable was identified as extraneous because the respective estimate $\tilde{\gamma}_h > 0.49$. Appendix A contains more detailed information on the values of the $\chi^2_{M-1}$ statistics and the estimates $\tilde{\gamma}_h$.

**Table 1. Estimates from the Data-Weighted Prior (DWP) regressions.**

| | Matrix $S_B$ | | Matrix $S_T$ | |
|---|---|---|---|---|
| | $\beta_{iR}$ | $\beta_{iIR}$ | $\beta_{iR}$ | $\beta_{iIR}$ |
| Andalusia | 0.205** | 0.004 (e) | 0.195** | 0.003 (e) |
| Aragon | 0.096** | 0.101** | 0.096** | 0.106 ** |
| Asturias | 0.002 (e) | 0.001 (e) | 0.002 (e) | 0.002 (e) |
| Cantabria | 0.048** | 0.148** | 0.048** | 0.128** |
| Castilla-Leon | 0.001 (e) | 0.001 (e) | 0.001 (e) | 0.001 (e) |
| Castilla-La Mancha | 0.072** | 0.062** | 0.071** | 0.059** |
| Catalonia | 0.173** | 0.002 (e) | 0.177** | 0.002 (e) |
| Valencia | 0.176** | 0.201** | 0.175** | 0.220** |
| Extremadura | 0.032** | 0.001 (e) | 0.000 (e) | 0.001 (e) |
| Galicia | 0.170** | 0.002 (e) | 0.150** | 0.002 (e) |
| Madrid | 0.105** | 0.000 (e) | 0.105** | 0.001 (e) |
| Murcia | 0.126** | 0.082** | 0.126** | 0.090** |
| Navarra | 0.201** | 0.107** | 0.199** | 0.116** |
| Basque Country | 0.267** | 0.002 (e) | 0.244** | 0.238** |
| Rioja | -0.054** | 0.079** | -0.054** | 0.091** |

Note: The symbol (**) means that the estimate is significantly different from zero at the 5% level according to $\chi^2_{M-1}$ statistics. The symbol (e) indicates that the variable has been identified as extraneous because the respective estimate $\tilde{\gamma}_h$ > 0 .49.

These results allow us to extract some basic conclusions about the Spanish case. Firstly, our measure of the influence of R&D on productivity, based on the developed methodology, confirms the existence of a clear, positive and significant relationship between the R&D stock growth rate and productivity growth in most cases. A positive and significant relationship can be observed in twelve regions under the $S_B$ Matrix and eleven regions under the $S_T$ approach. The cases in which this relationship is not significant or is negative, which only occurs in the case of Rioja in both approaches, are regions with a high presence of farming activities in their economic structure, or, as in the case of Asturias, constitute an economy that underwent a long structural crisis during the analyzed period due to being specialized in traditional heavy manufacturing. It is worth observing that the results are quite similar in both approaches, thus giving more consistency to the conclusions.

Secondly, our results confirm the relevance of R&D spillovers in the Spanish case. These spillovers appear significant (always with a positive effect) in seven cases under the $S_B$ approach and eight under the $S_T$ approach. In several cases we can observe that the effect of R&D, summing both the internal effect and those derived from spillovers, is much more relevant in regions that make less of an R&D effort than others do. For example, Madrid is the region that makes the highest effort in R&D in Spain, but the total contribution of the sum of internal and external R&D over productivity is higher in cases like Murcia, Aragon or Cantabria, which reap the gains of being in an area with higher general R&D efforts.

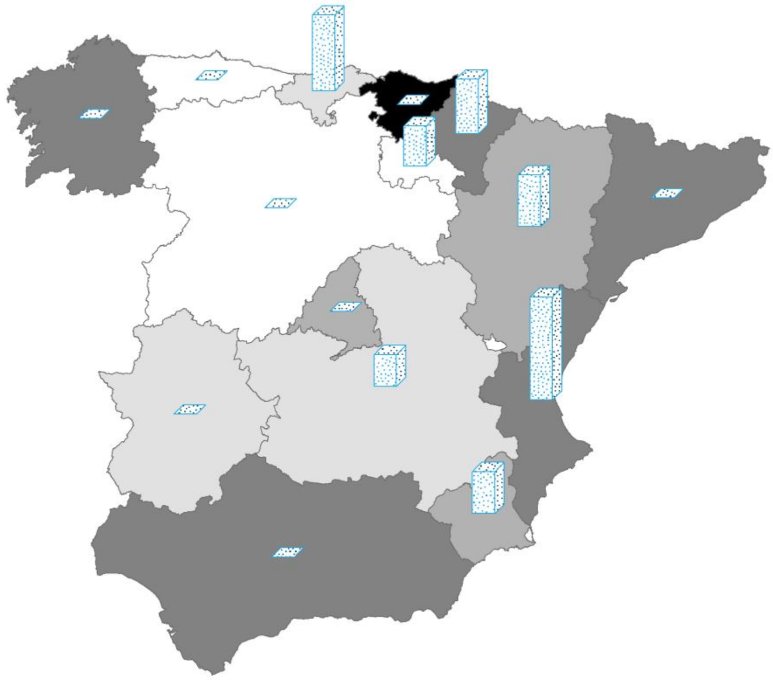
However, a third idea may be extracted from the results: the regions which obtain the largest gains from R&D spillovers are normally those that also make quite relevant efforts. There are some important exceptions to this finding. On the one hand, Madrid and Catalonia are the two regions that make the highest efforts in R&D in Spain; however, we did not find that they benefit from R&D spillovers. This is probably because they produce spillovers to other regions. On the other hand, Cantabria, which does not present high ratios of R&D effort, takes clear advantage of spillovers. However, successful behavior is normally that followed by Navarra, Valencia, Aragon or Castile-La Mancha, among others, consisting in making relevant R&D efforts of their own which enable them to capture even more relevant effects from R&D spillovers.

Finally, it is especially worth observing that there exist some geographical patterns in spillover behavior and in general R&D contributions to productivity. Firstly, the regions with a major R&D effect are the most urbanized areas of the country: Bilbao (Basque Country), Valencia, Barcelona (Catalonia) and Madrid. Secondly, Figure 2 shows R&D effects, both internal and from spillovers, in which it can be seen that the highest spillovers are from Navarra, Aragon and Cantabria, while a network of spillovers can also be identified between the Mediterranean coastal regions: from Catalonia to Valencia and Aragon and from Valencia to Murcia and Aragon. As may be observed, the spillover effects of R&D efforts are especially clear in the most developed and fastest growing area of the country: the northeast regions or, as they are usually called, the *Ebro-axis* area. Several studies (see Raymond (2002) or Villaverde (2004), among others) show that this area presents a convergence between them but distancing from the rest of the country.

Madrid causes similar positive effects, but only over one region, its closest one: Castile-La Mancha[5]. This means that the presence of a relevant nucleus of R&D close to a specific region does not imply a definitely positive effect (as occurs with Extremadura or Castile-Leon). To reap these gains from spillovers, intense commercial relationships must also be maintained and relevant efforts be made, as occurs in the cases of the Mediterranean coastal regions or between Castile-La Mancha and Madrid (through the intensive integration of Toledo or Guadalajara with Madrid's metropolitan area). These results are in consonance with previous research on regional growth patterns in the Spanish case (see De la Fuente (2002), among others).

[5] We say that this is the closest region because the two main cities not belonging to the Autonomous Community of Madrid belong to Castile-La Mancha: Guadalajara and Toledo.

**Figure 1. Map with results from the DWP regression with the $S_B$ matrix (*).**



(*)The intensity of grey on the map represent the own R&D effect over productivity. White regions are the ones that appear as exogenous. The height of the bars represents the R&D spillover effect.

The results are basically the same irrespective of the approach used. This is relevant because the first scenario used to obtain the results in Table 1 employs a geographic approach, while the second one is a commercial approach and the results obtained are basically the same. Geographical distance is still relevant in the spillover of R&D effects. New information and communication technologies mean that the globalization of economies and the knowledge creation process are more integrated, but they do not alter the fact that distances are still fundamental for the transmission of technological and knowledge advances. Spillover effects decrease faster with distances. This means that regions which made strong efforts in R&D, such as Galicia or Andalusia, do not reap the positive effects of the efforts of others regions, as occurs in other cases, like Castile-La Mancha, Murcia or Cantabria.

## 5. Conclusions.

The aim of this paper is to estimate economic models to explain productivity growths generated by R&D activities at a regional level considering the possible effects of the region's own R&D stock as well as the spillovers produced in other regions.

The paper proposes the use of an entropy-based technique to estimate these effects, which is the main contribution of the research study. The most relevant advantage of this estimator is that, even in situations characterized by a lack of large data samples, it is capable of identifying extraneous variables at the same time as it estimates the relevant parameters of the specified model. This estimation technique may be very useful when we wish to estimate models at a specific-regional level. Depending on the degree of heterogeneity of the set of regions analyzed, it is possible that some of these regions may present characteristics that more easily convert R&D efforts (generated within the region itself or obtained from other regions by R&D spillovers) into productivity gains, whereas in other regions the effect of (direct or spillover generated) R&D activities may be irrelevant.

We illustrate this idea with an empirical application for Spanish regions. Results show that R&D efforts are crucial in increasing productivity. However, we may conclude that spillovers are also highly relevant, even between regions. These spillovers lead to relevant increases in productivity, especially in those regions that make significant R&D efforts of their own and are located close to an R&D nucleus such as Madrid or Barcelona. Some spatial patterns of behavior can also be observed. Spillovers are more significant in the northeast area of the country, in which the regions are growing faster, are closer to the European Union and are more urbanized. These elements, which are relevant to understand general growth, seem to be relevant also in the understanding of knowledge transmission.

## References.

Aghion P, Howitt P, (1992) A model of growth through creative destruction. Econometrica, 60, 323-351.

Balmaseda FJ, Melguizo MJ, (2007) I+D como factor productivo en la economía española: un análisis empírico regional y sectorial, Hacienda Pública Española, 180, 9-34.

Beneito MJ, (2001) R&D productivity and spillovers at the firm level: Evidence from Spanish panel data, Investigaciones Económicas, 25, 289-313.

Bernardini R, (2008) A Composite Generalized Cross Entropy formulation in small samples estimation, Econometric Reviews, 27, 596-609.

Botazzi L, Pieri G, (2003) Innovation and spillovers in regions: Evidence from European patent data, European Economic Review, 47, 687-710.

Breschi, S. Lissoni F, (2001) Knowledge spillovers and local innovation systems: A critical survey, Industrial and Corporate Change, 10, 975-1005.

Brugger E. A, Stuckey B, (1987) Regional economic structure and innovative behavior in Switzerland, Regional Studies, 21, 241-254.

Brunsdon C Fotheringham, A.S, Charlton, M, (1996) Geographically Weighted Regression: A method for exploring spatial non-stationarity, Geographical Analysis, 28, 281-289.

Casetti, E. (1972) Generating models by the Expansion Method: applications to geographic research, Geographical Analysis, 4, 81-91.

Coe DT, Helpman E, (1995) International R&D spillovers, European Economic Review, 39, 859-887.

de la Fuente A (2002) On the sources of convergence: a close look at the Spanish regions, European Economic Review, 46 (1), 569-599.

Domar ED, (1946) Capital Expansion, Rate of Growth, and Employment, Econometrica, 14, 137-147.

Döring T, Schnellenbach J, (2006) What do we know about geographical knowledge spillovers and regional growth? A survey of the literature, Regional Studies, 40, 375-395.

Escriba FJ, Murgui MJ, (2007) El capital tecnológico como factor de producción en las regiones españolas, 1980-2000, Investigaciones Regionales, 10, 33-52.

Golan A, (2001) A simultaneous estimation and variable selection rule, Journal of Econometrics, 101, 165-193.

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

Grossman G, Helpman E, (1991) Innovation and growth in the global economy. Cambridge,-MA, MIT Press.

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

Harrod RF, (1939) An Essay Dynamic Theory, Economic Journal, 59, 14-33.

Jones, C. I. (1995), Time series tests of endogenous growth models, Quarterly Journal of Economics, 110 (2), 495–525.

Kapur JN, Kesavan HK, (1992) Entropy Optimization Principles with Applications. Academic Press. New York.

Kullback J, (1959) Information Theory and Statistics. Wiley. New York.

Lopez-Bazo E, Requena F, Serrano L, (2006) Complementarity between local knowledge and internationalization in regional technological progress, Journal of Regional Science, 46, 901-929.

Pukelsheim F, (1994) The three sigma rule, The American Statistician, 48, 88-91.

Raymond JL, (2002) Convergencia real de las regiones españolas y capital humano, Papeles de Economía Española, 93, 109-121.

Romer PM, (1990) Endogenous technological change, Journal of Political Economy, 98, 71-102.

Tödtling, F.(1990) Regional Differences and Determinants of Entrepreneurial Innovation. Empirical Results of an Austrian Case Study. In E. Ciciotti, N. Alderman and A. Thwaites (Eds) Technological Change in a Spatial Context - Theory, Empirical Evidence and Policy, 260-284.

Heidelberg: Springer.Villaverde J, (2004) Convergencia provincial en España: un análisis espacial, Papeles de Economía Española, 100, 210-219.

Varga A, Schalk HJ, (2004) Knowledge spillovers, agglomeration and macroeconomic growth: an empirical approach, Regional Studies, 38, 977-989.

Wakelin, K, (2001) Productivity growth and R&D expenditure in UK manufacturing firms. Research Policy, 30, 1079–90.

Zellner A, (1996) Models, prior information, and Bayesian analysis, Journal of Econometrics, 75, 51-68.

Zellner A, (1997) The Bayesian method of moments (BMOM): theory and applications. In: Fomby T, Hill R, (eds), Advances in Econometrics, 12, 85-105.

## APPENDIX A: $\chi^2_{M-1}$ STATISTICS AND ESTIMATES OF $\tilde{\gamma}_h$.

| | Matrix $S_B$ | | Matrix $S_T$ | |
|---|---|---|---|---|
| | $\beta_{iR}$ | $\beta_{iIR}$ | $\beta_{iR}$ | $\beta_{iIR}$ |
| Andalusia | 22.810 | 0.064 | 22.710 | 0.062 |
| | (0.023) | (0.738) | (0.023) | (0.738) |
| Aragon | 22.015 | 22.038 | 22.013 | 22.065 |
| | (0.025) | (0.025) | (0.025) | (0.025) |
| Asturias | 0.053 | 0.050 | 0.053 | 0.051 |
| | (0.739) | (0.739) | (0.739) | (0.739) |
| Cantabria | 21.840 | 22.328 | 21.841 | 22.194 |
| | (0.026) | (0.024) | (0.026) | (0.025) |
| Castilla-Leon | 0.049 | 0.050 | 0.049 | 0.050 |
| | (0.739) | (0.738) | (0.739) | (0.739) |
| Castilla-La Mancha | 21.912 | 21.880 | 21.912 | 21.872 |
| | (0.025) | (0.026) | (0.025) | (0.026) |
| Catalonia | 22.524 | 0.053 | 22.552 | 0.055 |
| | (0.023) | (0.738) | (0.023) | (0.739) |
| Valencia | 22.545 | 22.768 | 22.541 | 22.961 |
| | (0.023) | (0.023) | (0.023) | (0.022) |
| Extremadura | 21.808 | 0.049 | 0.048 | 0.049 |
| | (0.026) | (0.739) | (0.739) | (0.739) |
| Galicia | 22.492 | 0.054 | 22.344 | 0.055 |
| | (0.026) | (0.739) | (0.024) | (0.739) |
| Madrid | 22.058 | 0.048 | 22.057 | 0.049 |
| | (0.026) | (0.739) | (0.025) | (0.739) |
| Murcia | 22.181 | 21.952 | 22.179 | 21.989 |
| | (0.025) | (0.025) | (0.025) | (0.025) |
| Navarra | 22.768 | 22.071 | 22.752 | 22.122 |
| | (0.023) | (0.025) | (0.023) | (0.025) |
| Basque Country | 23.481 | 0.055 | 23.218 | 23.156 |
| | (0.021) | (0.736) | (0.021) | (0.022) |
| Rioja | 21.856 | 21.938 | 21.856 | 21.990 |
| | (0.026) | (0.025) | (0.026) | (0.025) |

The estimates of $\tilde{\gamma}_h$ are shown in brackets.