Received: 18 November 2016 | Revised: 13 June 2017 | Accepted: 21 June 2017
DOI: 10.1002/hec.3557

SUPPLEMENT PAPER

WILEY Health Economics

Determinants of transient and persistent hospital efficiency: The case of Italy

Roberto Colombi[1] | Gianmaria Martini[1] | Giorgio Vittadini[2]

[1]Department of Management, Information and Production Engineering, University of Bergamo, Bergamo, Italy
[2]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy

Correspondence
Gianmaria Martini, Department of Management, Information and Production Engineering, University of Bergamo, Bergamo, Italy.
Email: gianmaria.martini@unibg.it

JEL Classification: I11; I18; D24

Summary

In this paper, we extend the 4-random-component closed skew-normal stochastic frontier model by including exogenous determinants of hospital persistent (long-run) and transient (short-run) inefficiency, separated from unobserved heterogeneity. We apply this new model to a dataset composed by 133 Italian hospitals during the period 2008–2013. We show that average total inefficiency is about 23%, higher than previous estimates; hence, a model where the different types of inefficiency and hospital unobserved characteristics are not confounded allows us to get less biased estimates of hospital inefficiency. Moreover, we find that transient efficiency is more important than persistent efficiency, as it accounts for 60% of the total one. Last, we find that ownership (for-profit hospitals are more transiently inefficient and less persistently inefficient than not-for-profit ones, whereas public hospitals are less transiently inefficient than not-for-profit ones), specialization (specialized hospitals are more transiently inefficient than general ones; i.e., there is evidence of scope economies in short-run efficiency), and size (large-sized hospitals are better than medium and small ones in terms of transient inefficiency) are determinants of both types of inefficiency, although we do not find any statistically significant effect of multihospital systems and teaching hospitals.

KEYWORDS

closed skew-normal stochastic frontier, determinants of in efficiency, hospital persistent and transient efficiency

1 | INTRODUCTION

Exponential growth of health care costs and forecasts of soaring future trends have driven the attention of academics, health care managers, and policy makers on measures to contain health expenditure (World Health Report, 2000).[1] Several attempts have been made to improve the efficiency of providers (Jacobs, Smith, & Street, 2006): hospital competition, incentive payment system, and so on. Consequently, many studies have measured hospital efficiency and its variation over time and after

[1]The Organisation for Economic Co-operation and Development (OECD) countries' average per-capita health spending is about $3,500 in 2013, with peaks of $8,700 in the United States, $6,300 in Switzerland, about $5,000 in Germany and the Netherlands, and about $4,100 in France. In 2003, the OECD average was $2,200 (+59% in 10 years), whereas in the United States, it was $5,750 (+52%). (See the data available at the OECD website, https://data.oecd.org/healthres/health-spending.htm) In Italy, the total health spending is 8.8% of the gross domestic product; in the United States, it is 16.4% (OECD, 2013); 10 years before, it was, 7.9% and 14.5% for Italy and the United States, respectively, a substantial increase despite the global economic crisis that has instead reduced several other types of expenditures.

reforms.[2] This implies that a robust estimation of hospital productivity and cost is essential for both policy makers and managers (Hollingsworth & Street, 2006). Policy makers may include efficiency scores in hospital payment systems, for example, by providing premiums or penalties for those organizations that improve or worsen their performance over time.[3] Managers may instead be interested in comparing their organization with others, to assess their relative position and implementing corrections for improvements.

Hospital efficiency has been estimated using two alternative methods: data envelopment analysis (DEA) and stochastic frontier (SF) approach. DEA is a nonparametric approach based an linear programming.[4] An SF model, adopted here, is a parametric approach that has the advantage of considering the impact of random shocks and possible measurement errors; an SF model allows us to obtain interesting insights regarding hospital efficiency (e.g., input elasticities) and to discriminate between unobserved heterogeneity and inefficiency.

A typical SF model splits the error term into two random components: a normally distributed noise term (taking into account for random shocks and measurement errors) and a nonnegative inefficiency term. When dealing with panel data, SF models typically disentangle hospitals' unobserved heterogeneity from inefficiency. Unobserved characteristics are elements concurring to creating different hospital types (Dormont & Milcent, 2005): They may give rise to adverse selection in the funding process (e.g., a hospital may have a better road access system and so is treating more patients than do others). Moreover, inefficiency can be divided into persistent (time-invariant) and transient (time-varying) components. The former may be due to long-run moral hazards (e.g., an obsolete equipment that is not replaced for several periods); transient inefficiencies are related to short-run moral hazards (e.g., inefficient supplier selection and suboptimal resource allocation) or to trial-and-error processes in unknown situations.

As shown by Dormont and Milcent (2005), considering both unobserved hospital characteristics, persistent and transient inefficiencies are essential to understanding the differences in average health costs and productivity among regions and states, and among hospitals operating in the same environment. Although most SF models fail to consider simultaneously unobserved heterogeneity and persistent and transient inefficiencies, Colombi, Kumbhakar, Martini, and Vittadini (2014) recently presented an SF model that includes all these components in the error term, together with the normally distributed random shock. The possibility to disentangle the two types of inefficiencies, hospital unobserved heterogeneity and random shocks, comes from previous results (Dominguez-Molina, Gonzáles-Farías, & Ramos-Quiroga, 2004; Gonzáles-Farías, Dominguez-Molina, & Gupta, 2004). These contributions show that when combining the error terms of an SF model's normally distributed random variables and normal random variables left truncated at zero (inefficiencies cannot be negative), we obtain a generalization of the normal-skew distribution denoted as closed skew-normal (CSN) distribution.[5] This has an important consequence for econometric analysis: It is possible to provide a maximum likelihood (ML) estimation method for an SF model with this specific distribution. Colombi et al. (2014) applied these results to a model with four random components in the error term (which has therefore a CSN distribution), wrote the log-likelihood for applying the ML estimation method, and showed how to compute the persistent and transient inefficiency scores. This enables us to disentangle the two types of inefficiency, unobserved heterogeneity and random shocks.

This paper is a first attempt, to the best of our knowledge, to estimate transient and persistent hospital inefficiencies, disentangling both of them from unobserved heterogeneity. We apply the Colombi et al. (2014) SF model to a dataset composed of 133 Italian hospitals located in Lombardy, the most populated (about 10 million inhabitants) and richest (its gross domestic product is 20% of the national one) Italian region. Furthermore, we also present an extension of the Colombi et al. (2014) model: a maximum log likelihood estimation method for a 4-random-component SF model that includes exogenous variables affecting inefficiency levels. This implies modeling heteroskedasticity, which may be relevant because both model parameters and inefficiency estimates are adversely affected when these features are neglected. The model is also important from the empirical perspective as it allows us to test specific policy implications, avoiding biased two-step procedures. For instance, policy makers might be interested in evaluating whether large-sized hospitals are performing better in terms of persistent inefficiencies (because they may have better infrastructures) but not in terms of transient inefficiencies. Hence, this contribution may be an attempt to answer the need recently identified by Hollingsworth and Street (2006, p. 1058), that is, to better identify the nature and form of inefficiency.

We find that overall inefficiency in the hospital regional system is higher than the estimates obtained in previous contributions. Hence, an SF model where the different types of inefficiency and hospital unobserved characteristics are not confounded gives

[2]Good surveys on hospital efficiency contributions are provided by Hollingsworth (2003), Jacobs et al. (2006), and Rosko and Mutter (2008, 2011).
[3]Hollingsworth and Street (2006) provide examples on Australia and Norway. In Lombardy, Italy, the regional government acknowledges a share based on performance indicators (related to both health outcomes and productivity) of the annual budget allocated to each hospital.
[4]DEA does not require specifying ex ante any functional form regarding the efficiency frontier (the benchmark estimated in the efficiency exercise), but it considers any distance from it as inefficiency and does not take into account any measurement error.
[5]The term *closed* means that linear transformations or sums of CSN variables generate random variables belonging, in turn, to the CSN distribution family.

less biased inefficiency scores. Moreover, we find that persistent inefficiency is more important than transient inefficiency, as it accounts for 60% of total hospital inefficiency. Last, for-profit (FP), specialized, medium-sized, and small-sized hospitals have more transient inefficiency than not-for-profit (NFP), public (P), general, and large-sized hospitals.

The paper is organized as follows. Section 2 revises previous contributions on hospital efficiency. Section 3 presents the Colombi et al. (2014) SF model and its extension. Section 4 introduces the dataset, whereas Section 5 shows our results. Section 6 concludes the paper with some policy implications.

2 | LITERATURE REVIEW

The issue of hospital efficiency has been widely investigated in the literature (Castelli et al., 2015). The number of papers on this topic has grown almost exponentially until a few years ago (Hollingsworth & Street, 2006), whereas the number of contributions has increased at a lower pace during recent years (Castelli et al., 2015).[6] Table 1 presents previous contributions that studied hospital efficiency using SF models, ordered by publication year. We analyze 37 contributions, the vast majority (25 out of 37) of which are on U.S. hospitals; there are many papers on European countries (11) and one paper on Japan. A high percentage (31 papers out of 37) provides estimates of hospital efficiency using a cost frontier. Notably, because they are related to our contribution, both previous contributions on Italian hospitals analyze a production frontier, given that data on hospitals costs are not publicly available in Italy. Only 13 out of 37 of previous studies investigate cross-sectional data, whereas the majority exploits the possibility to include unobserved heterogeneity using panel data; the average observed period is 6 years. The average estimated hospital inefficiency (when available) is about 17%.

The most important information provided by Table 1 is that related to the SF model. It describes the error components included in the analysis; by inspection, the vast majority of papers (25 out of 37) has investigated hospital efficiency by considering only a time-invariant component. Hence, estimated efficiency scores may be biased because hospital unobserved heterogeneity is confounded with persistent inefficiency. A smaller number of contributions (eight) has instead adopted an SF model with a time-varying inefficiency component; in this case, estimated scores may be affected by both unobserved heterogeneity and persistent inefficiency.[7] Two of the most recent papers (Carey et al., 2008; Farsi & Filippini, 2008) have analyzed hospital efficiency by estimating the Greene (2005b) SF model. In this way, the transient estimated efficiency scores are not affected by unobserved heterogeneity, but all the inefficiency is transient whereas persistent inefficiency is confounded with unobserved heterogeneity. Hence, by inspection, this is the first paper where hospital efficiency is investigated with four random components separating unobserved heterogeneity, transient and persistent inefficiencies, and random shocks, providing scores not affected by the omission of one of these factors.

Many papers have studied the determinants of inefficiency. The results obtained range from the importance of internal factors (e.g., ownership, multi-hospital management, and specialty hospitals) to the importance of external factors (e.g., public payment policy, unemployment rate, and hospital competition). The empirical evidence depends upon the country or region investigated, the available dataset, and the estimation method adopted. As shown by Rosko and Mutter (2011), the evidence on ownership is mixed: For instance, Rosko and Mutter (2008) find that private NFP hospitals are more efficient than private FP ones, whereas McKay and Deily (2005) obtain an opposite result. Daidone and D'Amico (2009) present some evidence for Italy, using a sample of hospitals belonging to an Italian region (Lazio), and show that P hospitals are more efficient than NFP and FP ones. Bernet, Rosko, and Valdmanis (2008) display that independent hospitals are less efficient than multi-hospital systems, pointing out the possible presence of scale economies. Carey et al. (2008) present some evidence that general hospitals are more efficient than specialized hospitals (i.e., possible economies of scope in multiproduct hospitals).[8] Last, the evidence on the relation between hospital competition and efficiency is mixed: Rosko et al. (2007) find increases in efficiency, whereas the opposite result is found by Mutter and Rosko (2008).

Because we estimate a production frontier, it is necessary to take into account that the raw number of patients admitted into a hospital does not capture the complexity of different acute treatments (Newhouse, 1994). Discharges have not the same weight: More complex surgical operations (e.g., a marrow transplant) should have more weight than other treatments (e.g., carpal tunnel decompression). Hence, we give more weight to complex treatments; that is, we adjust each acute discharge with an indicator that takes into account the amount of resources allocated to each treatment, using the diagnosis-related group weight, which takes into account the treatment complexity.

[6]Excellent reviews are provided by Hollingsworth (2003), Hollingsworth and Street (2006), Rosko and Mutter (2008, 2011), and Hollingsworth (2008).
[7]Two papers have considered first time-invariant inefficiency and then time-varying inefficiency, without tackling the problems just highlighted.
[8]Rosko and Proenca (2005) find that hospitals are more efficient the higher the share of patients covered by U.S. Medicare and Medicaid public programs and that the unemployment rate (i.e., a proxy for uncompensated care) is positively associated with efficiency.

TABLE 1 Previous contributions on hospital efficiency with a stochastic frontier (SF) model

Author(s) (year)	Country	Sample size	Period	Efficiency frontier	Estimated inefficiency (%)	SF model
Wagstaff (1989)	Spain	49	1979	Cost	28	Time invariant
Zuckerman, Hadley, and Iezzoni (1994)	United States	1,600	1987	Cost	15–19	Time invariant
Vitaliano and Toren (1994)	United States	219	1991	Cost	18	Time invariant
Wagstaff and Lopez (1996)	Spain	43	1988–1991	Cost	58	Time varying
Koop, Osiewalski, and Steel (1997)	United States	382	1987–1991	Cost	14–16	Time invariant
Chirikos (1998)	United States	186	1982–1993	Cost	15	Time invariant
Mobley (1998)	Untied States	266	1984, 1990	Cost	Not available	Time invariant
Linna (1998)	Finland	43	1988–1994	Cost	7–14	Time varying
Rosko (1999)	United States	3,262	1994	Cost	20–26	Time invariant
Rosko and Chilingerian (1999)	United States	185	1989	Cost	4–17	Time invariant
Chirikos and Sear (2000)	United States	186	1982–1993	Cost	8–18	Time invariant
Deily, McKay, and Dorner (2000)	United States	790	1986–1991	Cost	15	Time invariant
Frech and Mobley (2000)	United States	378	1983–1991	Cost	20	Time invariant
Grosskopf, Margaritis, and Valdmanis (2001)	United States	792	1994	Production	6	Time invariant
Li and Rosenman (2001)	United States	90	1988–1993	Cost	35	Time invariant
Rosko (2001a)	United States	1,631	1990–1996	Cost	15	Time varying
Rosko (2001b)	United States	1,966	1997	Cost	13	Time invariant
Folland and Hofler (2001)	United States	791	1985	Cost	20	Time invariant
Jacobs (2001)	United Kingdom	232	1995–1996	Cost	12–17	Time invariant
McKay, Deily, and Dorner (2002)	United States	4,075	1986	Cost	14–17	Time invariant
Carey (2003)	United States	1,209	1998	Cost	18	Time invariant
Sari (2003)	United States	125	1990–1997	Cost	20–38	Time invariant or varying
Brown (2003)	United States	613	1992–1996	Production	Not available	Time varying
Street (2003)	United Kingdom	226	1999	Cost	10–13	Time invariant
Rosko (2004)	United States	616	1990–1999	Cost	11–14	Time varying
Rosko and Proenca (2005)	United States	1,368	1998	Cost	15	Time invariant
Deily and McKay (2006)	United States	16	1999–1901	Cost	13	Time invariant
Rosko, Proenca, Zinn, and Bazzoli (2007)	United States	1,144	2001	Cost	8	Time invariant
Barbetta, Turati, and Zago (2007)	Italy	531	1995–2000	Production	20	Time invariant
Carey, Burgess, and Young (2008)	United States	355	1998–2004	Cost	28	Time varying + unobserved heterogeneity
Herr (2008)	Germany	1,565	2001–2003	Production	17	Time invariant or varying
Farsi and Filippini (2008)	Switzerland	148	1998–2003	Cost	6–9	Time varying + unobserved heterogeneity
Daidone and D'Amico (2009)	Italy	108	2001–2005	Production	15 (public hospitals)	Time invariant
Rosko and Mutter (2010)	United States	543	1997–2004	Cost	10–16	Time varying
Herr, Schmitz, and Augurzky (2011)	Germany	541	2002–2006	Production	6	Time invariant
Besstremyannaya (2011)	Japan	617	1999–2007	Cost	5–13	Time varying
Barros, de Menezes, and Vieira (2013)	Portugal	51	1997–2008	Cost	13	Time varying

Concerning methods to estimate SF models, many contributions have been continuously developed over time since the first publications of Aigner, Lovell, and Schmidt (1977), Meeusen and van den Broeck (1977), and Battese and Corra (1977).[9] They cover both cross-sectional and panel data. The latter have initially focused on either persistent inefficiency (Battese & Coelli, 1988; Kumbhakar, 1987; Pitt & Lee, 1981; Schmidt & Sickles, 1984) or transient inefficiency (e.g., Battese & Coelli 1992; Cornwell et al., 1990; Kumbhakar, 1990; Lee & Schmidt, 1993). Later models have added other components to the error term: a firm effect treated as persistent inefficiency, a transient inefficiency component (Kumbhakar & Heshmati, 1995; Kumbhakar & Hjalmarsson, 1993, 1995), and an unobserved heterogeneity component together with a transient inefficiency term (Greene, 2005a, 2005b; Kumbhakar & Wang, 2005; Wang & Ho 2010).

The above models fail to take into account simultaneously persistent and transient inefficiencies, unobserved heterogeneity, and random shocks.[10] Colombi et al. (2014) introduced a four-random-component SF model and an ML estimation method.[11]Filippini and Greene (2016) propose to estimate the Colombi et al. (2014) model using a maximum simulated log-likelihood as an estimation method exploiting the possibility to characterize the four-random-component model as a pair of two-part disturbances in which each element of the pair has its own skew-normal distribution. This approach might be useful when the ML estimation method becomes computationally demanding, that is, for a long time interval (e.g., $T > 20$–25).

3 | A CLOSED SKEW-NORMAL SF MODEL WITH DETERMINANTS OF TRANSIENT AND PERSISTENT INEFFICIENCY

Our aim is to estimate transient and persistent hospital inefficiencies in a production frontier. Starting from the Colombi et al. (2014) SF model, we introduce a new element of analysis: the determinants of inefficiency. We consider the following hospital production frontier model:

$$y_{it} = \beta_0 + \boldsymbol{x}'_{it}\boldsymbol{\beta} + b_i - u_{it} - u_i + e_{it}, \tag{1}$$

where the index i, $i = 1, 2, \ldots, n$, denotes n hospitals in the sample and $t, t = 1, 2, \ldots, T$, T periods at which each hospital is observed. The dependent variable y_{it} is the logarithm of hospital i's number of patients treated in period t, $\mathbf{x}'_{it}$ is a row vector of p inputs transformed in hospital i's health care production process, and $\boldsymbol{\beta}$ is a column vector of p unknown parameters. The random hospital effect b_i captures unobserved heterogeneity, u_{it} is a nonnegative random variable for transient inefficiency of hospital i at period t, u_i is a nonnegative random variable for persistent inefficiency of hospital i, and e_{it} is a normal random variable representing the exogenous shock affecting hospital i's number of treated patients in period t. We assume the following:

- For $i = 1, 2, \ldots, n$, the random variables u_i, b_i and u_{it}, e_{it}, $t = 1, 2, \ldots, T$, are independent in probability. This means that, within every hospital, the random components in model (1) are independent.
- The random vectors $(b_i, u_i, u_{i1}, u_{i2}, \ldots, u_{iT}, e_{i1}, e_{i2}, \ldots, e_{iT})$, $i = 1, 2, \ldots, n$, are independent in probability; that is, the errors are independent between hospitals.
- For $i = 1, 2, \ldots, n$, u_i is a normal random variable with null expected value and variance σ_u^2, left truncated at 0, and b_i is a normal random variable with null expected value and variance σ_b^2.
- For $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, u_{it} is a normal random variable with null expected value and variance σ_{ut}^2, left truncated at 0, and e_{it} is a normal random variable that has a null expected value and variance σ_e^2.
- For $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, $\boldsymbol{x}'_{it}$s are row vectors of exogenous variables.

The deterministic component $\beta_0 + \boldsymbol{x}'_{it}\boldsymbol{\beta}$ is the production function mapping the inputs transformed in hospitals to provide acute treatments to admitted patients. The components u_{it} have expected values $\mu_{it} = \sqrt{\frac{2}{\pi}\sigma^2_{uit}}$ that depend on a set of covariates (exogenous determinants of the transient inefficiency) through the linear model:

$$\ln(\sigma^2_{uit}) = \alpha_0 + z'_{it}\boldsymbol{\gamma}, \tag{2}$$

where $\ln(\sigma^2_{uit})$ is the logarithm of the transient inefficiency variance, z'_{it} is a row vector of q exogenous determinants of transient inefficiency, and $\boldsymbol{\gamma}$ is a column vector of q unknown parameters. Moreover, the persistent inefficiency components u_i have

[9]See Coelli, Rao, O'Donnell, and Battese (2005), Greene (2009), and Kumbhakar and Lovell (2000) for surveys on these models.
[10]They have considered each of these elements individually (e.g., Pitt & Lee, 1981) or in combination with only one other element (e.g., Greene, 2005a, 2005b).
[11]A similar SF model has also been proposed by Kumbhakar, Lien, and Hardaker (2014) but using a three-step model that gives less efficient and more biased estimates. Tsionas and Kumbhakar (2014) present a Bayesian Markov chain Monte Carlo estimation method applied to the four-error-component model. However, as noted by Filippini and Greene (2016), the Bayesian approach is largely sensitive to the priors on the main estimation objects.

expected value $\mu_i = \sqrt{\frac{2}{\pi}\sigma^2_{ui}}$ that depends on exogenous determinants through the following linear model:

$$\ln(\sigma^2_{ui}) = \delta_0 + \boldsymbol{w}'_i\boldsymbol{\tau}, \tag{3}$$

where $\ln(\sigma^2_{ui})$ is the logarithm of the persistent inefficiency variance, $\boldsymbol{w}'_i$ is a row vector of q' exogenous determinants of transient inefficiency, and $\boldsymbol{\tau}$ is a column vector of q' parameters.

As shown by Colombi et al. (2014, Proposition 1), under the above assumptions, the vectors of outputs $\boldsymbol{y}_i = (y_{i1}, y_{i2} \ \dots \ , y_{iT})'$, $i = 1, 2, \ \dots \ , n$, are independent and have a CSN density. Accordingly, model (1) can be labeled as CSN-SF.

The maximization of the log-likelihood of model (1) and ML estimators are discussed by Colombi et al. (2014, Proposition 2), who also showed (Proposition 3) how to compute the efficiency scores $E[\exp(-u_i)|\boldsymbol{y}_i)]$ and $E[\exp(-u_{it})|\boldsymbol{y}_i)]$ for each hospital i and period t.

We fit model (1) with the additional Equations (2 and 3) for the determinants of efficiencies under two functional specifications: (1) Cobb–Douglas and (2) translog.[12] The equation representing the translog hospital production function is

$$\ln(y_{it}) = \beta_0 + \sum_{k=1}^{p} \beta_k \ln(x_{it}) + \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{p} \beta_{kj} \ln(x_{kit}) \ln(x_{jit}) + b_i - u_{it} - u_i + e_{it}, \tag{4}$$

where $\beta_{kj} = \beta_{jk}$. The translog production function collapses to the Cobb–Douglas production function if $\beta_{kj} = 0, j = 1, 2, \ \dots \ , p, k = 1, 2, 3, \ \dots \ , p$. One of the main assumptions of the CSN-SF model is that unobserved heterogeneity is uncorrelated with the frontier regressors. In order to have control over this assumption, we implement the Mundlak (1978) approach.[13] We add to Equation (4) the means over time of the time-varying input variables, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} \ln x_{it}$, so that we can rewrite Equation (4) as follows:

$$\ln(y_{it}) = \beta_0 + \sum_{k=1}^{p} \beta_k \ln(x_{it}) + \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{p} \beta_{kj} \ln(x_{kit}) \ln(x_{jit}) + \sum_{k=1}^{p} \delta_k \bar{x}_{ki} + b_i - u_{it} - u_i + e_{it}. \tag{5}$$

We test the joint significance of the Mundlak terms on the basis of a likelihood ratio test.

The Cobb–Douglas production function has output–input elasticities given by the first-order coefficients; that is, $\epsilon_{y,k} = \beta_k$. In the translog production function, these elasticities depend instead on the level of the inputs; that is, $\epsilon_{y,k} = \beta_k + \sum_{j=1}^{p} \beta_{kj} \ln(x_{jit})$.

In the available data, each hospital has five inputs (physicians, nurses, other workers, beds, and surgery rooms hours); regarding the possible determinants of the two inefficiency terms in (2) and (3), we investigate five factors: hospital ownership, specialty hospital, single hospital, teaching hospital, and size. Ownership is given by belonging to three possible categories: P, private NFP, and private FP. Hospital size is measured in terms of operating wards and then categorized into three classes, so that it is not strongly related to beds, avoiding multicollinearity problems.

4 | THE DATA

Our data are related to 133 acute hospitals located in Lombardy during the period 2008–2013, a panel balanced dataset covering 6 years.[14] Table 2 presents the variables included in the analysis and their description.

The variable $WEIGHTDIS_{it}$ represents the weighted annual output of hospital i in year t, using the diagnosis-related group weight. Regarding inputs, we consider three labor variables ($PHYS_{it}$, physician annual working hours; $NURSE_{it}$, nurse hours; and $OTHERS_{it}$ hours, i.e., administrative and technical staff), beds for acute discharges (BED_{it}, a proxy of capital), and the annual hours of activities in operating theaters ($HOURSURGE_{it}$). All input and output variables are expressed in logarithms.

The factors possibly affecting persistent and transient inefficiencies are hospital ownership (the dummies *PUB* for P, i.e., equal to 1 in case of P hospital, *FP* for private FP and *NFP* for private NFP hospitals), teaching status (the dummy *TEACHING* equal to 1 if there is a School of Medicine), specialization (the dummy *SPECIALIZED* equal to 1 if the hospital provides only a specific treatment) to capture for presence of economies of scope, and hospital system (the dummy variable *MULTIHOSP* is equal to 1 if the same management controls more than one hospital). Last, we consider the possible impact of hospital size on

[12]The Cobb–Douglas production function is popular and easier to estimate (less parameters involved). However, it has low flexibility because the input elasticity of substitution (i.e., the ratio between two inputs and their marginal products) is fixed.
[13]We are grateful to an anonymous referee for suggesting this change in our empirical strategy.
[14]In Italy, the organization of the health care sector is in charge of the regions, that is, local governments with heterogeneous administrated population and territorial extension. Lombardy is the largest region, with a 10 million population and an extension of 24,000 km^2; Valle d'Aosta, located on the Alps, has about a 128,000 population and an extension of 3,300 km^2.

TABLE 2 Definition of variables

Variable	Description
$WEIGHTDIS_{it}$	Hospital annual acute discharges corrected by treatment cost
$PHYS_{it}$	Annual working hours of physicians
$NURSE_{it}$	Annual working hours of nurses
$OTHERS_{it}$	Annual working hours of other workers
BED_{it}	Total beds for acute discharges
$HOURSURGE_{it}$	Annual hours of activity of surgery rooms
PUB_i	Dummy equal to 1 if the hospital is public
FP_i	Dummy equal to 1 if the hospital is private for profit
NFP_i	Dummy equal to 1 if the hospital is private not for profit
$TEACHING_i$	Teaching (non-teaching) hospital, dummy equal to 1 (0)
$SPECIALIZED_i$	Monospecialistic (general) hospital, dummy equal to 1 (0)
$MULTIHOSP_i$	Dummy equal to 1 if the hospital belongs to a multi-hospital system
$SMALL_i$	Small hospital size: dummy equal to 1 if hospital *i* has less than 12 operating wards (median of hospital ward distribution)
$MEDIUM_i$	Medium hospital size: dummy equal to 1 if hospital *i* has between 12 and 21 operating wards (third quartile of hospital ward distribution)
$LARGE_i$	Large hospital size: dummy equal to 1 if hospital *i* has more than 21 operating wards

TABLE 3 Descriptive statistics on output, input, and exogenous factor variables

Variables	Mean	Standard deviation	Minimum	Maximum
WEIGHTDIS (number)	20,491.8	10,338.2	378.2	58,645.3
PHYS (hr)	206,535	224,763.8	2,129	1,807,788
NURSE (hr)	375,769	389,080.4	13,615	2,574,852
OTHERS (hr)	278,143	338,417	6,588	2,499,123
BED (number)	241	205.9	15	1,223
HOURSURG (hr)	7,491	8,689.1	0	64,993
PUB (dummy)	60.90%			
FP (dummy)	27.07%			
NFP (dummy)	12.03%			
TEACHING (dummy)	15.04%			
SPECIALIZED (dummy)	5.26%			
MULTIHOSP (dummy)	52.63%			
SMALL (dummy)	22.55%			
MEDIUM (dummy)	27.07%			
LARGE (dummy)	50.38%			

inefficiency levels: We divide the hospitals into three groups, according to the operating wards in our sample. The first group identifies small hospitals: A hospital has a small size (dummy variable *SMALL* equal to 1) if it has a number of operating wards lower than the sample median (equal to 12). In the second group, there are medium-sized hospitals (dummy variable *MEDIUM* equal to 1) belonging to the third quartile of the ward sample distribution (21 wards), and the third group is for large-dimension hospitals (dummy variable *LARGE* equal to 1) with a number of operating wards greater than the third quartile.

Table 3 shows the descriptive statistics of output, inputs, and determinants of inefficiency variables. Regarding the latter, about 61% of the hospitals in the sample are P, 27% are private FP, and only 12% are private NFP. Teaching hospitals are only 15% of the sample total, specialized hospitals only 5%, whereas almost 53% belong to a hospital system. Nurses and other staff workers have an annual hour workload that is higher than that of physicians. The average hospital has 241 beds and is engaged in surgical activities for about 7,500 annual hours. All inputs are decreasing over the period, as there is a trend for savings in the regional health care sector.

5 | EMPIRICAL RESULTS

In this section, we report the empirical evidence obtained by applying the SF model (Equation 1) to our dataset. First, we present the econometric results regarding the estimated production function and the determinants of both inefficiencies. Then we analyze the estimated efficiency scores and the different hospital types.

5.1 The CSN-SF hospital production function with transient and persistent inefficiencies

Table 4 presents the econometric results regarding the production frontier obtained by applying the four-random-component SF model. The top part of the table reports the estimated coefficients and *t* ratios for the inputs (columns 2 and 3 for Cobb–Douglas and columns 4 and 5 for translog function); the bottom part shows instead the estimated coefficients and *t* ratios for the determinants of persistent and transient efficiencies, always for Cobb–Douglas (columns 2 and 3) and translog (columns 4 and 5). A positive (negative) coefficient decreases (increases) the efficiency. The estimated coefficients are obtained by applying the ML estimation method to the Mundlak-adjusted frontier (Equation 5); the likelihood ratio test applied to the Mundlak-adjusted model and to the frontier without it (i.e., Equation 1) has a *p* value close to 0. Hence, we can reject the null hypothesis that the Mundlak coefficients are not statistically significant.

Regarding the estimated hospital production functions, it is interesting to notice that in the Cobb–Douglas specification, all inputs are statistically significant, with the exception of physicians.[15] In detail, *lbed* has the highest coefficient (+0.49), whereas *lnurse* and *lother* have the same sign and small estimated effects (−0.08 and −0.05, respectively). Hence, a +1% in the amount of available beds generates a 0.49% increase in the hospital annual weighted output. The estimated coefficients for other nurses and other staff workers are statistically significant but negative. Daidone and D'Amico (2009) found a similar result for hospitals in another Italian region (Lazio) and have interpreted it as an excessive amount of these types of workers in hospital activities. This paper confirms this finding. The coefficient of *lhoursurg* is positive and statistically significant (+0.05%). In the Cobb–Douglas model, physicians have no effect on hospital weighted acute discharges; this may be due to a dominant effect of beds and of activity hours in surgery rooms (the latter may capture the physicians' surgical activity more than the annual hour of workload). The Mundlak adjustments have positive and statistically significant coefficients for physicians ($\overline{lphys}$, +0.28%), nurses ($\overline{lnurse}$, +0.25%), and surgery room hours ($\overline{lhoursurg}$, +0.23%), whereas beds and administrative workers are not statistically significant.

In the translog model, the only positive first-order significant estimated coefficient is for administrative workers ($lother = +1.01$), whereas the second-order statistically significant estimated coefficients are for physicians ($lphys^2$, +0.1%) and working hours in operating theaters ($lhoursurg^2 = 0.03$, weakly significant). Administrative worker ($lother^2$) coefficient is significant but negative. Interaction terms are not statistically significant. Physician ($\overline{lphys}$, +0.23%, weakly) and nurse ($\overline{lnurse}$, +0.44%) Mundlak adjustments are statistically significant and positive.

By applying the first derivatives, we get the output–input elasticities, shown in Table 5, together with the standard errors, computed with the delta method. We have that with the more flexible translog production function the elasticity of beds is +0.36% (lower than that obtained for the Cobb–Douglas case, +0.49%), that the elasticity of hours of activity in surgery rooms is +0.14% (higher than that under the Cobb–Douglas model, +0.05%), and that the physician elasticity is +0.08% (again higher than that with Cobb–Douglas, +0.003). The nurse elasticity is negative (−0.13% in the translog model and −0.08 under the Cobb–Douglas specification), as well as the other staff elasticity (−0.04% and −0.05%, respectively). Under the Cobb–Douglas specification, the return to scale is equal to 0.39 (statistically significant at 1%); that is, the hospitals in the sample are operating under decreasing returns to scale.

Table 4 reports also the evidence regarding the determinants of inefficiencies. Under the translog specification, ownership affects both persistent and transient inefficiencies: The estimated coefficient for private FP hospitals is significant and negative in case of persistent inefficiencies (−3.76), whereas it is positive for transient inefficiencies (+0.33), but it is not statistically significant. Because the dummy variable *FP* has NFP hospitals as the baseline case, this implies that private FP hospitals have less persistent inefficiencies than private NFP ones. P hospitals show no differences in terms of persistent inefficiencies with private NFP ones; however, they have less transient inefficiency ($PUB = -1.72$, weakly statistically significant). Further evidence on the ownership effect on both types of inefficiencies is provided by the Cobb–Douglas model. We find that FP hospitals have more transient inefficiency than private NFP ones ($FP = +1.63$) and less persistent inefficiency ($FP = -6.02$). Hence, we may interpret these results as an evidence that private FP hospitals have a higher structural (persistent) efficiency than other ownership types, whereas they have worse performances in terms of transient inefficiency. Following Dormont and Milcent's (2005) analysis, private FP hospitals have lower long-run moral hazards, maybe due to better incentive and monitoring schemes regarding long-run assets in an organization with well-defined property rights.

Regarding the other investigated determinants, we find that under the Cobb–Douglas model specialized hospitals have more transient inefficiency than general hospitals ($SPECIALIZED = +0.83$), whereas their estimated coefficient in the translog model is +0.12, not statistically significant. We find also that hospital systems have no more transient inefficiency than single hospitals (*MULTIHOSP* is not statistically significant under both specifications), as well as teaching hospitals. Similarly, we find

[15]Some inputs coefficients may be small or insignificant due to possible multicollinearity with beds, that have instead a significant and high coefficient.

TABLE 4 Hospital output frontier and determinants of persistent and transient inefficiencies (Mundlak adjusted)

Variables	Cobb–Douglas Estimate	Cobb–Douglas t ratio	Translog Estimate	Translog t ratio
Inputs				
Constant	0.72	1.43	−6.47***	−3.22
lphys	0.003	−0.12	0.20	0.41
lnurse	−0.08***	−2.59	−0.50	−0.84
lother	−0.05***	−4.63	1.01***	4.34
lbed	0.49***	12.85	0.39	0.73
lhoursurg	0.05***	3.91	0.47	1.52
$\overline{lphys}$	0.28***	3.19	0.23*	1.82
$\overline{lnurse}$	0.25**	2.05	0.44***	2.83
$\overline{lother}$	−0.03	−0.34	−0.08	−0.80
$\overline{lbed}$	−0.18	−1.59	−0.08	−0.48
$\overline{lhoursurg}$	0.23***	3.62	0.12	1.52
$lphys^2$			0.10***	4.46
$lnurse^2$			0.14	1.34
$lother^2$			−0.08***	−2.71
$lbed^2$			0.06	0.42
$lhoursurg^2$			0.03*	1.80
lphys × *lnurse*			−0.03	−0.55
lphys × *lother*			−0.03	−0.89
lphys × *lbed*			−0.03	−0.43
lphys × *lhoursurg*			−0.04	−1.03
lnurse × *lother*			−0.02	−0.34
lnurse × *lbed*			−0.11	−1.06
lnurse × *lhoursurg*			−0.03	−0.49
Log-likelihood	501.79		531.40	
Log-likelihood ratio *p* value	.0		.0	
Determinants of transient and persistent efficiencies				
lother × *lbed*			0.09**	2.35
lother × *lhoursurg*			0.00	−0.10
lbed × *lhoursurg*			0.04	0.82
Factors affecting transient inefficiency u_{it}				
Constant	10.50	0.00	−3.65	0.00
FP	1.63***	4.84	0.33	1.20
PUB	−0.98	−0.42	−1.72*	−1.73
SPECIALIZED	0.83**	2.22	0.12	0.26
MULTIHOSP	1.04	0.45	−0.004	−0.005
TEACHING	−0.39	−1.02	0.42	1.21
MEDIUM	0.94**	2.20	0.09	0.27
SMALL	1.63***	3.42	0.91***	2.51
Factors affecting persistent inefficiency u_i				
Constant	−3.46*	−1.83	−3.20	−1.48
FP	−6.02***	−9.26	−3.76***	−8.63
PUB	−0.02	0.00	0.59	0.23
SPECIALIZED	−4.35	−0.22	1.08	0.49
MULTIHOSP	−0.92	−0.26	0.19	0.08
TEACHING	0.05	0.02	0.55	0.40
MEDIUM	−0.28	−0.12	−0.97	−0.57
SMALL	1.30	0.71	0.50	0.31
Unobserved heterogeneity				
b_i (mean)	−3.19***	−20.54	−3.17***	−13.18

Note. −*m* is the label for input variable with Mundlak adjustment.

***1% significance. **5% significance. *10% significance.

no evidence that hospital specializations, hospital systems, and teaching hospitals affect persistent inefficiency (the estimated coefficients for these determinants are never statistically significant).

Hospital size deserves more attention instead, as we find that it affects transient inefficiency: Medium- and small-sized

TABLE 5 Output–input elasticities

	Physicians	Nurses	Others	Beds	Hours surgery
Cobb–Douglas	0.003 (0.02)	−0.08 (0.03)	−0.05 (0.01)	0.49 (0.04)	0.05 (0.01)
Translog	0.08 (0.04)	−0.13 (0.04)	−0.04 (0.01)	0.36 (0.45)	0.14 (0.28)

Note. Standard errors in parentheses.

TABLE 6 Efficiency scores

	Year						Ownership type		
Efficiency	2008	2009	2010	2011	2012	2013	Not for profit	For profit	Public
Persistent ($\exp^{u_i}$)	0.853	0.853	0.853	0.853	0.853	0.853			
Transient ($\exp^{u_{it}}$)	0.896	0.889	0.921	0.905	0.897	0.910			
Total	0.761	0.759	0.784	0.769	0.765	0.775			
Persistent ($\exp^{u_i}$)							0.830	0.974	0.804
Transient ($\exp^{u_{it}}$)							0.865	0.843	0.937
Total							0.720	0.822	0.754
Hospital classification based on persistent efficiency u_i									
Inefficient							9 (56%)	0 (0%)	58 (72%)
Moderately efficient							7 (44%)	2 (6%)	23 (28%)
Efficient							0 (0%)	34 (94%)	0 (0%)
Hospital classification based on transient efficiency u_{it}									
Inefficient							81 (84%)	179 (83%)	138 (28%)
Moderately efficient							12 (13%)	26 (12%)	162 (33%)
Efficient							3 (3%)	11 (5%)	186 (38%)
Hospital classification based on total efficiency									
Inefficient							68 (71%)	54 (25%)	276 (57%)
Moderately efficient							17 (18%)	33 (15%)	150 (31%)
Efficient							11 (11%)	129 (60%)	60 (12%)

hospitals have higher inefficiency than large-sized hospitals ($MEDIUM = 0.94$ and $SMALL = 1.63$, Cobb–Douglas model; $MEDIUM = 0.09$, not significant, and $SMALL = 0.91$, under the translog output function). This also implies that medium-sized hospitals have lower transient inefficiency than small-sized ones. Hence, we find evidence of a scale effect in hospital output function.

To sum up, the determinant analysis shows that it is important to disentangle the two types of inefficiency; for instance, we have observed that ownership has different transient and persistent effects, as well as size. Moreover, we find no evidence of teaching and hospital system effects on both types of inefficiency.

5.2 | Analysis of estimated efficiency scores

Table 6 shows the arithmetic averages of both types of hospital estimated efficiency scores. There are some interesting results:

- Persistent efficiency (85%) is lower than transient efficiency: The latter varies between 89.6% and 92.1% with a yearly average equal to 90.2%.
- Transient efficiency tends to (mildly) increase over time, from 89.6% at the beginning of the observed period to 91% at the end (+1.4%); hence, hospitals seem to improve their per-period performances.
- The average of estimated total efficiency scores (given by $E[\exp(-u_i - u_{it})|\mathbf{y}_{it}]$) in the Lombardy hospital system is about 76.8%; that is, the sector is still about 23.2% distant from the estimated hospital output frontier. This inefficiency estimate (about 23%) is higher that those presented in Table 1, showing that being able to disentangle transient and persistent efficiencies from unobserved heterogeneity may lead to higher and less biased estimates.

Taking the log of total efficiency scores, we obtain that transient efficiency accounts for 60% of total hospital efficiency; that is, it accounts for a larger part of total estimated efficiency. This confirms, in turn, that long-run inefficiency acts as an important constraint limiting hospital productivity.

Table 6 also displays a classification of hospitals according to two dimensions: (a) efficiency level and (b) ownership type. The former dimension is identified by dividing hospitals into three groups: those having an efficiency score lower than the median (inefficient group), between the median and the third quartile (moderately efficient group), and higher than the third quartile (efficient group). NFP hospitals tend to be inefficient across all the three (persistent, transient, and total) efficiency dimensions: 9 times (56% of their 16 time-invariant scores) they fall in the inefficient group in terms of persistent efficiency, 81 times (84%

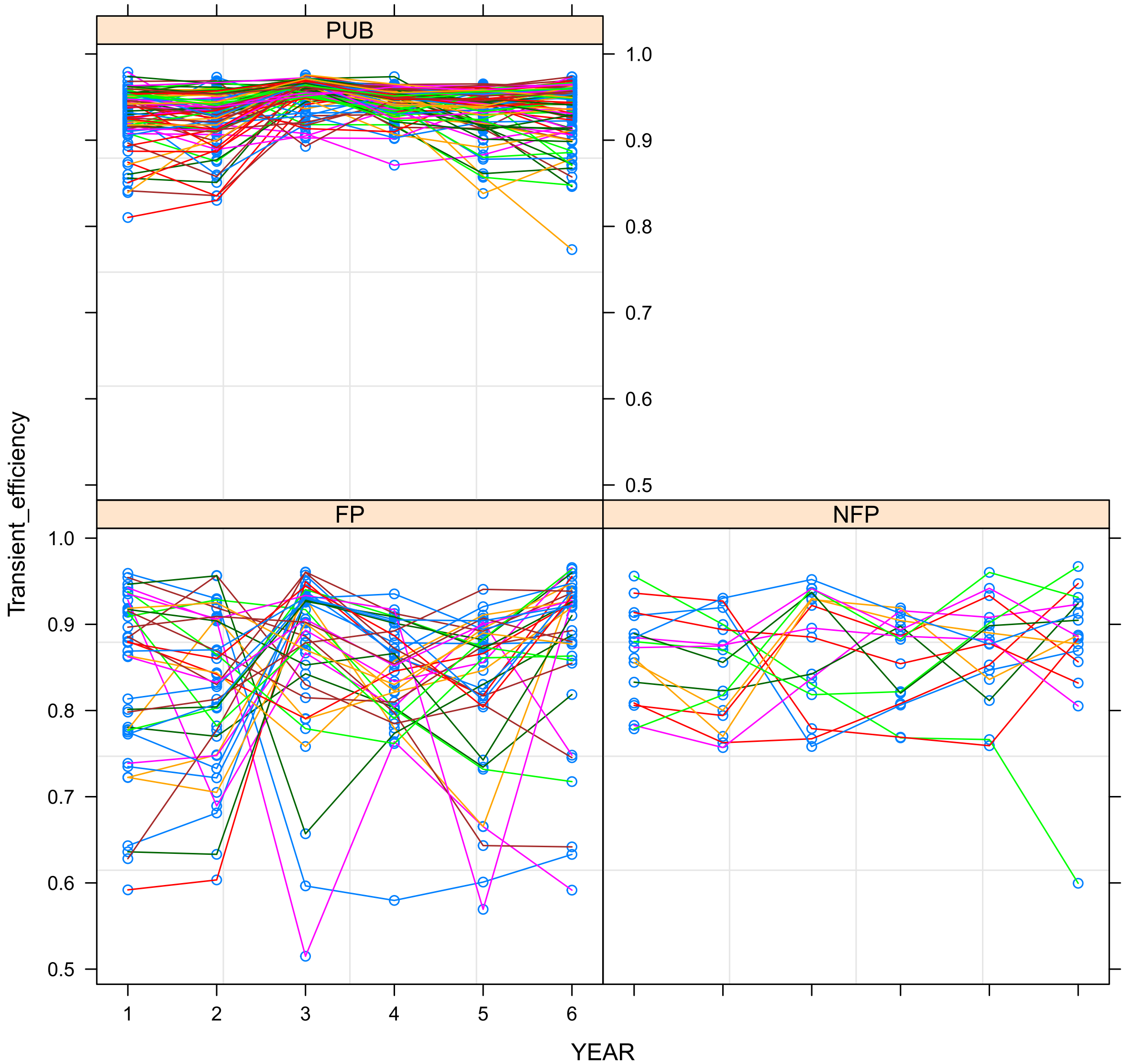

FIGURE 1 Transient efficiency scores by hospital ownership [Colour figure can be viewed at wileyonlinelibrary.com]

of their 96 time-varying scores) they belong to the inefficient group in terms of transient efficiency, while 68 times (71% of their 96 time-varying total scores) they fall in the inefficient group in terms of total efficiency. Many times, FP hospitals are in the efficient group in terms of persistent efficiency (34 times, 94% of their 36 time-invariant scores), whereas their majority (179 times, 83% of their 216 time-varying transient scores) belongs to the inefficient group in terms of transient efficiency. However, because transient efficiency accounts for a larger part of total efficiency, only a small part of FP hospitals' time-varying total efficiency scores belongs to the inefficient group (54 times, 25% of the total scores). The majority of P hospitals (58, 72% of the 81 time-invariant scores) fall in the inefficient group in terms of persistent efficiency, whereas they are more evenly distributed across the three groups when looking at transient efficiency: Moreover, the majority (186, 38% of the 486 time-varying scores) of P hospital transient scores belong to the efficient group. Again, as transient efficiency contributes more to total efficiency (and P hospitals belong much less in the transient efficient group), their majority (276 times, 57% of 486 time-varying scores) belongs to the inefficiency group according to total efficiency scores.

Figure 1 confirms, by inspection, the previous evidence on hospital ownership and transient efficiency. It shows the dynamics of the time-varying estimated scores over the observed period (2008–2013) among the three ownership types. P hospitals tend to be more transient efficient, while private FP hospitals have more observations with very low transient efficiency and they also show much more variability, whereas NFP hospitals display more variation in efficiency scores than P ones.

Figure 2 presents the dynamics of persistent efficiency for a subsample of 16 hospitals (the code in the top row of each panel represents a hospital); it identifies the ownership by a different color (e.g., P hospitals are in green). It is possible to observe the hospital individual behavior in terms of long-run efficiency and to rank each of them in the regional health system. The hospital

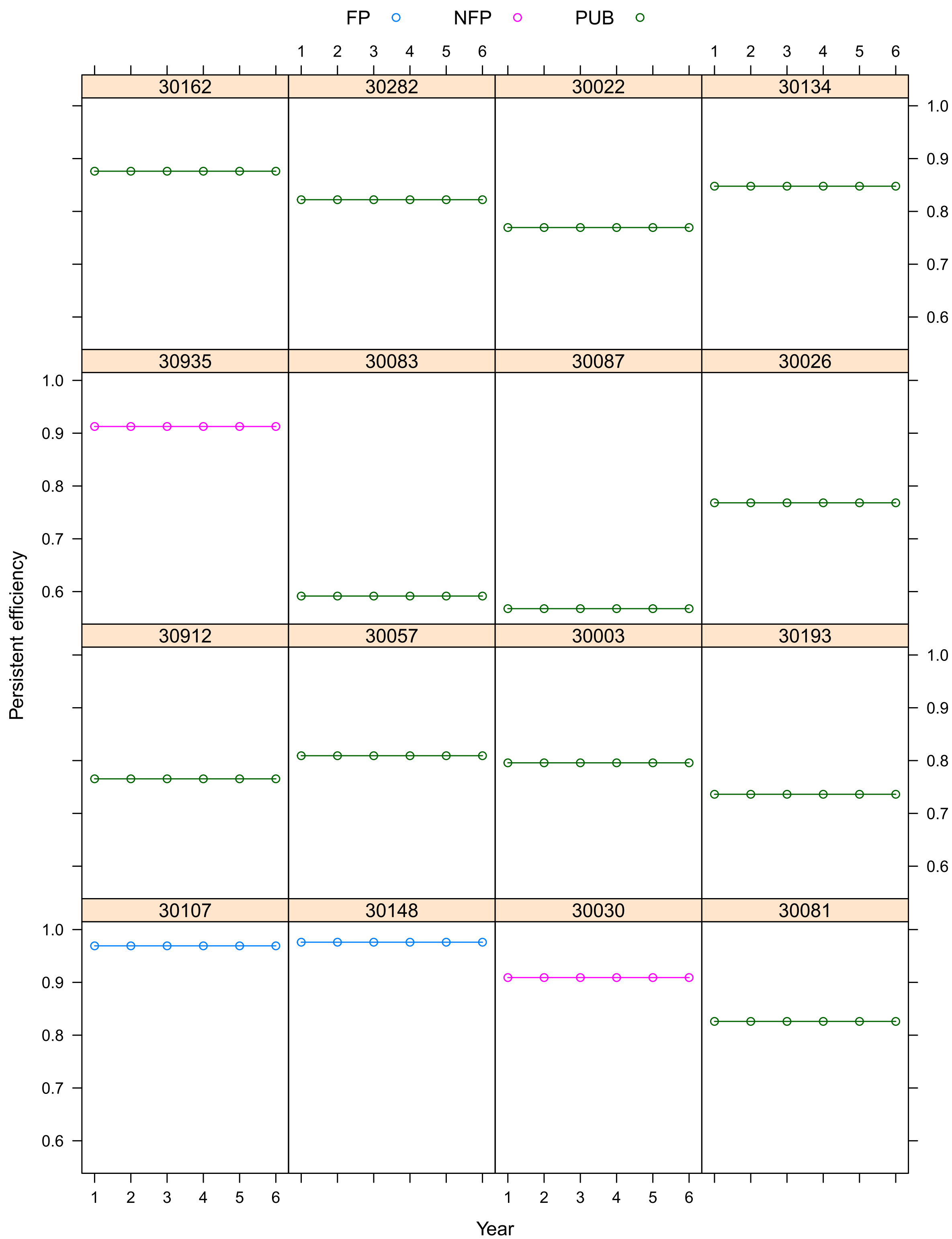

FIGURE 2 Hospital-specific persistent efficiency over time: a sample [Colour figure can be viewed at wileyonlinelibrary.com]

in the left panel, first row of the figure, with code 30162, is a public hospital with persistent efficiency that is about 4% lower than that of the hospital in the second row, left position with code 30935 (about 88% vs. 92%), an NFP one.

Figure 3 displays transient efficiency for the same hospital sample presented in Figure 2. In this case, the hospital in the left panel (first row of the graph, code 30162), which is P (in green), has decreased its transient efficiency during the period, from about 85% to about 75%. On the contrary, the hospital with code 30148 (fourth row, in blue), a private FP hospital, has improved its transient efficiency from about 87% in 2008 to about 91% in 2013.

FIGURE 3 Hospital-specific transient efficiency over time: a sample [Colour figure can be viewed at wileyonlinelibrary.com]

The hospital-specific behavior according to transient and persistent efficiencies for a subsample of 16 hospitals is shown in Figure 4. There are hospitals (e.g., hospital 30202, first row) with persistent efficiency (in red) higher than transient efficiency (in blue) over all the period and other hospitals (e.g., hospital 30029, third row) starting with a transient efficiency lower than the persistent one at the beginning of the period and ending with a transient efficiency higher than the persistent one. This implies that hospital managers have been able to improve their short-run performances during the observed period; this is an important

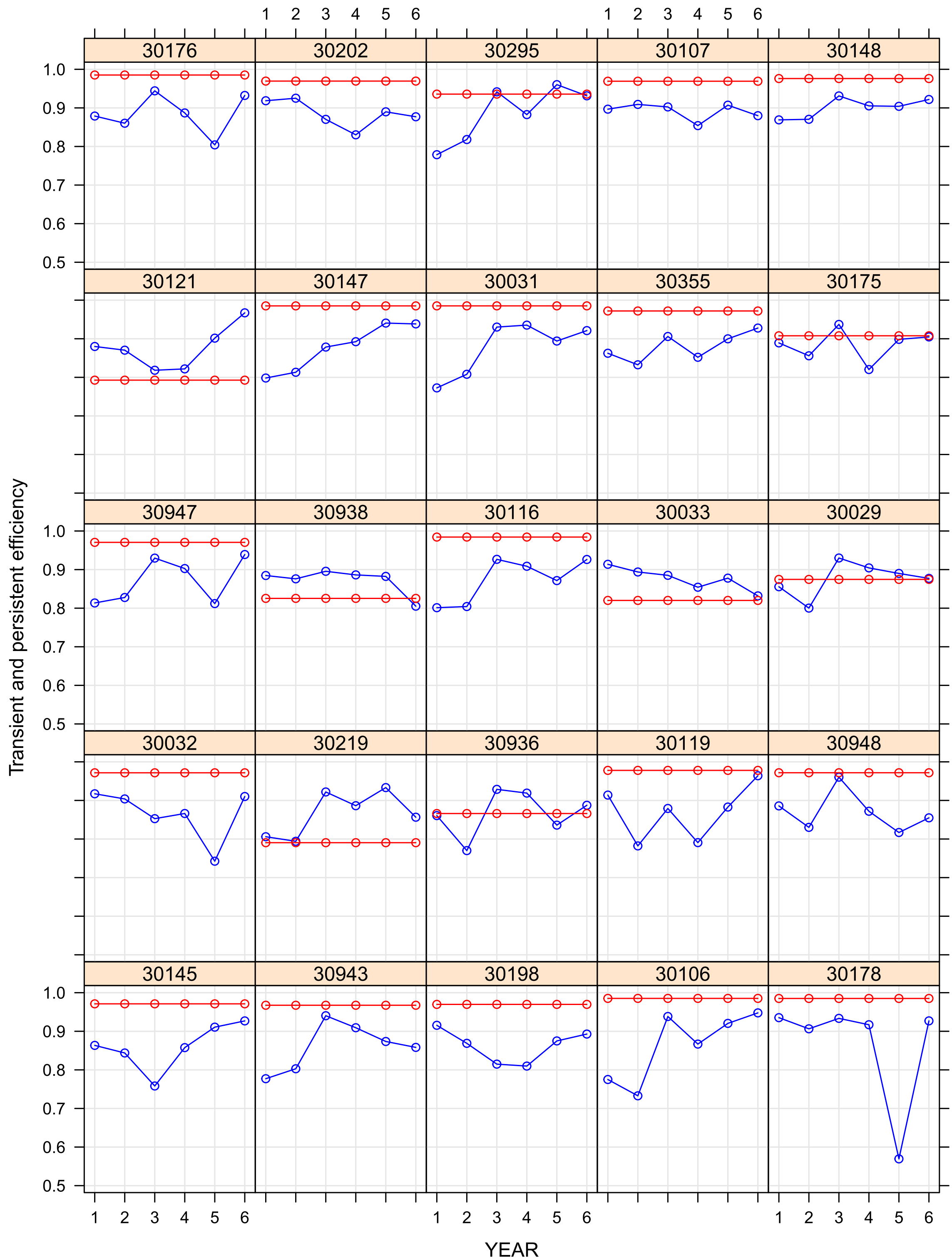

FIGURE 4 Hospital-specific persistent and transient efficiency over time: a sample [Colour figure can be viewed at wileyonlinelibrary.com]

result that deserves attention by regional regulators. Moreover, we have hospitals (e.g., hospital 30938, third row) with transient efficiency higher than the persistent one at the beginning of the period and lower instead at the end: The managers of these hospitals have bad short-run performance and should be carefully monitored by the regional government. Hence, hospitals' evaluation based on the two types of efficiency may become an element for incentive-based regulation: For instance, it may be feasible to ask hospital managers to improve their transient efficiency in a given period (e.g., 2–4 years) and to design incentives providing prizes if the target is achieved.

6 | CONCLUSIONS

In this paper, we investigate the issue of hospital efficiency using an extension of the Colombi et al. (2014) four-random-component SF model, which allows observation of the determinants of persistent (long-run) and transient (short-run) inefficiencies. We have applied the model to a dataset of 133 Italian hospitals located in Lombardy over the period 2008–2013. We provide the following empirical evidence. First, we find that average total inefficiency is about 23%, an estimate higher than previous ones. This confirms that being able to disentangle transient and persistent inefficiencies from unobserved heterogeneity may lead to higher and less biased estimates of hospital inefficiency. Second, transient efficiency seems to be more important than persistent efficiency, as it accounts for a large part (60%) of the total one. Third, average transient efficiency increases over time (more than +0.2% per year), showing that the regional system is improving its performances. Fourth, we provide some evidence that ownership matters in assessing hospital productivity, as private FP hospitals have lower persistent inefficiency and higher transient inefficiency than NFP ones, whereas P hospitals have lower transient inefficiency than NFP hospitals. Fifth, large-sized hospitals have lower transient inefficiency than medium- and small-sized hospitals, providing some insights on the existence of scale economies. Sixth, hospitals specialized in single treatments have higher transient inefficiency than general hospitals; that is, there is evidence of scope economies. Last, we do not find any evidence that multi-hospital systems and teaching activities are determinants of transient and persistent inefficiencies.

We can draw some policy implications from these insights. First, being able to distinguish between persistent and transient inefficiencies, it would be interesting to adopt a regulation model that provides incentives to improve transient inefficiency in the short run (e.g., −1% within 2–3 years) and persistent inefficiency over a longer period (5–10 years). Possible targets for improving transient inefficiency may involve increasing key performance indicators (e.g., discharges per working hours, discharges per bed, and higher activity rates for surgery rooms); that is, targets have to be designed to achieve better short-run performances through higher factors' productivity. Lower persistent inefficiency, being a long-run outcome, may be achieved through strategies involving a longer time horizon, for example, technology adoption that allows reducing the patients' length of stay and increasing the beds' utilization rate, renovation of hospital's physical organization, and so on.

Second, as persistent inefficiency seems to be more important than transient inefficiency, it is essential to consider which are its possible determinants. They may be linked to obsolete machinery and equipment (which exhibit low productivity and frequent breakdowns), to rigidities in the workforce (e.g., unupdated physicians' and nurses' human capital so that new technologies cannot be fully exploited or adopted), to gaps in the administrative staff activities (bad coding, bureaucracy burden imposed on health care staff, etc.), and to inefficient internal organization (e.g., duplicity of tasks in different wards such as operating and surgery rooms, need of building a new hospital, and reorganization of primary and secondary care in a local health system). This implies that policy makers need to adopt instruments to remove these possible sources of persistent inefficiency, ranging from personnel training to technology and infrastructure investments. Last, policy makers have to consider that there might be a scale effect in hospital productivity and balance the higher efficiency that may be reached by concentrating the acute discharge treatments in less large-sized hospitals, with the higher social costs imposed especially to the elderly population with mobility problems if some small hospitals are closed.

The above results may be extended by enlarging the time interval and by including more outputs in the analysis. Moreover, comparing different regional systems may help to deepen the understanding of factors that may affect efficiency. This is left for future research.

ORCID

Gianmaria Martini http://orcid.org/0000-0001-9503-4999

REFERENCES

Aigner, D. J., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37.

Barbetta, G. P., Turati, G., & Zago, A. M. (2007). Behavioral differences between public and private NFP hospitals in the Italian national health service. *Health Economics*, *16*, 75–96.

Barros, C. P, de Menezes, A. G., & Vieira, J. C. (2013). Measurement of hospital efficiency using a latent class stochastic frontier model. *Applied Economics*, *45*, 47–54.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, *38*, 387-399.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis*, *3*, 153-169.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, *20*, 325–332.

Battese, G. E., & Corra, G. S. (1977). Estimation of a production frontier model: With application to the pastoral zone of Eastern Australia. *Australian Journal of Agricultural Economics*, *21*, 169–179.

Besstremyannayam, G. (2011). Managerial performance and cost efficiency of Japanese local public hospitals: A latent class stochastic frontier model. *Health Economics*, *20*, 19–34.

Bernet, P., Rosko, M., & Valdmanis, V. (2008). The relationship between hospital cost inefficiency and debt ratings. *Journal of Health Care Finance*, *34*, 66–88.

Brown, H. S. (2003). Managed care and technical efficiency. *Health Economics*, *12*, 149–158.

Carey, K. (2003). Hospital cost efficiency and system membership. *Inquiry*, *40*, 25–38.

Carey, K., Burgess, J., & Young, G. (2008). Specialty and full-service hospitals: A comparative cost analysis. *Health Services Research*, *43*, 1869–1887.

Castelli, A., Street, A., Verzulli, R., & Ward, P. (2015). Examining variations in hospital productivity in the English NHS. *European Journal of Health Economics*, *16*, 243–254.

Chirikos, T. N. (1998). Identifying efficiently and economically operated hospitals: The prospects and pitfalls of applying frontier regression techniques. *Journal of Health Politics, Policy and Law*, *23*, 879–904.

Chirikos, T. N., & Sear, A. M. (2000). Measuring hospital efficiency: A comparison of two approaches. *Health Services Research*, *34*, 1389–1408.

Coelli, T., Rao, D., O'Donnell, C., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. New York, NY: Springer.

Colombi, R. (2013). Closed skew normal stochastic frontiers for panel data. In Torelli, N. (Ed.), *Advances in theoretical and applied statistics. Studies in theoretical and applied statistics* (pp. 177–186). Berlin, Germany: Springer.

Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-Skew normality in stochastic frontiers with individual effects and long/short run efficiency. *Journal of Productivity Analysis*, *42*, 123–136.

Cornwell, C., Schmidt, P., & Sickles, R. (1990). CProduction frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, *46*, 185–200.

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.

Deily, M., & McKay, N. (2006). Cost inefficiency and mortality rates in Florida hospitals. *Health Economics*, *15*, 419–431.

Deily, M., McKay, N., & Dorner, F. (2000). Exit and efficiency: The effects of ownership type. *Journal of Human Resources*, *35*, 734–747.

Dominguez-Molina, A., Gonzáles-Farías, G., & Ramos-Quiroga, R. (2004). Skew normality in stochastic frontier analysis. In Genton, M. (Ed.), *Skew elliptical distributions and their applications*.(pp. 223–241). London, UK: Chapman and Hall.

Dormont, B., & Milcent, C. (2005). How to regulate heterogeneous hospitals?*Journal of Economics & Management Strategies*, *3*, 591–621.

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

Filippini, M., & Greene, W. (2016). Persistent and transient productive inefficiency: A maximum simulated approach. *Journal of Productivity Analysis*, *45*, 187–196.

Folland, S. T., & Hofler, R. A. (2001). How reliable are hospital efficiency estimates? Exploiting the dual to homothetic production. *Health Economics*, *10*, 683–698.

Frech, H. E., & Mobley, L. R. (2000). Efficiency, growth, and concentration: An empirical analysis of hospital markets. *Economic Inquiry*, *38*, 369–84.

Gonzáles-Farías, G., Dominguez-Molina, A., & Gupta, A. (2004). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, *126*, 521–34.

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

Greene, W. (2005b). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, *23*, 7–32.

Greene, W. (2009). The econometric approach to efficiency analysis. In Fried, H. O., Lovell, C. A. K., & Schmidt, S. S. (Eds.), *The measurement of productive efficiency—Techniques and applications* (pp. 92–251). Oxford, UK: Oxford University Press.

Grosskopf, S., Margaritis, D., & Valdmanis, V. (2001). Comparing teaching and non-teaching hospitals: A frontier approach (teaching vs. non-teaching hospitals). *Health Care Management Science*, *4*, 83–90.

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

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

Hollingsworth, B. (2003). Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science*, *6*, 203–218.

Hollingsworth, B., & Street, A. (2006). The market for efficiency analysis of health care organizations. *Health Economics*, *15*, 1055–1059.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, *17*, 1107–1128.

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: Data envelopment analysis and stochastic frontier analysis. *Health Care Management Science*, *4*, 103–115.

Jacobs, R., Smith, P. C., & Street, A. (2006). *Measuring efficiency in health care*. Cambridge, UK: Cambridge University Press.

Koop, G., Osiewalski, J., & Steel, M. F. J. (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics*, *76*, 77–105.

Kumbhakar, S. C. (1987). The specification of technical and allocative inefficiency in stochastic production and profit frontiers. *Journal of Econometrics*, *34*, 335–348.

Kumbhakar, S. C. (1990). Production frontiers, panel data and time-varying technical inefficiency. *Journal of Econometrics*, *46*, 201–211.

Kumbhakar, S. C., & Heshmati, A. (1995). Efficiency measurement in Swedish dairy farms: An application of rotating panel data, 1976-1988. *American Journal of Agriculture Economics*, *77*, 660–674.

Kumbhakar, S. C., & Hjalmarsson, L. (1993). Technical efficiency and technical progress in Swedish dairy farms. In Fried, H. O., Lovell, C. A. K., & Schmidt, S. S. (Eds.), *The measurement of productive efficiency—Techniques and applications* (pp. 256–270). Oxford, UK: Oxford University Press.

Kumbhakar, S. C., & Hjalmarsson, L. (1995). Labour-use efficiency in Swedish social insurance offices. *Journal of Applied Economics*, *10*, 33–47.

Kumbhakar, S. C., Lien, G., & Hardaker, J. B. (2014). Technical efficiency in competing panel data models: A study of Norwegian grain farming. *Journal of Productivity Analysis*, *41*, 321–337.

Kumbhakar, S. C., & Lovell, C. A. K. (2000). *Stochastic frontier analysis*. Cambridge, MA: Cambridge University Press.

Kumbhakar, S. C., & Wang, H. J. (2005). Estimation of growth convergence using a stochastic production frontier approach. *Economic Letters*, *88*, 300–305.

Lee, Y., & Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical inefficiency. In Fried, H., & Lovell, K. (Eds), *The measurement of productive efficiency: techniques and applications*. New York, NY: Oxford University Press.

Li, T., & Rosenman, R. (2001). Cost inefficiency in Washington hospitals: A stochastic frontier approach using panel data. *Health Care Management Science*, *4*, 73–81.

Linna, M. (1998). Measuring hospital cost efficiency with panel data models. *Health Economics*, *7*, 415–427.

McKay, N., & Deily, M. (2005). Comparing high- and low-performing hospitals using risk adjusted excess mortality and cost inefficiency. *Health Care Management Review*, *30*, 347–360.

McKay, N., Deily, M., & Dorner, F. (2002). Ownership and changes in hospital inefficiency, 1986–1991. *Inquiry*, *39*, 388–99.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb–Douglas production functions with composed error. *International Economic Review*, *18*, 435–444.

Mobley, L. R. (1998). Effects of selective contracting on hospital efficiency, costs and accessibility. *Health Economics*, *7*, 247–261.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*, 69–85.

Mutter, R., & Rosko, M. (2008). The impact of ownership on the cost-efficiency of U.S. hospitals. In Blank, J., & Valdmanis, V. (Eds.), *Evaluating hospital policy and performance: Contributions from hospital policy and productivity research* (pp. 113–138). New York, NY: Elsevier.

Newhouse, J. P. (1994). Frontier estimation: How useful a tool for health economics?*Journal of Health Economics*, *13*, 317–322.

Pandey, M. D. (1998). An effective approximation to evaluate multinormal integrals. *Structural Safety*, *20*, 51–67.

Pitt, M., & Lee, L. (1981). Measurement of sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics*, *9*, 43–64.

Rosko, M. D. (1999). Impact of internal and external environmental pressures on hospital inefficiency. *Health Care Management Science*, *2*, 63–74.

Rosko, M. D. (2001a). Cost efficiency of US hospitals: A stochastic frontier approach. *Health Economics*, *10*, 539–551.

Rosko, M. D. (2001b). Impact of HMO penetration and other environmental factors on hospital X-inefficiency. *Medical Care Research and Review*, *58*, 430–454.

Rosko, M. D. (2004). Performance of U.S. teaching hospitals: A panel analysis of cost inefficiency. *Health Care Management Science*, *7*, 7–16.

Rosko, M. D., & Chilingerian, J. A. (1999). Estimating hospital inefficiency: Does case mix matter?*Journal of Medical Systems*, *23*, 57–71.

Rosko, M. D., & Mutter, R. L. (2008). Stochastic frontier analysis of hospital inefficiency: A review of empirical issues and an assessment of robustness. *Medical Care Research and Review*, *65*, 131–166.

Rosko, M. D., & Mutter, R. L. (2010). Inefficiency differences between critical access hospitals and prospectively paid rural hospitals. *Journal of Health Politics, Policy and Law*, *35*, 95–126.

Rosko, M. D., & Mutter, R. L. (2011). What have we learned from the application of stochastic frontier analysis to U.S. hospitals. *Medical Care Research and Review*, *68*(1 suppl), 75S–100S.

Rosko, M. D., & Proenca, J. (2005). Impact of network and system use on hospital X-inefficiency. *Health Care Management Review*, *30*, 69–79.

Rosko, M. D., Proenca, J., Zinn, J., & Bazzoli, G. (2007). The impact of membership in different types of systems on hospital cost-inefficiency. *Inquiry*, *44*, 335–349.

Sari, N. (2003). Efficiency outcomes of market concentration and managed care. *International Journal of Industrial Organization*, *21*, 1571–1589.

Schmidt, P., & Sickles, R. (1984). Production frontiers and panel data. *Journal of Business Economics and Statistics*, *2*, 367–374.

Street, A. (2003). How much confidence should we place in efficiency estimates?*Health Economics*, *12*, 895–907.

Tiemann, O., & Schreyögg, J. (2012). Changes in hospital efficiency after privatization. *Health Care Management Science*, *15*, 310–326.

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.

Tsionas, E. G., & Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effect model. *Journal of Applied Econometrics*, *29*, 110–132.

Vitaliano, D. F., & Toren, M. (1994). Costs and efficiency in nursing homes: A stochastic frontier approach. *Journal of Health Economics*, *13*, 281–300.

Wagstaff, A. (1989). Estimating efficiency in the hospital sector: A comparison of three statistical cost frontier models. *Applied Economics*, *21*, 650–72.

Wagstaff, A., & López, G. (1996). Hospital costs in Catalonia: A stochastic frontier analysis. *Applied Economics Letters*, *3*, 471–74.

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

Zuckerman, S., Hadley, J., & Iezzoni, L. (1994). Measuring hospital efficiency with frontier cost function. *Journal of Health Economics*, *13*, 255–80.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Colombi R, Martini G, Vittadini G. Determinants of transient and persistent hospital efficiency: the case of Italy. *Health Economics*. 2017;26(S2):5–22. https://doi.org/10.1002/hec.3557