Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 48-2012

Vahidin Jeleskovic and
Benjamin Schwanebeck

Assessment of a spatial panel model for the efficiency analysis of the heterogonous healthcare systems in the world

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Assessment of a spatial panel model for the efficiency analysis of the heterogonous healthcare systems in the world*

Vahidin Jeleskovic[+]

und

Benjamin Schwanebeck[♠]

September 9, 2012

Abstract

Various panel models were presented to resolve the ranking of global health care systems according to efficiency. However, in terms of the spatial distribution of statistical units, spatial dependence as a result of various forms of spatial interactions caused biased estimators in classical regression. To our knowledge, this is the first paper which analyzes the healthcare systems of WHO members with regard to spatial dependencies while distinguishing between heterogeneity and inefficiency. It was possible to determine a significant spatial autocorrelation. Therefore one have to consider these spatial spillovers when assessing the performance of healthcare systems. The most meaningful way of implementing these effects appears to be by regressing the health output on various explanatory variables through a combination of the fixed effects spatial lag and the fixed effects cross regressive model. This allows spatial spillovers due to level of education, healthcare expenditure, and the quality of the healthcare system itself, to be diagnosed. Modeling these spatial effects allows previous results to be given more precision with regard to the quality of the healthcare systems of WHO members.

JEL classification: C12, C21, I12

Keywords panel data, fixed effects, production of health, efficiency measurement, heterogeneity, spatial effects, spatial autocorrelation

---

* Acknowledgements: We gratefully acknowledge helpful comments from Reinhold Kosfeld, Jochen Michaelis, Andreas Ziegler and Ulrich Zierahn.
+ University of Kassel, Nora-Platiel-Str. 4, D-34109 Kassel, Email: jeleskovic@uni-kassel.de
♠ University of Kassel, Nora-Platiel-Str. 4, D-34109 Kassel, Email: schwanebeck@uni-kassel.de

1. Introduction

The World Health Report (WHR), published annually by the WHO, evaluates world health, including health care provision and thus also healthcare systems. In the WHR 2000 the systems were ranked according to efficiency based on an estimate of “health production”. The efficiency analysis was carried out by Evans et al. (2000a,b) using the fixed effects model (hereafter FEM). This consisted of data from 191 WHO members in the period from 1993-1997.

This approach was subsequently criticized by various authors. The majority of criticism was aimed at the objectivity, quality and validity of the measured effectiveness or efficiency of the healthcare systems. In addition, the two input factors used (healthcare expenditure per capita and average number of years in education) were correlated (cf. Williams 2001). Hollingsworth and Wildman (2002) used a different model. One point of criticism was that the assumption of the time invariance of the inefficiency as a condition for the application of the FEM might be compromised. Moreover, five years were considered to be possibly too long a time period. In spite of this, results similar to Evans et al. (2000a,b) were reached. Gravelle et al. (2002a,b) criticized that the data set was incomplete for 51 countries, and that it was completed by the model. Greene (2004) criticized the lack of distinction drawn between the inefficiency and heterogeneity of the countries, and presented various panel models to differentiate between the two. Nevertheless, the approach by Evans et al. (2000a,b) was innovative and a first step towards evaluating the efficiency of various healthcare systems, so providing an evaluation tool for policy makers.

To our knowledge, there exists a lack of research in considering spatial dependence when analyzing the efficiency of national healthcare systems. A review of the literature about assessing health system performance can be found in Kruk and Freedman (2008). There is also a wide field of spatial analysis about constructing and examining disease maps (cf., among others, Rushton 2003), but not for the efficiency of a whole healthcare system. Not to mention on a global level. So this paper should make a contribution to close this cap.

In their studies, Evans et al. (2000b) produced a depiction of the spatial distribution of inefficiency (see Fig. 1). This suggests that a high spatial autocorrelation exists.[1] In this context it is sensible to assume that disease, climatic conditions, economic systems etc. do not

[1] With so-called cold-spots in North America and Western Europe and so-called hot-spots in Central Africa measured on performance.

end at country borders (cf. Rushton 2003), and that strong or weaker webs of relationships between different systems in neighboring or nearby countries exist, so that healthcare systems also present an international spatial problem. The existence of spatial effects leads to two consequences: an economic and an econometric effect. On the one hand, this means that neighboring countries probably influence each other more strongly, which means that these effects should be measured and meaningfully interpreted. It would therefore make economic sense to model these effects as well, and interpret them economically. On the other hand, not taking the spatial effects into account leads to biased estimates in the classic regression model (LeSage und Pace 2009). For these two reasons we would like to extend the analysis by Evans et al. (2000b) and Greene (2004) by adding the modeling of spatial influences. Our analysis therefore focuses on the modeling and interpretation of the results of spatial effects. The classic analysis which we refer to here, without a modeling of spatial effects, is based on the same analysis by Evans et al. (2000b) and Greene (2004), and will therefore only be mentioned superficially for the purposes of this paper. In order to compare and contrast both analyses, with and without spatial effects, we use more or less the same data as Evans et al. (2000b) and Greene (2004).

In a first part for this paper, the results of the FEM, with its differentiation between heterogeneity and inefficiency (Evans et al. 2000b and Greene 2004), will be repeated for 126 countries, taking the points of criticism made by Gravelle et al. (20002a,b) into account. The actual scientific insights will be achieved in the second part in our analysis, in which the identification and implementation of spatial effects is carried out in relation to an already existing FEM. We will briefly present and then apply this analysis regarding the FEM, with consideration of heterogeneity and based on the three most currently accepted spatial regression models, namely spatial cross regressive, spatial lag, and spatial error model, in order to examine spatial spillovers. According to our knowledge, this is the first time that the spatial effects have been taken into consideration in this “WHO-healthcare-efficiency-context”, so that this paper represents an innovation in the analysis of global healthcare systems. This allows for previous results from other authors on the ranking of healthcare systems of the WHO countries to be given more precision regarding the countries included in the analysis.

Therefore, our paper is structured in the following way. In section 2, the methods used and the implementation of spatial lags are presented. Section 3 follows with an overview of the data used, the countries in question, and the characteristics involved in the creation of the weights matrix. The subsequent section presents the results of the estimates. Finally, a conclusion is drawn in chapter 5.

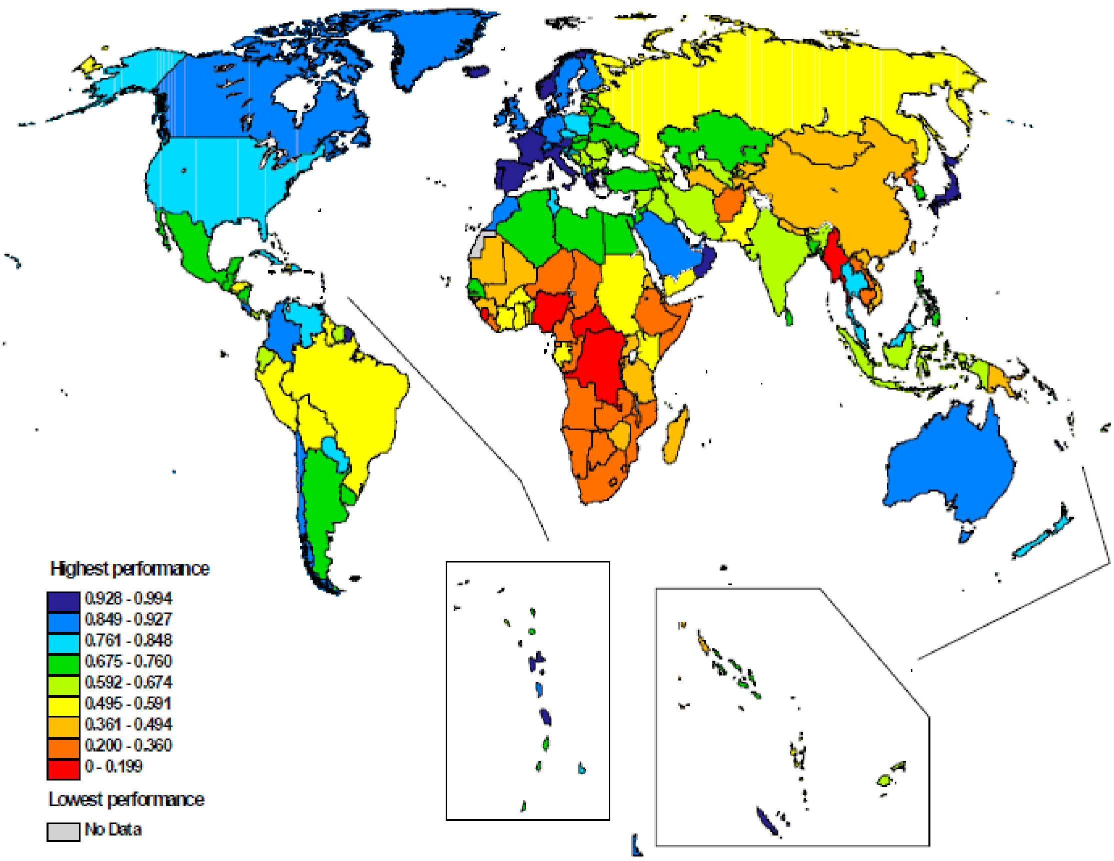

Figure 1: Global Efficiency Distribution, Estimate for 1997

Source: Evans et al. (2000b, 15)

2. Methodology

The approach by Evans et al. (2000b) – also initially used by Greene (2004) – to determine the effectiveness and efficiency of a healthcare system or health policy and reforms, and so to provide a possible evaluative approach for policy makers, is reproduced. Accordingly, a production function is estimated in the effectiveness study by the WHO (2000). It is pointed out, however, that a healthcare system cannot be represented by a classic production function.[2] When comparing healthcare systems, it is assumed that there is a differentiation from zero in terms of minimum health, and that maximum health is dependent on perfectly efficiently applied input factors. The effectiveness or efficiency of a healthcare system can therefore be viewed as deviation from the maximum. Figure 2 summarizes this situation.

[2] Because, in spite of everything, the input of no resources would not cause the output to drop to zero. Regarding this, Evans et al. (2000a) remarked that the population could not possibly be “dead”.

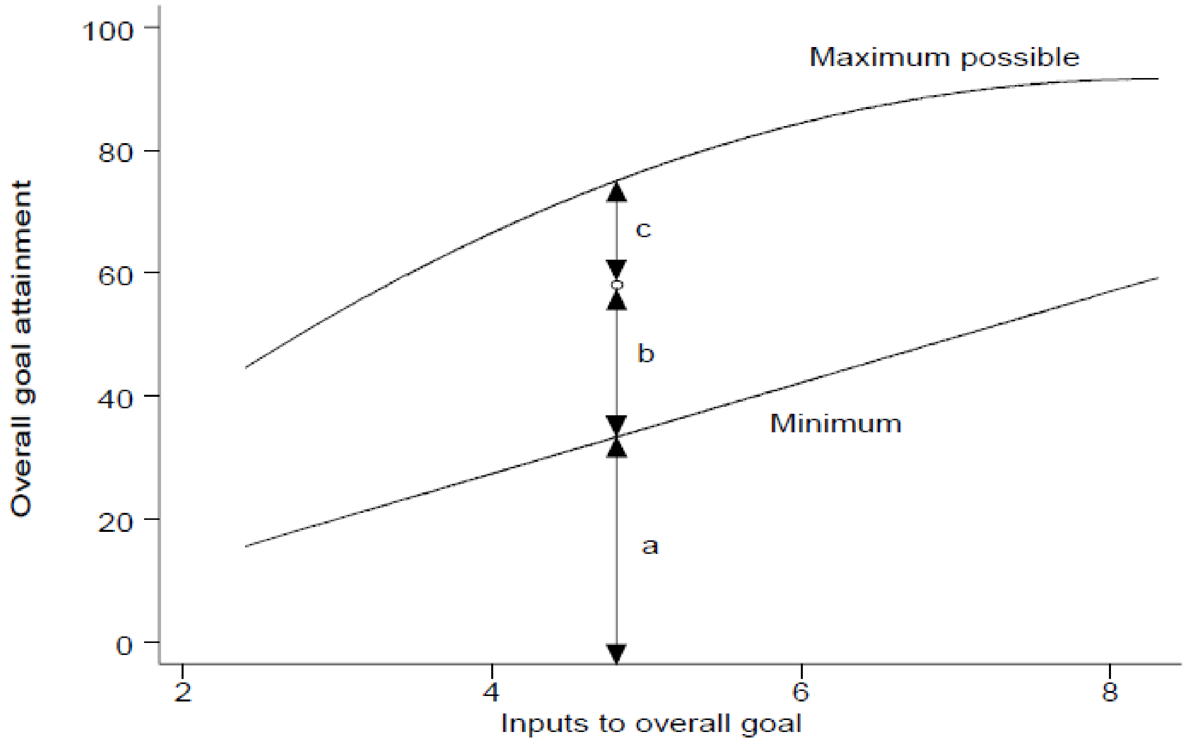

Figure 2: Quality of the health care system

Source: Evans et al. (2000b, 3)

The problem in this estimate, however, lies in the fact that the maximum cannot be observed, but only estimated. Moreover, one must also take into account that other factors can influence healthcare systems, in addition to the resources applied. In this respect, a production function is estimated using a fixed effects model. The FEM was chosen rather than the random effects model as a result of the Hausmann test (cf. Evans et al. 2000a,b). The advantage of this method lies in the fact that no assumptions regarding the distribution of technical efficiency need to be made (Evans et al. 2000a). A more detailed view and explanation for choosing the FEM can be found in Evans et al. (2000a) and Greene (2003).

2.1. Fixed effects model

The model is used in the form developed by Schmidt and Sickles (1984) and Cornwell et al. (1990) (in addition cf. Greene 2002). The production function is accordingly

(1) $$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_i,$$

whereby y_{it} describes the (logarithmic) output (here: COMP) of healthcare system i at time t, α a constant for all countries, x_{it} a vector of independent (logarithmic) input factors for country i also at time t, and u_i describes the country specific inefficiency of healthcare system i. The error term v_{it} as, random noise with a mean value of zero, comprises not only stochastic elements but also country specific and time variant heterogeneity, and is not correlated with other model components. The indices i and t respectively stand for each country and year.

The representation of (1) results in:

(2) $$y_{it} = (\alpha - u_i) + x'_{it}\beta + v_{it} = \alpha_i + x'_{it}\beta + v_{it},$$

with α_i as a new intercept containing the technical inefficiency as country specific constant. The parameters are now estimated using within estimators through OLS. Moving the function upwards makes it possible to determine the inefficiency as a deviation of the constant from the benchmark. Given this, u_i is always positive. In addition, a country becomes the "best", with maximum (100%) efficiency, if its estimated production function lies higher than any other. Equation (3) summarizes this:

(3) $$\hat{u}_i = max(\hat{\alpha}_i) - \hat{\alpha}_i \geq 0.$$

Technical inefficiency is defined as

(4) $$TE_i = \frac{E[y_{it}|x_{it}, u_i]}{E[y_{it}|x_{it}, u_i = 0]},$$

and would, in Figure 2, correspond with the ratio $(a + b)/(a + b + c)$.

The general, or total, efficiency results from the difference between technical inefficiency and minimal output (M_{it}), generated without input factors:

(5) $$E_i = \frac{E[y_{it}|x_{it}, u_i] - M_{it}}{E[y_{it}|x_{it}, u_i = 0] - M_{it}}.$$

In Figure 2 this corresponds with a ratio $b/(b + c)$ (cf. Greene 2003).

As formal function of the FEM, Evans et al. (2000b) chose the translog model with two input variables (HEXP and EDUC) as second-degree Taylor series of an unknown function:

(6) $$y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 \frac{1}{2} x_{1it}^2 + \beta_4 \frac{1}{2} x_{2it}^2 + \beta_1 x_{1it} x_{2it} + v_{it}.$$

Both Evans et al. (2000b) and Greene (2004), however, use a shortened version of the Translog Model, in which the final two terms are omitted. We also use this version here in order to maintain comparability, which, given the relevant input variables, produces the following function:

(7) $$logCOMP_{it} = \alpha_i + \beta_1 logHEXP_{it} + \beta_2 logEDUC_{it} + \beta_3 \frac{1}{2} log^2 EDUC_{it} + v_{it}.$$

Each of these variables is presented in chapter 3.

2.2. Heterogeneity

The problem with the model approach set out above is that it includes two important, and rather restrictive, assumptions. On the one hand, all time-variable heterogeneity goes into α_i and finally into the estimation of the inefficiency $\hat{u}_i$. The data set comprises completely different cultures, economies, political forms, climatic conditions, etc. This could significantly influence, and so distort, the estimate of inefficiency. Therefore, one must draw a distinction between inefficiency and heterogeneity. On the other hand, the assumption is made that inefficiency is time-invariant. The five years under consideration here could prove to be too long a time period for this assumption (Hollingsworth and Wildman 2002). Greene (2004) therefore discusses various methods based both on the FEM and the random effects model. The choice of the fixed effects model, however, has the advantage that no distribution assumptions need to be made with regards to inefficiency.

In the end, however, $\alpha + v_{it} - u_i$ contains both country specific heterogeneity and inefficiency, as well as time variable and invariable elements which are hard to separate. Greene (2004) should be consulted for a more precise discussion of which model is “best suited” for the consideration of the present problem, or what the advantages and disadvantages of each model are, and how heterogeneity and inefficiency can be viewed separately.

The option proposed by Greene (2004) to differentiate between heterogeneity and inefficiency in the present model approach will be followed here. Thus, in a first step, an estimate of the production function is made and so also of the time invariant inefficiency $\hat{u}_i$. In a second step, these estimated values are regressed as dependent variables onto various country specific variables, which measure, or identify, the heterogeneity. This is formally set out in equation (8):

(8) $$\hat{u}_i = \delta_{0i} + h_i'\delta + \varepsilon_i,$$

where h_i represents the vector of the independent variable, and ε_i the error term. Only data from 1997 are included in the regression, as, according to the assumption, $\hat{u}_i$ only contains time-invariant elements. This then allows the OLS regression estimate to follow. The variables and data are presented in chapter 3.

This enables the proportion of the time invariant inefficiency, which is due to the heterogeneous countries, to be identified, and the inefficiency to be adjusted accordingly (cf. Greene 2004).

There remains the possibility, however, that distortions persist due to variables and time variants that have not been factored in. Greene (2004) therefore developed further panel models to resolve this issue. The prime focus of this paper, however, is to take spatial effects into consideration. For this reason, the approach presented above has been chosen and applied, based on an FE panel model, in order to identify and take spatial effects into account[3].

2.3. Spatial autocorrelation

The aim of this present paper is to improve the results of Evans et al. (2004b) and Greene (2003) and (2004) by examining spatial effects. These can be taken into consideration in both steps of the regression. This can take place principally with the aid of three different regression models and combinations thereof. For the diagnosis of spatial effects we will first present Moran's I coefficient in order to check for disturbances errors in the residuals of a model (cf. Cliff and Ord 1973). Since the Moran's I is not the only test, several Lagrange Multiplier (LM) tests follow. With regard to the formal presentations in this section, and including the statements relating to them, we refer the reader to Anselin and Florax (1995), Klotz (1998), Eckey et al. (2005 and 2006), LeSage and Pace (2009), Elhorst (2010a), and Debarsy and Ertur (2010).

Moran's I measures the degree of linear association between the vector x of observed values of a geo-referenced variable X and its spatial lag $L(x)$, i.e. a weighted average of the

[3] Please note that we are dealing with a two-step regression here. In each step it is possible to apply one of three different spatial regression models both separately, and in combination with the three spatial regressions in the other regression step. These three spatial regression models are presented in the next section. This allows for an application of various combinations of different spatial models. However, this would exceed the reasonable limits of this analysis, which would also be the case, if the same analysis had been applied to the random effects models. Furthermore, this would also lack a continuous benchmark analysis based on RE models, as Evans et al. (2000a,b) also only restricted their analysis to FE models, due to the Hausmann test, so that we likewise only focus on the FEM here. Applying the RE model in this context shall be reserved for future research.

neighboring values. The Moran's I statistic can be calculated with a standardized, or an unstandardized spatial weights matrix. The standardized spatial weights matrix are preferable on several grounds and can be formulated as follows (LeSage and Pace 2009):

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{*} * (x_i - \overline{x}) \cdot (x_j - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = \frac{n}{S_0} \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \sum_{j=1}^{n} w_{ij}^{*} \cdot (x_j - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{*}$$

It relies on cross products to measure the spatial association by means of the spatial weights matrix W_{ij}. The significance of the spatial autocorrelation will be assessed by a null hypothesis H_0: no spatial autocorrelation. The rejection of this hypothesis indicates the existence of spatial association and provides a clue to further analysis.

If the regions are spatially autocorrelated, the absolute value of Moran's I will be significantly greater than zero (LeSage and Pace 2009). A positive Moran's I indicates positive spatial autocorrelation, i.e. the contiguous regions exhibit similarities, whereas a negative Moran's I implies that there might be competition between the neighboring regions, and that regions with high values tend to be located next to low-value neighboring regions. A value between 0.4 and 0.5 reveals high spatial autocorrelation, and a value of zero hence indicates no spatial autocorrelation (Anselin 1988). Therefore, a significant Moran's I statistic indicates spatial effects in the data (LeSage and Pace 2009). We calculate the Moran's I statistics for the residuals of an applied model year by year. Furthermore we calculate the statistics for the second step regressions. A significant Moran's I in the residuals means that there are spatial effects to be modeled and to be interpreted economically. Moreover, a significant Moran's I in the residuals also means that the estimator of parameters may - not necessarily - be biased and inconsistent (LeSage and Pace 2009). Inefficiency may also occur. We can therefore assume that the applied model is not able to deal with spatial effects.

But one problem remains: the Moran's I was developed for the OLS regression. So that test should only be used as an approximation for other models. Therefore further methods were designed. The joint statistic (LM test and likelihood ratio test, in the following LR test) simultaneously tests whether the endogenous variable and the error component are spatial lagged. This null hypothesis is analogous to the Moran's I test but on a global level: H_0: $\rho = \lambda = 0$, i.e. no spatial autocorrelation. The rejection of this hypothesis indicates that spatial dimension matters. Further LM tests should indicate which type of spatial autocorrelation prevails. Here we use marginal tests "before" the regression, i.e. to test for one form while assuming that the other is absent (H_0: $\rho = 0$, while $\lambda = 0$ and H_0: $\lambda = 0$, while $\rho = 0$), and LM tests "after" the regression, i.e. to test whether spatial effects still exists albeit spatial models were used. Furthermore we take a look at the second step. Inefficiencies estimated in the first step using several models are now regressed as dependent variables onto various country specific variables. Here we also use LM tests - analogous "before" the second step regression (H_0: $\rho = 0$, while $\lambda = 0$ and H_0: $\lambda = 0$, while $\rho = 0$) - to identify spatial autocorrelation depending on the specific model in the first step. This approach should help to detect the "best" spatial model for both steps (cf. Debarsy and Ertur 2010, and Elhorst 2010a).

Following this, the spatial effects are taken into consideration in the FEM when it is possible to reject the hypotheses above and to discriminate between the different forms of spatial autocorrelation. In the first step of the analysis, i.e. in the fixed effects model, spatial effects can be taken into account in three ways. [4]

The fixed effects spatial cross regressive model takes the autocorrelation in the exogenous variable into account, so that (7) can be extended to

(9a) $$logCOMP_{it} = \alpha_i + \beta_1 logHEXP_{it} + \beta_2 logEDUC_{it} + \beta_3 \frac{1}{2} log^2 EDUC_{it} + \beta_4 W logHEXP_{it} + \beta_5 W logEDUC_{it} + \beta_6 W \frac{1}{2} log^2 EDUC_{it} + v_{it},$$

where W represents a weights matrix. Here it is assumed that the endogenous variable not only depends on the exogenous variables in its own region, but also on the exogenous variables of the neighboring countries. This model continues to be estimated using the OLS estimation.

In contrast, the fixed effects spatial lag model assumes that the dependent variables of neighboring regions influence the endogenous variable. This is shown in equation (9b):

(9b) $$logCOMP_{it} = \alpha_i + \beta_1 logHEXP_{it} + \beta_2 logEDUC_{it} + \beta_3 \frac{1}{2} log^2 EDUC_{it} + \rho W logCOMP_{it} + v_{it}.$$

The spatial effects are taken into consideration in the fixed effects spatial error model:

(9c) $$logCOMP_{it} = \alpha_i + \beta_1 logHEXP_{it} + \beta_2 logEDUC_{it} + \beta_3 \frac{1}{2} log^2 EDUC_{it} + v_{it}$$
$$with\, v_{it} = \lambda W v_{it} + \varepsilon_{it},$$

[4] We refer the reader to chapter 3 for details on the composition of the necessary weights matrix W.

where ε_{it} is a neighbor-independent error term. Coincidental shocks, the non-consideration of relevant influential variables, etc. in neighboring regions are included in the model as a further influence on the endogenous variable. The maximum likelihood estimate is applied for both the spatial lag and spatial error model. In addition, the combination of fixed effects spatial cross regressive model and fixed effects spatial lag model, i.e. the spatial effects in the exogenous and endogenous variables, is also estimated using maximum likelihood.

Analogous to the first step regression, the spatial effects can also be taken into consideration in the second step. The following equations result for each respective model:

(10a) $\hat{u}_i = \delta_{0i} + h_i'\delta + Wh_i'\delta + \varepsilon_i$

(10b) $\hat{u}_i = \delta_{0i} + h_i'\delta + \rho W\hat{u}_i + \varepsilon_i$

(10c) $\hat{u}_i = \delta_{0i} + h_i'\delta + \varepsilon_i$

with $\varepsilon_i = \lambda W\varepsilon_i + e_i$.

The interpretation follows analogously to the first step. Also here, (10a) and (10c) are additionally combined. The final two, and also their combinations, are again estimated using maximum likelihood, and (10a) using OLS.

Spatial effects are taken into consideration in the first, second, or in both steps of the regression. The estimate and tests were carried out in Matlab, using, among others, codes developed by Elhorst (2010b), and Debarsy and Ertur (2010).

3. Data

The data used here originally stems from the WHO, but were downloaded[5] from Greene's website at the Stern School of Business at New York University, and have already been analyzed by various authors (including Evans et al. 2000a,b, Gravelle et al. 2002a,b and Greene 2004). The complete panel data set contains observations of 191 WHO members including the data gathered from some communities or government districts in certain countries for the years from 1993 to 1997. Only one year of data was gathered in case of 50 countries and smaller spatially-administrative units, and Algeria did not gather a full year's data. This means that these are not taken into account here. Furthermore, 14 "island states" as

[5] Originally (28.05.2009) available at: http://pages.stern.nyu.edu/wgreene/Econometrics/PanelDataSets.htm
Temporarily offline and since recently (status: 25.03.2012) available at:
http://people.stern.nyu.edu/wgreene/Econometrics/PanelDataSets.htm

well as countries with no neighbors, and countries for which data from all neighbors are missing, have been removed from the data set, as estimates for countries without neighbors would become biased in the first step, because in such cases relevant variables would have to be omitted in the model.[6] This means that the data set used here covers 126 countries at five points in time. The World Health Report 2000 and the many ancillary publications on the WHO website provide a more detailed description of the data, and information on how it was gathered.

Two outcome variables were gathered by the WHO: the disability adjusted life expectancy value "DALE" and the "COMP" measurement, a composite of five health goals. The first of these has already been analyzed by many (among others Evans et al. 2000a, Hollingsworth and Wildmann 2002, Gravelle et al. 2002a,b, and Greene 2004 and 2004) and will not be examined further here, as qualitatively analogue results for the COMP variable have been achieved.[7] The five goals gathered in COMP comprise of the by year health (or overall health), the distribution of health, the responsiveness of the system, the responsiveness in distribution and fairness in financing.[8] These goals received the same weighting. We refer the reader to the World Health Report 2000 (WHO 2000) and Evans et al. (2000b) for a more precise description of the composition of the variable. The natural logarithm of COMP is used, as also done by Evans et al. (2000b) and Greene (2004).

Two variables serve as input factors and are also included in the production function in logarithmic form. HEXP represents the health expenditure per capita in 1997 $ppp, and EDUC the average number of years of schooling. For reasons of comparability, and analogous to Greene (2004), the simplified form applied by Evans et al. (2000b) is used instead of generating the complete translog function.

Eight further variables then serve as indicators for international heterogeneity. These, however, only contain observations for the year 1997 (analogous to Greene 2004). The Gini coefficient GINI serves as an indicator of income distribution and moves between 0 (complete equality) and 1 (complete inequality). VOICE and GEFF are measurements by the World

[6] No assumptions with regard to neighbors were made in case of the following countries, which means that they were not included in the data set: the Bahamas, Barbados, Cape Verde, Comoros, Fiji, Island, Jamaica, Maldives, Malta, Mauritius, Philippines, Samoa, Tonga, Tunisia.

[7] We can provide these results upon request.

[8] In our opinion the COMP variable is better suited to display the output variable: "the quality of a health system". Furthermore, also in this case a parallel analysis regarding an additional dependent variable would exceed the reasonable limits of this paper. We therefore focus on analysis of the dependent variable COMP, but would point out that we have achieved qualitatively similar results in case of DALE, which we can provide on request.

Bank, whereby the former represents an indicator of democratization and political freedom, and the latter an indicator of the effectiveness of government. The dummy variables TROPICS and OECD stand for tropical climate and membership of the OECD. POPDEN describes the density of population in km^2, PUBFIN the percentile share of health expenditure borne by the government, and GDPC the gross domestic product per capita in 1997 $ppp. The density of population and GDP per capita enter into the model in logarithmic form. We refer the reader to Greene (2003, appendix A) and Greene (2004) for a more detailed examination of the variables, information on how they were gathered, estimates carried out for missing data, and updates. Table 1 provides an overview of the descriptive statistic for 1997. As mentioned above, the data were sourced from Greene's website, so that a comparison with previous estimates that do not include modeling of the spatial effects can easily be made.

	Non OECD members		OECD members		All	
	Mean value	Std. Dev.	Mean value	Std. Dev.	Mean value	Std. Dev.
COMP	69.98	10.72	89.37	4.03	74.44	12.61
HEXP	235.94	286.30	1489.09	773.80	524.36	691.05
EDUC	5.28	2.42	9.05	1.55	6.15	2.76
GINI	0.401	0.085	0.300	0.065	0.378	0.091
VOICE	-0.251	0.733	1.253	0.542	0.095	0.939
GEFF	-0.326	0.662	1.154	0.633	0.014	0.904
TROPICS	0.557	0.499	0.034	0.186	0.437	0.498
POPDEN	1085.27	3578.31	470.13	1020.81	943.69	3183.51
PUBFIN	51.46	19.76	72.51	14.20	56.31	20.59
GDPC	4323.33	4528.51	18056.39	7056.83	7484.11	7783.87
Sample	97		29		126	

Table 1: Descriptive statistic of variables for 1997

Source: own representation based on Greene (2004, 962)

For an identification using Moran's I, LM tests, and to take the spatial autocorrelation into account, a binary weights matrix was created and subsequently standardized with the respective sum of the row dividing the elements (cf. Anselin and Florax 1995, and Cliff and Ord 1973, 87ff.), with a main focus on country borders. Neighboring countries were added, however, in the case of nine "island states" and countries without neighbors, based on economic, political, historical and cultural interrelation, in order to avoid a too significant reduction of the data set.[9] In addition, Great Britain borders not only Ireland, but also France in this model.[10]

A discussion of the problems which occurred in this context, for instance the insufficient consideration of short sea routes, economic interdependence, commuter flows, differences in length of borders, overseas territories, etc. (cf., among others, Eckey et al. 2006, and Klotz 1998), shall not be addressed further here.[11]

4. Results

Table 2 displays the test statistics "before" the regression. The columns LM_J and LR_J describes the joint tests and the last two ones show the statistics of the marginal LM tests (first one $H_0: \rho = 0$, latter one $H_0: \lambda = 0$). Due to the fact that one can only test for an endogenous spatial lag and/or for spatially correlated errors, we accomplished these tests with the "normal" exogenous variables (X = HEXP, EDUC, EDUC²) and in addition with the spatially correlated exogenous variables (***W***X = HEXP, EDUC, EDUC², ***W***HEXP, ***W***EDUC, ***W***EDUC²). The results show a high spatial autocorrelation, however, no type of spatial autocorrelation prevails. It seems that one should include spatially lagged exogenous variables.

Tests with	LM_J	LR_J	LM_ρ	LM_λ
X	55.4371***	59.2539***	55.1365***	48.9036***
WX	47.6399***	51.9114***	44.1388***	40.5945***

Table 2: Tests for the existence of spatial autocorrelation in the first step

***, **, * describes the significance at the 99%, 95%, 90% level.

Table 3 displays the results of the FEM with and without spatial effects, which were achieved through application of the methods described above. The "FEM" column describes the results of the panel model without consideration of the spatial autocorrelation, based on equation (7). The results diverge, though not qualitatively and not to a significant level, from those of Evans et al. (2000b) and Greene (2004), even though only 129 countries were examined. All

[9] These were Australia (bordering Indonesia and New Zealand), Bahrain (bordering Saudi Arabia and Qatar), Cyprus (bordering Turkey), Japan (bordering China and the Republic of Korea), New Zealand (bordering Australia), the Republic of Korea (bordering China and Japan), Singapore (bordering Indonesia and Malaysia), Sri Lanka (bordering India) and Trinidad and Tobago (bordering Venezuela).
[10] For example, because of the Eurotunnel.
[11] These problems were partially factored only in case of the nine island states and Great Britain. Nevertheless, this circumstance offers starting points for further research we are planning on carrying out in the future.

input factors contribute to an explanation of the endogenous variable. However, the residuals (year 1-5) show a high spatial autocorrelation (as measured by Moran's I). Therefore and due to the joint and marginal LM tests, implementing spatial modeling appears to be necessary.

Variable	FEM	FESCRM	FESLM	FESEM	FESLCRM
HEXP	0.0085***	0.0068***	0.0059***	0.0050***	0.0054***
EDUC	0.0517**	0.0985***	0.0478**	0.0886***	0.0999***
EDUC²	0.0349**	-0.0142	0.0109	0.0044	-0.0151
WHEXP	-	0.0139***	-	-	0.0094***
WEDUC	-	-0.1153***	-	-	-0.1032***
WEDUC²	-	0.0913***	-	-	0.0563**
WCOMP	-	-	0.3850***	-	0.3570***
$\boldsymbol{W}\vartheta$	-	-	-	0.4020***	-
R^2	0.2067	0.2492	0.9990	0.9988	0.9990
Adj./corr.	0.2042	0.2432	0.2118	0.1974	0.2534
σ	0.0061	0.0059	0.0056	0.0055	0.0055
AIC	-10.2033	-10.2440	-10.3750	-10.3807	-10.3897
BIC	-10.1821	-10.2016	-10.3467	-10.3595	-10.3403
			Moran's I		
year1	0.2450***	0.2018***	-0.0156	0.2644***	-0.0241
year2	0.2235***	0.2265***	-0.0355	0.2372***	-0.0248
year3	0.1097*	0.1850***	-0.0693	0.1749***	-0.0918
year4	0.2263***	0.1930***	-0.0291	0.2546***	-0.0352
year5	0.2143***	0.2042***	-0.0417	0.2264***	-0.0401
			LM tests		
LM_ρ	3.0267*	13.3314***	0.7607	2.363	6.4925**
LM_λ	0.4914	0.1495	0.0824	76.2057***	0.0712

Table 3: Regression results of the first step, all variables as logarithms

***, **, * describes the significance at the 99%, 95%, 90% level.

The subsequent columns show the results of this consideration through application of the various spatial methods. For example, in the fixed effects spatial cross regressive model "FESCRM" the estimate value for ***W***HEXP means that achieving the health aims (COMP) within a county is positively dependent on the healthcare spending in the neighboring country, thus an expected result.

This means that if a country raises its spending on improvement of its healthcare system, the neighboring countries appear to profit from it. However, a higher level of education (measured in school years) while the number of school years is low, would have a negative effect. The higher the number of school years, the more the positive effect of higher education prevails (see ***W***EDUC²). However, there continues to be a high spatial autocorrelation in the residuals. Therefore and due to the LM tests "after" the regression (last rows), this model does not appear to be adequate for estimating “health production”. Equally, the fixed effects spatial error model “FESEM” also seems to be inadequate here for the purpose of taking spatial effects into consideration.

The fixed effects spatial lag model “FESLM” and the combination of the fixed effects spatial lag and fixed effects spatial cross regressive model “FESLCRM” eliminate the spatial autocorrelation measured year by year (Moran’s I), albeit the LM tests (last rows) prefer the FESLM. Here, no spatial autocorrelation is observable. In addition, a meaningful interpretation of the coefficients of the variables used can be made here. Thus achieving health aims (COMP) inside a country is positively dependent on a country’s own expenditure on healthcare (HEXP) and domestic education (EDUC), as well as the expenditure on healthcare (***W***HEXP), the general “health” (***W***COMP), and higher education in the neighboring countries. Albeit there must be an adequately high number of school years for the positive effect to prevail (cf. ***W***EDUC and ***W***EDUC²). One must note here that there are, in part, large differences in the estimated coefficients in models with spatial effects, which underlines the need for spatial analyses. This circumstance could result in grave repercussions, in particular in practical applications.

According to	LM_ρ	LM_λ	Tests with
FEM	20.2935***	16.9131***	
FESLM	9.6800***	3.4411*	
FESCRM	13.3924***	15.9941***	h
FESEM	20.6839***	17.6361***	
FESLCRM	6.2433**	1.8299	
FEM	11.4432***	8.4724***	
FESLM	3.4864*	2.1073	
FESCRM	11.7852***	9.4616***	$\boldsymbol{W}h$
FESEM	11.8903***	8.6898***	
FESLCRM	2.0262	1.2622	

Table 4: Tests for the existence of spatial autocorrelation in the second step

***, **, * describes the significance at the 99%, 95%, 90% level.

Further LM tests help to identify the “best” model. These are also marginal tests “before” the second step regression (first one H_0: $\rho = 0$, latter one H_0: $\lambda = 0$). We also accomplished these tests with the “normal” exogenous variables (h) and in addition with the spatially correlated exogenous variables ($\boldsymbol{W}h$) of the second step regression. Inefficiencies estimated in the first step using several models are now tested to identify spatial autocorrelation in the second step depending on the specific model in the first step. According to theses tests it seems that only the FESLCRM in the first step and a spatial cross regressive model (SCRM) in the second step eliminate the spatial autocorrelation in the second step (cf. last row).

So both models (FESLM and FESLCRM) appear to be adequate according to the tests, the information criteria and based on the R quadratics, although the combination of the spatial lag and spatial cross regressive models is minimally preferable.

WHO (Evans et al. 2000b)		FEM		FESLCRM		
Position	Country	Position	Country	Position	Country	previous
1	France	1	France	1	Japan	5
2	Italy	2	Italy	2	Singapore	4
3	San Marino	3	Spain	3	Israel	21
4	Andorra	4	Singapore	4	Greece	9
5	Malta	5	Japan	5	Cyprus	18
6	Singapore	6	Austria	6	Thailand	37
7	Spain	7	Norway	7	Dominican Rep.	42
8	Oman	8	Portugal	8	Columbia	19
9	Austria	9	Greece	9	Chile	27
10	Japan	10	Oman	10	Australia	24
25	Germany	25	Saudi Arabia	25	US	32
50	Poland	50	Trinidad&Tobago	50	Bahrain	35
100	St. Kitts and Nevis	100	Zimbabwe	100	Netherlands	96
189	Central African R.	124	Myanmar	124	Myanmar	124
190	Myanmar	125	Nigeria	125	Nigeria	125
191	Sierra Leone	126	Central African R.	126	Central African R.	126
	UK (24)		Germany (20)		France (17)	
	US (72)		UK (13)		Germany (31)	
			US (32)		UK (23)	

Table 5: Ranking of total efficiency of selected countries (own presentation based on Greene (2004). The column labeled "previous" refers to the FEM.

When spatial effects are taken into account in the estimate of inefficiency, a different efficiency ranking results compared to the FEM: one which has been corrected for spatial effects. This condition is presented in Table 5. The results from Evans et al. (2000b) are included here to facilitate comparability. There is a tendency for countries with fewer neighbors to perform better here, as they apply fewer input factors and are therefore more efficient. [12] Israel is an interesting case. The neighboring states show relatively "bad" performance, yet Israel seems to have an efficient healthcare system in spite of this. The

[12] For example, Thailand produced such "good" results, as there were no data available for a majority of its neighbors. As a result, fewer factors for "health production" were applied which leads to a seemingly higher efficiency. If all neighboring countries are taken into account (e.g. Myanmar), a more realistic picture would result.

explanation for this could either be that it has absorbed the “bad” influences of neighboring healthcare systems “well”, or that it has completely isolated itself from them. At this point it is important to refer once again to the challenge of creating the “correct” weights matrix, as discussed in the previous chapter, which probably necessitates a deeper analysis of the modeling of the weights matrix.

One must remember that no distinction is drawn between time-invariant inefficiency and heterogeneity in the FEM. The highly significant estimated value for $\boldsymbol{W}\vartheta$ in the FESEM could point to the fact that relevant spatial input factors – which could, in addition, implement heterogeneity in the model – are not included in the production function, and that inputs to date do not represent an adequate depiction of health production (cf. in addition Eckey et al. 2006). Therefore one have to distinguish between heterogeneity and inefficiency. This is carried out, analogous to Greene (2004), by way of a further regression. The estimated inefficiency is regressed in the second step for variables depicting heterogeneity. The results for this second regression step can be found in Tables 6 and 7. The estimated inefficiencies resulting from the fixed effects, fixed effects spatial lag and fixed effects spatial lag & cross-regressive model were used as endogenous variables in Table 6 and estimated using the OLS estimate, according to equation (8).

Variable	According to FEM	According to FESLM	According to FESLCRM
constant	0.7382***	0.6279***	0.6086***
GINI	0.3104***	0.2385***	0.2628***
TROPICS	0.0050	-0.0008	-0.0020
PUBFIN	-0.0000	0.0002	0.0002
LPOPDEN	-0.0047	-0.0028	-0.0035
LGDPC	-0.0794***	-0.0649***	-0.0629***
GEFF	-0.0010	-0.0030	-0.0004
VOICE	-0.0099	-0.0172	-0.0194*
OECD	0.0397*	0.0503**	0.0572**
R^2	0.7045	0.6134	0.5887
Adj. R^2	0.6843	0.5870	0.5606
σ	0.0701	0.0649	0.0655
AIC	-5.1732	-5.3284	-5.3063
BIC	-4.9706	-5.1258	-5.1037
Moran's I	0.3092***	-0.1395	-0.1017

Table 6: OLS regression results of the second step according to the FEM, FESLM and FESLCRM in the first step

***, **, * describes the significance at the 99%, 95%, 90% level

One should remember that inefficiency is per se time-invariant (assumption). The “according to FEM” column in Table 6 thus represents the (consistent) reproduction of Greene’s (2004) results and only presents small – though not significant – divergences, due to the difference in size of sample used here. Accordingly, both income (LGDPC) and income distribution (GINI) could form a significant contribution to explaining inefficiency in all models. Here, however, one should also bear in mind that the size of the estimated coefficients varies when considering the spatial effects. The higher the level of income per capital, and the more evenly it is distributed, the higher the efficiency of the healthcare system. Interestingly, membership of the OECD raises the inefficiency[13] – if only in slightly significant terms. The spatial correlation in the first model is relatively high in the second step. If it has already been taken into consideration in the first step, implementation of spatial effects is no longer necessary in the second step: the Moran’s I becomes insignificant. But referred to the LM tests (cf. Table 4), the “best” models seems to be FESLCRM in the first step and SCRM in the second step. The results of these models can be found in the last column in Table 7.

According to FESLCRM (and also to FESLM), it is interesting that the VOICE variable becomes weakly significant and exhibits the desired sign, i.e. the more democratic a state is, the more efficient its healthcare system is. This again shows here that both the fixed effects spatial lag and also the fixed effects spatial lag & cross regressive model are adequate methods to control for spatial autocorrelation. In order to demonstrate the importance of the spatial analysis, we present results from the second regression step where no consideration had previously been taken of the spatial effects in the first regression step. So where these are only initially taken into consideration in the second step, the results displayed in the left part of Table 7 are produced. The estimates were carried out according to the spatial regression models presented in chapter 2. In this, no model contributes to reducing the spatial autocorrelation, but the significant spatial effects exhibit the desired signs; the level of inefficiency sinks the more efficient a neighboring state is (spatial lag model “SLM” and spatial lag & cross regressive model “SLCRM”), and the higher its per capita income is (first SCRM).

It seems that in order to eliminate distortions as a result of spatial effects, these already have to be built into the first step regression. This is the only way to significantly reduce the spatial autocorrelation and thus to achieve a consistent estimate of the coefficients (cf. Table 2-4 and 6,7). Comparing the last columns in Table 6 and Table 7 indicates that it is not clear which

[13] Note that the effect of a tendentially higher income of the OECD members is already displayed through the GDPC variable. The same occurs in Greene (2004).

sequence (FESLCRM and OLS or FESLCRM and SCRM) appears to generate more precision (cf. the information criteria and the R quadratics). Due to the LM tests, one should decide on the latter one.

According to	FEM				FESLCRM
Variable	SCRM	SLM	SEM	SLCRM	SCRM
constant	0.9588***	0.5730***	0.6487***	0.6473***	0.6538***
GINI	0.2658**	0.2486***	0.2722***	0.2732***	0.2998***
TROPICS	0.0117	0.0028	0.0237	0.0228	0.0198
PUBFIN	0.0000	0.0001	0.0001	0.0001	0.0000
LPOPDEN	-0.0041	-0.0034	-0.0028	-0.0030	-0.0031
LGDPC	-0.0645***	-0.0645***	-0.0692***	-0.0636***	-0.0699***
GEFF	-0.0030	-0.0060	-0.0111	-0.0057	-0.0024
VOICE	-0.0157	-0.0135	-0.0123	-0.0143	-0.0222*
OECD	0.0370	0.0427**	0.0285	0.0293	0.0332
WGINI	0.1956	-	-	0.0328	-0.0222
WTROPICS	-0.0566	-	-	-0.0451	-0.0379
WPUBFIN	-0.0006	-	-	-0.0002	-0.0001
WLPOPDEN	-0.0063	-	-	-0.0031	-0.0032
WLGDPC	-0.0385**	-	-	-0.0076	0.0063
WGEFF	0.0144	-	-	0.0157	0.0169
WVOICE	0.0095	-	-	0.0065	0.0186
WOECD	-0.0023	-	-	-0.0045	-0.0126
WIneff	-	0.2910***	-	0.3090***	-
Wε	-	-	0.3710***	-	-
R^2	0.7277	0.7169	0.7510	0.7297	0.6271
Adj. R^2	0.6877	0.6975	0.7340	0.6900	0.5724
σ	0.0697	0.0624	0.0620	0.0612	0.06467
AIC	-5.0571	-5.3903	-5.4185	-5.3008	-5.2071
BIC	-4.6744	-5.1652	-5.2159	-4.8956	-4.8244
Moran's I	0.2188***	0.2660***	0.3092***	0.2227***	-0.0845

Table 7: Estimated results from the second step according to the FEM and FESLCRM

***, **, * describes the significance at the 99%, 95%, 90% level

5. Conclusion

The studies by Evans et al. (2000a,b) were an innovative approach to determining inefficiencies in healthcare systems. The data set includes almost every country in the world and so covers almost all of the world's population. The fixed effects model was used to determine efficiencies and produce rankings. The studies were subsequently analyzed further by other authors. These include Greene (2004), who criticized the lack of differentiation between inefficiency and heterogeneity, and presented various possible ways of implementing this. In the FEM, the only possible way to achieve differentiation is through an additional regression. In this, income and income distribution are important factors for explaining inefficiency and so contribute to differentiating between healthcare systems. Membership of the OECD also seems to provide a contributing explanatory factor, albeit only a weakly significant one.

The approach by Greene (2004) and the results of the study by Evans et al. (2000b) were reproduced here. The data set used in this paper only includes observations for 126 states, yet the results are consistent with those of previous analyses.

The aim of this paper was to take spatial effects into consideration when determining the efficiency of global healthcare systems. The assumption seems plausible that a state surrounded by industrialized nations with, for example, relatively high per capita income and efficient healthcare systems, benefits from the spatial spillovers. This is why various approaches were chosen and analyzed here to implement spatial effects. This was carried out both in the first and also the second step of the regression.

It was possible to determine a high spatial autocorrelation, so that it must be concluded that significant spatial spillovers exist. Healthcare systems, therefore, are not spatially closed systems. The implementation of spatial effects seems to be most meaningful in the first step using a combination of the fixed effects spatial lag and fixed effects cross regressive model. This allows spatial spillovers to be diagnosed, resulting from the level of education, healthcare expenditure and the quality of the healthcare system itself. Thus, in the second step, it seems that eventually no further examination of spatial effects is required or one should decide on the spatial cross regressive model. The general results of Evans et al. (2000b) and Greene (2004) are, however, not refuted through spatial examination, but rather given more precision.

The other methods proposed by Greene (2004) provide a starting point for further analysis of spatial effects, above all based on a random effects panel model. The study could also equally be carried out using the other output variable (DALE). Similarly, the construction of the weights matrix should be examined further, so that more precise estimates can be made. In this context, the strong web of relationships between states with sea borders should be taken into account. All of these will provide the starting points for our research in future.

References

Anselin, L. (1988): “Spatial econometrics: Methods and models”. Dordrecht, Kluwe academic publishers.

Anselin, L. and R. J. Florax (1995): “Small Sample Properties of Tests for Spatial Dependence in Regression Models”, in: L. Anselin und R. J. Florax (Ed.): *New Directions in Spatial Econometrics*, Berlin: 21-52.

Cliff, A. D. and J. K. Ord (1973): *Spatial Autocorrelation,* London.

Cornwell, C., P. Schmidt and R. Sickles (1990): “Production Frontiers with Cross Sectional and Time Series Variation in Efficiency Levels”, in: *Journal of Econometrics*, 46: 185-200.

Debarsy, N. and C. Ertur (2010): “Testing for spatial autocorrelation in a fixed effects panel data model”, in: *Regional Science and Urban Economics*, 40 (6): 453-470.

Eckey, H.-F., R. Kosfeld and M. Türk (2005): “Interregionale und internationale Spillover-Effekte zwischen EU-Regionen”, in: *Jahrbücher für Nationalökonomie und Statistik*, 225: 600-621.

Eckey, H.-F., R. Kosfeld and M. Türk (2006): “Räumliche Ökonometrie”, in: *WiSt Wirtschaftswissenschaftliches Studium*, 10: 548-554.

Elhorst, P. (2010a): “Spatial Panel Data Models”, in: Fischer M.M. and A. Getis (eds.): *Handbook of Applied Spatial Analysis*, Berlin: 377-407.

Elhorst, P. (2010b): “Matlab Software for Spatial Panels”, Paper presented at the IVth World Conference of the Spatial Econometrics Association (SEA), Chicago, June 9-12, 2010.

Evans, D., J. Lauer, C. Murray and A. Tandon (2000a): The Comparative Efficiency of National Health Systems in Producing Health: An Analysis of 191 Countries, GPE Discussion Paper, No. 29, EIP/GPE/EQC, WHO.

Evans, D., J. Lauer, C. Murray and A. Tandon (2000b): Measuring Overall Health System Performance for 191 Countries, GPE Discussion Paper, No. 30, EIP/GPE/EQC, WHO.

Gravelle, H., R. Jacobs, A. Jones and A. Street (2002a): Comparing the Efficiency of National Health Systems: Econometric Analysis Should be Handled with Care, University of York, Health Economics, UK, Manuscript.

Gravelle, H., R. Jacobs, A. Jones and A. Street (2002b): Comparing the Efficiency of National Health Systems: A Sensitivity Approach, University of York, Health Economics, Manuscript.

Greene, W. (2002): Fixed and Random Effects in Stochastic Frontier Models, Stern School of Business, Department of Economics, Working Paper 02-16.

Greene, W. (2003): Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems, Stern School of Business, Department of Economics, Working Paper 03-10.

Greene, W. (2004): "Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems", in: *Health Economics*, 13: 959-980.

Hollingsworth, J. and B. Wildman (2002): "The Efficiency of Health Production: Re-Estimating the WHO Panel Data Using Parametric and Nonparametric Approaches to Provide Additional Information", in: *Health Economics*, 11: 1-11.

Klotz, S. (1998): "Ökonometrische Modelle mit raumstruktureller Autokorrelation. Eine kurze Einführung", in: *Jahrbücher für Nationalökonomie und Statistik*, 218: 168–196.

Kruk, M. E. and L. Freedman (2008): "Assessing health system performance in developing countries: A review of the literature", in: *Health Policy*, 85: 263-276.

LeSage, J. and Pace, R.K. (2009): *Introduction to spatial econometrics*, Boca Raton, FL.

Rushton, G. (2003): "Public health, GIS, and spatial analytic tools", in: *Annual Review of Public Health*, 24: 43-56.

Schmidt, P. and R. Sickles (1984): "Production Frontiers and Panel Data", in: *Journal of Business and Economic Statistics*, 2 (4): 367-374.

Williams, A. (2001): "Science of Marketing at WHO? A Commentary on World Health 2000", in: *Health Economics*, 10: 93-100.

World Health Organization (2000): "The World Health Report, 2000, Health Systems: Improving Performance", Geneva.