UNIVERSITY of York

HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 22/28

Is austerity good for efficiency, at least? A counterfactual assessment for the Italian NHS

Calogero Guccio; Giacomo Pignataro;

Domenica Romeo and Francesco Vidoli

November 2022

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Is austerity good for efficiency, at least? A counterfactual assessment for the Italian NHS

Calogero Guccio[a], Giacomo Pignataro[a,b,*], Domenica Romeo[a], Francesco Vidoli[c]

[a] *University of Catania, Department of Economics and Business*
[b] *Politecnico di Milano, Department of Management, Economics and Industrial Engineering*
[c] *University of Urbino Carlo Bo, Department of Economics, Society and Politics*

Abstract

In recent decades, austerity measures have been widely adopted in public healthcare systems, so as to cope with financial constraints. This paper assesses the impact of a specific policy implemented in some Italian regions since 2007 with the purpose of reducing their healthcare spending deficit, the so called Recovery Plans (Piani di rientro), on the technical efficiency of their hospitals. Using a unique sample of administrative data relative to a large panel of hospitals in the period 2003-2010, and employing, as identification strategy, the exogenous introduction of the austerity policy in some regions, we find that the policy had a detrimental effect on the efficiency of the hospitals operating in the regions subjected to the policy. The results show that the efficiency loss grows over time, suggesting the existence of negative cumulative effects of the austerity policy.

Keywords: Hospitals, Recovery plans, Technical efficiency, Austerity, Spending cuts
JEL: I10, I18, D24

*Corresponding author: giacomo.pignataro@unict.it

1. Introduction

Several countries, mainly in Europe, in the aftermath of the 2008 financial crisis, were deeply involved in implementing policies to reduce their public deficits. Most of these policies were focused on spending cuts[1], and many studies have investigated their impact, not only in terms of their immediate objectives, but also with regard to social welfare consequences. This issue has been quite widely explored in the health field, since public healthcare was largely hit by spending cuts.

The main concern in the analysis of the effects of healthcare spending cuts on the overarching goals of any healthcare system relates, specially, to health and equity, since public spending reduction determines an unavoidable restriction of coverage and hence of access to healthcare (Reeves et al., 2015). A study by the European Observatory on Health Systems and Policies (Maresso et al., 2015), surveying the experience of over 45 countries in reacting to pressure created by the financial and economic crisis began in 2008, did not find any conclusive evidence about the effects of the economic crisis and the related austerity measures on the general health conditions of the different populations under investigation, also because "*the full scale of the effects of the crisis on health may not be apparent for years*" (Maresso et al., 2015, p. 171). However, focusing on one of the consequences of crisis for several countries, that is public spending cuts and consequent coverage restrictions, the authors find that these are among the pathways that may have undermined financial protection and equitable access to healthcare. When more specific health outcomes are considered, for which the time span between exposure and outcome is relatively short, like in the case of mental health, for which studies are convergent in showing a deterioration during the crisis, Mattheys et al. (2016) and Akhter et al. (2018) show that the material and psycho-social factors are the most important determinants of the mental health gap between the most and least deprived areas of the local authority under exam. The authors deduce that austerity, widening the inequalities in these determinants may also widen the inequalities in health.

Despite the wide number of studies investigating the impact of spending cuts on several dimensions of social welfare, at least as far as healthcare is concerned, very few ones are concerned with their effects on efficiency. This

[1] According to Stuckler et al. (2017), "*the majority of deficit reduction policies (>80%) involved budget cuts rather than tax increases*".

is, indeed, a crucial issue for the assessment of the effectiveness of such policies. While the short-term objective of saving on the global cost of service may be straightforwardly achieved, the spending cuts policies may be inefficient in the long run. Wenzl et al. (2017) explain the potential trade-off between short-term cost savings and long-run efficiency: "*if cuts are excessive, however, making some activity economically unviable, they may also be associated with reductions in service volumes. A similar dynamic may apply to cutting health worker incomes – if cuts are substantial, service volumes and quality may decrease*". In terms of policy assessment, then, following Maresso et al. (2015), it can be said that "*these types of response reflect a tendency to put the short-term need for quick savings above the need for efficiency and longer-term expenditure control*".

Our paper represents a contribution aiming at filling the knowledge gap related to the effects of austerity, or spending cuts, policies on the efficiency in the production of healthcare services, so as to enlighten the potential trade-off between short-term and long-term policy objectives. We focus on a specific set of austerity policies implemented in Italy since 2007, aiming at reducing the budget deficit of regional governments originated by healthcare expenditure, under the control of the central government. The specific objective of our analysis is to assess the impact of such policies, in the regions where they were implemented, on the technical efficiency of hospitals, by means of a counterfactual approach, using as a control group similar hospitals in structural terms in the regions where the policies were not enacted. Our results show quite clearly that hospitals in regions where the austerity policies were implemented experienced an increase in technical inefficiency, since the spending cuts first reduced the quantity of inputs they employed and, subsequently, the volume of output was reduced more than proportionally relative to inputs.

Quite few papers may be related to the topic dealt with in this paper. A couple of papers assessing the technical efficiency of Greek hospitals can be mentioned here, since they use data covering the time period during which Greece experienced the implementation of severe spending cuts policies. Polyzos (2012) estimates efficiency, over three years, (2009-2011) for 117 Greek hospitals (divided in three groups, by size). He simply compares efficiency scores along the three years and observes that efficiency generally improved along the time period considered. Xenos et al. (2017) proceed with a DEA estimation of efficiency over four years (2009-2012) for 108 Greek general hospitals; in a second stage, with a Tobit, they regress efficiency scores on what they regard as contextual variables, even if they are endogenously related to the production process of hospitals (average length of stay, bed occupancy rate, size of the hospital, number of patients and of diagnostic procedures). They find that, at first (2009-2010), there was a substantial reduction in productivity, followed by a gain in the second period (2010-2011), while remaining constant in the third period (2011-2012). Without entering into the discussion of their results, these papers do not make any attempt of examining the performance of hospitals during the time period considered in connection with the policies implemented in the same period, above all the ones enacting spending cuts. A different result can be found in a paper that examine a larger sample of hospitals, in terms of their geographical location. Samut and Cafrı (2016) carry out a two-stage analysis of technical efficiency of hospitals in 29 OECD countries, in the time period 2000-2010. In the first stage, technical efficiency is assessed by DEA while, in the second stage, a panel Tobit analysis was used to examine the determinants of efficiency. The authors observe that efficiency declined in 2009 and 2010, and attribute this pattern to the cuts in health spending occurred in many countries in their sample, even if this conclusion is not supported by their second stage analysis. To the best of our knowledge, therefore, our paper is, if not the first one, among the first ones to examine the causal link between spending cut policies and the efficiency in the production of healthcare services, in particular hospital care. Since we consider policies based on the intervention on specific regional situations characterised by high budget deficits, which, therefore, do not originate from a general economic context of crisis, we are thus able to isolate the impact of policies, without dealing with the problem of disentangling it from any potential overlapping effect of the general economic context. Moreover, since the policy is implemented only for regions, facing a deficit problem, this allows us to use a counterfactual approach for a rigorous exploration of the causality link between the policy and the technical efficiency of hospitals treated with that policy, by comparing hospitals in regions treated with the policy and hospitals in regions where the policy was not implemented.

We believe that this paper offers several contributions to the existing literature. First, as emphasised above, it is among the earliest papers to attempt a rigorous assessment of the impact of austerity policies on hospital sector efficiency. Second, to the best of our knowledge, it is the first paper to employ the recent developments of the literature on endogenous SFAs (Karakaplan and Kutlu, 2017; Karakaplan, 2022) to assess the technical efficiency of hospitals. Finally, through the use of a large panel of hospitals and the employment of a matching technique, it significantly contributes to the limited literature that employs frontiers in a counterfactual setting (Lindlbauer et al., 2016).

The paper proceeds as follows. In section 2, we will provide some information on the policy implemented in Italy, since 2007, aiming at recovering the deficit arising from healthcare spending in regions with a considerable deficit, and we briefly survey the main contributions on the analysis of the impact of this policy. In section 3, we depict our empirical strategy, for the assessment of technical efficiency of hospitals and for the use of the counterfactual approach. The section also includes the presentation of the data used in our application. In section 4, the main results are presented, and in section 5, some concluding remarks are drawn.

2. The control of regional healthcare deficits in Italy

2.1. The institutional features of the policy

Since its establishment, the Italian National Health Service (NHS) has been characterised by a progressive devolution process aimed at improving the efficiency and the quality of healthcare services. It was only with the constitutional reform adopted in 2001 that administrative and fiscal autonomy was formally granted to regions and autonomous provinces, leading to the federalisation and decentralisation of the healthcare system (Arcà et al., 2020). Since then, healthcare funding have been collected by the central government and redistributed at the regional level according to the local population size and its composition by age. The decentralisation process resulted in significant regional discrepancies in outputs, resources allocation and healthcare spending, jeopardising the sustainability of the system (Giancotti et al., 2020). In particular, some regions failed, more than others, to balance their budgets, due to their limited managerial capacity and poor health service performance. As a consequence, in 2006 the cumulative deficit from healthcare spending reached six billion euro (Depalo, 2019). The central government was forced to impose specific Recovery Plans (RPs) (the so-called *Piani di Rientro*), which are still active, in a small number of regions with a considerable health budget deficit. RPs, actually introduced with the national Budget Law in 2007, represent an extraordinary mechanism to re-centralise the control of healthcare spending to the State and reorganise healthcare services, intervening on the factors responsible for the economic and financial imbalances (Giancotti et al., 2020). A RP is a formal agreement between a region and the central government, which commits the region to lay out a consolidation path to be implemented over a three-year period to restore regional accounts. The actual realisation of the plan will be closely monitored by the central government. Regions with deficits larger than 5% of the overall level of funding are required to submit a three-year operational program to be approved by the Ministry of Health together with the Ministry of Economy and Finance (Bordignon et al., 2020). Then, both the authorities evaluate the provision of healthcare services by the region, through both an ex-ante and an ex-post quarterly monitoring activity, to guarantee that the measures implemented do not affect the provision of the essential level of care (Depalo, 2019). As previously mentioned, RPs are structured over three years under the condition of meeting the plans objectives, otherwise it will be renewed for an additional three-year period. Additionally, if the concerned region fails to reach the goals planned for the first year, the central government is entitled to appoint a commissioner in charge of the effective implementation of the program. The commissioner acts on behalf of the central government and oversees the regions in health-related decisions, mainly those related to the plan fulfilment. Such stringent measure comes with a further increase of regional taxes and the stop of central government non-mandatory transfers (Bordignon et al., 2020).

RPs pursue two main objectives. First, they attempt to contain costs in order to reach a balance budgetary condition. Second, they must ensure the provision of a given level of healthcare services, otherwise regions are asked to increase regional taxation. Cost containment strategy is based on several measures: institutional reorganisation through hospital mergers (*e.g.* reduction of the 40% in the number of local health authorities, OASI, 2021); strict standards in terms of hospital beds allocation and hospitalisation rates (*i.e.* de-hospitalisation policy); labour force rationing through freezing of personnel turn-over and a block on hiring; control over pharmaceutical consumption through direct distribution of drugs; reduction in the volume of services provided by private accredited facilities; introduction of centralised purchase to avoid further rise in spending; use of health insurance card system to ensure the appropriateness of community prescribing (General State Accounting Office, 2009).

Plans were first signed on February 2007. For the first round (2007-2009) five regions were enrolled in RP because of their large deficits: Abruzzo, Campania, Lazio, Liguria and Molise. In mid-2007 also Sicilia and Sardegna were introduced to RPs preceding Calabria at the end of 2009. Piemonte and Puglia entered in RP in 2011. Lazio was the first to be commissioned in 2008, followed by Abruzzo in 2009, Campania and Molise in 2010 and Calabria in 2011. Liguria and Piemonte were successful in implementing RPs, restoring their balances and left the plans in 2010 and 2016, respectively. Sardegna also exited from the RP thanks to its special statute (Bordignon et al., 2020). Abruzzo is still under RP though without the external commissioner. The remaining regions continue to face RPs.

2.2. Empirical findings on the impact of RPs

There is large consensus on the effectiveness of RPs on regions' economic and finance balance (Italian Ministry of Health, 2014; Atella et al., 2019). According to Aimone Gigio et al. (2018) the measures adopted, mainly cuts in medical staff and reduction in the number of hospital beds per thousands of inhabitants, allowed to align cost structures of regions in RPs to non-RPs regions. However, RPs' effects on healthcare outcomes are still under question. Depalo (2019) shows that the containment of health spending resulting from RPs' adoption came at a cost, in terms of specific efficiency indicators. In particular the author provides robust evidence for a small increase in mortality rates, at least for Lazio, Abruzzo, Campania and Sardegna. Additionally and strictly correlated to the above-mentioned indicator, the study highlights a drop in total hospitalisation for the second cycle of the policy. Similarly, Arcà et al. (2020) find that cuts in annual spending, following RPs, increased avoidable deaths, mostly cancer-related, by 3%. According to the authors, RPs also had an impact on regional migration, as witnessed by the rise in hospital care seeking in regions without RPs, questioning the equity in the access to services. The financial benefits from the plan arising from the rationalisation of the supply structure, are also stressed by Bordignon et al. (2020). However, contrarily to the above-mentioned studies, in their analysis Bordignon et al. (2020) do not detect substantial consequences on health outcomes and on the use of health care services, at least for the services under investigation. As a confirmation of this, Giancotti et al. (2020) show that not only RPs did not impact on overall hospital efficiency but even improved technological progress and total factor productivity. However, their results cannot be generalised given the restricted number of hospitals included in the analysis.

Though certainly successful in erasing regional deficits, we cannot exclude that the implementation of recovery plans carried several drawbacks which deserve deeper attention. First, given the higher proportion of workers reaching retirement age (due to the ageing of the workforce), a complete block of personnel turnover may give rise to concerns over medium-term sustainability. Second, quality indicators, though slightly improved, are still lower in RPs regions than in regions without RPs. Such between-regions inequality which is highly perceived by patients fosters inter-regional mobility with significant financial implications (*e.g.* the need to compensate for the negative balance of mobility, Aimone Gigio et al., 2018).

As mentioned in section 1, also for Italy, with respect to the implementation of RPs, there is no investigation of their impact on efficiency, with the exception of Giancotti et al. (2020), whose evidence on the matter is, however, limited for the reasons stated above. In the next section, we will provide the empirical strategy for addressing such an issue in the most rigorous way.

3. Empirical strategy

In this section, we first provide a brief reference to the general features of the methods used in the paper for estimating technical efficiency of hospitals (section 3.1) and for the application of the counterfactual approach (section 3.2), at the basis of the empirical strategy for the analysis of our data, presented in section 3.3.

3.1. Production efficiency estimation methods

There is an extensive literature estimating the technical efficiency of healthcare providers through two main alternative approach, Data envelopment analysis (DEA) and Stochastic frontier analysis (SFA), respectively (see *e.g.* the reviews by Worthington, 2004; Hollingsworth, 2008; Tiemann et al., 2012; Kiadaliri et al., 2013; Mahdiyan et al., 2019), though some trade-off between the two techniques exists. Being nonparametric, DEA requires minimal assumptions about the frontier but does not account for statistical noise in the data (Cavalieri et al., 2018). Contrarily, SFA, using statistical regressions, splits the usual standard error term into efficiency and noise (Hollingsworth, 2016). However, such parametric method needs strong assumptions about the frontier functional form (Jacobs, 2001). Despite less used in the past compared to the nonparametric technique, the use of SFA in the healthcare setting has become a common practice in recent years, thanks to the advancements in the modelling techniques as well as the increased computational capacities (Hollingsworth, 2016).

Wei et al. (2018) investigate the relationship between Chinese hospitals' efficiency and the use of high technology (*i.e.* tomography and magnetic resonance). Cost inefficiency estimated through SFA increases with technology employment. Using both SFA and OLS regressions, Mateus et al. (2015) compare hospitals' discharge plans from England, Spain, Portugal and Slovenia and find that for all countries but Slovenia, beds availability and human resources are key drivers for the production process. With the aim of assessing the impact of geographic dependence on Italian hospitals' efficiency Cavalieri et al. (2020) apply a spatial stochastic frontier analysis and show that neighbouring effects on ranking scores are negligible compared to that of regional disparities and institutional factors. The importance of institutional quality in the Italian context is confirmed in the study by Boffardi (2022) who employ a stochastic frontier framework to a panel data of 20 regions. Still in the same national context, Barra et al. (2022) build a composite indicator using the Benefit of the Doubt approach to be included in a stochastic analysis as hospital outcome. Results show that managerial inefficiencies are the main responsible for hospitals' inefficiencies especially for southern regions. Finally, focusing the research to Lombardy, Colombi et al. (2017) assess persistent (time-invariant) and transient (time-varying) inefficiency of 133 hospitals using a 4-random component SF model which accounts for unobserved heterogeneity. Authors demonstrate that ownership, specialisation, and size significantly affect both efficiency measures.

While there are numerous and wide-ranging applications of production efficiency estimation in healthcare, few attention, in our opinion, has been devoted in many empirical papers to the correct identification between the shape of the frontier (and thus specification of the production function) and determinants of the distance of individual DMUs from the frontier (and thus determinants of inefficiency).

Two approaches are usually chosen. The first one modifies the specification of the frontier by introducing the determinants of inefficiency into the shape itself, thus estimating a unique function. In doing so, this approach incurs the problem of not being able to uniquely identify the effect of the determinants. The second approach uses a two-stage technique, by regressing the efficiency scores obtained through a first estimation stage on contextual variables, or through truncated regressions (see *e.g.* Simar and Wilson, 2007) or through classical ordinary least squares (Banker and Natarajan, 2008). The main problem with the latter approach lies on a well-known separability condition that may not be satisfied in many real-world situations, namely, the assumption that these factors have no influence on the shape of the frontier function but only influence the probability of being more or less efficient.

Karakaplan and Kutlu (2017); Karakaplan (2022) bypass these limitations within an endogenous panel SFA estimation framework[2], which allows for instrumenting – in a single stage – separately the inefficiency and the shape of the frontier.

Equation (1), therefore, represents, in our opinion, an optimal application tool for testing the treatment DID coefficient directly on inefficiency:

$$\begin{cases} y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it} - u_{it}, & i = 1, \ldots, n\,; 1.\ t = 1, \ldots, T \\ u_{it} = h(\boldsymbol{Z}'_{uit}\boldsymbol{\phi}_{\boldsymbol{u}})u_i^* \end{cases} \tag{1}$$

where $y_{it} \in \mathbb{R}_+$ is the output of unit i at time t, $\boldsymbol{x}_{it} \in \mathbb{R}_+^p$ is the vector of inputs, v_{it} is the symmetric two-sided error representing random effects and $u_{it} > 0$ is the one-sided error term which represents technical inefficiency[3]. In this framework, $u_{it} > 0$ is, therefore, a one-sided error term capturing the inefficiency depending by a vector of exogenous variables $\boldsymbol{Z}'_{uit}$ (in this case, the treated dummy, the year of treatment and the DID term) and by a producer-specific random component u_i^*.

3.2. Counterfactual, matching and alignment methods

The counterfactual approach is nowadays a golden standard for evaluating the effects of public policies, in order to check for their effectiveness in changing the behaviour or the conditions of a certain target population in the desired direction, *i.e.* to determine to what extent the intervention – rather than other factors – contributed to the achievement of a certain outcome.

The matching approach may solve the problem of a correct identification of the causal effect, by constructing statistical twins, *i.e.* by finding in the group of untreated subjects those units that appear most similar to the treated units in all relevant pre-treatment characteristics such that the causal effect can be inferred from the difference in the mean of the two outcomes. This empirical

[2]The endogeneity implications do not affect our application and are therefore omitted in the specification of the more general form proposed by the authors.

[3]The two-sided residual term is usually assumed to be normally distributed: $v \sim N(0, \sigma_v^2)$ while u is distributed as a half-normal and is always positive: $u \sim N^+(0, \sigma_u^2)$. The classical model also assumes that v and u are each identically independently distributed (*iid*) and the covariates in the model.

approach has become an increasingly popular method in many fields, including health studies (Rubin, 1997; Christakis and Iwashyna, 2003; Büchner et al., 2016; Peña-Longobardo et al., 2021).
However, even if the assumptions and information requirements are clear, there is no unanimous consensus on how to implement such a procedure and, in particular, on which estimators to use to measure the fit of the matching procedure, so that it is sufficiently robust (Rambachan and Roth, 2019). To be useful and reliable for empirical purposes, in fact, the matching procedure must be based on a correctly specified propensity score that asymptotically balances the observed covariates and asymptotically removes the bias conditioned by these covariates (Rosenbaum, 2002), *i.e.* the treatment and control groups exhibit – after matching – the same joint distribution of the observed covariates.

Unfortunately, in general, the "correct" propensity score model is unknown. No full consensus exists among scholars and practitioners on the most appropriate way to deal with this problem. Rosenbaum and Rubin (1984) originally suggested that a practical approach may be to iteratively check the specification of the propensity score model. They provided an algorithm for estimating a propensity score that involves iteratively checking if matching on the estimated propensity score produces balance, estimating many candidate models and sequentially learn from one specification to the next. Genetic Matching (Diamond and Sekhon, 2013) may be a useful generalisation of the propensity score method, given that it eliminates the need to manually "*iteratively check the propensity score by using a search algorithm to iteratively check and improve covariate balance*" (Diamond and Sekhon, 2013). As underlined by Lindlbauer et al. (2016) "*a major advantage of this method is that not only the variables' means in the intervention and the control group are aligned, but that both groups have the same joint distribution of observed covariates after matching*".
Other methods such as the Nearest Neighbour (Thoemmes and Kim, 2011), the Full matching (Hansen, 2004) and the Optimal Pair (Hansen and Klopfer, 2006; Austin, 2014) can be used to test the robustness of the genetic matching and to verify the optimality of the obtained balance.

3.3. Data and estimation strategy

In our empirical exercise, data provided by the Italian Ministry of Health (specifically the Department of Healthcare) related to hospitals' discharge and resources has been used. Our starting dataset consists of a balanced panel of 547 hospitals operating in the Italian NHS in the period 2003-2010. Using a balanced sample of hospitals allows us to control for possible confounding factors due to hospital mergers, closures or reorganisations. In our sample, hospitals are distinguished according to two main categories. The first group includes public hospitals directly managed by LHAs (Hospital Units - Ospedali a Gestione Diretta or Presidi Ospedalieri), Hospital Trusts (Aziende Ospedaliere) and other public hospitals. The second group consists of accredited private for-profit hospitals (Case di Cura Accreditate). In fact, differentiating hospitals according to the ownership structure might be crucial for the efficiency estimation (Barbetta et al., 2007; Daidone and D'Amico, 2009).

Data referred to the hospitals' capacity level and workforce are commonly considered as input variables (see *e.g.* Rezaei et al., 2016; Azreena et al., 2018; Nepomuceno et al., 2020). In particular, in this paper the number of hospital beds proxies the capital factor, while the number of full-time medical and non-medical staffs (*i.e.* physicians, nurses and others) approximates the labour factor. Moving to the output variable, with no direct information on quality measures (*i.e.* readmission rates, risk adjusted discharge mortality) and in order to allow for technology differences in the provision of hospital services, drawing from Cavalieri et al. (2020), we use casemix-weighted discharge considering acute patients only (see also Daidone and D'Amico, 2009). Specifically, our output variable becomes the monetary revenue for all discharged acute patients, using, for each discharge, the national Diagnosis related group (DRG) tariff for inter-regional mobility[4].

4. Results

Our empirical exercise has been carried out through two steps. First, we identify the counterfactual setting (see section 4.1) and, second, we estimate the technical efficiency differentials specifically due to the RPs policy (see section 4.2).

4.1. Matching and alignment

This first step aims at mimicking the construction of the randomised study that facilitates direct comparison between treated and untreated groups.

[4]For a comprehensive motivation of such empirical choice, see Cavalieri et al. (2020).

For this purpose, propensity score may be used to construct matched hospitals that can be compared directly.
As stated in section 3.2, it is crucial to verify the optimality of the propensity score both in terms of the explanatory variables, but also by testing the results of different matching models. Table A.1 and Table A.2 show the balance on covariates after matching (in terms of standardised mean difference) by matching method (Nearest Neighbour, Full matching, Optimal Pair and Genetic), as well as the ones for the unmatched set of units. It is clear that the benchmark (the non-matched case) is in any case outperformed by any matching method, with the genetic method being the most effective one. Genetic matching algorithm has then been chosen to construct the corresponding group of untreated hospitals.
Figure 1 reports, in a graphical format, the empirical evidence on the standardised differences in the means of our variables between the treated group and the untreated one, showing a strong improvement over the total unmatched group.

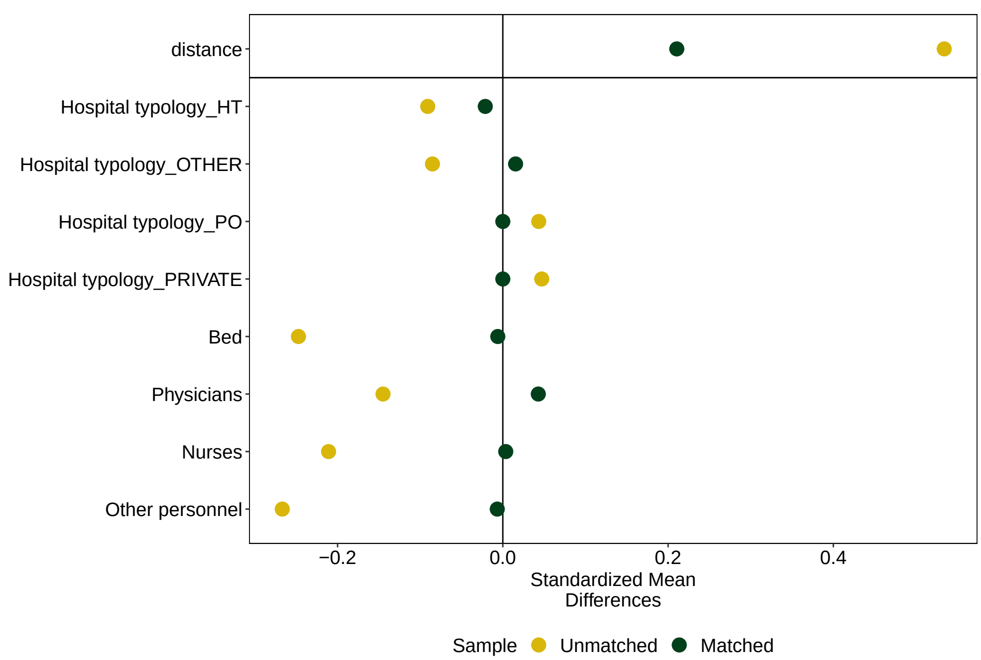

Figure 1: Covariate balance before and after adjusting, Genetic matching method

Reducing the effects of confounding factors in observational studies is thus the ultimate goal of matching methods so as to allow ceteris paribus comparison between two groups. It is, therefore, necessary to assess the distribution of the probability of treatment assignment, conditional on the observed baseline characteristics, through propensity score. Figure 2 provides clear evidence of the propensity score distributions by group after the matching phase. First of all, there are no treated units that are not matched. Secondly, the group of treated and matched untreated appears very similar in terms of distribution, whereas the hospitals that are discarded from matching show a very different and almost complementary propensity score distribution to the treated.

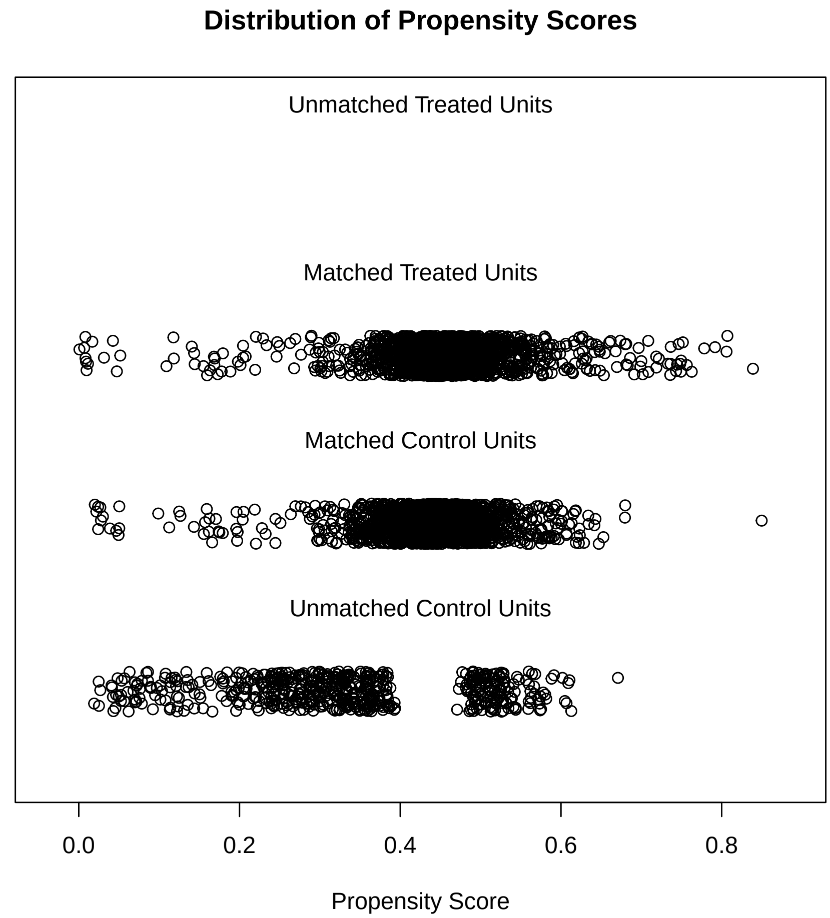

Figure 2: Propensity score distributions by group, Genetic matching method

Another check concerning the balance between the treated group and the matched group of untreated can be carried out by calculating the Student t-test between the two populations before and after matching. Table 1 clearly shows the alignment in average terms of the non-dichotomous variables used; for each of them – after matching – a very good alignment is evident.

Variable		Statistic	p-value	Mean in group 0	Mean in group 1
Bed	Before	7.454	0.000	225.79	172.63
	After	0.020	0.984	172.76	172.63
Physicians	Before	4.664	0.000	131.99	109.33
	After	-1.391	0.164	102.78	109.33
Nurses	Before	6.525	0.000	285.87	214.50
	After	-0.170	0.865	212.76	214.50
Other personnel	Before	8.091	0.000	263.34	178.34
	After	0.376	0.707	182.02	178.34

Table 1: Student t-test before and after matching (0=untreated, 1= treated)

After verifying alignment and comparability between the two groups of hospitals by mimicking a randomised controlled trial, it is, therefore, possible to check for the effect of the RPs policy in a differential way between the 247 treated hospitals (for the years 2003-2010) and the 247 untreated matched ones.

4.2. Estimation of the impact of RPs on technical efficiency of hospitals

To assess the technical efficiency of hospitals, as a first step, the estimation of the relationship between the inputs and the output has been carried out, following a Cobb-Douglas log-log specification. Table 2 reports the maximum likelihood estimates for the parameters of the time-varying decay SFA model (baseline model, Battese and Coelli, 1992). In this specification, the time-varying decay parameter η allows to check whether the level of inefficiency increases relative to the base level. Some findings emerge: *(i)* the degree of inefficiency ($\eta < 0$) increases over time in global terms and compared to the year 2007; *(ii)* all the input are statistically significant and the sum of the elasticities is slightly below unity showing slightly decreasing returns to scale; *(iii)* the value of the γ parameter (the ratio of the variance of inefficiency to total variance), equal to 0.98, shows how the contribution of the random term to the estimate is very low, evidencing a very good explanatory contribution of the inputs in the estimate of the output.

	Coefficient	Std. err.	z	P>z	[95% conf.	interval]
Bed	0.6675	0.0155	42.970	0.000	0.6370	0.6979
Physicians	0.0418	0.0115	3.650	0.000	0.0194	0.0643
Nurses	0.1139	0.0130	8.790	0.000	0.0885	0.1393
Other personnel	0.1284	0.0128	10.030	0.000	0.1033	0.1534
Constant	4.0155	0.0489	82.060	0.000	3.9196	4.1114
μ	-4.4452	4.8596	-0.910	0.360	-13.9699	5.0795
η	-0.0506	0.0032	-15.940	0.000	-0.0568	-0.0444
σ^2	1.9992	1.8215			0.3352	11.9231
γ	0.9833	0.0151			0.9060	0.9972
σ_u^2	1.9659	1.8214			-1.6040	5.5357
σ_v^2	0.0333	0.0008			0.0317	0.0349

Table 2: Baseline model: time-varying panel Stochastic Frontier

Once the baseline model has been verified, it is necessary to go further by checking – on the inefficiency term and not on the frontier specification – if the treatment has generated differences between the group of treated and untreated hospitals. As discussed in section 3.1, we use the Karakaplan and Kutlu (2017); Karakaplan (2022) framework (see equation (1)), for this purpose.

	Coefficient	Std. err.	z	P>z	[95% conf. interval]	
			Frontier estimation			
Bed	0.6577	0.0151	43.430	0.000	0.6280	0.6874
Physicians	0.0239	0.0116	2.060	0.039	0.0012	0.0466
Nurses	0.1301	0.0132	9.890	0.000	0.1043	0.1559
Other personnel	0.1318	0.0132	9.960	0.000	0.1059	0.1578
Constant	4.0877	0.0510	80.100	0.000	3.9876	4.1877
			Inefficiency			
Treated (0/1)	-0.3317	0.1276	-2.600	0.009	-0.5817	-0.0817
Year of treatment	0.1730	0.0336	5.150	0.000	0.1072	0.2388
DID	0.4125	0.0511	8.070	0.000	0.3123	0.5127
Constant	-1.5284	0.0952	-16.050	0.000	-1.7150	-1.3418

Table 3: Endogenous panel stochastic cost frontier model

Table 3 shows a DID term positive and significant. It is, therefore, verified, with a proper counterfactual setting, that the treatment has led to an increase in technical inefficiency in hospitals under RPs. This is an interesting result and shows that the policy has been effective for the short-term objective of achieving cost savings, but it has not supported the improvement of productive efficiency for hospitals.

Exploiting the proposed empirical setting, we can go further and test whether the impact of the policy was increasing over time and whether it affected all the different types of hospitals in the same way.

	Coefficient	Std. err.	z	P>z	[95% conf. interval]	
			Frontier estimation			
Bed	0.6526	0.0151	43.270	0.000	0.6230	0.6822
Physicians	0.0263	0.0115	2.290	0.022	0.0038	0.0488
Nurses	0.1307	0.0130	10.030	0.000	0.1052	0.1563
Other personnel	0.1308	0.0131	10.020	0.000	0.1052	0.1564
Constant	4.1033	0.0507	80.940	0.000	4.0039	4.2026
			Inefficiency			
Treated (0/1)	-0.3338	0.1275	-2.620	0.009	-0.5836	-0.0840
Year of treatment	0.1750	0.0332	5.280	0.000	0.1100	0.2400
Constant	-1.5273	0.0951	-16.050	0.000	-1.7137	-1.3408
*Coefficient DID (Treated * Year)*						
Year=2007	0.2634	0.0649	4.060	0.000	0.1362	0.3906
Year=2008	0.2104	0.0658	3.200	0.001	0.0815	0.3393
Year=2009	0.5097	0.0627	8.130	0.000	0.3868	0.6326
Year=2010	0.6531	0.0610	10.700	0.000	0.5335	0.7727

Table 4: Endogenous panel stochastic cost frontier model, treated * year.

In Table 4, we can observe the interaction term between DID and the dummy variables related to the post treatment years[5]. As it might have been expected, the impact of the policy did not propagate immediately over time and simultaneously to all hospitals. It is essentially from the year 2009 (see Figure 3) onward that financial cuts first affect inputs and then, consequently, affect the level of outputs, which is compressed more than proportionally relative to the inputs.

[5] As expected a disaggregated specification of the determinants on the inefficiency term has no significant impact on both the estimates of the production frontier and the distribution of efficiency estimates (see Figure A.1, too).

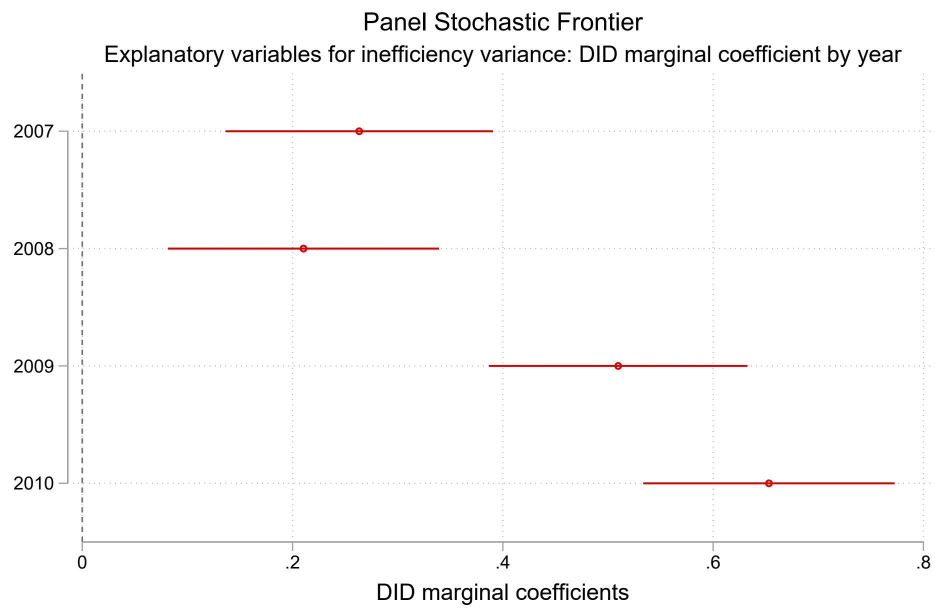

Figure 3: Did marginal coefficients by year (baseline: DID=0, year=2010)

An impact, therefore, which was not very short-term and which led to an increase in inefficiency of the treated hospitals, which was not homogeneous across the different types of hospitals. Table 5 shows the counterfactual frontier model in which the DID term has been made to interact with the post-treatment years and with a dummy relating to the public-private[6] hospitals typology.

[6]Public=1 if Hospital Units, Hospital Trusts and Other Hospitals; Public=0 if Private For Profit Hospitals.

	Coefficient	Std. err.	z	P>z	[95% conf. interval]	
		Frontier estimation				
Bed	0.6895	0.0156	44.080	0.000	0.6588	0.7201
Physicians	0.0459	0.0115	4.010	0.000	0.0235	0.0684
Nurses	0.0943	0.0132	7.130	0.000	0.0684	0.1202
Other personnel	0.1239	0.0127	9.770	0.000	0.0991	0.1488
Constant	4.0233	0.0476	84.600	0.000	3.9301	4.1165
		Inefficiency				
Treated (0/1)	-0.1511	0.1314	-1.150	0.250	-0.4086	0.1063
Year of treatment	0.1807	0.0344	5.250	0.000	0.1133	0.2481
Constant	-2.4608	0.1322	-18.610	0.000	-2.7199	-2.2017
*Coefficient DID (Treated * Public dummy * Year)*						
Pub=0, Year=2007	0.1206	0.1106	1.090	0.276	-0.0963	0.3374
Pub=0, Year=2008	0.1646	0.1087	1.510	0.130	-0.0484	0.3777
Pub=0, Year=2009	0.2634	0.1057	2.490	0.013	0.0561	0.4706
Pub=0, Year=2010	0.4936	0.0994	4.970	0.000	0.2988	0.6884
Pub=1, Year=2007	1.6798	0.1572	10.690	0.000	1.3717	1.9879
Pub=1, Year=2008	1.5723	0.1592	9.880	0.000	1.2603	1.8842
Pub=1, Year=2009	1.9891	0.1564	12.720	0.000	1.6826	2.2957
Pub=1, Year=2010	2.0894	0.1552	13.470	0.000	1.7853	2.3935

Table 5: Endogenous panel stochastic cost frontier model, treated * public dummy * year.

It is immediately noticeable that public hospitals are the ones that were most affected by the cost containment policy, even if the private ones, still in 2010, show a higher level of inefficiency (see Figure 4).

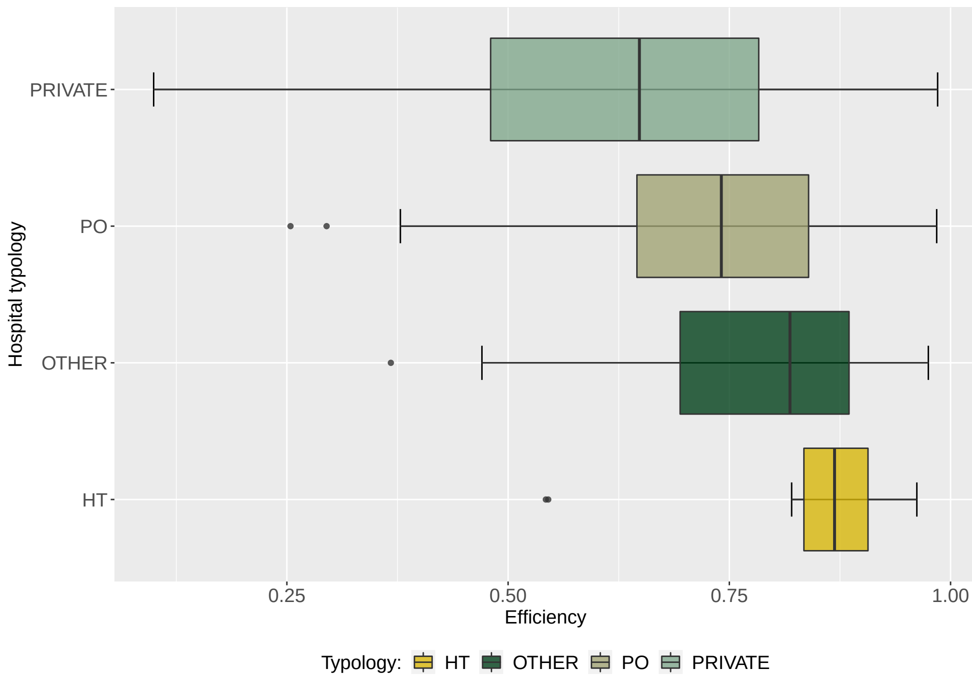

Figure 4: Estimated efficiency by hospital typology, year 2010

5. Final remarks

Austerity policies have been widely used in several countries as a tool for remedying the financial crisis of their public sectors. The Italian public healthcare sector has also been interested by such policies, with mixed results. While it appears that austerity policies have been quite effective in bringing public spending under control (*e.g.* Atella et al., 2019; Bordignon et al., 2020), at the same time they have had unintended effects on health outcomes and on the whole quality of public healthcare systems (*e.g.* Arcà et al., 2020; Depalo, 2019). One of the key elements of austerity policies was related to the idea that there was a large inefficiency in the public sector that needed to be recovered, such that it was virtually possible to introduce budget-cutting policies without affecting the level of healthcare services. However, to the best of our knowledge, the effects of austerity policies on public sector efficiency had not yet been sufficiently investigated.
In this paper we have considered the case of Italy, which is an interesting example for assessing the effects of austerity policies in the healthcare sector.

The regional structure of the NHS has in fact created over time an asymmetry between the responsibility of collecting funds (mainly) from the central government, and the expenditure managed by the regions, which has led to large regional budget deficits that have required specific recovery plans, implemented since 2007. The primary objective of the RPs was to restore the economic and financial sustainability of regional healthcare systems, while preserving the levels of care based specifically on the idea that there were large margins of inefficiency that the regions could use to buffer the budget cuts.

From a methodological point of view, the counterfactual framework together with the methodological Karakaplan (2022) approach allowed us to test the difference in efficiency due to the RPs directly on the inefficiency term by keeping the shape of the production frontier fixed. Two main results emerge: the first one is that the purely financial consolidation policy increases the production inefficiency of hospitals by impacting not only on the input side but also on the output side, by decreasing it more than proportionally; the second issue is that this policy has affected the public sector more than the private sector.

References

Aimone Gigio, L., Alampi, D., Camussi, S., Ciaccio, G., Guaitini, P., Lozzi, M., Mancini, A.L., Panicara, E., Paolicelli, M., 2018. La sanità in Italia: il difficile equilibrio tra vincoli di bilancio e qualità dei servizi nelle regioni in piano di rientro. Banca d'Italia.

Akhter, N., Bambra, C., Mattheys, K., Warren, J., Kasim, A., 2018. Inequalities in mental health and well-being in a time of austerity: Follow-up findings from the stockton-on-tees cohort study. SSM - Population Health 6, 75–84.

Arcà, E., Principe, F., Van Doorslaer, E., 2020. Death by austerity? the impact of cost containment on avoidable mortality in italy. Health Economics 29, 1500 – 1516.

Atella, V., Belotti, F., Bojke, C., Castelli, A., Grašič, K., Kopinska, J., Piano Mortari, A., Street, A., 2019. How health policy shapes healthcare sector productivity? Evidence from Italy and UK. Health Policy 123, 27 – 36.

Austin, P.C., 2014. A comparison of 12 algorithms for matching on the propensity score. Statistics in Medicine 33, 1057–1069.

Azreena, E., Juni, M.H., Rosliza, A., 2018. A systematic review of hospital inputs and outputs in measuring technical efficiency using data envelopment analysis. International Journal of Public Health and Clinical Sciences 5, 17–35.

Banker, R., Natarajan, R., 2008. Evaluating contextual variables affecting productivity using data envelopment analysis. Operations research 56, 48–58.

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

Barra, C., Lagravinese, R., Zotti, R., 2022. Exploring hospital efficiency within and between italian regions: new empirical evidence. Journal of Productivity Analysis 57, 269–284.

Battese, G., Coelli, T., 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in india. Journal of Productivity Analysis 3, 153–169.

Boffardi, R., 2022. How efficient is the italian health system? evidence on the role of political-institutional dynamics. Socio-Economic Planning Sciences , 101388.

Bordignon, M., Coretti, S., Piacenza, M., Turati, G., 2020. Hardening subnational budget constraints via administrative subordination: The italian experience of recovery plans in regional health services. Health Economics 29, 1378–1399.

Büchner, V.A., Hinz, V., Schreyögg, J., 2016. Health systems: changes in hospital efficiency and profitability. Health care management science 19, 130–143.

Cavalieri, M., Di Caro, P., Guccio, C., Lisi, D., 2020. Does neighbours' grass matter? testing spatial dependent heterogeneity in technical efficiency of italian hospitals. Social Science & Medicine 265, 113506.

Cavalieri, M., Guccio, C., Lisi, D., Pignataro, G., 2018. Does the extent of per case payment system affect hospital efficiency? evidence from the italian nhs. Public finance review 46, 117–149.

Christakis, N.A., Iwashyna, T.J., 2003. The health impact of health care on families: a matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses. Social science & medicine 57, 465–475.

Colombi, R., Martini, G., Vittadini, G., 2017. Determinants of transient and persistent hospital efficiency: The case of Italy. Health economics 26, 5–22.

Daidone, S., D'Amico, F., 2009. Technical efficiency, specialization and ownership form: Evidences from a pooling of italian hospitals. Journal of Productivity Analysis 32, 203 – 216.

Depalo, D., 2019. The side effects on health of a recovery plan in italy: A nonparametric bounding approach. Regional Science and Urban Economics 78.

Diamond, A., Sekhon, J.S., 2013. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. Review of Economics and Statistics 95, 932–945.

General State Accounting Office, 2009. Relazione Unificata sull'Economia e la Finanza Pubblica per il 2009. Technical Report. Ministry of Economy and Finance.

Giancotti, M., Sulku, S.N., Pipitone, V., Mauro, M., 2020. Do Recovery Plans Improve Public Hospitals Efficiency and Productivity? Evidence from Italy. International Review of Business Research Papers 16.

Hansen, B.B., 2004. Full matching in an observational study of coaching for the sat. Journal of the American Statistical Association 99, 609–618.

Hansen, B.B., Klopfer, S.O., 2006. Optimal full matching and related designs via network flows. Journal of Computational and Graphical Statistics 15, 609–627.

Hollingsworth, B., 2008. The measurement of efficiency and productivity of health care delivery. Health economics 17, 1107–1128.

Hollingsworth, B., 2016. Health system efficiency: measurement and policy, in: Health system efficiency: How to make measurement matter for policy and management. European observatory on health systems and policies, pp. 99–114.

Italian Ministry of Health, 2014. L'erogazione dei LEA nelle Regioni in Piano di rientro Trend 2007-2012. Technical Report. SiVeAS office, Italian Ministry of Health.

Jacobs, R., 2001. Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. Health care management science 4, 103–115.

Karakaplan, M., Kutlu, L., 2017. Endogeneity in panel stochastic frontier models: an application to the japanese cotton spinning industry. Applied Economics 49:59, 5935–5939.

Karakaplan, M.U., 2022. Panel stochastic frontier models with endogeneity. The Stata Journal 22, 643–663.

Kiadaliri, A.A., Jafari, M., Gerdtham, U.G., 2013. Frontier-based techniques in measuring hospital efficiency in iran: a systematic review and meta-regression analysis. BMC health services research 13, 1–11.

Lindlbauer, I., Schreyögg, J., Winter, V., 2016. Changes in technical efficiency after quality management certification: A dea approach using difference-in-difference estimation with genetic matching in the hospital industry. European Journal of Operational Research 250, 1026–1036.

Mahdiyan, S., Dehghani, A., Tafti, A.D., Pakdaman, M., Askari, R., 2019. Hospitals' efficiency in iran: A systematic review and meta-analysis. Journal of Education and Health Promotion 8.

Maresso, A., Mladovsky, P., Thomson, S., Sagan, A., Karanikolos, M., Richardson, E., Cylus, J., Evetovits, T., Jowett, M., Figueras, J., Kluge, H., 2015. Economic crisis, health systems and health in Europe: Impact and implications for policy. European Observatory on Health Systems and Policies.

Mateus, C., Joaquim, I., Nunes, C., 2015. Measuring hospital efficiency—comparing four european countries. The European Journal of Public Health 25, 52–58.

Mattheys, K., Bambra, C., Warren, J., Kasim, A., Akhter, N., 2016. Inequalities in mental health and well-being in a time of austerity: Baseline findings from the stockton-on-tees cohort study. SSM - Population Health 2, 350–359.

Nepomuceno, T.C., Silva, W.M., Nepomuceno, K.T., Barros, I.K., 2020. A dea-based complexity of needs approach for hospital beds evacuation during the covid-19 outbreak. Journal of healthcare engineering 2020.

OASI, 2021. OASI Report 2021. Technical Report 3. Observatory On Healthcare Organizations and Policies In Italy.

Peña-Longobardo, L., Rodríguez-Sánchez, B., Oliva-Moreno, J., 2021. The impact of widowhood on wellbeing, health, and care use: A longitudinal analysis across europe. Economics & Human Biology 43, 101049.

Polyzos, N., 2012. A three-year performance evaluation of the nhs hospitals in greece. Hippokratia 16 4, 350–5.

Rambachan, A., Roth, J., 2019. An honest approach to parallel trends. Unpublished manuscript, Harvard University .

Reeves, A., McKee, M., Stuckler, D., 2015. The attack on universal health coverage in Europe: recession, austerity and unmet needs. European Journal of Public Health 25, 364–365.

Rezaei, S., Zandian, H., Baniasadi, A., Moghadam, T.Z., Delavari, S., Delavari, S., 2016. Measuring the efficiency of a hospital based on the econometric stochastic frontier analysis (sfa) method. Electronic physician 8, 2025.

Rosenbaum, P.R., 2002. Observational Studies. Springer-Verlag New York: 2nd edition.

Rosenbaum, P.R., Rubin, D.B., 1984. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American statistical Association 79, 516–524.

Rubin, D.B., 1997. Estimating causal effects from large data sets using propensity scores. Annals of internal medicine 127, 757–763.

Samut, P.K., Cafrı, R., 2016. Analysis of the efficiency determinants of health systems in oecd countries by dea and panel tobit. Social Indicators Research 129, 113–132.

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

Stuckler, D., Reeves, A., Loopstra, R., Karanikolos, M., McKee, M., 2017. Austerity and health: the impact in the UK and Europe. European Journal of Public Health 27, 18–21.

Thoemmes, F.J., Kim, E.S., 2011. A systematic review of propensity score methods in the social sciences. Multivariate Behavioral Research 46, 90–118.

Tiemann, O., Schreyögg, J., Busse, R., 2012. Hospital ownership and efficiency: a review of studies with particular focus on germany. Health policy 104, 163–171.

Wei, Y., Yu, H., Geng, J., Wu, B., Guo, Z., He, L., Chen, Y., 2018. Hospital efficiency and utilization of high-technology medical equipment: a panel data analysis. Health policy and technology 7, 65–72.

Wenzl, M., Naci, H., Mossialos, E., 2017. Health policy in times of austerity—a conceptual framework for evaluating effects of policy on efficiency and equity illustrated with examples from europe since 2008. Health Policy 121, 947–954.

Worthington, A.C., 2004. Frontier efficiency measurement in health care: A review of empirical techniques and selected applications. Medical Care Research and Review 61, 135 − 170.

Xenos, P., Yfantopoulos, J., Nektarios, M., Polyzos, N., Tinios, P., Constantopoulos, A., 2017. Efficiency and productivity assessment of public hospitals in greece during the crisis period 2009–2012. Cost Effectiveness and Resource Allocation 15.

Appendix A. Robustness checks for the matching algorithm

Variable	Unmatched	Nearest Neighbour	Full Matching	Optimal Pair	Genetic
HT hospitals	-0.0196	0.0051	0.0031	0.003	-0.003
Other public hospitals	-0.0252	0.0127	0.0279	0.0112	0.003
PO hospitals	0.0215	0.0005	-0.0301	0.0091	0
Private hospitals	0.0233	-0.0183	-0.0008	-0.0234	0
Bed	-0.2473	0.0326	0.018	0.0216	-0.0006
Physicians	-0.145	0.0491	0.0101	0.0407	0.0419
Nurses	-0.2109	0.0326	0.0146	0.0284	0.0051
Other personnel	-0.2669	0.0235	0.0003	0.0233	-0.0115

Table A.1: Balance on covariates after matching, standardised mean difference by matching method

	Control	Treated
Unmatched	2560	1968
Nearest Neighbour	1968	1968
Full Matching	696.52	1968
Optimal Pair	1968	1968
Genetic	1968	1968

Table A.2: Effective sample sizes by matching method

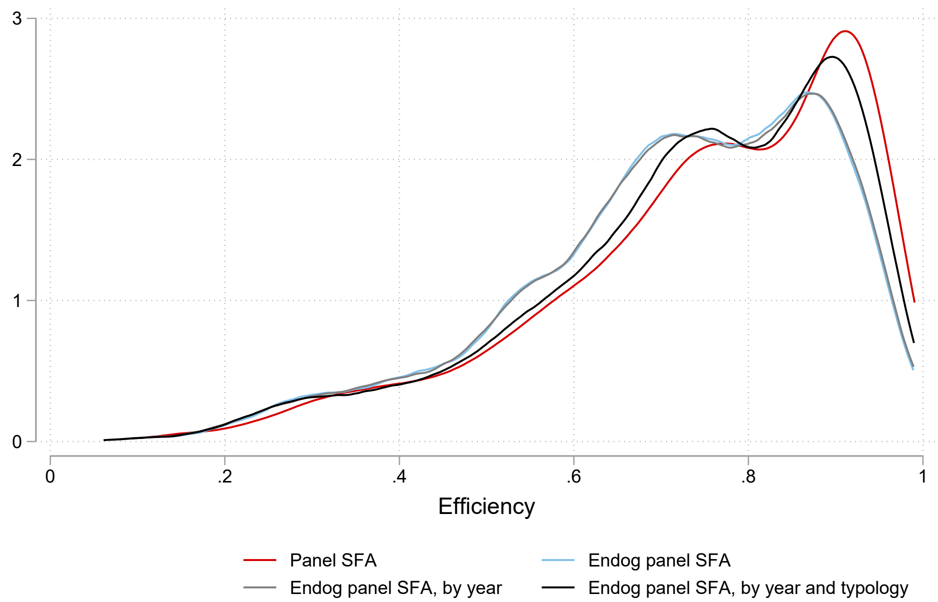

Figure A.1: Kernel distribution relative to estimated efficiency by model