Hospital efficiency under prospective reimbursement schemes: An empirical assessment for the case of Germany

Helmut Herwartz* Christoph Strumann‡

February 13, 2013

Abstract

The introduction of prospective hospital reimbursement based on diagnosis related groups (DRG) has been a conspicuous attempt to decelerate the steady increase of hospital expenditures in the German health sector. In this work the effect of the financial reform on hospital efficiency is subjected to empirical testing by means of two complementary testing approaches. On the one hand, we apply a two-stage procedure based on non-parametric efficiency measurement. On the other hand, a stochastic frontier model is employed that allows a one-step estimation of both production frontier parameters and inefficiency effects. To identify efficiency gains as a consequence of changes in the hospital incentive structure we account for technological progress, spatial dependence and hospital heterogeneity. The results of both approaches do not reveal any increase in overall efficiency after the DRG reform. In contrast, a significant decline in overall hospital efficiency over time is observed.

JEL-Classification: C21, D61, I11, I18

Keywords: Hospital efficiency, stochastic frontier analysis, data envelopment analysis, spatial analysis, diagnosis related groups

---

*Department VWL, Georg-August-Universität zu Göttingen, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany, hherwartz@uni-goettingen.de, tel.: +49 551 397 350, fax: +49 551 397 279

†Corresponding author. Institut für Statistik und Ökonometrie, Christian-Albrechts-Universität zu Kiel, Olshausenstr. 40, D-24118 Kiel, Germany, cstrumann@stat-econ.uni-kiel.de, tel.: +49 431 880 2379, fax: +49 431 880 7605

‡We thank two anonymous referees, Uwe Jensen, Andrew Street, the participants of the Konferenz "10 Jahre Forschungsdatenzentren" der Statistischen Ämter des Bundes und der Länder 2012 in Berlin, dggö - Jahrestagung 2011 in Bayreuth and Jahrestagung des Vereins für Socialpolitik 2010 in Kiel for helpful comments and discussions on earlier versions of this manuscript. We also thank the employees of the Forschungsdatenzentrum der Statistischen Landesämter - Standort Kiel/Hamburg for their cooperation.

1 Introduction

The introduction of prospective hospital reimbursement based on diagnosis related groups (DRG) can be regarded as the major structural reform in health care for almost 3 decades in Germany (Braun et al. 2007). One of the main goals of the reform has been to reduce the steady increase of hospital expenditures, which have doubled from 1991 to 2007, reaching almost 80 billion Euro in 2010, i.e. 3.2% of the German gross domestic product (GDP) (Statistisches Bundesamt, 2012). Prior to the reform, German hospitals have been reimbursed in terms of per diem payments. This system has created incentives for keeping patients hospitalized as long as possible, likely resulting in an inefficient use of resources. In December 1999 the federal German government announced the introduction of a prospective payment system in 2004 as an attempt to increase hospital efficiency (e.g. Hensen et al., 2008 and Lungen and Lapsley, 2003). Under DRG based financing, hospitals receive a fixed rate for each admission depending on a patient's diagnosis. If the costs for a particular case are lower (higher) than the reimbursement, the hospital realizes profits (losses). As a consequence, hospitals face an increased pressure on their financial performance and a higher risk of insolvency.

As intended by the reform, under prospective payment it is profitable for hospitals to reduce the lengths of stay and to raise simultaneously the number of treated cases (e.g. Böcking et al., 2005). However, the new incentive structure might also encourage opportunistic practices which affect - positively or negatively - hospital performance. Under prospective payment hospitals could enhance their profits by (i) classifying a patient in a DRG that produces a higher reimbursement (up-coding, e.g. Simborg, 1981), (ii) readmitting the patients (Böcking et al., 2005), (iii) selecting the more lucrative patients and treatments (cream skimming, e.g. Levaggi and Montefiori, 2003). For the case of Italy, Berta et al. (2010) have shown that cream skimming and up-coding negatively affect hospital efficiency, while readmissions have a positive effect. As shown by Herwartz and Strumann (2012) German hospital performance is characterized by a significant increase of negative spatial spillovers after the shift to prospective payment reflecting a rise of competition between hospitals for so called low cost patients. The empirical evidence of the effect of competition on hospital efficiency is rather inconclusive. For the English National Health Service the results of Cooper et al. (2010) suggest that hospital competition within markets with fixed prices can increase hospital efficiency. Chua et al. (2011) find a positive relationship between hospital efficiency in Australia and competition as measured by the Hirschman-Herfindahl Index, and a negative relationship when the number of competing private hospitals is used to indicate competition.

In summary, the overall effect of prospective payment on hospital efficiency is not obvious. This has been underlined for various countries by the inconclusive empirical evidence for efficiency improvements after the shift to prospective payments. For the cases of Norway (Biörn et al., 2010 and Biörn et al., 2003), Portugal (Dismuke and Sena, 1999) and Finland (Linna, 2000), positive effects of the DRG introduction on hospital efficiency have been found. In the same time no effect is detected for the cases of Italy (Barbetta et al., 2007), Austria (Sommerguters-Reichmann, 2000), and the US (Virginia: Chern and Wan, 2000 and New Jersey: Borden, 1988). A failure of prospective payment to improve hospital performance is mainly explained by a decrease of the number of admitted patients (Barbetta et al., 2007), which is observed for different countries, e.g. the US (e.g. Coulam and Gaumer, 1991), Italy (France et al., 2005) and Germany (e.g. Braun et al., 2007). The incentive to shift admissions to outpatients treatment under prospective payment serves as an explanation of the theoretically counterintuitive decline in treated cases. Another reason for a lack of a positive DRG introduction effect on efficiency could be a dominance of opportunistic behavior that adversely affects hospital performance.

The aim of this study is to subject the potential effect of the financial reform on overall hospital efficiency in Germany to empirical testing. In particular, we examine if hospitals have realized efficiency gains after the DRG reform. For this purpose, we analyze technical efficiency for an unbalanced cross-section of about 1600 German hospitals over 12 years (1995 to 2006) covering the DRG announcement (at the end of 1999) and introduction (2004) period. If there were any effects of the reform, we expect efficiency gains after these particular dates for at least two reasons. Firstly, we account for changes of technology and scale adjustments. Secondly, there have been no other major exogenous shocks affecting hospital performance during the sample period.

To address the robustness of empirical findings we employ two complementary testing approaches based on alternative hospital efficiency measurements, i.e. non-parametric *Data Envelopment Analysis* (DEA) and parametric *Stochastic Frontier Analysis* (SFA). On the one hand, we apply a two-stage procedure based on the Malmquist index decomposition of DEA efficiency scores in pure technical efficiency change. On the other hand, a fixed effects panel stochastic frontier model is employed that allows a one-step estimation of both production frontier parameters and inefficiency effects. As shown by Herwartz and Strumann (2012) German hospital performance is characterized by time varying spatial interdependence. Previous studies on the effect of prospective payment on hospital efficiency do not account for potential interaction between hospitals. Ignoring this form of dependence cannot only affect estimation accuracy, but in the presence of spatial spillovers estimation results might also be biased (Anselin, 1988). Moreover, if spatial clusters exist for both the dependent and explanatory variables, estimated relationships might appear stronger than they actually are (Bech and Lauridsen, 2008). In order to enable efficient estimation of the DRG reform effects, we apply a spatially autoregressive fixed effects panel model with spatially autoregressive disturbances (SARAR) in the two-stage analysis. For SFA modeling spatial dependence is taken into account by means of specifying time varying region-specific random effects. In contrast to the analysis in Herwartz and Strumann (2012) the one-step SFA treatment avoids the inherent inconsistency of two-stage approaches based on estimated SFA efficiency scores (Wang and Schmidt, 2002).

Section 2 sketches the alternative empirical testing strategies and introduces the data. Empirical results are discussed in Section 3 and Section 4 concludes. An appendix collects methodological details and complementary empirical results.

2 Methodology

In this Section we provide the complementary empirical strategies, which are based either on the non-parametric DEA or on the parametric SFA. By means of both approaches, we test if overall hospital efficiency has significantly increased after the DRG announcement or introduction. As pointed out by Jacobs (2001) each approach has its particular merits and weaknesses, and could, in principle, measure distinct aspects of hospital efficiency. Both techniques have been employed to analyze hospital performance in Germany (DEA: Tiemann and Schreyögg, 2012, 2009, Werblow et al., 2010, Werblow and Robra, 2007, Staat, 2006, Staat and Hammerschmidt, 2003 and Helmig and Lapsley, 2001; SFA: Herr et al., 2011 and Herr, 2008; DEA and SFA: Herwartz and Strumann, 2012). Based on the non-parametric DEA efficiency scores we follow a two-stage analysis. In the first stage, the Malmquist index decomposition of DEA efficiency scores in pure technical efficiency change is determined. Controlling for observable and hidden hospital heterogeneity, efficiency gains are identified in the second stage as a consequence of changes in the hospital incentive structure (e.g. Dismuke and Sena, 1999). The parametric analysis consists of a one-step estimation of both production frontier parameters and inefficiency effects in a fixed effects panel stochastic frontier model proposed by Wang and Ho (2010). Furthermore, the data and the determination of variables are also described in this Section.

2.1 Two-stage approach

To disentangle efficiency change from technological progress the Malmquist index decomposition in pure technical efficiency change is determined by means of DEA efficiency scores (e.g. Burgess and Wilson, 1995 and Sommersguter-Reichmann, 2000). A particular merit of DEA is its inherent flexibility. In this framework, the production or cost function does not require an explicit specification. Thus, one avoids assumptions about profit-maximization or cost-minimization behavior that might be inappropriate for (non-profit) hospitals (Zweifel et al., 2009). Potential effects of the DRG reform on hospital efficiency are investigated by means of a regression of logarithmic pure technical efficiency change. We implement an unbalanced spatial panel data model with time dummy variables, while controlling for hidden and observable heterogeneity across hospitals in form of fixed effects and explanatory variables, respectively. An improvement of efficiency is identified by testing for increasing time effects. As shown by Herwartz and Strumann (2012), hospital performance in Germany is characterized by a significant increase of negative spatial spillovers after the DRG reform. Additionally, German hospitals exhibit positive spatial dependence due to similar unobservable factors of nearby observations. For efficient assessment of the DRG reform effects we account for two distinct channels of spatial dependence simultaneously in the fixed effects panel model.

2.1.1 First-stage analysis

The input-based DEA efficiency score of hospital i under constant returns to scale, denoted $\theta^c_{i,t_1|t_2}$, is obtained by means of a comparison of its set of inputs and outputs in period t_1 with that of all hospitals in period t_2, where $t_1, t_2 \in \{t-1, t\}$. The measure is defined as the radial distance of the i-th hospital at time t_1 to the frontier function at time t_2 that is determined from a linear combination of the best practicing (efficient) units in t_2. To compare hospitals with distinct exogenously fixed input variables we account for non-discretionary input variables (Banker and Morey, 1986).

The input-based index of productivity is the geometric mean of the change in efficiency under both frontier functions in $t-1$ and t. For the i-th hospital the Malmquist index, along with its decomposition in efficiency change (EC) and technological change (TC), is given by

$$MI_{i,t} = \left[\frac{\theta^c_{i,t-1|t-1}}{\theta^c_{i,t|t-1}} \cdot \frac{\theta^c_{i,t-1|t}}{\theta^c_{i,t|t}}\right]^{1/2} = \underbrace{\frac{\theta^c_{i,t-1|t-1}}{\theta^c_{i,t|t}}}_{EC_{i,t}} \cdot \underbrace{\left[\frac{\theta^c_{i,t|t}}{\theta^c_{i,t|t-1}} \cdot \frac{\theta^c_{i,t-1|t}}{\theta^c_{i,t-1|t-1}}\right]^{1/2}}_{TC_{i,t}}. \tag{1}$$

In (1), $EC_{i,t}$ measures the movement over time of hospital i towards the frontier function and represents a change in efficiency (Färe et al., 1992, 1994). Moreover, $TC_{i,t}$ is the geometric mean of the change in efficiency under changing technology given the production bundles in $t-1$ and t. Thus, it indicates a shift in the constant returns to scale technology. The efficiency change component can be further decomposed in pure technical efficiency change (PEC) and scale efficiency adjustments (SEA)

$$EC_{i,t} = \underbrace{\frac{\theta^v_{i,t-1|t-1}}{\theta^v_{i,t|t}}}_{PEC_{i,t}} \cdot \underbrace{\left[\frac{\theta^v_{i,t|t}}{\theta^c_{i,t|t}} \cdot \frac{\theta^c_{i,t-1|t-1}}{\theta^v_{i,t-1|t-1}}\right]}_{SEA_{i,t}},$$

where $\theta^v_{i,t_1|t_2}$ is the respective efficiency score under variable returns to scale (Banker et al., 1984) (for more details see Appendix A). The pure efficiency change, $PEC_{i,t}$, measures the relative efficiency enhancement and is invariant to changes in the technology and scale adjustments (Burgess and Wilson, 1995 and Sommersguter-Reichmann, 2000). For the ease of interpretation, we consider the inverse of $PEC_{i,t}$ as the pure technical efficiency change of hospital i from period $t-1$ to t

$$\gamma_{it} = \theta^v_{i,t|t} / \theta^v_{i,t-1|t-1}. \tag{2}$$

By construction, values of γ_{it} above (below) unity indicate an improvement (regress) in efficiency.

Hospitals characterized by an outstandingly good or poor performance might affect the test results in two different ways. On the one hand, estimated efficiency scores could be severely distorted by measurement errors in observations that support the deterministic frontier (e.g. Wilson, 1995). On the other hand, hospitals performing particularly poor might invalidate the second stage regression results. To guard against these shortcomings, we follow Johnson and McGinnis (2008) to detect efficient and inefficient outliers. Outlying hospitals are excluded from the analysis.[1]

2.1.2 Second-stage analysis

For the logarithmic pure technical efficiency change, $\ln(\gamma_{it})$, we implement a SARAR model that accounts for two distinct channels of spatial dependence simultaneously. On the one hand, negative spatial spillovers might occur due to the competition for low cost patients and, on the other hand, positive spatial dependence could be the result of similar unobservable factors of nearby hospitals (Herwartz and Strumann, 2012). The model reads in time t as

$$y_t = \lambda_t \boldsymbol{W}_t y_t + \boldsymbol{Z}_t \beta + \omega_t + \delta_t + e_t, \quad \text{with} \quad e_t = \rho_t \boldsymbol{M}_t e_t + \epsilon_t,\ t = 1, \ldots, T, \tag{3}$$

[1] Hospitals are treated as an inefficient outlier if a convex-combination of worst performing hospitals can produce the same level of output using half the inputs. An efficient outlier is detected if it is possible to double the inputs without becoming inefficient. As it turns out, diagnostic results are qualitatively very similar across alternative threshold values for outlier detection.

where y_t is an $N \times 1$ vector comprising the logarithm of pure technical efficiency change based on DEA efficiency scores, $y_t = (\ln(\gamma_{1t}), \ldots, \ln(\gamma_{Nt}))'$, $\boldsymbol{Z}_t$ is an $N \times K$ matrix of observations of K explanatory variables, β a $K \times 1$ vector of parameters, ω_t is an $N \times 1$ vector comprising the individual effects of the N hospitals and δ_t denotes the time effect. Due to the computation of the Malmquist Index decomposition as the first difference of the DEA efficiency scores, the sample period comprises 11 years (1996 to 2006). The pattern of spatial dependence is captured by the $N \times N$ spatial weights matrices $\boldsymbol{W}_t$ and $\boldsymbol{M}_t$ with zero diagonal elements and row normalized constants (such that each row sums to unity). The number of hospitals, N, varies with t, since observations for some hospitals do not cover the full sample period. The spatial lag coefficient λ_t measures the direct effect of the weighted neighboring observations on the elements in y_t (Anselin, 1988). Spatial dependence due to similar unobservable factors of nearby observations is quantified by means of the spatial autocorrelation coefficient ρ_t. Both spatial parameters are restricted to be less than unity in absolute value. Finally, ϵ_t is an $N \times 1$ vector of location specific i.i.d. disturbances, $\epsilon_t \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_N)$, where $\boldsymbol{I}_N$ is the N-dimensional identity matrix. The model is estimated by means of a Maximum Likelihood approach (see Appendix B for a formal representation of the log-likelihood function).

2.1.3 Hypotheses testing

The effect of prospective payment on hospital efficiency is examined by means of testing for significant increases of the time dummy coefficients of 5 subperiods. In particular, to verify an announcement effect (AE), the pre-announcement period (SP_1: 1996 to 1999) is compared with the post-announcement-pre-reform period (SP_2: 2000 to 2003). An introduction effect is evaluated by means of two alternative strategies. The first approach (IE_1) takes the AE into account and compares SP_2 with the post-reform period (SP_3: 2004 to 2006). Secondly, the AE is neglected and the pre-reform period (SP_4: 1996 to 2003) is compared with SP_3 (IE_2). Finally, an overall effect (OE) is examined by comparing SP_1 with the post-announcement period (SP_5: 2000 to 2006). The empirical testing strategy can be summarized by the following hypotheses with the tested effects in parentheses

$$H_0^1 : \bar{\delta}_1 = \bar{\delta}_2 \qquad \text{vs.} \qquad H_1^1 : \bar{\delta}_1 < \bar{\delta}_2 \qquad (AE) \tag{4}$$

$$H_0^2 : \bar{\delta}_2 = \bar{\delta}_3 \qquad \text{vs.} \qquad H_1^2 : \bar{\delta}_2 < \bar{\delta}_3 \qquad (IE_1) \tag{5}$$

$$H_0^3 : \bar{\delta}_4 = \bar{\delta}_3 \qquad \text{vs.} \qquad H_1^3 : \bar{\delta}_4 < \bar{\delta}_3 \qquad (IE_2) \tag{6}$$

$$H_0^4 : \bar{\delta}_1 = \bar{\delta}_5 \qquad \text{vs.} \qquad H_1^4 : \bar{\delta}_1 < \bar{\delta}_5 \qquad (OE), \tag{7}$$

where $\bar{\delta}_1 = \frac{1}{4}\sum_{t=1996}^{1999}\delta_t$, $\bar{\delta}_2 = \frac{1}{4}\sum_{t=2000}^{2003}\delta_t$, $\bar{\delta}_3 = \frac{1}{3}\sum_{t=2004}^{2006}\delta_t$, $\bar{\delta}_4 = \frac{1}{8}\sum_{t=1996}^{2003}\delta_t$, $\bar{\delta}_5 = \frac{1}{7}\sum_{t=2000}^{2006}\delta_t$ with the year 1996 serving as reference, i.e. $\delta_{1996} = 0$. The hypotheses are tested by means of one-sided t-tests based on the covariance matrix of estimated time effects.[2]

2.2 One-step approach

The flexibility of DEA goes along with the (often) unrealistic assumption of a deterministic production frontier. All deviations from the frontier are considered as technical inefficiency, although they might also reflect measurement errors or other stochastic influences. By means of SFA it is possible to distinguish between inefficiency and noise components, at the cost of using a more restrictive parametric approach. However, the application of two-stage procedures based on estimated SFA efficiency scores may lead to invalid conclusions (Wang and Schmidt, 2002). We avoid this problem by means of a one-step estimation of both production frontier parameters and inefficiency effects (Kumbhakar et al., 1991). In contrast to the non-parametric DEA, a production function needs to be specified within the SFA framework. To capture variable returns to scale and, thus, allow hospitals to operate at an inefficient scale size, a translog production function is assumed (Tiemann and Schreyögg, 2009 and Jacobs et al., 2006). We specify time specific frontier parameters to account for heterogeneity over time in the production process of

[2]Regarding the applied two-stage approach some remarks are in order. In the second stage regression analysis the logarithmic transformation of the pure technical efficiency change, γ_{it}, ensures an unbounded dependent variable and, thus, enables a consistent Maximum Likelihood estimation (Simar and Wilson, 2007). However, Simar and Wilson (2007) mention that in finite samples the estimated DEA efficiency scores are biased and serially correlated in a complicated fashion. This invalidates standard approaches to inference, e.g. based on the inverse of the negative Hessian of the log-likelihood. To analyze the robustness of inferential results, we apply an adaptation of the bootstrap procedure suggested by Simar and Wilson (2007). However, the difference between bootstrap based and asymptotic results is negligible and, therefore, we only document the latter.

German hospitals. Non-discretionary input variables are incorporated in the model as exogenous factors determining the inefficiency. We apply an adaptation of the estimation procedure of a fixed effects panel stochastic frontier model in Wang and Ho (2010) to account for hospital specific unobservable factors, which are likely captured by the inefficiency term and thus leading to biased results (Greene, 2005). To our knowledge, a one-step treatment of SFA models with spatial error terms and lag dependence (SARAR) has not been proposed yet. Instead, spatial dependence in form of spatial clusters of hospital performance is incorporated in the model by specifying region-specific random effects with time varying variance. The stochastic production frontier panel model for hospital i located in region j at time t reads as

$$\ln q_{ijt} = \omega_{ij} + \tau_t + \sum_k \alpha_{tk} \ln x^D_{ijtk} + \sum_k \sum_{k \geq l} \alpha_{tlk} \ln x^D_{ijtk} \ln x^D_{ijtl} + \nu_{ijt} - u_{ijt}, \tag{8}$$

$$u_{ijt} = h_{ijt} \cdot u^*_{ij}, \quad h_{ijt} = \exp\left\{ \sum_k \kappa_{tk} \ln x^N_{itk} + \beta' z_{ijt} + \delta_t + \eta_{jt} \right\}, \tag{9}$$

$$u^*_{ij} \sim \mathcal{N}^+(\mu, \sigma^2_u), \quad \nu_{ijt} \sim \mathcal{N}(0, \sigma^2_\nu), \quad \text{and} \quad \eta_{jt} \sim \mathcal{N}(0, \sigma^2_{\eta t}), \tag{10}$$

where q_{ijt} is the output, x^D_{ijtk} and x^N_{ijtk} are the k-th discretionary and non-discretionary input variables, respectively. The inefficiency term u_{ijt} is given as a product of a time-invariant inefficiency term u^*_{ij}, and a function h_{ijt} that formalizes the variation of inefficiency over time. The time-invariant inefficiency term u^*_{ij} is truncated at zero and is assumed to have a normal distribution with mode μ and variance σ^2_u. Stochastic deviations from the frontier are captured by ν_{ijt}. The fixed unobservable effect of hospital i located in region j is denoted by ω_{ij}, τ_t and δ_t are time effects. While the former captures technological change, the latter describes the conditional temporal pattern of inefficiency and is used for testing the hypotheses about the effect of the DRG reform on hospital performance. This is done analogously as described in (4) to (7), only substituting the direction of the alternative hypotheses and defining the year 1995 as reference. The random effect of region j at time t is denoted by η_{jt}. Its time specific variance $\sigma^2_{\eta t}$ allows for time varying spatial clusters among hospitals.

As proposed by Wang and Ho (2010) the fixed effects ω_{ij} are removed from the model by a first-difference transformation. To obtain the marginal likelihood function of the hospitals of region j the region-specific random effects, $\eta_{j1}, \ldots, \eta_{jT}$, have to be integrated out. Due to the non-linearity the likelihood function does not attain a closed-form solution, and the model is estimated by means of a Simulated Maximum Likelihood approach with antithetic sampling and 200 simulations (for details see Appendix C).

2.3 Data and variables

2.3.1 The data

Data are drawn from two distinct sources. Hospital data are extracted from the annual hospital statistics collected by the statistical offices of the federal states ("Statistische Landesämter"). District-level data are obtained from the "Regionaldatenbank Deutschland - GENESIS", which is administered by the statistical office of North Rhine-Westphalia ("Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen"). Annual data cover the period from 1995 until 2006 and have been provided by the "Forschungsdatenzentrum der Statistischen Landesämter - Standort Kiel/Hamburg". University hospitals are not considered in the analysis. Hospitals, which have missing values for relevant variables or obvious data inconsistencies, are also excluded from the sample. On average, 0.3% (1998) to 2.7% (2003) and 0.1% (1998) to 0.9% (2002) of the hospitals are detected, respectively, as efficient and inefficient outliers and removed from the analysis. Moreover, to facilitate the interpretation of the time effects, hospitals with less than 2 data points are excluded (0.7% of the hospitals). Due to the first-difference transformation in the one-step approach only successive data points of hospitals are considered for estimation (11% of the observations have to be removed).

2.3.2 Variables

The selection of variables used to quantify the input and output measures and serving as explanatory factors determining the efficiency change and inefficiency is related to the existing literature on hospital efficiency (e.g. Herwartz and Strumann, 2012, Herr et al., 2011, Tiemann and Schreyögg, 2009, Herr, 2008, Farsi and Filippini, 2008, Lee et al., 2008, Helmig and Lapsley, 2001, Chang, 1998).

Input variables controlled by the hospitals are the amount of material expenses (in 2005 prices) (*exp*), the number of employed physicians (*phys*), nurses (*nurses*) and non-medical employees (*nonmed*). The number of beds (*beds*) is treated as a non-discretionary input. A hospital's output is measured in terms of treated cases weighted by the respective average resource usage (*wcases*), which is approximated by the nationwide average length of stay of patients treated in a particular clinical department (Appendix D provides an explicit representation of case mix weights).

The list of explanatory variables consists of dummy variables controlling for private for profit (*private*) and non-profit private hospitals (*non-profit*), the market share (*ms*, number of patients of a hospital relative to competitors located in the same district), occupancy rate (*occrate*), the mortality rate (*mort*), the hospitals' budget (*budget*, total expenses per bed in 2005 prices), an index of specialization (*spec*, Evans and Walker, 1972), the fraction of people aged over 65 years and living in a hospitals' district (*age*65), the degree of urbanization (*popdens*, population per km^2) and the district's GDP per capita (*gdp*, in 2005 prices). Some of the regressors are measured in natural logarithms (namely: *ms*, *mort*, *occrate*, *budget*, *gdp* and *popdens*).

2.3.3 Spatial specification

In the second stage regression analysis, two alternative spatial weights matrices are used to specify $\boldsymbol{W}$ and $\boldsymbol{M}$ in (3) to address robustness of the empirical results. The elements $w_{ij} = w_{ij}^*/\sum_{j=1}^N w_{ij}^*$ and $m_{ij} = m_{ij}^*/\sum_{j=1}^N m_{ij}^*$ are built from binary matrices, with $w_{ij}^* = 1$ and $m_{ij}^* = 1$, if the i-th and the j-th hospital are contiguous, respectively. The definition of contiguity differs across alternative weights matrices. The first concept, $\boldsymbol{W}_d$ and $\boldsymbol{M}_d$, is to define hospitals as contiguous if they are located in the same district. For the second set of weights matrices, $\boldsymbol{W}_n$ and $\boldsymbol{M}_n$, two hospitals are considered contiguous if they are either located in the same district, or if their respective districts of residence are neighbors. In the one-step SFA estimation the district of the hospital defines its region.

3 Results

In the following, we discuss estimation and inferential results for the potential effect of prospective payment on hospital efficiency. Detailed results for efficiency change, estimated time and spatial effects are documented and discussed in Appendix E.

3.1 Two-stage analysis

Estimation results for different model specifications are reported jointly with average estimated time effects and test statistics of the DRG hypotheses (4) to (7) in Table 1. The log-likelihood values of the SARAR specifications significantly exceed the corresponding statistic of a model neglecting spatial dependence (OLS), $\rho = \lambda = 0$. The highest log-likelihood statistic is achieved under the district based spatial weights matrices, $\boldsymbol{W}_d$ and $\boldsymbol{M}_d$. Thus, the district based spatial layout appears most appropriate to model hospital efficiency changes.[3]

The estimated parameters do not vary markedly across distinct spatial specifications. However, estimated effects of specialization, budget, GDP and population density appear to be weaker for the spatial models in comparison with OLS (e.g. Bech and Lauridsen, 2008). Nevertheless, there are no qualitative differences between the distinct model specifications. In particular, the estimation results indicate a positive relationship between market shares and efficiency changes. Specialized hospitals are characterized, on average, by lower efficiency gains in comparison with non-specialized hospitals. This might be explained by smaller possibilities of specialized hospitals to realize further efficiency gains, since these hospitals are found to be relatively more efficient (e.g. Daidone and D´Amico, 2009). The occupancy rate has a significantly positive parameter estimate implying that hospitals which are fully stretched are more likely to enhance their efficiency. Furthermore, high mortality and the budget size are correlated with lower efficiency, while hospital efficiency change appears invariant with regard to the age structure of the district's population. Noting that the estimated marginal impacts are mostly in line with findings of related studies (e.g. Herwartz and Strumann, 2012, Herr et al., 2011,

[3] We also estimate more parsimoniously parameterized model variants, i.e. the spatial lag model, $\rho = 0$, and the spatial error model, $\lambda = 0$ (Anselin, 1988). However, respective log-likelihood statistics are significantly smaller than the statistics of the corresponding SARAR models.

Table 1: Second-stage results of alternative regression models

	OLS	SARAR			
		$\boldsymbol{W}_d\&\boldsymbol{M}_d$	$\boldsymbol{W}_n\&\boldsymbol{M}_n$	$\boldsymbol{W}_d\&\boldsymbol{M}_n$	$\boldsymbol{W}_n\&\boldsymbol{M}_d$
time effects	yes	yes	yes	yes	yes
private	−0.017	−0.012	−0.022	−0.020	−0.019
non-profit	0.009	0.009	0.009	0.009	0.012
$\ln(ms)$	0.108**	0.111**	0.112**	0.112**	0.111**
spec	−0.051**	−0.037**	−0.047**	−0.047**	−0.049**
$\ln(mort)$	−0.037**	−0.031**	−0.039**	−0.038**	−0.038**
$\ln(occrate)$	0.081**	0.083**	0.081**	0.085**	0.075**
$\ln(budget)$	−0.237**	−0.220**	−0.233**	−0.231**	−0.228**
$age65$	0.003	0.004	0.008*	0.010**	0.003
$\ln(gdp)$	0.094*	0.064	0.058	0.046	0.080
$\ln(popdens)$	0.152**	0.058	0.097	0.143**	0.144**
spatial parameters	no	yes	yes	yes	yes
$LOGLIKE^a$	−809	−365**	−737**	−716**	−762**
average time effect estimates					
$\bar{\delta}_1$ (96-99)	0.116	0.207	0.110	0.132	0.086
$\bar{\delta}_2$ (00-03)	−0.251	−0.335	−0.216	−0.272	−0.232
$\bar{\delta}_3$ (04-06)	−0.300	−0.302	−0.314	−0.320	−0.303
$\bar{\delta}_4$ (96-03)	−0.067	−0.064	−0.053	−0.070	−0.073
$\bar{\delta}_5$ (00-06)	−0.272	−0.321	−0.258	−0.292	−0.262
test statistics					
$\bar{\delta}_1 < \bar{\delta}_2$ (AE)	46.003	58.721	34.143	43.494	39.562
$\bar{\delta}_2 < \bar{\delta}_3$ (IE$_1$)	4.729	−3.044**	8.843	4.357	6.753
$\bar{\delta}_4 < \bar{\delta}_3$ (IE$_2$)	17.915	17.762	18.864	17.810	17.418
$\bar{\delta}_1 < \bar{\delta}_5$ (OE)	35.025	44.567	29.820	34.223	31.013

Significance levels: ** 5%; * 10%; [a]significance levels for log-likelihood ratio tests against OLS; estimation is based on 16989 observations.

Tiemann and Schreyögg, 2009, Herr, 2008, Farsi and Filippini, 2008, Lee et al., 2008, Helmig and Lapsley, 2001, Chang, 1998), we believe that the explanatory factors control appropriately for heterogeneity of hospital performance. This offers an accurate quantification of potential efficiency gains in response to the financial reform.

Finally, the diagnostic results about the impact of the financial reform on overall hospital efficiency are discussed in detail. Opposite to an expected positive announcement effect, the pre-reform period is characterized by the highest efficiency gains as indicated by average time effects estimates. In contrast to the other model specifications, the SARAR model under the district spatial layout ($\boldsymbol{W}_d$, $\boldsymbol{M}_d$) indicates a significant improvement in efficiency after the DRG introduction. Thus, it appears that the negative introduction effect indicated by the other models could be due to an insufficient account of the spatial correlation pattern between hospitals. In other words, spatial spillover effects might be responsible for the indicated decrease in overall efficiency after the DRG introduction. However, the negative announcement effect dominates the positive effect of the DRG introduction. Thus, the overall effect of the financial reform is negative. This result is invariant to the underlying spatial specification.[4]

3.2 SFA one-step estimation

In order to highlight the spatial dependence of hospital performance we also estimate a more parsimoniously parameterized model variant, i.e. $\sigma_{\eta t}^2 = 0$. Table 2 displays the estimation and test results of both SFA approaches. To facilitate the interpretation of the estimated translog input coefficients, output elasticities with respect to inputs are reported for selected years. The estimated output elasticities are mostly positive and vary substantially over time and model specifications. Estimated effects on inefficiency are characterized by (significant) sign differences depending on the applied model. If spatial dependence is taken into account the estimated marginal impacts on inefficiency are mostly in line with findings of related studies (e.g. Herr, 2008 and Herwartz and Strumann, 2012). Thus, we believe that the spatial SFA offers an accurate quantification of potential efficiency gains in response to the DRG reform. Similar to the DEA based two-stage approach the pre-reform period is characterized by the highest efficiency gains. As opposed to the expected reform effect, overall inefficiency is increasing over time and, thus, indicating the absence of any positive DRG reform effect.

[4] Alternatively, we estimate a model with joint dummy variables for the DRG announcement and introduction period. However, the corresponding model diagnostics suggest a strong recommendation for the specification of separate year dummy variables, indicating substantial heterogeneity of hospital efficiency over time. This result holds also for the SFA one-step estimation.

Table 2: One-step SFA estimation results

	$\text{SFA}\vert_{\sigma^2_{\eta t}=0}$	$\text{SFA}\vert_{\sigma^2_{\eta t}>0}$		$\text{SFA}\vert_{\sigma^2_{\eta t}=0}$	$\text{SFA}\vert_{\sigma^2_{\eta t}>0}$
output elasticities for selected years			*effects on inefficiency*		
exp (1995)	0.284**	0.232*	*private*	−0.747**	2.535*
exp (1999)	0.234**	0.206**	*non-profit*	−0.232**	2.186
exp (2003)	0.123**	−0.010	$\ln(ms)$	−0.423**	−0.518
exp (2006)	0.112**	0.025	*spec*	0.664**	−1.792**
phys (1995)	0.155**	0.087	$\ln(mort)$	−0.045	−1.329**
phys (1999)	0.182**	0.149	$\ln(occrate)$	0.644**	−8.128**
phys (2003)	0.202**	0.093	$\ln(budget)$	−0.109	0.284
phys (2006)	0.169**	0.055	*age*65	−0.014	0.595*
nurses (1995)	0.101	0.068	$\ln(gdp)$	−0.462**	−2.004
nurses (1999)	0.165**	0.026	$\ln(popdens)$	−0.205**	0.326
nurses (2003)	0.388**	0.407**			
nurses (2006)	0.468**	0.443**	*LOGLIKE*	−16073	−15820
nonmed (1995)	0.184**	0.148	σ^2_u/σ^2_ν	0.197	1.155
nonmed (1999)	0.142**	0.130	σ^2_ν	0.504	0.333
nonmed (2003)	0.033	0.066	σ^2_u	0.020	0.445
nonmed (2006)	0.023	0.059	μ	−3.623	14.296
average time effect estimates			*test statistics*		
$\bar{\delta}_1$ (96-99)	−0.052	−0.028	$\bar{\delta}_1 < \bar{\delta}_2$ (AE)	0.389	−0.111
$\bar{\delta}_2$ (00-03)	−0.395	0.075	$\bar{\delta}_2 < \bar{\delta}_3$ (IE_1)	2.430**	−2.108
$\bar{\delta}_3$ (04-06)	−1.334	3.713	$\bar{\delta}_4 < \bar{\delta}_3$ (IE_2)	1.547*	−1.950
$\bar{\delta}_4$ (96-03)	−0.204	0.018	$\bar{\delta}_1 < \bar{\delta}_5$ (OE)	0.779	−1.282
$\bar{\delta}_5$ (00-06)	−0.797	1.634			

Significance levels: ** 5%; * 10%; estimation is based on 16835 observations.

4 Conclusions

This study analyzes technical hospital efficiency in Germany for the period 1995 to 2006 that includes the announcement (2000) and introduction (2004) of the DRG based financing system. In particular, we examine if the overall efficiency has increased after the DRG reform.

To address the robustness of diagnostics and inferential results we provide comparative applications of the non-parametric *Data Envelopment Analysis* and parametric *Stochastic Frontier Analysis*. On the one hand, we apply a two-stage procedure based on the Malmquist index decomposition of DEA efficiency scores in pure technical efficiency change. On the other hand, a fixed effects panel stochastic frontier model is employed that allows a one-step estimation of both production frontier parameters and inefficiency effects. Accounting for spatial dependence and observable and hidden hospital characteristics, both approaches do not indicate an increase in overall efficiency neither after the announcement nor the introduction of the DRG reform. In contrast, a significant decline in overall hospital efficiency is diagnosed. The results are robust against several alternative model variants, i.e. distinct spatial layouts, threshold values for outlier detection or the use of bootstrap standard errors.

The lack of a positive announcement effect could be explained by the incentive of hospitals to build up financial reserves after the reform was announced. This could be achieved by means of increasing the reimbursements under the per diem payment system. The reserves could help to finance investments in new technologies and facilities, and might strengthen the financial performance and lower the risk of insolvency under the new prospective payment system. As mentioned above, increasing the revenues in the previous financing system could be related to inefficiently long hospital stays. The absence of a positive DRG introduction effect might indicate that the transformation process towards the new incentive structure has not been completed till 2006. Moreover, the observed decline in treated patients after the DRG introduction (e.g. Braun et al., 2007) may also serve as an explanation. Furthermore, the dominance of opportunistic practices might explain adverse effects on efficiency. The identification of potential opportunistic practices and their effects on hospital efficiency could guideline future political reforms to improve overall hospital performance and, perhaps, enable a deceleration of the steady increase of hospital expenditures.

References

Anselin, L. (1988). Spatial econometrics: methods and models (Boston: Kluwer Academic Publishers).

Banker, R. D. and R. C. Morey (1986). Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research* 34(4), 513-521.

Banker, R. D., A. Charnes and W.W. Cooper (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science* 30, 1078-1092.

Barbetta, G. P., G. Turati and A. M. Zago (2007). Behavioral differences between public and private not-for-profit hospitals in the Italian national health service. *Health Economics* 16(1), 75-96.

Bech, M. and J. Lauridsen (2008). Exploring the spatial pattern in hospital admissions. *Health Policy* 87(1), 50-62.

Berta, P., G. Callea, G. Martini and G. Vittadini (2010). The effects of upcoding, cream skimming and readmissions on the Italian hospitals efficiency: A population-based investigation. *Economic Modelling* 27, 812-821.

Biorn, E., T. P. Hagen, T. Iversen and J. Magnussen (2003). The effect of activity-based financing on hospital efficiency: a panel data analysis of DEA Efficiency Scores 1992-2000. *Health Care Management Science* 6(4), 271-283.

Biorn, E., T. P. Hagen, T. Iversen and J. Magnussen (2010). How different are hospitals' responses to a financial reform? The impact on efficiency of activity-based financing. *Health Care Management Science* 13(1), 1-16.

Böcking, W., U. Ahrens, W. Kirch and M. Milakovic (2005). First results of the introduction of DRGs in Germany and overview of experience from other DRG countries. *Journal of Public Health* 13(3), 128-137.

Borden, J.P. (1988). An assessment of the impact of diagnosis-related group (Drg)-based reimbursement on the technical efficiency of New Jersey hospitals using data envelopment analysis. *Journal of Accounting and Public Policy* 7(2), 77-96.

Braun, T., F. Rau and K.H. Tuschen (2008). Die DRG-Einführung aus gesundheitspolitischer Sicht. Eine Zwischenbilanz. In: Schellschmidt, H. (Ed.), *Krankenhaus-Report '07*. Stuttgart: Schattauer, 3-22.

Burgess, J. and P. Wilson (1995). Decomposing hospital productivity changes, 1985-1988: A nonparametric Malmquist approach. *Journal of Productivity Analysis* 6, 343-363.

Chang, H. (1998). Determinants of hospital efficiency: the case of central government-owned Hospitals in Taiwan. *Omega* 26(2), 307-317

Chern, J.-Y. and T.T. Wan (2000). The impact of the prospective payment system on the technical efficiency of hospitals. *Journal of Medical Systems* 24(3) 159-172.

Chua, C., A. Palangkaraya and J. Yong (2011). Hospital competition, technical efficiency, and quality. *Economic Records* 87(277), 252-268.

Cooper, Z.N., S. Gibbons, S. Jones and A. McGuire (2010). Does Hospital Competition Improve Efficiency? An Analysis of the Recent Market-based Reforms to the English NHS, London: Centre for Economic Performance.

Coulam, R.F. and G.L. Gaumer (1991). Medicare's prospective payment system: a critical appraisal. *Health Care Financing Review*, Supplement 45-77.

Daidone and D´Amico (2009). Technical efficiency, specialization and ownership form: evidences from a pooling of Italian hospitals. *Journal of Productivity Analysis* 32(3), 203-216.

Dismuke, C. E. and V. Sena (1999). Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. *Health Care Management Science* 2(2), 107-116.

Evans, R. G. and H. D. Walker (1972). Information theory and the analysis of hospital cost structure. *The Canadian Journal of Economics* 5(3), 398-418.

Färe, R., S. Grosskopf, M. Norris and Z. Zhang (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1), 66-83.

Färe, R., S. Grosskopf, B. Lindgren, P. Roos (1992). Productivity changes in Swedish pharmacies 1980-1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis* 3, 85-101.

Farsi, M. and M. Filippini (2008). Effects of ownership, subsidization and teaching activities on hospital costs in Switzerland. *Health Economics* 17(3), 335-350.

France, G., F. Taroni and A. Donatini (2005). The Italian health-care system. *Health Economics* 14(1), 187-202.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126(2), 269-303.

Helmig, B. and I. Lapsley (2001). On the efficiency of public, welfare and private hospitals in Germany over time - A sectoral DEA-study. *Health Services Management Research* 14(4), 263-274.

Hensen, P., S. Beissert, L. Bruckner-Tuderman, T. A. Luger, N. Roeder and M. L. Müller (2008). Introduction of diagnosis-related groups in Germany: evaluation of impact on in-patient care in a dermatological setting. *The European Journal of Public Health* 18(1), 85-91.

Herr, A. (2008). Cost and technical efficiency of German hospitals: does ownership matter? *Health Economics* 17(9), 1057-1071.

Herr, A., H. Schmitz and B. Augurzky (2011). Profit efficiency and ownership of German hospitals. *Health Economics* 20(6), 660-674.

Herwartz, H. and C. Strumann (2012). On the effect of prospective payment on local hospital competition in Germany. *Health Care Management Science* 15(1), 48-62.

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis. *Health Care Management Science* 4(2), 103-115.

Jacobs, R., P.C. Smith, and A. Street (2006). Measuring efficiency in health care: analytic techniques and health policy. Cambridge University Press: Cambridge.

Johnson, A. L. and L. F. McGinnis (2008). Outlier detection in two-stage semiparametric DEA models. *European Journal of Operational Research* 187(2), 629-635.

Kumbhakar, S.C., S. Ghosh and J.T. McGuckin (1991). A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business & Economic Statistics, American Statistical Association* 9(3), 279-286.

Lee, K., K. Chun and J. Lee (2008). Reforming the hospital service structure to improve efficiency: Urban hospital specialization. *Health Policy* 87(1), 41-49.

Levaggi, R., M. Montefiori (2003). Horizontal and vertical cream skimming in the health care market. DISEFIN Working Paper No. 11/2003.

Linna, M. (2000). Health care financing reform and the productivity change in Finnish hospitals. *Journal of Health Care Finance* 26(3), 83-100.

Lungen, M. and I. Lapsley (2003). The reform of hospital financing in Germany: an international solution? *Journal of Health Organization and Management* 17(5), 360-372.

Simar, L and P. W. Wilson (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* 136(1), 31-64.

Simborg, D.W. (1981). DRG creep: a new hospital-acquired disease. *The New England Journal of Medicine* 304(26), 1602-1604.

Sommersguter-Reichmann, M. (2000). The impact of the Austrian hospital financing reform on hospital productivity: empirical evidence on efficiency and technology changes using a non-parametric input-based Malmquist approach," *Health Care Management Science* 3(4), 309-321.

Staat, M. (2006) Efficiency of hospitals in Germany: a DEA-bootstrap approach. *Applied Economics* 38(19), 2255-2263.

Staat, M. and M. Hammerschmidt (2003). Benchmarking the Health Sector in Germany - An Application of Data Envelopment Analysis. Research paper, Institute for Market-Oriented Management, University Mannheim.

Statistisches Bundesamt (2012). Gesundheit - Kostennachweis der Krankenhäuser. Fachserie 12 Reihe 6.3, Wiesbaden 2012.

Tiemann, O. and J. Schreyögg (2012). Changes in hospital efficiency after privatization. *Health Care Management Science*, forthcoming.

Tiemann, O. and J. Schreyögg (2009). Effects of ownership on hospital efficiency in Germany. *BuR Business Research Journal* 2(2), 115-145.

Wang, H.-J. and C.-W. Ho (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics* 157(2), 286-296.

Wang H.-J., and P. Schmidt (2002). One-Step and Two-Step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis* 18(2), 129-144.

Werblow A., A. Karmann, B.P. Robra (2010) Effizienz, Wettbewerb und regionale Unterschiede in der stationärenVersorgung. In: Klauber, J., M. Geraedts, J. Friedrich*Krankenhaus-Report 2010 - Schwerpunkt: Krankenhausversorgungin der Krise?* (Stuttgart, Schattauer), 41-71.

Werblow A., B.P. Robra (2007) Einsparpotenziale im medizinfernen Bereich deutscher Krankenhäuser -eine regionale Effizienzfront-Analyse. In: Klauber, J., H.Schellschmidt, B.P. Robra, *Krankenhaus-Report 2006 -Schwerpunkt: Krankenhausmarkt im Umbruch* (Stuttgart, Schattauer), 133-151.

Wilson, P.W. (1995). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis* 6(1), 27-45.

Zweifel, P., F. Breyer and M. Kifmann (2009). Health Economics (Berlin: Springer).

A DEA efficiency scores

The efficiency score, θ_i^v, is obtained under the assumption of variable returns to scale (Banker et al., 1984) by solving the following linear program

$$
\begin{aligned}
\theta_i^v = \arg\min_{\theta_i^v,\tau}\{\theta_i^v > 0 | \sum_l \tau_l q_{pl} \geq q_{pi} \quad & \forall \quad p \in \{1, ..., s\} \\
\theta_i^v x_{ki}^D \sum_l \tau_l x_{kl}^D \quad & \forall \quad k \in \{1, ..., m^D\} \\
x_{ji}^N \geq \sum_l \tau_l x_{jl}^N \quad & \forall \quad j \in \{1, ..., m^N\} \\
\sum_l \tau_l = 1,\ \tau_l > 0 \quad & \forall \quad l = 1, ..., N\},
\end{aligned}
$$

where q_{ri}, x_{ji}^N and x_{ki}^D denote output, non-discretionary and discretionary input variables of hospital i. The numbers of outputs, non- and discretionary inputs, and reference hospitals are s, m^N, m^D, and N, respectively.

B Second stage Maximum Likelihood estimation

The (unbalanced) model can be written in matrix notation as

$$
y = \lambda \begin{pmatrix} \boldsymbol{W}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{W}_T \end{pmatrix} y + \begin{pmatrix} \mathfrak{o}_1 & \cdots & \mathfrak{o}_1 \\ \iota_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \iota_T \end{pmatrix} \delta + \omega + \boldsymbol{Z}\beta + e, \quad e = \rho \begin{pmatrix} \boldsymbol{M}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{M}_T \end{pmatrix} e + \epsilon, \quad (11)
$$

where $y = (y_1', \ldots, y_T')'$, $\boldsymbol{Z} = (\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_T')'$, $e = (e_1', \ldots, e_T')'$ and $\epsilon = (\epsilon_1', \ldots, \epsilon_T')'$. The coefficients of the time dummy variables, δ_t, are collected in $\delta = (\delta_2, \ldots, \delta_T)'$, where $t = 1$ is the benchmark, $\mathfrak{o}_t$ and ι_t is an $N_t \times 1$ vector of zeros and ones, respectively, where N_t is the number of hospitals sampled in time t. Fixed effects are summarized in $\omega = (\omega_1', \ldots, \omega_T')'$, where ω_t is an $N_t \times 1$ vector comprising the individual effects of the N_t hospitals. These are dropped out by means of the within transformation. The panel and cross-sectional models are estimated by means of a Maximum Likelihood (ML) approach.

Model (11) can be written as

$$\mathbf{B}\mathbf{A}\widetilde{y} = \mathbf{B}\begin{pmatrix}\widetilde{\mathbf{1}} & \widetilde{\boldsymbol{Z}}\end{pmatrix}\begin{pmatrix}\delta \\ \beta\end{pmatrix} + \epsilon,$$

where $\widetilde{y}$, $\widetilde{\mathbf{1}}$ and $\widetilde{\boldsymbol{Z}}$ are the time demeaned variables of y, $\mathbf{1}$ and $\boldsymbol{Z}$, respectively, where

$$\mathbf{1} = \begin{pmatrix}\mathfrak{o}_1 & \cdots & \mathfrak{o}_1 \\ \iota_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \iota_T\end{pmatrix}, \ \mathbf{B} = \begin{pmatrix}\boldsymbol{B}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{B}_T\end{pmatrix}, \ \mathbf{A} = \begin{pmatrix}\boldsymbol{A}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{A}_T\end{pmatrix}$$

and $\boldsymbol{B}_t = \boldsymbol{I}_{N_t} - \rho \boldsymbol{M}_t$, $\boldsymbol{A}_t = \boldsymbol{I}_{N_t} - \lambda \boldsymbol{W}_t$. Assuming a multivariate normal distribution of the error terms, the log likelihood function is given by

$$\ln L = \sum_{t=1}^{T}\left(-\frac{N_t}{2}\ln(2\pi\sigma^2) + \ln|\boldsymbol{A}_t| + \ln|\boldsymbol{B}_t| - \frac{\epsilon_t'\epsilon_t}{2\sigma^2}\right), \tag{12}$$

where

$$\epsilon_t = \begin{cases}\boldsymbol{B}_t\left(\boldsymbol{A}_t\widetilde{y}_t - \widetilde{\boldsymbol{Z}}_t\beta\right) & \forall \quad t = 1 \\ \boldsymbol{B}_t\left(\boldsymbol{A}_t\widetilde{y}_t - \widetilde{1}_t\delta_t - \widetilde{\boldsymbol{Z}}_t\beta\right) & \forall \quad t = 2, \ldots, T\end{cases}$$

and $\sigma^2 = \sum_{t=1}^{T}(\epsilon_t'\epsilon_t/N_t)$. The ML estimator is

$$\begin{pmatrix}\hat{\delta}_{ML} \\ \hat{\beta}_{ML}\end{pmatrix} = \left[\begin{pmatrix}\widetilde{\mathbf{1}}' \\ \widetilde{\boldsymbol{Z}}'\end{pmatrix}\widehat{\mathbf{B}}'\widehat{\mathbf{B}}\begin{pmatrix}\widetilde{\mathbf{1}} & \widetilde{\boldsymbol{Z}}\end{pmatrix}\right]^{-1}\begin{pmatrix}\widetilde{\mathbf{1}}' \\ \widetilde{\boldsymbol{Z}}'\end{pmatrix}\widehat{\mathbf{B}}'\widehat{\mathbf{B}}\widehat{\mathbf{A}}\widetilde{y},$$

where $\widehat{\mathbf{B}} = \begin{pmatrix}\widehat{\boldsymbol{B}}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widehat{\boldsymbol{B}}_T\end{pmatrix}$, $\widehat{\boldsymbol{B}}_t = \boldsymbol{I}_{N_t} - \hat{\rho}_{ML}\boldsymbol{M}_t$, $\widehat{\mathbf{A}} = \begin{pmatrix}\widehat{\boldsymbol{A}}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widehat{\boldsymbol{A}}_T\end{pmatrix}$ and $\widehat{\boldsymbol{A}}_t = \boldsymbol{I}_{N_t} - \hat{\lambda}_{ML}\boldsymbol{W}_t$.

C One-step Simulated Maximum Likelihood estimation

The model after the first-difference transformation reads as

$$\begin{aligned}\Delta \ln q_{ij} &= \Delta\tau + \Delta f(X_{ij},\alpha) + \Delta v_{ij} - \Delta u_{ij}, \\ \Delta u_{ij} &= \Delta h_{ij} u^*_{ij}, \qquad \Delta\nu_{ij} \sim \mathcal{N}(0,\Sigma_{\nu ij}),\end{aligned} \tag{13}$$

where $\Delta \ln q_{ij} = \begin{pmatrix} \ln q_{ij2} - \ln q_{ij1} \\ \vdots \\ \ln q_{ijT_{ij}} - \ln q_{ijT_{ij}-1} \end{pmatrix}$, $\Delta h_{ij} = \begin{pmatrix} h_{ij2} - h_{ij1} \\ \vdots \\ h_{ijT_{ij}} - h_{ijT_{ij}-1} \end{pmatrix}$, $\Delta \nu_{ij} = \begin{pmatrix} \nu_{ij2} - \nu_{ij1} \\ \vdots \\ \nu_{ijT_{ij}} - \nu_{ijT_{ij}-1} \end{pmatrix}$,

$\Delta\tau = \begin{pmatrix} \tau_2 - \tau_1 \\ \vdots \\ \tau_{T_{ij}} - \tau_{T_{ij}-1} \end{pmatrix}$ and $\Delta f(X_{ij},\alpha) = \begin{pmatrix} \Delta f(X_{ij2},\alpha_2) \\ \vdots \\ \Delta f(X_{ijT_{ij}},\alpha_{T_{ij}}) \end{pmatrix}$ are $T_{ij}\times 1$ vectors with $\Delta f(X_{ijt},\alpha_t) =$ $f(X_{ijt},\alpha_t) - f(X_{ijt-1},\alpha_{t-1})$, $f(X_{ijt},\alpha_t) = \ln x_{ijt}\alpha_t + \sum_k \sum_{k\geq l} \alpha_{tkl} \ln x_{ijtk} \ln x_{ijtl}$. The error terms of successive time points of the i-th hospital, i.e. ν_{ijt} and ν_{ijt-1} are correlated due to the first-difference transformation. Thus, $\Delta\nu_{ij}$ is multivariate normal distributed with covariance matrix Σ_{vij}. The matrix has $2\sigma^2_\nu$ on the main diagonal. The off-diagonals contain either $-\sigma^2_\nu$ for successive correlated observations and zeros otherwise. For example a hospital with data at time $t = 1,2,3,5,6$ obtains

$$\Delta\nu_{ij} = \begin{pmatrix} \nu_{ij2} - \nu_{ij1} \\ \nu_{ij3} - \nu_{ij2} \\ \nu_{ij6} - \nu_{ij5} \end{pmatrix} \quad \text{with} \quad \Sigma_{vij} = \begin{pmatrix} 2\sigma^2_\nu & -\sigma^2_\nu & 0 \\ -\sigma^2_\nu & 2\sigma^2_\nu & 0 \\ 0 & 0 & 2\sigma^2_\nu \end{pmatrix}.$$

The estimated log-likelihood function is given by

$$\widehat{\ln L} = \sum_{j=1}^{J} \ln \left[\frac{1}{S} \sum_{s=1}^{S} \exp\left(\sum_{i=1}^{N_j} \ln \tilde{f}(\varepsilon_{ij}|\eta^s_j) \right) \right],$$

where J is the number of regions, N_j is the number of hospitals in region j, η^s_j is a $(T\times 1)$ vector of simulated random effects and $\ln \tilde{f}(.)$ is an unbiased simulator for the conditional log-likelihood function of the i-th hospital (Wang and Ho, 2010)

$$\begin{aligned}\ln f(\varepsilon_{ij}|\eta_j) &= -\frac{1}{2}(T_{ij}-1)\ln(2\pi) - \frac{1}{2}|\Sigma_{vij}| - \frac{1}{2}\Delta\varepsilon_{ij}'\Sigma_{vij}^{-1}\Delta\varepsilon_{ij} \\ &+ \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right),\end{aligned}$$

where $\mu_* = \dfrac{\mu/\sigma_u^2 - \Delta\varepsilon_{ij}'\Sigma_{vij}^{-1}\Delta h_{ij}}{\Delta h_{ij}'\Sigma_{vij}^{-1}\Delta h_{ij} + 1/\sigma_u^2}$, $\sigma_*^2 = \dfrac{1}{\Delta h_{ij}'\Sigma_{vij}^{-1}\Delta h_{ij} + 1/\sigma_u^2}$ and $\Delta\varepsilon_{ij} = \Delta\ln q_{ij} - \Delta\tau - \Delta f(X_{ij},\alpha)$.

D Construction of case mix weights

The more time the treatments of cases belonging to the j-th clinical department take relative to all other treatments, the higher the weight, π_j, of the respective cases. Let c_{ij} be the number of cases in the j-th clinical department of the i-th hospital. Then, the weighted cases of hospital i are calculated as

$$wc_i = \sum_{j=1}^{J} \pi_j\, c_{ij},$$

where $\pi_j = los_j/los_G$, $los_j = (\sum_{i=1}^{N} days_{ij}/c_{ij})/N$ is the mean length of stay for the cases belonging to the j-th clinical department over all hospitals and $los_G = (\sum_{j=1}^{J} los_j)/J$ is the mean length of stay over all clinical departments and all hospitals.

E Complementary empirical results

Table 3 shows the results of the estimated variance of the random regional effects of the SFA model and the spatial parameter estimates of the SARAR model under the district based spatial weights matrices, $\boldsymbol{W}_d$ and $\boldsymbol{M}_d$. Moreover, the Malmquist decomposition in pure efficiency change and the estimated time effects of both model specifications are provided. The spatial parameter estimates of λ, ρ and σ_η^2 are characterized by substantial heterogeneity over time. In the two-stage analysis of technical efficiency change the spatial autocorrelation parameter, ρ, is mostly positive while the spatial lag parameter, λ, is negative. This result is similar

Table 3: Complementary estimates

	two-stage analysis (SARAR($\boldsymbol{W}_d$ & $\boldsymbol{M}_d$))				one-step SFA estimation	
year	γ^a	δ	λ	ρ	δ	σ^2_η
1995	-	-	-	-	-	0.022
1996	1.218	-	-0.018	0.032	-0.623	0.083
1997	2.881	1.250^{**}	-0.469^{**}	0.533^{**}	1.347	0.001
1998	1.023	-0.184^{**}	-0.023	0.109^{**}	3.748	2.351
1999	0.968	-0.239^{**}	0.000	0.070	-4.612	1.741
2000	0.997	-0.212^{**}	-0.005	0.021	2.373	0.019
2001	0.914	-0.293^{**}	0.024	0.050	2.409	0.002
2002	0.569	-1.142^{**}	-0.836^{**}	0.696^{**}	-6.087	0.519
2003	1.552	0.307^{**}	-0.052	0.009	1.606	0.681
2004	0.688	-0.570^{**}	0.036	-0.007	4.194^{**}	4.720^{**}
2005	1.014	-0.179^{**}	-0.028	0.071^{*}	5.467^{**}	0.217
2006	1.042	-0.157^{**}	-0.018	0.118^{**}	1.477	0.289

^{a}geometric mean; significance levels: ** 5%; * 10%;

to the findings of Herwartz and Strumann (2012), who applied a spatial two-stage analysis of estimated efficiency scores. However, the magnitudes of the estimated spatial effects are much smaller for efficiency change as for efficiency scores. And there is no hint for an increase in negative spatial spillovers of efficiency change due to the DRG reform as detected for the level of efficiency in Germany by Herwartz and Strumann (2012). The largest magnitude of both spatial parameters, λ and ρ, is obtained for periods with considerable overall efficiency improvements (1997) or deteriorations (2002). Thus, the interaction between nearby hospitals is particularly strong in periods that are characterized by an outstandingly large change in overall hospital performance.