Health expenditure in Italy:  
a comprehensive intergovernmental model

Fabio Pammolli\textsuperscript{a}, Francesco Porcelli\textsuperscript{b}, Francesco Vidoli\textsuperscript{c}, Guido Borà\textsuperscript{d}

\textsuperscript{a}IMT, Institute for Advanced Studies Lucca, Piazza San Ponziano 6, 55100 Lucca, Italy  
\textsuperscript{b}CAGE, Centre for Competitive Advantage in the Global Economy, Warwick, UK  
\textsuperscript{c}University of Roma Tre, Via G. Chiabrera, 199 Rome, Italy  
\textsuperscript{d}University of Siena and CERM Foundation - Competitività, Regole, Mercati

Abstract

The paper provides an analysis of the determinants of health spending differentials among the Italian Regions for the years 1998-2010. An original standardised spending framework is proposed that takes into account the appropriateness of the services provided and price/technical efficiency, as well as the main socioeconomic demand variables. Italian economic duality in terms of differences between historical and standard expenditure needs is reflected in regional health systems. Results suggest that spending reduction by the Regions since 2006 for budgetary needs has mainly been achieved through the quantity of health services supplied rather than by reducing inefficiency.

Keywords: Expenditure needs, Health care expenditure, Intergovernmental relationships, Composite indicator

1. Introduction and research questions

The current structure of the Italian National Health System, (defined by law 833/1978, legislative decrees 502/1992 and 517/1993 and more recently by legislative decrees 229/1999 and 68/2011), assigns to regional governments the responsibility of redressing territorial imbalances achieving a set of guaranteed service levels (first known as “uniforms” and later as “essential” services level). As a result of the Italian dual economy, intergovernmental grants are the main source of funding in poor

Email addresses: pammolli@gmail.com (Fabio Pammolli), fporcelli78@gmail.com (Francesco Porcelli), fvidoli@gmail.com (Francesco Vidoli), guido.bora@unisi.it (Guido Borà)

Preprint submitted to Elsevier  
June 20, 2016
southern regions and local fiscal revenues are the main source of funding in richer northern and central regions.

In this setting, Italian regions share a common budgeting mechanism and a common institutional framework, but they maintain administrative autonomy (even with respect to private expenditure co-financing) facing different local demand and supply factors (for a more complete description of the Italian health care system see Levaggi & Zanola, 2003).

Given this financing scheme, it was therefore crucial to draw up a proper funding scheme between the Central and the regional governments in order to overcome the “soft budgeting constraint” issue well specified by Kornai et al. (2003) and Bordignon & Turati (2009) for the Italian case.

Given that the “public health policy is the result of the interaction of several layers of government” (Bordignon & Turati, 2009), since the 1990s the Regions presumably inflated their expenditure and expected the Central government to pay off their annual deficits in the belief that it would cover the remaining costs.

The evolution of public health expenditure since the end of the nineties confirms this trend, showing two major trends. Until 2006 a strong upward trend is evident; expenditure increased to 6.6% of GDP, a significant increase with respect to 5.0% in 1998. In 2006, this trend seems to stop and national expenditure remains substantially constant in monetary value\(^1\).

![Figure 1: Current national health expenditure. Source: Istat (Health for All), period 1998-2013](image)

\(^1\)Note that Italy, with current health expenditure at 7.0% of GDP, ranks slightly above the OECD average of 6.4% of GDP, and behind countries such as France (8.7%), Germany (8.4%) and the US (8.3%).
The “strategic game” based on local government expectations played essentially at political level by regional and central governments, at least until 2006, was that Regions claimed that the Central government deliberately under-finances them while the Central government claimed that Regions overspend. This led to implementation of a new set of funding rules.

Expenditure restraint since 2006 is the result of implementation of two joint agreements (“Patti per la Salute” in Italian) between Central government and Regions for the years 2007-2009 and 2010-2012\(^2\). The new set of rules focuses on the transition from a system based on “soft budget constraint” to one based on the “strong empowerment principle” through the introduction of spending cut plans (“Piani di rientro” in Italian) for the Regions with excessive deficit spending.

Against this background and given this robust mechanism for the containment of new debt and stronger regional accountability, analysing regional health expenditure as a microfunded strategic game is relatively uninformative. Identifying structural regional imbalances at the macro level that cannot be attributed to local factors and may be due to efficiency and different regional supply may be more informative.

In particular, when expenditure is expressed as a percentage of GDP, Italy is divided into two large clusters, typical of Italian economic dualism: SD and SDS Regions\(^3\) increased their mean expenditure from 7% of GDP in 1998 to almost 10% of GDP in 2011, while the other regions\(^4\) (NRS, NR, CR) spend, on average, 5% to 7% of GDP.

The comparison between per capita expenditure and expenditure as a percentage of GDP confirms the key need for proper equalization criteria for public health resources in order to ensure a relatively homogeneous level throughout Italy (in terms of expenditure and service level). The joint agreements of 2014-2016 raised the issue of introducing standard expenditure needs calculation in the health sector, as originally conceived in 2011 (Act. 68/2011), albeit in a basic form.

The Italian National Audit Office (Corte dei Conti, 2015, p. 197) noted that the current quantification criterion of regional health needs was still mainly based on regional population; it therefore proposed finding a set of indicators to assess efficiency and appropriateness for each region with reference to an aggregate level for

\(^2\text{Confirmed for the period 2013-2015.}\)

\(^3\text{Southern regions (SD): Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria; Special administrative regions of the South (SDS): Sicilia, Sardegna.}\)

\(^4\text{Special administrative Northern regions (NRS): Valle d’Aosta, Trentino Alto Adige, Friuli Venezia Giulia; Northern regions (NR): Piemonte, Lombardia, Veneto, Liguria, Emilia Romagna; Central regions (CR): Toscana, Umbria, Marche, Lazio.}\)
As conducted in a previous research paper Francese & Romanelli (2014) our aim was to identify the main drivers of expenditure differentials at a regional level; the main improvement concerns explicit calculation of technical inefficiency (via output and input composite indicators) and estimation of the output gap\(^6\) for a better understanding of the global performance of each Region. Moreover, this analysis creates a bridge between the literature on intergovernmental fiscal equalisation and that on local government performance indicators and incentives in the public sector (see e.g. Mizell, 2008; Burgess & Ratto, 2003; Lockwood & Porcelli, 2011).

Indeed, evaluation and continuous improvement in terms of output, quality, access to care and cost of national health systems have become a key issue in the policies of more developed countries (see e.g. Fisher et al., 2009). Much work has recently gone into improving estimation models, both from an administrative point of view (national authorities, such as the United Kingdom National Health Service and the Canadian Institute for Health Information, have been established) and a methodological point of view Jencks et al., 2000, Kwon, 2003 or Nuti et al., 2011 and Pammolli & Salerno, 2011 for the Italian national health system).

Evaluation of national health systems can be conducted according to a plurality of subjects and objectives: breadth of performance and financial or service indicators can be considerable, ranging from analysis of the whole system to specific experiences of individual patients.

In the international literature, the evaluation of performance and efficiency or supply level analysis has been carried out at the supranational, national, regional and local level (see e.g. Ibrahim, 2001).

In such complex systems, what needs to be measured is a key question. In general, health care systems can be evaluated with respect to factors such as the quality/quantity of care, access to care and cost/expenditure; although of interest, these dimensions are not the only analytical dimensions. For example, Paakkonen & Seppala (2014), evaluated the accessibility, efficiency and equality of treatment. But the real challenge of evaluation methods, at least from the statistical and econometric standpoint, is the multidimensionality of national health systems. From a practical point of view, to evaluate such subjects, robust and reliable techniques are necessary that measure complex phenomena in a synthetic way. For example, to compare improvements in performance, quality or service level, it is crucial to

\(^5\)Hospital care (44% of health expenditure), district care (51%) and community care (5%).

\(^6\)Difference between the historical and the standard service level.
have: (i) a set of comparable indicators built on a reliable information system and
(ii) robust methods that make it possible to integrate the information into a single
measure so as to set or indicate benchmarks or reference parameters.
In particular, Smith (2002) highlights three critical methodological aspects concern-
ing composite indicators in health sector analysis: (i) calculation of the set of weights,
(ii) sterilization of the effects of external variables on performance and (iii) the assump-
tions underlying the calculation models. With regard to these aspects, he notes
the absence of a broad consensus on a shared methodology to identify the optimal
set of weights to be used for the composite index.
Finally, Bankauskaite & Dargent (2007) observe that aggregate measures of demand
and supply for health services lack “precision and combine uncertain weighting sys-
tems, imprecision arising from the potential non-comparability of component mea-
sures, and misleading reliability in the form of whole-population averages that mask
distribution issues”.
Against this background, new methods to estimate regional supply and technical
efficiency in a robust way are proposed in this paper: the optimal set of weights
is chosen endogenously to the data and is consequently independent of the choices
made by individual researchers.

2. Methodological proposal

The proposed methodological approach can be divided into four logically distinct
phases:

1. in the first step the aggregate level of performance (output) and input employed
   in the healthcare system are estimated for each region and each year (paragraph
   2.1);
2. in the second phase, the level of technical efficiency is calculated on the basis
   of the input and output composite indicator (paragraph 2.2);
3. the third phase is devoted to estimation of demand in reduced form (named
   output function); in particular, this phase aims to calculate the standard service
   level that, compared with the historical service level, makes it possible to locate
   the output gap for each region by measuring how each regional system meets its
   demand. Regions producing more services (in terms of quality and quantity)
   than the standard will have a positive output gap, while Regions with a negative
   output gap will have performance that is lower than the potential demand from
   their territories 2.3.1);
4. the last step of our approach concerns estimation of the cost function in a
   reduced form (expenditure function) that makes it possible to identify the


standard requirements of each region; technical inefficiency and the output gap are included as covariates, and the fixed effects approach makes it possible to estimate the price inefficiency of production inputs for each region. In this way, standard needs can easily be calculated for each region, isolating the share of historic expenditure due to inefficiency at the required level to fill the performance deficit (paragraph 2.3.2).

2.1. Output and input composite indicators

The aggregate level of input and output was calculated using a specific composite indicators (CI) technique named “Benefit of the Doubt” (BoD) by Melyn & Moesen (1991). Several authors (e.g. Sørensen, 2015 and Lauer et al., 2004) consider this to be one of the most promising techniques developed in the last two decades, especially by virtue of its theoretical properties; it is particularly useful in avoiding subjective choices in the multidimensional health sector.

Unlike other weighting methods based on mean measures, BoD makes it possible to find the optimal set of weights for the elementary indicators of each unit endogenously\(^7\). In this way, the resulting indicator is the highest possible for each unit: a property particularly “useful in policy arena, since policy-makers could not complain about unfair weighting: any other weighting scheme would have generated lower composite scores”. (Nardo et al., 2005).

The application of production efficiency techniques to the CIs field is relatively straightforward, as suggested by Witte & Rogge (2009), because ”the Benefit of the Doubt approach is formally tantamount to the original input-oriented CCR-DEA model of Charnes et al. (1978), with all questionnaire items considered as outputs and a dummy input equal to one for all observations”.

The basic productive efficiency framework, in fact, designates a production technology in which the activity of each decision making unit is characterized by a set of inputs \(x \in \mathbb{R}_+^p\) used to produce a set of outputs \(y \in \mathbb{R}_+^q\). The production set is the set of technically feasible combinations of \((x, y)\):

\[
\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} | x \text{ can produce } y\}
\]

where \(\Psi\) is the so-called support of \(H(x, y)\).

Given this premise, the Farrell-Debreu efficiency scores (input oriented) for a given production scenario \((x, y) \in \Psi\) when \(x\) is constant and equal to 1 for every unit (as

---

\(^7\)In other words “not in an arbitrary way”, since in many fields of application it is not possible \(ex-ante\) to assign a given weight to each indicator or, more generally, the same weight to each unit.
in CIs) may be written as:

$$\theta(x, y) = \inf \{\theta | (\theta, y) \in \Psi\}$$  \hspace{1cm} (2)

and consequently the FDH estimator is provided by the particular disposal hull of the sample points:

$$\hat{\Psi}_{FDH} = \{ (1, y) \in \mathbb{R}^{1+q}_+ | y < Y_i, i = 1, ..., n \}. \hspace{1cm} (3)$$

Hypothesizing the convexity of $\Psi$, the convex hull of $\hat{\Psi}_{FDH}$ can be named $\hat{\Psi}_{BoD}$ in accordance with Cherchye & Kuosmanen (2002):

$$\hat{\Psi}_{BoD} = \{ (1, y) \in \mathbb{R}^{1+q}_+ | y < \sum_{i=1}^{n} \gamma_i y_i \text{ for } (\gamma_1, ..., \gamma_n) \text{ such that } \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n \}. \hspace{1cm} (4)$$

For technical details of the BoD methodology and for robustness enhancements, see for example e.g. Vidoli et al. (2015).

The most important BoD properties are: (i) the set of weights is determined endogenously through the observed performance of each unit and the benchmark is not based on constraints or on theoretical choices, but is the linear combination of observed performances; (ii) the CI is weak monotone and scale invariant and (iii), as already mentioned, the set of weights is the highest possible for the single unit.

2.2. Technical efficiency

In the second step of our approach, the input and output composite indicators estimated by the BoD technique described above are used to measure the level of technical efficiency achieved by each region in the health service sector. Efficiency is usually measured with reference to the relationship between resources employed and products obtained. From an applicative point of view, this means:

- in output space, for a given technology and inputs, outputs actually produced match maximum potentially producible outputs;
- in input space, for a given technology and output level, the amount of input used matches the minimum potentially usable quantity.

In the last few decades frontier productive efficiency methods have been widely developed\(^8\) and usually estimated by two different (parametric and nonparametric) techniques.

\(^8\)For a complete survey, see Fried & Lovell (2008).
More specifically, methods can be named “parametric” when a particular form of production (cost) function is assumed, while they are called “nonparametric” when the form is not assumed and the frontier is only identified by some specific properties. Besides classical estimation methods, in the last few years several interesting methods, which can be seen as a “bridge” between the two groups, have been developed: semi-nonparametric and semiparametric methods. Table 1 suggests a non exhaustive taxonomy.

<table>
<thead>
<tr>
<th>Production technology</th>
<th>Frontier techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonparametric</td>
<td>*DEA, Farrell (1957); Charnes <em>et al.</em> (1978)</td>
</tr>
<tr>
<td></td>
<td>*FDH, Deprins <em>et al.</em> (1984); Grosskopf (1996)</td>
</tr>
<tr>
<td></td>
<td><em>Order-m, Daraio &amp; Simar (2007b)</em></td>
</tr>
<tr>
<td>Semi-nonparametric</td>
<td><em>StoNED, Kuosmanen &amp; Kortelainen (2012)</em></td>
</tr>
<tr>
<td></td>
<td><em>StoNEZD, Johnson &amp; Kuosmanen (2011)</em></td>
</tr>
<tr>
<td>Semiparametric</td>
<td><em>Park &amp; Simar (1994), Park <em>et al.</em> (2007)</em></td>
</tr>
<tr>
<td>Parametric</td>
<td>*SFA, Aigner <em>et al.</em> (1977); Meeusen &amp; van den Broeck (1977)</td>
</tr>
<tr>
<td></td>
<td><em>DFA, Aigner &amp; Chu (1968)</em></td>
</tr>
</tbody>
</table>

Table 1: Estimation techniques by production technology

Another distinction between models may lie in error term specification: in deterministic models, it is assumed that all observations \((x, y)\) belong to the production set \(\Psi\), while in stochastic models there may be noise in the data and it may happen that \((x_i, y_i) \not\in \Psi\) for some \(i\).

All classical methods are affected by some basic imperfections; nonparametric DEA, for instance, seems particularly flexible and generalizable, but it is not possible to recognize whether the differences in efficiency - namely the distance between observed and maximum possible output - are due to technical inefficiency or to noise/outlier effects (Greene, 2008). The SFA parametric frontier, instead, allows explicit distinction in the error term between inefficiency, ascribable to inefficiency factors, and measurement errors, due to accidental noise and therefore not directly attributable to the local policy maker; however, the most important drawback associated with the SFA approach is lack of flexibility due to a priori specification of the production function and the error term.

In the practical application (paragraph 3), we chose DEA and Order-\(m\) nonparametric efficiency techniques in line with the approach followed in the calculation of composite indicators. Considering a production set as in equation (1), the Farrell-Debreu efficiency scores
for a given production scenario \((x, y) \in \Psi\) (here in a output orientation setting\(^9\)) can be defined as:

\[
\lambda(x, y) = \sup \{ \lambda | (x, \lambda y) \in \Psi \}
\]  
(5)

In practice \(\Psi\) is unknown and has to be estimated from a random sample of production units \(\chi = \{(X_i, Y_i) | i = 1, \ldots, n\}\), assuming \(Prb((X_i, Y_i) \in \Psi)\) (so called deterministic frontier models); the matter is therefore related to estimation of the support of the random variable \((X, Y)\) where \(\Psi\) is supposed to be compact.

If \(\Psi\) is convex:

\[
\tilde{\Psi}_{DEA} = \{(x, y) \in \mathbb{R}^{p+q} | y < \sum_{i=1}^{n} \gamma_i y_i \text{ for } (\gamma_1, \ldots, \gamma_n) \text{ such that } \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, \ldots, n\}
\]  
(6)

makes it possible to calculate the output oriented efficiency score \(\hat{\theta}(x, y)\) as:

\[
\hat{\theta}_{DEA}(x_0, y_0) = \sup \{ \lambda | (x_0, \lambda y_0) \in \Psi_{DEA} \}
\]  
(7)

To bypass one of the most critical issues of the deterministic DEA approach, i.e. lack of robustness, Daraio & Simar (2005) proposed a probabilistic formulation of efficiency (for a more complete theoretical summary see Fried & Lovell, 2008 and Daraio & Simar, 2007a). Considering a sample of \(m\) random variables with replacement \(S_m = \{Y_{i1}\}_{i=1}^{m}\) drawn from the density of \(Y\), the random set \(\tilde{\Psi}_m\) can be defined as:

\[
\tilde{\Psi}_m = \bigcup_{j=1}^{m} \{(x, y) \in \mathbb{R}^{p+q} | x \leq X, y \geq Y_j\}
\]  
(8)

Therefore the effect of an abnormal or outlier unit is damped because the single unit is not compared to all the others, but to a sample subset of size \(m\).

This generalization makes it possible to iteratively calculate a sample subset of size \(m\) (for \(b = 1, \ldots, B\) times) and for each \(b\) iteration the distance of the single unit from the maximum values; more formally:

\[
\tilde{D}_m^b(x, y) = \sup \{ \lambda | (x, \lambda y) \in \tilde{\Psi}_m \}, \forall b = 1, \ldots, B
\]  
(9)

Then we can calculate the (robust) Order-\(m\) efficiency \(\hat{\theta}_m\), following Cazals et al. (2002) and even in a Shepard formulation, as the empirical mean over \(B\):

\[
\hat{\theta}_m(x, y) = \frac{1}{\hat{E}(\tilde{D}_m^b(x, y))}
\]  
(10)

---

\(^9\)The output oriented approach was chosen to estimate the amount of services that Regions could provide with the same input.
2.3. Demand and expenditure function

The proposed model is focused on estimation of the public demand and supply functions as reported in equations (11) and (12) respectively:

\[ q = d(R, D, c) \]  
\[ c = s(P, S, q) \]

where \( q \) are the services offered; \( c \) the unit cost of services; \( R \) average regional income (GDP); \( D \) demand covariates (for example, population structure); \( P \) input prices and \( S \) contextual supply variables (such as household expenditure).

The dependent variable is considered per capita\(^{10}\).

2.3.1. Demand function

In order to simplify the estimation process, the reduced form of the demand function is obtained by substituting equation (12) in (11) and it is therefore equivalent to the model in equation (13) named output function:

\[ q = h(R, D, P, S) \]

Switching from the output to the empirical theoretical model first requires definition of dependent variable \( q \) that must correctly enclose the level of health services of each Region. To this end, a composite indicator \( CI_{it} \) for output was calculated for each region \( i \) and year \( t \) in order to capture the multi-output characteristic of the production function typical of the health sector\(^{11}\).

Taking advantage of the panel structure of our dataset, the empirical model can then be written as a linear panel fixed effects model as shown in equation (14):

\[ CI_{it} = \alpha_i + \eta_t + \beta_1 R_{it} + \beta_2 D_{it} + \beta_3 S_{it} + \epsilon_{it} \]

where \( i \) is the reference region; \( t \) the reference year; \( CI_{it} \) the output composite indicator; \( \beta \) the coefficients; \( \alpha \) the regional fixed effects; \( \eta \) the annual fixed effects and \( \epsilon \) the idiosyncratic error. Note that in the absence of detailed information on input (labour and capital) prices, their impact on spending is only approximated by

\(^{10}\)Other normalization criteria (e.g. the equivalent population for the consumption of hospital services, Francese & Romanelli, 2014) were not considered due to implicit inclusion of a demand variable in the cost structure.

\(^{11}\)Note that in the absence of the explicit local preferences set, the BoD procedure makes it possible to find the best weighting structure, which is equal to the other units without implicitly requiring a uniform distribution of local preferences.
regional fixed effects.
Finally the model in equation (14) is better specified by including a set of structural variables:

1. mobility flows in terms of net balance between entry and exit patients between Regions ($M_{it}$) in order to measure if regional output meets domestic demand or also the demand of patients from other regions. This indicator can also be interpreted, albeit indirectly, as a proxy for regional health service quality;
2. the average effect of the joint agreements (*Patti per la Salute*) ($z_{it}$) on service levels of Regions\textsuperscript{12} that signed them in the period 2006-2010 (last year of our sample).

Hence the final model of the output function can be specified as:

$$CI_{it} = \alpha_i + \eta_t + \beta_0 M_{it} + \sum_{t=2006}^{2010} \beta_t z_{it} + \beta_1 R_{it} + \beta_2 D_{it} + \beta_3 S_{it} + \epsilon_{it} \quad (15)$$

and effectively estimated with the *Within-the-Group* (WG) estimator; the estimated values $\hat{CI}_{it}$ can thus be defined as the standard output of each Region, taking as benchmark the Region that, for the same contextual variables, produces the highest performance per capita (i.e. the greatest fixed effect):

$$\hat{CI}_{it} = E[CI_{it}|\alpha_{max}, \eta_t, R_{it}, D_{it}, S_{it}] \quad (16)$$

Note that the effects of mobility and joint agreements are excluded from equation (16).

The difference between historical output $CI$ and standard output, namely the output gap ($w_{it}$), can be calculated by year $t$ and Region $i$ as:

$$w_{it} = CI_{it} - \hat{CI}_{it} \quad (17)$$

Finally, it is important to estimate the share of output gap due to different mobility of patients between regions ($w_{it,mob}$); in this way, for Regions with positive mobility, it is possible to more precisely identify a Region’s “own” output gap ($w_{it,own}$) as the output produced for its residents. Equations (18) and (19) show the analytical formulae.

$$w_{it,mob} = E[CI_{i,t}|M_{i,t}] \quad (18)$$

$$w_{it,own} = CI_{i,t} - \hat{CI}_{i,t} - w_{it,mob} \quad (19)$$

\textsuperscript{12}Lazio, Abruzzo, Molise, Campania, Calabria and Sicily.
2.3.2. The expenditure function

To simplify the estimation process, the cost function of the health services in reduced form can be obtained substituting equation (11) into (12); the cost function in equation (20) usually known in the literature as expenditure function.

\[ c = f(P, S, R, D) \]  

(20)

The longitudinal structure of our dataset makes it possible to use a linear panel fixed effects empirical model as shown in equation (21):

\[ H_{it} = \phi_i + \tau_t + \gamma_1 S_{it} + \gamma_2 R_{it} + \gamma_3 D_{it} + u_{it} \]  

(21)

where \( i \) is the reference region; \( t \) the reference year; \( H_{it} \) the current health expenditure; \( \gamma \) the coefficients; \( \phi_i \) the regional fixed effects; \( \tau_t \) the annual fixed effects and \( u_{it} \) the idiosyncratic error. Again, in this case, because of the lack of detailed information on input prices, the relative impact on spending is approximated through the regional fixed effects.

The public expenditure function model specified in equation (21) is very similar to the classical one; the originality of our approach lies in inclusion of two estimated variables among the regressors: nonparametric technical inefficiency \( \theta_{it} \) and by way of the estimated demand function, the output gap \( w_{it} \). Thus the final formulation of the expenditure function can be written as:

\[ H_{it} = \phi_i + \tau_t + \delta_1 \theta_{it} + \delta_2 w_{it} + \sum_{t=2006}^{2010} \gamma_t z_{it} + \gamma_1 S_{it} + \gamma_2 R_{it} + \gamma_3 D_{it} + \psi_{it} \]  

(22)

where the dummy \( z_{it} \) is again included in order to measure the contribution of the joint agreements on spending levels.

The first target of the proposed estimation strategy is to obtain consistent and unbiased estimates for: \( \delta_1 \), the share of expenditure due to technical inefficiency; \( \delta_2 \), the amount of expenditure needed to fill the deficit in terms of performance in the different Regions; \( \gamma_t \), the effectiveness of the joint agreements in reducing spending being equal to the socio-economic context and output level.

The second objective pertains to the estimation of price inefficiency for each regional system through regional fixed effects \( \phi_i \); taking the Region with the smallest fixed effect as benchmark, it is possible to measure the distance between the fixed effects and the minimum value for each Region.

The total value of inefficiency \( I_{it} \), therefore has two parts(equation (23)):

\[ I_{it} = \delta_1 \theta_{it} + [\hat{\phi}_i - \hat{\phi}_{\text{min}}] \]  

(23)
where the first component ($\theta_i$) measures technical inefficiency for the purpose of estimating possible savings linked to better use of input $[\dot{\phi}_i - \text{min}(\dot{\phi}_i)]$ measures price inefficiency or the possibility of using the same quantity of inputs with a lower unit cost. It is important to note that technical inefficiency is (by construction) a time variable, whereas price inefficiency is not; price inefficiency should therefore be interpreted as the average level for the entire period of analysis.

The last step of the analysis involves calculation of standard expenditure needs according to two specifications:

1. in the first specification, the standard needs ($\hat{H}_{it}^a$) take into account inefficiency $I_{it}$, but not differences in service levels throughout the territory, recognising standard expenditure compatible with historical services to each region.

$$\hat{H}_{it}^a = E[H_{i,t}|\phi_{\text{min}}, \delta_2 w_{it}, \tau_t, R_{i,t}, D_{i,t}, S_{i,t}]$$ (24)

2. in the second specification, in addition to inefficiency, standard needs ($\hat{H}_{it}^b$) take into consideration the share of expenditure required to fill the output gap ($\delta_2 w_{it}$); this second measure allows every Region to offer (at the efficient cost level) the maximum service level needed to fully satisfy the demand from its territory.

$$\hat{H}_{it}^b = E[H_{i,t}|\phi_{\text{min}}, \tau_t, R_{i,t}, D_{i,t}, S_{i,t}]$$ (25)

3. Application

3.1. Output composite measure

Estimation of the aggregate regional health level in terms of output produced in the years 1998-2010 by the Italian Regions is the main purpose of this section. Specifically, the output CI is an aggregate measure of the “quantity” of services provided. It must not be regarded as an indicator of appropriateness of the services or linked to the outcome of treatments because:

1. one of the main advantages of the BoD method is the possibility of obtaining the most objective set of weights; introducing additional considerations about value or appropriateness of each simple output would expose to unacceptable risk of distortion of the results;

---

13It is quite normal that results be strongly linked to the basic data available and its quality and uniformity throughout the national territory.
2. The correct definition of care appropriateness cannot be entered by economic researchers, but should primarily be set by the policy maker based on medical and/or epidemiological criteria; otherwise there is risk of confusing historical with desired service level;

3. In a cost framework, the output indicator should reflect the “quantity” of output produced independently of other criteria (appropriateness, sustainability or social relevance).

Selection of the output variables was therefore the first step of our application. To exclude spurious correlations between the simple indicators, Principal Component Analysis (PCA) was conducted to obtain independent and uncorrelated factors. Table 2 reports the principal orthogonal and independent factors\textsuperscript{14} for the simple output indicators. The first three PCA eigenvalues explain about 77\% of the total variance, readily explaining the single factors in the light of the variables most associated with them: factor 1 appears to be linked to the dimension variables; factor 2 to quality of service, such as interregional mobility and care, while factor 3 is linked to long term care and rehabilitation.

Figure 2 shows output CI dynamics for each Region from 1998 to 2010. Though with different magnitudes, an average increase in service level can be seen in almost all Regions. Lombardy, Veneto, Emilia Romagna and Piedmont show the highest level of performance.

3.2. Input composite measures

Using a similar analytical process, the input variables involved in the production process were analysed in order to summarize the information; Table 3 shows the PCA input factors (labour and capital). Some considerations about capital goods chosen for the analysis are necessary: as highlighted in the literature, especially for specialized health sectors, capital goods are not easily compared in terms of quality or substitution rate (between new and old goods). It is therefore desirable to include a subset of capital goods in the analysis that are comparable and shared by all regional units. From an “industrial” point of view these goods can help identify hospitals with “better” equipment. Although this issue is beyond the scope of the present paper, it may be a useful tool for improving estimate reliability. We therefore chose only basic hospital capital goods available in all Regions\textsuperscript{3}.

\textsuperscript{14} The printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 60 are marked with an asterisk. Values less than 30 are not printed; Rotation method: Varimax with Kaiser normalization
<table>
<thead>
<tr>
<th>Variables</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospitalization_total</td>
<td>96</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_acute</td>
<td>96</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_private</td>
<td>95</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_acute</td>
<td>94</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital</td>
<td>94</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_public_acute</td>
<td>94</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_public</td>
<td>94</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_private_acute</td>
<td>94</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_acute_private</td>
<td>93</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_acute_public</td>
<td>91</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_public</td>
<td>91</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_private</td>
<td>88</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_rehabilitation_private</td>
<td>82</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_rehabilitation</td>
<td>80</td>
<td>* 50</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_rehabilitation</td>
<td>80</td>
<td>* 41</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_rehabilitation_public</td>
<td>72</td>
<td>* 58</td>
<td>.</td>
</tr>
<tr>
<td>Days_hospital_rehabilitation_private</td>
<td>70</td>
<td>* 40</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_rehabilitation_public</td>
<td>70</td>
<td>* 58</td>
<td>.</td>
</tr>
<tr>
<td>Interregional_mobility_PRC</td>
<td>.</td>
<td>73 *</td>
<td>.</td>
</tr>
<tr>
<td>Interregional_mobility2_PRC</td>
<td>.</td>
<td>73 *</td>
<td>.</td>
</tr>
<tr>
<td>Assisted_per_doctor</td>
<td>.</td>
<td>48</td>
<td>.</td>
</tr>
<tr>
<td>Hospital_beds_rate</td>
<td>.</td>
<td>48</td>
<td>.</td>
</tr>
<tr>
<td>Assisted_per_pediatrician</td>
<td>.</td>
<td>48</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_mean_rehabilitation</td>
<td>.</td>
<td>. 91 *</td>
<td>.</td>
</tr>
<tr>
<td>Hospitalization_mean_rehabilitation_private</td>
<td>.</td>
<td>. 83 *</td>
<td></td>
</tr>
<tr>
<td>Hospitalization_mean_rehabilitation_public</td>
<td>.</td>
<td>. 74 *</td>
<td></td>
</tr>
<tr>
<td>Hospitalization_mean_private</td>
<td>.</td>
<td>52</td>
<td>64 *</td>
</tr>
<tr>
<td>Hospitalization_mean_acute_private</td>
<td>.</td>
<td>48 59</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Principal component analysis - output factors (per capita)
Figure 2: Robust output composite indicator dynamics by Region and year
Table 3: Principal component analysis - input factors (per capita)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technicians_employees_private</td>
<td>88</td>
<td>*</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Nurses_employees_public</td>
<td>86</td>
<td>*</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Technicians_employees_public</td>
<td>86</td>
<td>*</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Nurses_employees_private</td>
<td>84</td>
<td>*</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Rehabilitation_employees_public</td>
<td>81</td>
<td>*</td>
<td>-42</td>
<td>.</td>
</tr>
<tr>
<td>Rehabilitation_employees_private</td>
<td>74</td>
<td>*</td>
<td>-39</td>
<td>.</td>
</tr>
<tr>
<td>Nurses_employees_SSN</td>
<td>72</td>
<td>*</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Computed_tomography</td>
<td>.</td>
<td>77</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Blood_cell_counter</td>
<td>.</td>
<td>75</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Anesthesia_machine</td>
<td>.</td>
<td>70</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Tables_Fixed_Radiological_Systems</td>
<td>.</td>
<td>69</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Operating_tables</td>
<td>.</td>
<td>69</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Hyperbaric_chambers</td>
<td>-38</td>
<td>50</td>
<td>*</td>
<td>.</td>
</tr>
<tr>
<td>Doctors_employees_public</td>
<td>44</td>
<td>.</td>
<td>76</td>
<td>*</td>
</tr>
<tr>
<td>Doctors_employees_SSN</td>
<td>.</td>
<td>.</td>
<td>74</td>
<td>*</td>
</tr>
<tr>
<td>Doctors_employees_private</td>
<td>.</td>
<td>.</td>
<td>36</td>
<td>64</td>
</tr>
<tr>
<td>Outpatient_surgery_beds</td>
<td>.</td>
<td>.</td>
<td>53</td>
<td>38</td>
</tr>
<tr>
<td>Outpatient_surgery</td>
<td>.</td>
<td>.</td>
<td>92</td>
<td>*</td>
</tr>
</tbody>
</table>

The simple variables make it possible to better identify the first four factors (that explain more than 70% of the total variance). More specifically:

- Factor1 is positively correlated with nurses, technicians and staff rehabilitation;
- Factor2 is positively correlated with hospital capital goods;
- Factor3 is positively correlated with doctors (public and private);
- Factor4 is positively correlated with outpatient surgery.

To make the input factors more readable from an economic point of view, two input composite indicators were calculated:

- Input CI#1 = labour, composed of factor1 (nurses and technicians) and factor3 (doctors);
- Input CI#2 = capital, composed of factor2 (capital goods) and factor4 (beds).

Figure 3 is very explanatory: it shows the average dynamic of the input composite indicators (labour and capital) in the period 1998-2010. We see a strong increase in labour input until 2003, after which it remains almost constant, while capital input shows a constant decline, especially from 2003.
Figure 3: Average dynamic of the input CIs - labour and capital - per year
3.3. Technical production efficiency

Once service level (output) and resources (inputs) used in the production process have been analysed, the relative technical efficiency index (relative to the benchmarks) can easily be calculated for each region and each year; note that in the nonparametric formulation (DEA or Order-m), benchmarks are not identified a priori, but are the result of the linear optimization problem. More in particular, the technical efficiency score was calculated including two output - size (factor1, table 2) and quality/mobility (factor2, table 2) - and two input - labour CI and capital CI (see table 3 and figure 3).

The calculation\textsuperscript{15} was done with DEA and Order-m robust estimators, obtaining an excellent correlation between the two scores. Figure 4 shows the DEA technical efficiency dynamic by region and year (1998-2010). We see a limited increase in efficiency, especially after 2003, and a major increase (in absolute and relative terms) for Tuscany, Veneto, Lombardy, Emilia Romagna and Piedmont in the last years of analysis.

3.4. Demand function estimation and the output gap

In this section, the reduced form of the demand function (i.e. output function) for health services is estimated using the theoretical model shown in section 2.3.1 with output CI as dependent variable.

In line with the empirical model of reference (eq. 15), the following variables were included as regressors: (i) interregional mobility balances among hospitals as structural variables; (ii) regional GDP per capita (as income proxy), dividing the population into classes by age and covariates related to health condition as contextual demand variables; (iii) private health expenditure, a medical technological progress indicator and other covariates related to living conditions as contextual supply variables (used as instrumental variables for estimation of the demand function); finally, (iv) State-Regions joint agreements by year of introduction as institutional determinants.

The longitudinal panel structure of our dataset (20 Regions for 15 years from 1998 to 2012) makes it possible to estimate the linear panel “fixed effect” model using the Within-the-Group estimator, limiting territorial and time trend issues in line with equation (15).

Table 4 shows the coefficients estimated for the output and demand functions. Full model point estimates for the output function are reported in column 1 while columns 2 to 7 show the estimates related to partial models in order to evaluate

\textsuperscript{15}With R package FEAR, Wilson (2008).
Figure 4: Technical production efficiency dynamic per Region and year
the single contribution for categories of covariates. To assess coherence between
the empirical and theoretical models, columns 8 and 9 show the demand function
estimates; in this model, unlike in the output function, expenditure per capita is a
proxy for unit cost of service as opposed to supply contextual variables. In particular,
the estimation was carried out without instrument the current expenditure per capita
in column 8, and is repeated in column 9 using the supply contextual variables as
instrumental variables for current expenditure in order to test for endogeneity.

Table 4 highlights various interesting results: (i) interregional mobility is one
of the most significant factors for health care demand levels; (ii) the presence of
spending cut plans had a negative effect on output level, especially in the medium
term; it can therefore be deduced that part of the cost savings in the last years of
the period were not due to greater efficiency but reduced output level; (iii) the main
contextual demand covariates (GDP per capita, health condition and population
groups) show a coherent relationship with expenditure; (iv) coefficients estimated by
age group show that health service demand is concentrated in three groups: 0-4 years,
45-54 years and over 75 years; (v) despite the low level of statistical significance of
each individual coefficient, private expenditure and the other supply covariates show
the expected direction in terms of impact on demand.

When we take the estimated coefficients of the demand function into account (column
9), we observe that estimate strength increases after instrumenting expenditure per
capita with the supply contextual variables; this evidence highlights the endogeneity
issue of the demand function.

It is therefore necessary to use two-stage methods to correct endogeneity (in relation
to output). Another issue to be noted is that the reduced form of demand not only
makes it possible to overcome this problem, but also to avoid linking standard service
level and standard expenditure too directly.

Referring to the full model in column 1 of table 4, output standard is estimated
as expected service level. The difference between historical and standard output (the
output gap) can therefore be calculated for each region and year according to the
equation (17).

From an economic point of view, output gap indicates supply adequacy for each
region, taking a regional health system that maximizes performance per capita as
benchmark (in our application Lombardy) for given contextual variables. In other
terms, output gap shows the service levels per capita that other Regions with the

---

16Instrumental variable tests are also reported in column 9 and show a good correlation between
instrumental variables and output (Anderson canon. corr. LM statistic) as well as their substantial
exogeneity (Sargan statistic).
Table 4: Estimated demand function, output CI as dependent variable
same socio-economic context of reference could achieve if they operated in the same way as Lombardy.

The estimated the output gap of each Region\textsuperscript{17} is the combination of two components: (i) the output gap inside the Region, which measures the extent to which the Region is able to meet the demand of its citizens with respect to the benchmark; (ii) the output gap linked to mobility. If the mobility output gap is negative it increases the output gap inside the Region and it may not be necessary to isolate it; vice versa when it is positive, it must be identified and subtracted from the total output gap in order to correctly identify the within-region output gap.

3.5. Expenditure function estimation

As a consequence of the demand function estimation, the model shown in section 2.3.2 was used to estimate the reduced form of the health services cost function, i.e. the expenditure function.

In line with the equation (22), the applicative specification was designed like the output function except for the structural variables\textsuperscript{18}: in this case, the technical inefficiency index (section 3.3) and the output gap (section 3.4) are also included as structural variables.

Table 5 shows the cost and expenditure function coefficients, as in the previous table. Column 1 shows full model point estimates while columns 2-7 show partial pattern estimates. Cost function estimates are reported in columns 8 and 9; note that unlike the expenditure function, a second order polynomial of estimated output CI replaces the demand contextual variables. More in detail, as in table 4, the coefficients of the cost function were estimated without instruments (column 8), while these measures were repeated using the contextual demand covariates as instrumental variables (column 9) in order to test for endogeneity.

The results shown in table 5 confirm all the major results of the demand function with high levels of statistical significance. In addition to the previous results, it is important to emphasize that: (i) the technical inefficiency index and the output gap have positive signs and a high degree of statistical significance, meaning that a large part of historical expenditure is linked to input excess and/or a service level not in line with the standard reference; (ii) the

\textsuperscript{17}According to equations (18) and (19) and using the mobility balance coefficient reported in column 1, table 4.

\textsuperscript{18}This issue highlights the specularity of the two models in line with the theoretical reference framework.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Expenditure function</th>
<th>Cost function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS (1)</td>
<td>OLS (2)</td>
</tr>
<tr>
<td>Technical inefficiency</td>
<td>127.2</td>
<td>199.3</td>
</tr>
<tr>
<td>Output gap</td>
<td>372.2</td>
<td>570.0</td>
</tr>
<tr>
<td>SPENDING CUT JOINT AGREEMENTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spending cut joint agreements (t+1)</td>
<td>26.21</td>
<td>47.06</td>
</tr>
<tr>
<td>Spending cut joint agreements (t+2)</td>
<td>5.864</td>
<td>22.46</td>
</tr>
<tr>
<td>Spending cut joint agreements (t+3)</td>
<td>-38.18</td>
<td>-8.941</td>
</tr>
<tr>
<td>Spending cut joint agreements (t+4)</td>
<td>-46.99</td>
<td>-40.99</td>
</tr>
<tr>
<td>CONTEXTUAL DEMAND COVARIATES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP at market prices - (€) real per capita (base 2005)</td>
<td>0.0399</td>
<td>0.0294</td>
</tr>
<tr>
<td>Resident population 0-4 M+F - % total pop.</td>
<td>0.0668***</td>
<td>0.064***</td>
</tr>
<tr>
<td>Resident population 5-14 M+F - % total pop.</td>
<td>81.99</td>
<td>66.35</td>
</tr>
<tr>
<td>Resident population 15-24 M+F - % total pop.</td>
<td>163.4</td>
<td>158.5</td>
</tr>
<tr>
<td>Resident population 25-34 M+F - % total pop.</td>
<td>65.99</td>
<td>5.766</td>
</tr>
<tr>
<td>Resident population 45-54 M+F - % total pop.</td>
<td>202.8</td>
<td>164.1</td>
</tr>
<tr>
<td>Resident population 55-64 M+F - % total pop.</td>
<td>73.23</td>
<td>75.35</td>
</tr>
<tr>
<td>Resident population 65-74 M+F - % total pop.</td>
<td>103.9</td>
<td>145.2</td>
</tr>
<tr>
<td>% Heavy smokers 15+ M+F</td>
<td>-1.487</td>
<td>-3.139</td>
</tr>
<tr>
<td>% People consuming vegetables once a day 3+ M+F</td>
<td>-1.331</td>
<td>-1.262</td>
</tr>
<tr>
<td>% People consuming fish once a week 3+ M+F</td>
<td>-0.144</td>
<td>-1.825</td>
</tr>
<tr>
<td>% People consuming beef occasionally once a week 3+ M+F</td>
<td>-1.366</td>
<td>-0.962</td>
</tr>
<tr>
<td>% People consuming cheese at least once a week 3+ M+F</td>
<td>-0.549</td>
<td>-2.352</td>
</tr>
<tr>
<td>% People proper breakfast 3+ M+F</td>
<td>-0.393</td>
<td>-0.177</td>
</tr>
<tr>
<td>% People main meal dinner 3+ M+F</td>
<td>0.669</td>
<td>0.459</td>
</tr>
<tr>
<td>Malignant tumours incidence rate 0-84 M</td>
<td>5.619</td>
<td>0.146</td>
</tr>
<tr>
<td>Malignant tumours incidence rate 0-84 M (square)</td>
<td>-0.04865</td>
<td>0.0000683</td>
</tr>
<tr>
<td>HOUSEHOLD HEALTH EXPENDITURE (%)</td>
<td>-20.19</td>
<td>-16.54</td>
</tr>
<tr>
<td>Technological progress index in medical supply</td>
<td>-2.77</td>
<td>-4.79</td>
</tr>
<tr>
<td>Activity rate 15+ F</td>
<td>5.692</td>
<td>3.830</td>
</tr>
<tr>
<td>% pop. with university degree M+F</td>
<td>9.069</td>
<td>2.335</td>
</tr>
<tr>
<td>% families complain about noise pollution</td>
<td>0.784</td>
<td>0.878</td>
</tr>
<tr>
<td>% families complain about air pollution</td>
<td>-0.099</td>
<td>-0.101</td>
</tr>
<tr>
<td>Life expectancy 75 M</td>
<td>-32.30</td>
<td>-47.06</td>
</tr>
<tr>
<td>Life expectancy 75 F</td>
<td>-3.693</td>
<td>-2.77</td>
</tr>
<tr>
<td>OUTPUT COVARIATES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output composite indicator (Robust BoD)</td>
<td>1595.6</td>
<td>941.9</td>
</tr>
<tr>
<td>Output composite indicator (Robust BoD) - square</td>
<td>-0.064</td>
<td>-0.499</td>
</tr>
<tr>
<td>Num. observations</td>
<td>244</td>
<td>244</td>
</tr>
<tr>
<td>Regional fixed effects</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Yearly fixed effects</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Robust R²</td>
<td>0.982</td>
<td>0.972</td>
</tr>
<tr>
<td>Standardized coefficients</td>
<td>0.0323</td>
<td>0.1197</td>
</tr>
</tbody>
</table>

Table 5: Cost and expenditure function, dependent variable = current public health expenditure per capita
specific joint agreements are linked (on average) to a reduction in expenditure, but not until two years of application; the statistical non significance of the coefficients and the empirical evidence obtained by estimation of the demand function (agreements had a negative impact on service levels) confirm the limited effectiveness of these tools; (iii) the major demand contextual covariates (per capita GDP, health condition and population) show a coherent relation with expenditures especially due to the expected U-trend of the age groups; finally, (iv) household private spending shows an expected negative sign and high significance, in line with the principle of substitution between private and public spending.

Considering the cost function estimation (with instrumental variables\textsuperscript{19}) reported in column 9, it is interesting to note that the strong positive relationship between output and expenditure remains significant and that larger regional health care systems can take advantage of growing economies of scale.

3.6. Standard and historical expenditure assessment

Considering the expenditure function coefficient estimates (column 1, table 5), the share of current expenditure due to technical inefficiency can easily be measured at regional level and over the years. Figure 5 shows the distribution of total technical inefficiency as a percentage of historic expenditure from 1998 to 2010, highlighting a declining trend especially since 2003 (from 9% of historical expenditure in 1998 to 4% in 2010). This empirical evidence is essentially due to the joint agreements put in place in the second half of the year 2000. These agreements allowed a slowdown in growth rates of current expenditure and considerably improved efficiency, largely thanks to savings on the input side.

As described in section 2.3.2, the price inefficiency can be calculated taking into account the difference between the fixed effects of each region and the fixed effect of Lombardy (the benchmark, with the lowest fixed effect). Since the difference is time-invariant, it must be interpreted as the average value over the study period in each regional system. Figure 6 shows the distribution of technical and price inefficiency compared to historical expenditure, year 2012. Two main findings stand out: (i) technical inefficiency, in monetary terms, is minimal on average (2% of the historical expenditure) and is only significantly higher in the southern regions; (ii) in order to decrease inefficiency,

\textsuperscript{19}Instrumental variable tests are reported in column 9 of table 5, and show a good correlation between the instrumental variables and output (Anderson canon. corr. LM statistic) and substantial exogeneity of the instrumental variables (Sargan statistic).

25
Figure 5: Technical inefficiency distribution as a percentage of historical expenditure.
it is therefore necessary to act mainly on input prices rather than input quantity.

The estimated coefficients of the expenditure function (column 1, table 5) can also be used to evaluate the two components of the output gap ($w_{it, mob}$, eq. (18) and $w_{it, own}$, eq. (19)) also in monetary terms, for each regional health system. Figure 7 shows the trend of the global output gap as a percentage of historical expenditure. Again, the very strong increasing trend up to 2005 (from -16% in 1998 to -10% in 2005) loses strength from 2006 onwards (remaining constant at -10%).

Finally, the estimated coefficients of the expenditure function make it possible to estimate regional standard expenditure needs according to equations (24) and (25), section 2.3.2: the former ($\hat{H}_{iit}^a$) is calculated as the estimated expenditure deducted only from inefficiency without taking standard output level into account (that is, acknowledging the historical service level of each Region), while the latter ($\hat{H}_{iit}^b$) also considers inefficiency and the share of expenditure required to fill the output gap in order to satisfy regional demand, even in a cost efficient setting.

Figure 8 shows the differences between global standards ($\hat{H}_{iit}^a_{2012}$ and $\hat{H}_{iit}^b_{2012}$) and historical expenditure as a percentage of historical expenditure, year 2012. As regards $\hat{H}_{iit}^a_{2012}$, differences always appear negative and higher than $\hat{H}_{iit}^b_{2012}$; in this
Figure 7: Distribution of the output gap as a percentage of the historical expenditure.
case, standard needs may exceed historical expenditure (for Umbria, Marches and Tuscany) by virtue of lower inefficiency combined with a larger output gap. Lombardy, which takes the role of benchmark for output gap and price efficiency, has standard needs that are nearly identical in the two cases of standardization and in line with historical expenditure, year 2012.

Finally, Italian economic duality is also reflected in regional health systems in terms of differences between historical and standard expenditure needs. If on one hand Southern Regions show a higher output gap due to higher inefficiency, on the other in the last few years Northern Regions were able to obtain more than standard resources (see table 8 for complete results).

4. Efficiency, output gap, appropriateness of services and deficit

Finally, it is useful to evaluate the correlation between global output gap and global inefficiency (both expressed in monetary terms) with some health care appropriateness indicators and with health deficit variations. Estimates of the two
monetary values were made using a plurality of appropriateness indices and the annual variation in health deficit by means of a panel fixed effects model. Given these premises, one indicator of efficiency, one of complexity and two appropriateness indicators\textsuperscript{20}, characterized by persistent regional variability in time, were selected:

- the Comparative Performance Index (\textit{CPI}): \textit{CPI} is the ratio of standardized average hospitalization per case-mix to average hospitalization of the reference standard\textsuperscript{21}; its polarity is negative;

- the Case Mix Index (\textit{CMI}): the index is useful to evaluate the different complexity of treated cases; it is calculated as the ratio of average weighted hospitalization to average weighted admission of the national standard\textsuperscript{22}; its polarity is positive;

- the percentage of medical Diagnosis Related Groups for surgical wards (\textit{DRG}): The \textit{DRG}, is an organizational suitability indicator and was proposed in the joint agreement between Central government and Regions in 2010 to measure the appropriateness of surgical ward use; it is the ratio of the number of patients discharged from DRG surgical wards to the total number of patients discharged from surgical wards; its polarity is negative (lower indicator, greater appropriateness);

- the percentage of Caesarean birth on total (\textit{Caesarean}): the index is a clinical appropriateness indicator, calculated as the ratio of the number of caesarean births to total births; its polarity is negative (lower value, greater benefit for patients and less cost).

Table 6 clearly shows that both inefficiency and output gap have a high correlation with appropriateness indices and are therefore a good proxy for them.

\textsuperscript{20}Note that the difference between “clinical appropriateness” and “organizational appropriateness” was introduced in the Italian National Health Plan, year 1998-2000. The former relates to the provision of medical care and intervention of proven effectiveness in contexts characterized by a favourable benefit-risk profile for the patient. The second concerns the choice of the most suitable delivery mode to maximize safety and patient well-being and to optimize production efficiency and resource consumption.

\textsuperscript{21}\textit{CPI} makes it possible to measure and compare the efficiency and effectiveness of different units with respect to the standard: values above 1 indicate lower efficiency than the standard, and values below 1 reflect higher efficiency than the standard.

\textsuperscript{22}Higher values indicate higher case complexity than the standard.
More specifically, columns (1) to (5) of table 6 show the relationship of the appropriateness indices to inefficiency per capita (in monetary terms). Given the polarity of each indicator, it is clear that if CPI, CMI and Caesarean increase, inefficiency increases, while it tends to decrease with increasing DRG index. The relative impact of the indicators seems to be very similar, slightly more than 1 € per capita in relation to a deviation of 1% from the national average, except for CMI that has a stronger impact (approximately 2 € per capita). Columns (6) to (10) of table 6 show symmetrically the impact of the appropriateness indices on output gap per capita (also in monetary terms). The positive relationship between CPI and output gap (€0.90 per capita for every 1% deviation from the national average) and the negative correlation with the DRG index (€0.80 per capita for every 1% deviation from the national average) are evident; no statistically significant impact of CMI and Caesarean on the output gap was found.

The covariates were expressed as percentage deviation from the national average in order to more easily compare inefficiency impact; thus, each coefficient represents the change in inefficiency in € per capita when the appropriateness index deviates by 1% from the national average.

The covariates are again expressed as percentage deviation from the national average in order to more easily compare the impact on output gap.

Table 6: Appropriateness indices, inefficiency and output gap (€ per capita, covariates expressed as a % deviation from national average)

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Percent of technical and price inefficiency</th>
<th>Output gap (own and mobility linked)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>CPI</td>
<td>1.159</td>
<td>-0.0314</td>
</tr>
<tr>
<td>ICM</td>
<td>2.073</td>
<td>3.076</td>
</tr>
<tr>
<td>DRG</td>
<td>-1.389</td>
<td>-1.298</td>
</tr>
<tr>
<td>Caesarean</td>
<td>1.114</td>
<td>1.303</td>
</tr>
</tbody>
</table>

N 246 246 246 246 246 260 260 260 260 260
R² 0.978 0.973 0.975 0.972 0.977 0.966 0.962 0.964 0.965 0.963
Regional fixed effects Yes Yes Yes Yes Yes Yes Yes Yes Yes
Annual fixed effects Yes Yes Yes Yes Yes Yes Yes Yes Yes
In conclusion, table 7, shows the relationship between annual percentage variation in inefficiency, global output gap and annual percentage variation in deficit. The results clearly show that the deficit reduction, implemented by the Regions since 2006 for budgetary needs, does not seem to have any statistically significant impact on inefficiency, whereas it has a negative impact on the output gap. Spending reduction was therefore mainly achieved by reducing the quantity of health services supplied rather than by reducing inefficiency.

<table>
<thead>
<tr>
<th>Δ Deficit (Perc.)</th>
<th>Δ Inefficiency (Perc.)</th>
<th>Δ Output gap (Perc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.000235</td>
<td>-0.000573</td>
</tr>
<tr>
<td></td>
<td>[0.960]</td>
<td>[0.000]***</td>
</tr>
</tbody>
</table>

N 73 80

R² 0.333 0.618

Regional fixed effects Yes Yes

Annual fixed effects Yes Yes

Table 7: Impact of annual % change in deficit on inefficiency and on global output gap

5. Concluding remarks

In this paper, the levels of services provided (standard and historical), the inputs used and the level of efficient spending of each regional system were considered explicitly and in detail for the first time, along with estimation of health spending. The new concept of output gap, i.e. the difference between historical output and standard output measured with reference to the structure of the resident population, was used. Besides the output gap, it was possible to measure the level of expenditures attributed to the supply of services over or under the standard.

It is important to underline that without considering the standard level of services that Local Authorities must provide in relation to their standard level of expenditure, the whole process of intergovernmental fiscal equalisation may not achieve its final goal, namely, the production of local services that provide a minimum level of welfare to all citizens irrespective of the socio-economic characteristics of the jurisdiction in which they live.

The effort was therefore to combine in a single framework the need for more accurate allocation of intergovernmental transfers and a higher level of service thanks to the increased efficiency of production factors. Equity and efficiency are two aspects that must be pursued and evaluated together in order to increase collective well-being. In these terms, the proposed framework may help policy makers to better design regional
improvement plans so that they take regional specificities, areas of improvement and specific health demand into account.

The final results show that spending reduction was achieved mainly by modifying the quantity of health services supplied rather than by reducing inefficiency, but not uniformly throughout the territory: the technical inefficiency gap between the Centre-Northern and the Southern regions worsened over time. In 2010, the Southern regions were almost twice as inefficient as Northern and Central regions, not only due to lower levels of quality and quantity of services offered, but also due to incorrect input ratios.

Finally, this paper exploited recent methodological improvements, using the BoD composite indicators technique to estimate the aggregate level of input and output and Order-$m$ efficiency (Daraio & Simar, 2007a) to assess robust technical efficiency. Further improvements will focus mainly on better estimation of price inefficiency (no longer based on regional fixed effects) and on greater territorial subdivision (from Regions to local health boards).

References


36
<table>
<thead>
<tr>
<th>Region</th>
<th>Resident population 2012</th>
<th>Historical expend. 2012</th>
<th>Technical inefficiency</th>
<th>Price inefficiency</th>
<th>Total inefficiency</th>
<th>Own output gap</th>
<th>Mobility output gap</th>
<th>Total output gap</th>
<th>Standard expend. net of total inefficiency</th>
<th>Standard expend. net of total historical inefficiency</th>
<th>Difference standard - historical with output gap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
<td>(D)</td>
<td>(E = C+D)</td>
<td>(F)</td>
<td>(G)</td>
<td>(H = C+D)</td>
<td>(I = B-E)</td>
<td>(J = I-B)</td>
<td>(K = B-E-H)</td>
</tr>
<tr>
<td>Piemonte</td>
<td>4.366.968</td>
<td>8.223</td>
<td>12</td>
<td>398</td>
<td>410</td>
<td>-124</td>
<td>-20</td>
<td>-144</td>
<td>7.813</td>
<td>-410</td>
<td>7.057</td>
</tr>
<tr>
<td>Valle d’Aosta</td>
<td>127.215</td>
<td>280</td>
<td>0</td>
<td>63</td>
<td>63</td>
<td>-31</td>
<td>-7</td>
<td>-38</td>
<td>217</td>
<td>-63</td>
<td>255</td>
</tr>
<tr>
<td>Lombardia</td>
<td>9.747.312</td>
<td>18.130</td>
<td>252</td>
<td>0</td>
<td>252</td>
<td>-52</td>
<td>221</td>
<td>170</td>
<td>17.878</td>
<td>-252</td>
<td>17.708</td>
</tr>
<tr>
<td>Trentino Alto Adige</td>
<td>1.034.856</td>
<td>2.197</td>
<td>36</td>
<td>239</td>
<td>275</td>
<td>-198</td>
<td>-9</td>
<td>-207</td>
<td>1.922</td>
<td>-275</td>
<td>2.129</td>
</tr>
<tr>
<td>Friuli Venezia Giulia</td>
<td>1.220.058</td>
<td>2.506</td>
<td>57</td>
<td>296</td>
<td>352</td>
<td>-224</td>
<td>13</td>
<td>-211</td>
<td>2.154</td>
<td>-352</td>
<td>2.364</td>
</tr>
<tr>
<td>Emilia Romagna</td>
<td>4.359.958</td>
<td>8.188</td>
<td>75</td>
<td>983</td>
<td>1.057</td>
<td>-643</td>
<td>173</td>
<td>-470</td>
<td>7.131</td>
<td>-1.057</td>
<td>7.601</td>
</tr>
<tr>
<td>Toscana</td>
<td>3.681.254</td>
<td>6.814</td>
<td>0</td>
<td>454</td>
<td>454</td>
<td>-647</td>
<td>82</td>
<td>-565</td>
<td>6.360</td>
<td>-454</td>
<td>6.925</td>
</tr>
<tr>
<td>Umbria</td>
<td>884.533</td>
<td>1.647</td>
<td>6</td>
<td>75</td>
<td>81</td>
<td>-209</td>
<td>13</td>
<td>-197</td>
<td>1.566</td>
<td>-81</td>
<td>1.762</td>
</tr>
<tr>
<td>Marche</td>
<td>1.543.333</td>
<td>2.778</td>
<td>84</td>
<td>299</td>
<td>293</td>
<td>-367</td>
<td>4</td>
<td>-370</td>
<td>2.485</td>
<td>-293</td>
<td>2.855</td>
</tr>
<tr>
<td>Abruzzo</td>
<td>1.309.551</td>
<td>2.331</td>
<td>75</td>
<td>496</td>
<td>571</td>
<td>-330</td>
<td>-39</td>
<td>-369</td>
<td>1.760</td>
<td>-571</td>
<td>2.129</td>
</tr>
<tr>
<td>Molise</td>
<td>313.282</td>
<td>651</td>
<td>20</td>
<td>142</td>
<td>163</td>
<td>-57</td>
<td>14</td>
<td>-44</td>
<td>488</td>
<td>-163</td>
<td>532</td>
</tr>
<tr>
<td>Puglia</td>
<td>4.051.500</td>
<td>7.159</td>
<td>298</td>
<td>1.489</td>
<td>1.787</td>
<td>-714</td>
<td>-56</td>
<td>-770</td>
<td>5.372</td>
<td>-1.787</td>
<td>6.141</td>
</tr>
<tr>
<td>Basilicata</td>
<td>577.049</td>
<td>1.056</td>
<td>49</td>
<td>199</td>
<td>248</td>
<td>-127</td>
<td>-24</td>
<td>-151</td>
<td>808</td>
<td>-248</td>
<td>959</td>
</tr>
<tr>
<td>Calabria</td>
<td>1.958.285</td>
<td>3.380</td>
<td>173</td>
<td>828</td>
<td>1.001</td>
<td>-359</td>
<td>-123</td>
<td>-482</td>
<td>2.379</td>
<td>-1.001</td>
<td>2.860</td>
</tr>
<tr>
<td>Sicilia</td>
<td>5.000.000</td>
<td>8.740</td>
<td>324</td>
<td>1.558</td>
<td>1.883</td>
<td>-1.054</td>
<td>-104</td>
<td>-1.159</td>
<td>6.857</td>
<td>-1.883</td>
<td>8.016</td>
</tr>
<tr>
<td>Sardegna</td>
<td>1.639.012</td>
<td>3.319</td>
<td>133</td>
<td>763</td>
<td>836</td>
<td>-195</td>
<td>-23</td>
<td>-219</td>
<td>2.483</td>
<td>-836</td>
<td>2.701</td>
</tr>
</tbody>
</table>

| Italy (total)         | 59.543.622               | 110.266                 | 2.220                  | 12.866            | 15.086            | -7.098         | 9                    | -7.069         | 95.180                                    | -15.086                                        | 102.249                                         | -8.017                                         |