UNIVERSITÀ DEGLI STUDI DI ROMA "TOR VERGATA"

FACOLTA' DI ECONOMIA

DOTTORATO DI RICERCA IN

ECONOMETRIA ED ECONOMIA EMPIRICA

CICLO DEL CORSO DI DOTTORATO:
XXI

TITOLO:

Essays on Economics of Health and Social Care

CANDIDATO:
Francesco D'Amico

A.A. 2009/2010

Docente Guida: Prof. Vincenzo Atella

Coordinatore: Prof. Franco Peracchi

Dedico questo lavoro a tutti coloro* che hanno avuto la pazienza di sostenermi e sopportarmi.

(*tra gli altri in particolare: Alessandro, Andrea, Antonio, Carlo, Cate, Ciriaco, Doğan, Domenico, Eduardo, Federica, Federico, Fiammetta, Franca, Francesca, Gabriele, Giuseppe, Giuseppe, Jonathan, Lucia, Marco, Margherita, Natalia, Nebibe, Noemi, Rubén, Sergio, Silvio, Silvio, Tom, Valentina, Vincenzo).

Un sentito ringraziamento al Prof. Vincenzo Atella, al Prof. Franco Peracchi, a Silvio Daidone e Jose' Luis Fernandez.

Abstract

Technical progress, prevention and patient health outcomes: a new look of technical progress in health care

The aim of the paper is to disentangle the roles that patients, physicians and technology can have on patient health outcomes using a large and detailed dataset of Italian patients collected by the Italian College of General Practitioners (SIMG) over the period 2001-2006. As our data show the existence of heterogeneity in the time needed to reach an optimal level of health stock, we concentrate on this measure of health outcome rather than simply on the level of patient health stock. The empirical work will then be based on two different analyses. We first explore whether patients recovering faster (in terms of time needed to reach cholesterol levels suggested by international clinical guidelines) exhibit a lower hospitalization rate for cardiovascular diseases and then the determinants of the speed of recovery to a good health status. The results confirm that a 10% increase in the speed of recovery can reduce hospitalization rates by 0.8%. Furthermore, we show that recovering to a good health status is a multifaceted phenomenon and that each single actor plays an important role in reducing the time needed to achieve the therapeutic goal, with technology that explain at the best 62% of the combined effect. These results are then discussed in terms of policy.

Stochastic Frontiers and Technical Efficiency: evidences from a panel of Italian hospitals

We evaluate how the productive structure and level of specialization of a hospital affect technical efficiency by analyzing a six-year panel database (2000/2005) drawn from hospital discharge records and Ministry of Health data. We adopt a distance function approach, while measuring the technical efficiency level with stochastic frontier techniques. After controlling for environmental variables and hospital case-mix, inefficiency is negatively associated with specialization and positively associated with capitalization. Capitalization is typical of private structures which, on average, use resources less efficiently with respect to public and not-for-profit hospitals. Finally, by looking at scale elasticities, we find some evidence of unexploited economies of scale, leaving room for centralization.

A stochastic frontier approach to assess the efficiency of English councils with social services responsibility

Using a stochastic frontier approach, this paper explores efficiency in the commissioning of publicly funded social services among 148 English Councils through a six-year panel database (2002-2007). It creates a specific framework for the analysis of social services commissioning, covering institutional care, day-care centres and care at home services. Our estimates of economic inefficiency provide key policy evidence about how to organise and commission social care in a context in which optimization is critical both for social and financial purposes. The results suggest a slight decrease through time in the average inefficiency score, whose sample average moves from 1.080 in the first year to 1.076 in the last year of the panel. Residential and nursing care show a cost-output elasticity around 7 times greater than that of home care and 15 times greater than that of

other community care services. Significantly, greater savings are obtained when the market is open to private and voluntary providers.

Keywords: Technical Progress, Adjustment Process, Cholesterol, Statins; Stochastic Frontiers, Distance Functions, Technical Efficiency; Expenditure Efficiency, Long Term Care, Commissioning

Abstract

Technical progress, prevention and patient health outcomes: a new look of technical progress in health care

L'obiettivo di questo lavoro e' la stima quantitativa del diverso ruolo che pazienti, medici e tecnologia medica hanno sull'esito dei trattamenti sanitari. L'analisi e' basata sull'uso di una vasta e dettagliata banca dati di pazienti italiani raccolta dalla SIMG (Società Italiana di Medicina Generale) per il periodo 2001-2006. Nel caso particolare di pazienti in terapia statinica per malattie cardio-vascolari, i dati mostrano l'esistenza di eterogeneità per quanto riguarda la velocità con la quale essi raggiungono il livello di colesterolo ottimale. L'analisi empirica si concentra su due livelli: il primo di essi analizza se pazienti che mostrano una convergenza più rapida al livello di colesterolo ottimale, ottengono anche benefici in termine di riduzione dell'ospedalizzazione. I risultati confermano che un incremento del 10% della velocità di convergenza riduce dello 0,8% la probabilita' di ospedalizzazione. Ad un secondo livello, si dimostra empiricamente che tale convergenza e' un fenomeno nel quale ognuno degli attori in causa (paziente, medico e tecnologia) gioca un ruolo significativo, con la la tecnologia che riesce a spiegare fino al 62% dell'effetto totale. I risultati ottenuti forniscono spunti di discussione in termini di politica sanitaria.

Stochastic Frontiers and Technical Efficiency: evidences from a panel of Italian hospitals

In questo paper si studia come l'efficienza tecnica ospedaliera venga influenzata della struttura produttiva e del livello di specializzazione. I dati disbonibili provengono dal Ministero della Sanita' e si basano sulle schede di dimissione ospedaliera (SDO) raccolte nel periodo 2000-2005 nella Regione Lazio. L'analisi empirica effettuata utilizza un approccio di fontiera stocastica con funzione di distanza output e input oriented. Il modello adottato controlla per un gruppo di variabili ambientali e per il case-mix ospedaliero. L'inefficienza stimata risulta associata negativamente con la specizlizzazione e positivamente con la capitalizzazione delle strutture. La capitalizzaztione della struttura produttiva e' tipica degli ospedali privati i quali, in media, utilizzano le risorse in maniera meno efficiente rispetto a quelli pubblici e non-profit. Infine, dall'osservazione dell'elasticita' di scala, e' possibile notare come ci sia evidenza di economie di scala inesplorate, suggerendo l'ipotesi di un maggiore accorpamento delle strutture.

A stochastic frontier approach to assess the efficiency of English councils with social services responsibility

Attraverso l'uso di un modello di frontiera stocastica, in questa analisi vengono esplorate le dinamiche dell'efficienza relative al commissioning dei servizi sociali da parte di 148 municipalita' inglesi, per gli anni inclusi tra il 2002 ed il 2007. Nel lavoro viene creato un modello specifico per il caso dell'assistenza di tipo residenziale, nei centri polivalenti e a domicilio. Stimare l'efficienza di questa componente della spesa pubblica risulta fondamentale nell'ambito dell'ottimizzazione delle decisioni relative all'organizzazione e alla gestione dei servizi sociali. I risultati suggeriscono una leggera riduzione dell'inefficienza media nel tempo, la quale passa da un valore di 1,080 nel primo anno a un valore di 1,076 nell'ultimo. L'elasticita' stimata della spesa rispetto al livello dell'assistenza residenziale e' sette volte superiore a quella domiciliare e quindici volte superiore rispetto agli altri tipi di assistenza non-residenziale. Si evidenzia inoltre che maggiori risparmi di spesa sono ottenuti quando il mercato e' allargato in misura maggiore agli operatori privati e appartenenti al terzo settore.

Keywords: Progresso Tecnologico, Processo di Aggiustamento, Colesterolo, Statine; Frontiere Stocastiche, Funzione di Distanza, Efficienza Tecnica; Efficienza della Spesa, Long Term Care, Commissioning

Index

Chapter 1

Technical progress, prevention and patient health outcomes: a new look of technical progress in health care[1]

1.1 Introduction

Patient health outcomes are complex phenomena and disentangling empirically their determinants is an ever more complex task. This complexity mainly arises due to the interplays of three different actors that individually have an important role in determining patient final health outcomes: the patient itself, the physician and available technologies. In fact, a valid diagnosis from a physician could not sort the hoped effect just because treatments are not available or are not effective; similarly, the availability of effective treatment cannot produce the desired effect as long as the physician is not able to properly match patient disease and treatment, or because patients are not compliant to physician recommendations. Despite its importance from a policy perspective, to our knowledge no study has ever tried to empirically disentangle and measure the different roles that these actors jointly play on patient health outcomes.

Within the literature these themes have been approached by both economists and health professionals. Economists have mostly been engaged in exploring the sole role of technology in the health care sector. As result, today we know that technical progress is partly responsible for the substantial real growth in expenditure experienced by almost all countries over the past few decades (Cutler and McClellan, 2001, TECH Research Network, 2001, Atella *et al.*, 2003, Jones, 2005 and Valletta, 2007). However, it remains unresolved whether better medical care has a role in explaining the health outcome improvements of the worldwide population. In particular, what is still a matter of debate is whether the ongoing rising expenditure on health care is "worth it" (Cutler and McClellan, 2001, and Valletta, 2007). Often, this lack of knowledge is due to absence of evidence on the role that patients and physicians play in making available new medical technologies effective. Providing drugs to patients who are not compliant reduces aggregate drug efficacy. Similarly, making available new technologies that are misused by physicians is, again, a waste of resources that reduces new technology aggregate efficacy.

In the medical literature these aspects have been approached in a more accurate way using patient level data, although, quite often, the analyses have been limited in terms of patients enrolled and, furthermore, not all aspects of the problem have been analyzed jointly. In fact, apart from articles that evaluate drug (new technologies) efficacy by means of clinical trials, there exists an ample literature that looks at the interplay of patient compliance and drug treatment, where the role of physician is missing (E.

---

1 This chapter is based on "Who is responsible for your health: you, your doctor or new technologies?" by F. D'Amico and V. Atella

Johansson et al., 1996, or McCombs et al, 2004), or analyses based on physician management strategies, with patients and technologies that are missing (Hyman, Pavlik, and Vallbona, 2000 or Gurmankin et al, 2002).

In our view, the main reason that has prevented researchers to succeed in this task is the lack of a sufficient level of clinical and socio-demographic detail when using “micro” data, even for those studies that adopt a disease-specific approach (McClellan and Kessler, 2002, Atella *et al.*, 2003, Grootendorst et al, 2007). Furthermore, objective measures of health outcomes are needed. Finally, to be informative for policy makers the analyses have to be conducted at population level rather than on small samples of patients (such as those used in clinical trials). Clearly, such an approach is highly data demanding, as it requires information on patient health profiles at the beginning of the observation period (initial conditions), the treatments and events to which they have been exposed over time, their health outcomes has to be compared before and after any specific treatment, and physician treatment strategies must be controlled. This implies to adopt an investigation strategy similar to that used in randomized clinical trials, but extended to the population.

By adopting a disease specific approach, our aim with this paper is to try to fill this gap by disentangling and measuring the contribution that each of the above mentioned actors can have on the health outcomes of a large representative population of Italian patients suffering of hypercholesterolemia and treated only with a specific class of drugs, namely the statins.[2] Hypercholesterolemia is a particularly interesting condition to analyze in this context, for at least three different reasons. First of all, there has been a growing public concern over the last 10 years about high levels of cholesterol in the population, that may have changed patient awareness about the problem, influencing their behavior (mainly through changes in compliance rates) over time. In fact, excessive levels of cholesterol are among the main contributing factors to the insurgence of angina pectoris, ischemic ictus, myocardial infarction attack (MIA), and transitory ischemic attack (TIA). This implies that the faster the reduction of cholesterol to “guideline” levels, the lower the probability of incurring a future cardiovascular event, in particular for subjects with high cardiovascular risk. Secondly, although the entire drug industry has witnessed substantial technical progresses over the last 20 years, Statins represent a class of drugs which has shown a continuous improvement over time in its efficacy to lower cholesterol[3]. Last but not least, new guidelines to treat cholesterol have been introduced in the recent years, challenging doctors to continuously adapt their patient management strategies to the new evidence based medicine.

Compared to previous literature, we further innovate by answering a different and more important question within hypercholerolemic patients. As our data show the existence of patient heterogeneity in the time needed to reach an optimal level of cholesterol, we concentrate on exploring the determinants of this measure of health outcome rather than simply on patient cholesterol levels.

In fact, while analyses on the “level” of patient health stock have been performed by many researchers in the past (see among others Lichtemberg 1996 and 2007; Cutler and McClellan 2001; Cutler, Rosen et al., 2006; Long at Al., 2007), to our knowledge no one

[2] This restriction is without loss of generality, given that the vast majority of hypercholesterolemic patients are treated only with statins.

[3] Statins also represent the drug selling class with the highest turnover in most industrialized countries.

has ever provided an answer to the "time needed" to reach an optimal level of health. Providing evidence on this issue has important policy implications (beyond the specific case under investigation) given that a longer patient exposure to specific adverse health shocks may results in serious future negative health outcomes, such as hospitalization, invalidity and death. As we will see in the following sections, hypercholesterolemia is precisely one such case in which patients' recovering time from high cholesterol levels has changed remarkably over time and over cohort of treated patients.

The empirical work will then be based on two different analyses. We will first explore whether patients with faster speed of recovery (in terms of shorter time needed to reach optimal cholesterol levels suggested by international guidelines) exhibit lower hospitalization rates for Coronary Hearth Diseases (CHD) and then will analyze its determinants. The analysis will be based on data from the Health Search database, which collects information gathered by Italian GPs. Unlike standard registry datasets, it contains a rich set of information that allows us to disentangle the beneficial effects of new technologies (more effective drugs) from *i)* improvements in patient compliance to medication, *ii)* improvements in physicians' ability to manage the disease ("process" innovation) and *iii)* standard confounding problems present in several previous empirical analyses (patient past clinical history).

The paper is organized into seven sections. Section two presents some stylized facts about cholesterol trends around the world and in Italy. Sections three and four present the methodology used to define and construct our objective health indicator and empirical models, respectively. Section five presents the data, the sample selection process and the steps followed to construct the variables of interest. Section six shows the main results and their policy implications. In particular, we provide quantitative evidence on how the speed of recovery can affect hospitalization rates for CHD and then we measure the role of each single factor on the speed of recovery. Finally, conclusions are drawn in section seven.

1.2. Cholesterol trends and the role of statins: Some stylized facts and a puzzle.

1.2.1 The stylized facts

Coronary heart disease (CHD) is the single largest cause of death around the industrialized world. In 2005 more than 1 in every 5 deaths in the United States were due to CHDs. Across Europe, diseases of the heart and circulatory system (CVDs) caused over 4 million deaths in 2005, representing nearly half of all deaths in Europe during that year (55% of deaths in women and 43% deaths in men). The main forms of CVDs are coronary heart diseases (CHDs) and stroke. Just under half of all deaths from CVDs are the consequence of a CHD and nearly a third are caused by stroke. The death rate from CHDs in the UK remains one of the highest in Western Europe, at 40%. Lower levels are recorded for France (28%) and Spain (32%), with only Ireland and Finland having a higher mortality rate than the UK. Despite CVD and CHD, mortality has dropped significantly, the number of people living with CHDs has continued to rise[4].

[4] Data are from British Heart Foundation Statistics Database (2005): http://www.heartstats.org/.

Hypercholesterolemia is a major risk factor for CHD. In the United States in recent decades, more than 50% of adults (>20 years old) have total cholesterol concentrations of at least 5.18 mmol/L (200 mg/dL) (Ford et Al., 2003), the level that the National Cholesterol Education Program (NCEP) expert panel considers “borderline-high risk” (Antonopoulos, 2002). This evidence is of particular importance as recent studies suggest that downward trends in total serum cholesterol may be flattening (Ford et Al., 2003 and Arnett et Al., 2002, 2005). Among European countries[5], the prevalence of total cholesterol for both men and women in England and Scotland fell between 1994 and 1998, but increased slightly between 1998 and 2003 (Unal et Al., 2004). Within the northern population of Sweden there has been a continuous decline in total cholesterol levels of 10% from 1986 to 2004. These findings are uniform across all age and gender groups, with few exceptions (Eliasson et Al., 2006). Although mean serum total cholesterol concentrations have dropped constantly across all industrialized countries, there are still many patients with cholesterol well above safe threshold levels suggested by guidelines.[6]

Concerning Italy, there is no single comprehensive national study to our knowledge that has monitored trends in cholesterol levels. The only national data available are those collected in Health Search (HS), a dataset managed by the Society of the Italian College of General Practitioners (SIMG).[7] Panel a) in Figure 1 reports the trends in “Total” and “Low-Density Lipoprotein” (LDL) cholesterol levels for the patients included in the HS sample. According to these data we can see that both measures of cholesterol levels are decreasing. Furthermore, in panel b) we observe how the distribution of cholesterol levels has evolved over time. Interestingly, we can distinguish a movement towards the left (a reduction in average levels) as well as a shrinking of the distribution shape. This second point is even more important as it proves that in Italy, not only the LDL average, but also its variability is reducing.

Although differences in population samples make are hard the comparison of these results to other international experiences, the picture that emerges still seems to suggest that Italian patterns for cholesterol levels are in line with those observed in other countries.

1.2.2 The puzzle

A further inspection of the data highlight a more interesting phenomenon as reported in Figure 2 and in the Table 1a and 1b, where LDL cholesterol level trends are split by

[5] There are very few countries in Europe where consistent data on cholesterol levels have been collected over a relevant period of time.

[6] The EUROASPIRE II study (2001) found that only 51% of patients on lipid lowering therapy (statins) were achieving their goals, ranging from 70% in Finland down to 31% in the Czech Republic. This evidence has been corroborated by many studies, both pan-European and country specific. The REALITY study (2004) found a similar failure of goal achievement in large numbers of patients. A similar lack of adherence to goals has been found in the United States (Pearson et al., 2000).

[7] It is worth noting that these data refer only to patients under treatment and not to the whole population. As such, caution should be used when interpreting these results for inference purposes. For example, it can be the case that despite decreasing cholesterol levels for treated patients, a flattening or an increase in level would be recorded if we were to extend the analysis to the whole population, including unchecked and untreated patients.

patient cohorts according to their initial year of treatment. Each line in figure 2 refers to a cohort of incident patients who began statin treatment in a specific year and have been followed thereafter. The colored areas around the trends represent the confidence intervals. A number of interesting results emerge from the graph:

1. average starting level of LDL is lower across cohorts. This is consistent with the fact that cholesterol guidelines in Italy have been updated over time, progressively including patients with lower blood lipid levels in the treatment protocol but with higher cardiovascular risk, thus moving toward a more preventative approach to dyslipidemia[8];
2. the speed at which cholesterol levels reduce over time has increased. This is particularly true if we look at the absolute changes in cholesterol levels after the first year of treatment across cohorts. If we look in particular at Table 1b we can observe how the reduction in the average cholesterol level tend be greater cohort by cohort, when considering the first two years of treatment.
3. patients who started the therapy earlier are likely to converge to a "higher" level of cholesterol over a "longer" period of time. This phenomenon appears regularly throughout all of the cohorts in our sample (all trajectories intersect);
4. older cohorts, on average, do not reach the LDL cholesterol minimum target of 120 mg/dl.

While the first two results can be easily understood by simply referring to what occurred in the sector over the period of investigation, the last two results are somehow counterintuitive, since we would expect that longer treatment periods would be conducive to better health outcomes.

As clearly shown in Fig. 3, new and more effective chemical compounds (the so called "second-generation statins") have been marketed and prescribed in Italy over the period of our investigation, and the market shares of these compounds have also been affected. Another important change that has affected the market is represented by the introduction in 2004 of larger pack-sizes (28-30 vs. 14 tablets per package), which may have affected patient behavior by improving total compliance and persistence with the treatment, thus leading to better cholesterol management (see Fig. 4). However, these two factors alone can hardly explain the different cohort patterns reported in Fig. 2, given that new active ingredients and larger pack-sizes are supposed to be made available to all patients, irrespective of the cohort to which they belong.

According to our framework, what remains as plausible factor that could explain such uneven cholesterol reduction patterns across cohorts may be the physician behavior in treating cholesterol. In fact, we suspect that GPs who see their patients responding to therapies could decide not to reset or modify it even when new protocols or newer compounds are introduced in the market. This could be caused by prudential attitudes, which aim to avoid any potential side-effect as a consequence of the adoption of newer drugs or of different (and higher) dosages. At the same time, in absence of past evidence, "newcomer" patients are more likely to be treated with newer drugs. Under this

[8] By itself this phenomenon is not able to explain the faster achievement of lower cholesterol targets. In fact, while it is quite easy to lower from very high cholesterol level to medium levels, it is harder to lower from medium levels to target levels.

assumption, the presence of a technical change (in terms of both process and products) could then lead to a gap in health outcomes for earlier and longer-lasting treated patients compared to “newcomer” patients.

Based on these simple evidence, in the next paragraphs we first define our indicator of objective health status and build an empirical model whose aim is to provide a quantitative assessment of the effects that each single actor may have had on the cholesterol reduction patterns over the period of 2002-2006 in Italy.

1.3. The definition and measurement of the health indicator.

One of the implications of the Grossman model (Grossman, 1972) is that health is an inherited durable capital good that depreciates over time. Therefore, investment in health can be seen as an activity where medical care is combined with other inputs in order to produce new health to partly counteract the natural deterioration stemming from health shocks. Thus, the demand for health care can be considered as a derived demand of goods and services to preserve the inherited stock of health $(HS_{i,t})$ and/or to further achieve a desired stock of health $(HS^{*}_{i,t})$.

The presence of an inherited durable capital good generates a disequilibrium model with a quasi fixed input represented by the stock of health. At each time *t*, patients may not be in equilibrium and for this reason they demand health care. This implies that the i-*th* patient’s health status at time *t* can be represented by the following partial adjustment model:

$$HS_{i,t} = HS_{i,t-1} + \lambda\, (HS^{*}_{i,t} - HS_{i,t-1}) - \delta_i HS_{i,t-1} \qquad (1)$$

with $0 \leq \lambda \leq 1$ and $\delta_i > 0$. Therefore, the net investment in health status at time *t* for the i-*th* patient can be represented as:

$$I_{i,t} = \lambda (HS^{*}_{i,t} - HS_{i,t-1}) - \delta_i (HS_{i,t-1}) \qquad (2)$$

Net investment in health will be zero when the optimal level of health is achieved and no deterioration occurs, while it will be positive whenever deterioration is zero and patients demand positive values of health (i.e., investment in health occurs). Obviously, health status variation could be negative as long as deterioration is greater than health demand or, more specifically, deterioration occurs and no new investment is made.

Analogously to investment theory, the parameter λ in equation (2) represents the “speed” at which individuals are able to achieve their target value. It is defined as the fraction of the existing gap to the target, covered by new health investment in one unit of time. If patients reach their goal in one period then λ=1, while if their health status remains unchanged or even reduces, then λ=0 or it becomes negative.

In the empirical literature on investments, λ has always been considered as an “average” parameter to be estimated. Although it could be possible using micro data to make λ individual specific and time varying (i.e. $\lambda_{i,t}$), to our knowledge no one as ever attempted to obtain it as a measurable variable and then to further understand its determinants. In order to achieve this goal, we make two simplifying assumptions that, without loss of generality, help designing the empirical procedure. In particular let states the following two hypotheses:

HP 1: the health demand function is specific of individuals suffering from hypercholesterolemia and therefore at risk of incurring CVD-related events. Therefore, $HS_{i,t} = f(LDL_{i,t})$;

HP 2: that $f(LDL_{i,t})$ is a generic function that inversely relates patient health status to patient cholesterol levels ($\partial HS_{i,t} / \partial LDL_{i,t} < 0$).[9]

As already stated in the introduction, from our perspective the first assumption is important to define our concept of health status and to link it to a well defined and objectively measurable clinical indicator. The second assumption is useful to provide the specific explicit relationship existing between the clinical indicator and the health status level.[10] Therefore, in what follows we will refer to the health demand function of individuals suffering from hypercholesterolemia and therefore at risk of incurring CVD-related events. Based on these premises, then Eq. (1) can be rewritten in the following way:

$$f(LDL_{i,t}) = f(LDL_{i,t-1}) + \lambda(f(LDL^*_{i,t}) - f(LDL_{i,t-1})) - \delta_i f(LDL_{i,t-1}) \qquad (3)$$

where the optimal level of LDL cholesterol (LDL^*) is patient specific, time independent[11] and solely a function of his/her cardiovascular risk index[12] ($r_{i,t}$), that summarizes several patient characteristics (smoking habits, clinical conditions, genetic factors, previous experience of CVD events, age and gender):

$$LDL^*_{i,t} = f(r_{i,t}) \qquad (4)$$

In order to define λ as an empirical measurable variable, we can simply solve equation (3) for λ to obtain:

$$\lambda_{i,t} = (\Delta f(LDL_{i,t}) + \delta_i f(LDL_{i,t-1}))/\Delta f(LDL^*_{i,t}) =$$

[9] In what follows we will refer to "LDL cholesterol" and to "cholesterol" interchangeably.

[10] Within the framework of clinical trials, the level of cholesterol can be considered a surrogate endpoint (or marker), that is a measure of the effect of a certain treatment that may correlate with a *real* clinical endpoint but doesn't necessarily have a guaranteed relationship. The National Institutes of Health (USA) defines surrogate endpoint as "a biomarker intended to substitute for a clinical endpoint". The FDA and other regulatory agencies will often accept evidence from clinical trials that show a direct clinical benefit to surrogate markers. A good example is represented by cholesterol. While elevated cholesterol levels increase the likelihood for heart disease, the relationship is not linear - many people with normal cholesterol develop heart disease, and many with high cholesterol do not. "Death from heart disease" is the endpoint of interest, but "cholesterol" is the surrogate marker.

[11] This implies that for some given patient clinical characteristics, the optimal level of cholesterol is given and constant (i.e., suffering from diabetes sets the optimal level of LDL to a range between 70 and 100 mg/dl and this target does not change over time).

[12] The cardiovascular risk index represents the individual predicted risk at time *t* to incur in a CVDs in the following 10 years on the basis of the assessed current health and life-style profile (see S. Giampaoli, L. Palmieri, A. Mattiello et al., 2005).

$$= \Delta' f(LDL_{i,t}) / \Delta f(LDL^{*}_{i,t}) \qquad (5)$$

where the subscripts i and t to λ follow from the fact that all variables used to compute λ are patient specific and time varying. Furthermore, $\Delta f(LDL_{i,t}) = f(LDL_{i,t}) - f(LDL_{i,t-1})$, $\Delta f(LDL^{*}_{i,t}) = f(LDL^{*}_{i,t}) - f(LDL_{i,t-1})$. As we will see, from an empirical point of view the term $\Delta' f(LDL_{i,t}) = \Delta f(LDL_{i,t}) + \delta_I f(LDL_{i,t-1})$ is simply the absolute variation in the health stock expressed as a function of the LDL level.

However, the definition of $\lambda_{i,t}$ stemming from eq. (5) is interesting because it has a nice clinical interpretation. In fact, although researchers cannot usually observe the single components that characterize the numerator in eq. (5), they can be interpreted as patient "good behavior" ($\Delta f(LDL_{i,t})$) and "bad behavior" ($\delta_I f(LDL_{i,t-1})$) in achieving the therapeutic goals. This interpretation can be better understood by looking at the graph in Fig. 5, where we observe two different hypothetical paths of LDL cholesterol towards the target (LDL^{*}). The blue line represents the behavior of a patient whose net investment is characterized only by "good behavior", while the black line represents the behavior of a patient who alternates periods of "good behavior" to periods of "bad behavior" (identified by those periods in which $\Delta LDL_{i,t}>0$ and therefore $\Delta HS_{i,t} = HS_{i,t}-HS_{i,t-1} = \Delta f(LDL_{i,t}) \leq 0$). It is clear that the speed at which the first patient reaches the target is greater than the speed of the second patient.

In eq. (5) the denominator ($\Delta f(LDL^{*}_{i,t})$) is positive by definition and tends to zero while the numerator in the first term ($\Delta f(LDL_{i,t})$) could be either positive or negative. In terms of investment theory this introduces an inconsistency in our model, as investment can only be zero or positive. Whatever deteriorates the stock of capital should be included in the term $\delta_I f(LDL_{i,t-1})$. A possible way to solve this issue is to consider the term $\Delta f(LDL_{i,t})$ censored to zero from below ($\Delta f(LDL^{c}_{i,t})$), while setting the term $\delta_I f(LDL_{i,t-1})$ as function of the indicator 1(.) taking the value of zero when $\Delta f(LDL_{i,t})\geq 0$, and one when $\Delta f(LDL_{i,t})<0$, in order to capture only the "bad behavior" of patients. Given this interpretation, eq. (5) becomes:

$$\lambda_{i,t} = \Delta f(LDL^{c}_{i,t}) / \Delta f(LDL^{*}_{i,t}) +$$

$$1(\Delta f(LDL_{i,t}))* (\Delta f(LDL_{i,t})) / \Delta f(LDL^{*}_{i,t}) \qquad (6)$$

Under this new specification the first term is always positive as both numerator and denominator are positive or equal to zero[13] and it is bounded from above to 1 by construction; the second term is instead negative, given that its numerator is negative by construction. From an economic and clinical point of view this implies that in the presence of "bad behavior" the speed of convergence to the therapeutic goal is reduced. At the same time, the greater the "good behavior" part, the greater the speed of recovery will be.[14]

---

[13] It could be possible to have situations in which the ratio takes the form 0/0. However, from a theoretical point of view these cases can occur only when a patient reaches his/her target. In all these cases, from an empirical point of view we will replace the resulting "missing" occurrence with a one, indicating that at that point no investment is necessary.

[14] According to eq. (6), the requirement $0\leq\lambda_{i,t}\leq 1$ will occur only if "bad behavior" are absent ($1(\Delta(LDL_{i,t-1}))$=0) or whether it is less important than "good behavior" ($\Delta f(LDL^{c}_{i,t})>1(\Delta f(LDL_{i,t-1}))* (\Delta f(LDL_{i,t-1}))$).

In our empirical implementation we assume that $HS_{i,t}=f(LDL_{i,t})=1/LDL_{i,t}$, as it is a simple functional form which respects what we stated as a sufficient requirement in the HP 2. Naturally, other functional specifications are equally suitable (i.e. $HS_{i,t} = -LDL_{i,t}$). After the implementation of sensitivity tests on this respect, we found that the choice of alternative functional forms does not alter the final result nor the empirical evidence presented in the following paragraphs.

1.4. The data.

Our empirical analysis is based on data obtained from the Health Search Database (HSD), a longitudinal observational database run by the Italian College of General Practitioners (SIMG – Società Italiana di Medicina Generale) since 1998. The HSD contains data from computer-based patient records from General Practitioners (GPs) throughout Italy. Participation is on a voluntary base, but selection of the GPs has been made in order to match with the regional organization of the NHS and to include a number of patients proportional to the size of the Italian adult population (Fabiani et al., 2004).

The HSD collects patient-level data which are linked through a unique anonymous identifier to drug prescriptions, clinical events and diagnoses, hospital admissions and causes of death[15]. It contains patient-level information on prescriptions such as dispensing date, drug information (ATC code, quantity and type of active ingredient and number of pills) and the general practitioner recommended dosage (GPRD). It also includes hospitalization status by primary Diagnosis Related Groups (DRG), information on patients' clinical histories and co-payment exemptions and a set of socio-demographic indicators.

Up to December 31st 2006, the dataset contained information collected by 796 GPs for a total of 1,532,357 patients, 15,727,442 diagnoses, 108,441,541 diagnostic tests and 77,276,255 drug prescriptions[16]. To improve the quality of the information collected, we have restricted our analysis to information gathered by those 400 GPs who collect patient records according to a specific mathematical algorithm developed by Health Search that guarantees completeness and consistency of the data (Fabiani et Al., 2004; Sessa et al., 2004)[17]. For our purposes, a sub-sample of patients has been extracted from the HSD[18]. The selection has been conducted using two main inclusion criteria: *i)* patients aged

[15] These data are collected by GP using the MILLEWIN® software. All diagnoses are coded according to the ICD-9-CM (International Classification of Diseases, 9th Revision, Clinical Modification. Drug names are coded according to the Anatomical Therapeutic Chemical (ATC) Classification.

[16] Diagnostic tests include laboratory measures of cholesterol level, blood pressure levels (systolic and diastolic), frequency of the heart beat and other observational measures usually conducted by clinicians and recorded as signs or summarized in ratings that are mixed with the results from conversations with patients (cardio-vascular risk index, etc.).

[17] This selection takes place every 3 months. Information from physicians who fail to meet standard quality criteria are not considered for inclusion in epidemiological studies.

[18] While the HSD can be thought of as a random sample representative of the Italian population, our sub-sample (i.e. patients who have received at least one prescription of statins in the observed period) can be thought of as a random sample from the target population represented by Italian hypercholesterolemic patients.

between 39 and 70 at the time of their first inclusion in the HSD and *ii)* patients who received a prescription of statins at any point in time over the 2001-2006 period[19]. The sub-sample extracted using this procedure consists of 42,140 patients and 1,272,797 observations, of which 784,068 refer to prescriptions, and the remaining 488,729 to diagnostic tests[20]. A description of this selection process is contained in Table 2.

Since the initial sample consists of daily observations, we collapsed the data after constructing the variables in order to obtain quarterly observations at patient level. The use of a quarter as the time frame in our analysis seems a reasonable compromise in terms of reducing the number of zero occurrences in drug consumption. Furthermore, several sources of clinical evidence show how statins may produce most of their effects after one quarter of use[21].

Table 1 reports the major operational steps taken to obtain the final sample. It is important to highlight that we have dropped observations on the first quarter of 2001 since we want to discard non-incident patients from our analysis. We have also dropped observations of those patients who started the treatment in 2006 as we believe that a one year period in not enough to observe a clear pattern in patient cholesterol levels. Finally, as we use lagged values of cholesterol levels in our empirical analysis, the whole 2001 results discarded from the final sample. This does not imply that the cohort 2001 is discarded as well, since we include in the analysis all of the incident units which are observed from the year 2002 onwards. The final sample is then a quarterly unbalanced panel for the period 2002-2006, which consists of 4,293 patients, for a total of 21,200 observations[22].

We have further to specify that individuals can enter the regression sample in a moment which is posterior with respect to their cohort of origin. In fact cohort membership is determined on the basis of the first statin prescription, while the first LDL test could occur at a later point. Whether LDL value or another patient's information is missing, the unit cannot be included in the sample. We believe this discrepancy does not necessarily affect the estimates results. However, we make sensitivity checks by performing the same estimates on a reduced sample which only includes individual observed from the beginning of their respective cohort.

The key variables in our model are the individual LDL cholesterol level[23], observed through time, and its optimal value LDL^*. While the HS dataset provides information for the first variable, its optimal value needs to be estimated for each single patient. According to Eq. 4 and to international guidelines, we set the target (optimal) LDL cholesterol level as a function of the individual cardio-vascular risk ($r_{i,t}$), split in three

---

[19] This is the time span over which HS had a stable and reliable dataset at the time we started this project.

[20] Patients for whom we observe at least one prescription and one diagnostic test are 40,859.

[21] Prof. Claudio Cortese (Faculty of Medicine, University of Rome Tor Vergata), personal communication.

[22] It is important to note that the number of observations in the empirical analysis is reduced due to the use of the LDL cholesterol variable that enters with two lags.

[23] An important aspect for the construction of this variable is that patients in the HSD are heterogeneous in terms of frequency of their cholesterol checks through diagnostic tests. This may pose some problems as we aim at building a quarterly variable and has implicitly required us to impute a LDL value through a simple linear interpolation whenever patients did not get a cholesterol check in an in-between period. Clearly, we are aware that this could represent a strong assumption, which gets even stronger as the distance between diagnostic tests increases. From simple descriptive statistics, it appears that patients are checked more frequently at the beginning of their treatment. Related summary statistics are available upon request.

different risk classes: low risk (0%≤ $r_{i,t}$≤5% and no past events), medium risk (5%≤ $r_{i,t}$≤10% and no past events) and high risk ($r_{i,t}$>10% or past events). For low risk patients LDL^* needs to be lower than 120 mg/dl, for medium risk patients instead LDL^* should be kept below 100 mg/dl while for high risk patients LDL^* has to be maintained under the level of 70 mg/dl.[24]

Concerning the explanatory variables, the following is a short description of the main methodological problems and operational steps we have encountered in their construction. As patients may shift active ingredients within the quarter, we have decided to attribute a prevalent active ingredient to each quarter. We define a prevalent active ingredient as an active ingredient that covers 80% or more of the pills prescribed by the doctor in that quarter. On the basis of this assumption, we build a dummy variable for each of the 5 active ingredients included in our study within the class of statins (Simvastatin, Pravastatin, Fluvastatin, Atorvastatin, and Rosuvastatin) that takes a value of one whenever one of the active ingredients was prevailing in the quarter according to our criteria.[25]

In any medical therapy, a fundamental factor is the positive or negative patient attitude in following the treatment. There are various indicators which could be used to approximate this outcome. In Atella et Al. (2009) we had the opportunity to use both drug adherence and persistency indicator. In this context, adherence is defined as the patient capacity or willingness to adhere to the therapy prescribed by a physician and is usually measured as a possession ratio. Persistency is instead defined as the propensity of the patient to regularly attend GP's office and therapy: we define patients as persistent if they visited their GP for at least 90%[26] of the quarters in which he/she is observed. We use this last indicator as our proxy for the patient's collaborative attitude to the therapy[27].

A secondary prevention dummy variable records whether a patient has experienced a cardio vascular event, which we identify with TIA, ictus, angina or IMA. The same approach has been used for dummies selecting for co-morbidities such as hypertension and diabetes. In order to capture some socio-economic factors, we have used information on exemptions. The dataset includes information on different types of exemption (by age, pathology, income and invalidity). For each type of exemption we have constructed a specific dummy. It must be noted that in some cases the exemption characteristic may overlap with other patient characteristics (i.e. age or secondary prevention)[28].

Finally, we include cohort indicators that capture fixed effects related to the beginning period of the therapy. These indicators were simply built using as a reference

[24] Cholesterol levels could be also expressed in terms of mml/l. The conversion from mg/dl to mml/l can be obtained by simply dividing by 38,67.

[25] As an alternative measure we have run the models considering the proportional influence of any active ingredients on the quarterly cholesterol variation achievement. The results were totally robust, in accordance with our main estimates.

[26] We use a higher threshold compared to the 80% threshold used in the literature (see for example Atella et al. 2009), since we require more precision from this variable. In this case we cannot include therapy adherence information as this would result in an excessive loss in the number of the observations.

[27] Unfortunately, we have not been able to also use the adherence indicator due to a lack of information on physicians' recommended daily dosage on many of the prescriptions. Its inclusion in the analysis would then have caused a consistent loss of observations (of about two thirds). Furthermore, the persistence indicator looks more appealing since it is simple to build and does not suffer from the missing data problem.

[28] It might also be the case that some patients did not ask for exemption despite being entitled to it.

to the year in which the therapy started, according to the date of the first prescription of a statin.

The remaining set of variables is self-explanatory, merely containing socio-demographic characteristics and time-dummies.

Table 3 reports all of the descriptive statistics on all the observations. As we can clearly see, the average value of $\lambda_{i,t}$ over the whole period is close to 0.1, meaning that patients take 10 quarters on average to reach the optimal LDL value. Our sample is equally split between male and female patients, 70% of which are over 60 years old. Atorvastatin and Simvastatin make up more than 60% of statin prescriptions, the daily dosage prescribed being close to 25 mg. Fifty-nine per cent of the sample suffers from hypertension, about 29% have diabetes, 9% are in secondary prevention and 70% are reported as highly persistent. In terms of cohort dummies, most of our sample observations are made of patients who started the treatment in 2002, then being observed over a longer period. In terms of time dummies, on the other hand, we observe a peak in the number of the observations in 2004 and 2005.

Since our goal is to understand the difference that treatment had on the health outcomes of different cohorts of patients, Table 4 reports descriptive statistics by cohort at patient level. What clearly emerges from Table 4 is that the speed of recovery has sharply and monotonically increased across cohorts, with the 2002 cohort recording the lowest value (10% or 10 quarters) and the 2005 cohorts reporting the highest value (40% or about 2.5 quarters). This increase came irrespective of the fact that the average health status of these cohorts has worsened, given that the number of hypertensive, diabetic and secondary prevention patients has increased across cohort samples. In particular, hypertensive patients have increased from 49% to 67%, diabetic patients have increased from 23% to 40% and, finally, patients in secondary prevention have increased from 6.4% to 13.0%. In other words, the improvements in the speed of recovery have been obtained despite the fact that the target level of cholesterol has slightly reduced on average across cohorts. The increasing rates of patients which are reported to be hypertensive, diabetic or in secondary prevention, represent a signal of an augmented ability of GPs in selecting higher CVD risk population. Further, also the role of health information spread by mass-media and other informative sources, might have progressively induced more patients-at-risk to be checked and treated by their GP.

In terms of age and gender, the cohorts have remained almost constant. The only significant change we can observe is the regional composition of our sample, which sees an increase in the share of patients coming from the South and the Islands and a decrease in the number of patients residing in the Center of Italy.

As a further analysis we have decomposed the value of λ in its "good" and "bad" behavior components. As it can be easily seen, most of the increase in the speed of recovery across cohorts is due to improvements of the "good behavior" part rather than to reduction of its "bad behavior" component. In other words, this seems to indicate that whatever has determined an increase in the speed of recovery is due to something that only marginally has to do with patient behavior (compliance to medical advices), but could most probably due to technological change, either through product (new active ingredients) or process (physician cholesterol management) improvements.

Table 5 and Table 6 contains the same information which is included in Table 3 and Table 4 but for the reduced sample composed of individuals which are observed without a discrepancy between their cohort membership and their first appearance in the estimation sample. We note that the behavior of such a reduced sample is very similar to the full sample, including all the patients selected. This similarity give us some clue that the full-sample does not necessarily suffer of a bias.

Attrition patterns for the full-sample are reported in Table 7, while in Table 8 we include the distribution of the patients by cohort and by first appearance in the full-sample. As expectable, most of the patients are present from the beginning or from the second year of their cohort, but a non negligible proportion of them is observed only on a later time. Table 9 includes instead attrition patters for the reduced sample, as we previously defined it. When performing our estimates we will also control for attrition trough the use of weights inversely proportional to the probability of inclusion in the sample. This last step will make us able to control whether the effects found are driven by movements within the sample rather than to the behavior of the population under analysis.

1.5 Empirical results

1.5.1 The effect of the speed of recovery on hospitalization rates.

In this section we explore the statistical relationship between hospitalization rates for CHD and $\lambda_{i,t}$ (as a regressor) in a probit equation where the dependent variable is now a dummy for hospitalization. In fact, it must be stressed that from an empirical point of view the increase in the speed of recovery does not automatically translate into lower hospitalization rates, and this new hypothesis has to be submitted to empirical testing. From a methodological point of view, we have approached the problem by evaluating the probability of hospitalization in a two-year period since a patient first appearance in the dataset for each cohort of patients[29]. We then use the speed of recovery evaluated in the first year of treatment as a predictor of the hospitalization rate in the second period.[30] Cohorts are defined on the base of the year in which the treatment has started for each patient. As cohorts are different in terms of speed of recovery we expect to observe a negative parameter for our $\lambda_{i,t}$ variable, proving that faster recovery reduces hospitalization (after conditioning for patient-specific covariates).

In order to carry this new analysis, we had to base our estimates on a different group of observations, which includes only a sub-set of patients used in the previous analysis. In fact, we have restricted our analysis to only those patients for whom a full set of information was available for at least two consecutive years since their first appearance in the sample.[31] As our unit of time is the year, for each single patient we have obtained a mean lagged value of $\lambda_{i,t}$, a mean value of the current hospitalization status plus a set of time invariant patient specific mean controls (initial cholesterol level, gender, age class,

[29] A two-year time window is the only one that allows us to have a uniform time period for all cohorts, as we observe patients who started the treatment in 2005 also in 2006.

[30] In this way we aim also to reduce possible endogeneity problems.

[31] Obviously, the main concern with this selection is that we are including only patients who are somehow persistent with the therapy and therefore we should not observe differences across cohorts. However, as we will see, even for this subset, a higher value of lambda is able to negatively affect (reduce) their hospitalization rate.

co-morbidity factors and cohort effects). As a result, our final sample consists of 3,329 observations, each of them referring, by construction, to a different patient.

The empirical results are reported in Table 10. The most important finding we obtain is that an increase by 10% in the speed of recovery ($\lambda_{i,t}$) seems able to reduce the hospitalization rate by a percentage just below 0.8%. This effect does not seem endogenous since the two pieces of information come at different temporal lags: the reduced hospitalization risk can be then interpreted as the positive outcome deriving from a faster recovery to the optimal LDL level. Less significant results derive from the control covariates, there being a small gender effect (men are slightly more likely to go to the hospital) and a stronger geographical effects (people from the NW are less likely to be hospitalized). As expected, patients in secondary prevention were more likely to be hospitalized.

1.5.2 The determinants of the speed of adjustment[32]

1.5.2.1. The empirical model

In the previous section we explored the relationship between the speed of recovery to target cholesterol levels and actual hospitalization rates. Although in terms of health outcomes this is an important result by itself, a further question regards the role of the different determinants of the speed at which patients reduce their cholesterol level, thus reducing their exposure to CVD adverse events .

Obtaining information on the determinants of $\lambda_{i,t}$ represents a crucial factor in terms of therapy's effectiveness at micro level and for health policy at macro level. In fact, the main aim of the treatment is to avoid the occurrence of a CHD event, which could determine a hospitalization with severe inability problems in the future or, even worse, bring the patient to death. The faster the recovery to the optimal LDL value, the lower will be the probability that patients will experience a cardiovascular event. Therefore, knowing the determinants of this speed of adjustment may be crucial for policy makers in the health care sector.

Given our measure of $\lambda_{i,t}$ from eq. (6), the next step is to define an empirical model in which our dependent variable ($\lambda_{i,t}$) is regressed against a set of covariates that should allow to disentangle the specific role played by patients, physicians and technical innovation. A plausible model that can meet our objectives is represented by the following equation:

$$\lambda_{i,t} = \varphi(LDL_{i,t-1}, HP_{i,t}, LS_{i,t}, SEC_{i,t}, TR_{i,j,t}, t_t, c_i) + \varepsilon_{i,t} \qquad (7)$$

where $LDL_{i,t-1}$ is the one period lagged cholesterol level capturing the fact that a patient with a higher level of cholesterol may converge faster to the goal, $HP_{i,t}$ is a vector of variable defining patient health profile, $LS_{i,t}$ is a vector of variable that refers to patient lifestyle, $SEC_{i,t}$ is a vector of patient demographic and socio-economic characteristics,

[32] Results concerning the determinants of cholesterol "levels" are available upon request by the authors.

$TR_{i,j,t}$ refers to the *j-th* active ingredient taken by the *i-th* patient, t_t is a time trend to capture improvement in the speed of adjustment to the optimal target that is not otherwise captured by the other variables and $c_{i,t}$ is a starting treatment patient cohort dummy intended to capture the learning process by physicians in the management of cholesterol over time (more recent patients are treated better). Finally, $\varepsilon_{i,t}$ is a standard additive idiosyncratic error term distributed $N(0,\sigma^2)$.

The first four regressors in eq.(7) are intended to capture patient behaviour, while $TR_{i,j,t}$ and t_t account for technological change. From a different perspective, we can also note that the kind of treatment received through the specific active ingredient used ($TR_{i,j,t}$), which is pertinent to the biochemical composition of the drug, represents a "product" innovation, while the trend variable (t_t), can be considered as a measure of "process" innovation as it could also represents the learning process by physicians and refers to changes occurring over time in GPs' abilities to choose the appropriate combinations of drug, dosage and lifestyle advice. This learning process should occur over time through field experience, information and updates obtained from participation in conferences and from reading of medical literature.

Concerning patient behaviour variables, the $HP_{i,t}$ vector includes a dummy for the presence of past CVD events (primary vs. secondary prevention) and dummies for hypertension and diabetes. The vector $LS_{i,t}$ includes a variable controlling for smoking behaviour, while the $SEC_{i,t}$ vector includes variables such as age, gender, regional residence and different measures of exemption.

In terms of coefficient expected signs in eq. (7), we must remember that, in accordance with a principal-agent framework, only physicians can decide what to prescribe and whether to switch patients to a new therapy; whereas patient decisions are limited to fully adhering or not to the proposed treatments and lifestyle advice. These patient decisions may be independent of the therapy suggested, but influenced by other patient-specific factors.[33]

All treatment variables should have a positive effect since they improve patients' capacity to reach their goal. The existence of a bad health profile (i.e. those in secondary prevention or suffering from diabetes or hypertension) should increase the speed of adjustment. Similarly, elderly patients should record a higher adjustment speed as they are likely to be more aware and worried about their health conditions. As female patients are seemingly genetically more protected from this kind of pathology, they may be less induced to reach the goal. No clear sign is expected from the income exemption status, since we cannot control for educational status. In fact, it could be the case that people on lower incomes adjust faster since they are exempted and do not pay for drugs; but at the same time they are also likely to be less educated and therefore could underestimate the importance of reducing cholesterol levels. We expect monotonically increasing negative

[33] A good example is represented by economic variables such as disposable income ($Yd_{i,t}$) and drug price (p_t) that negatively affect patient adherence to treatment. In fact, high drug price may discourage using drugs (especially in presence of asymptomatic conditions) and high income levels reduce the probability of being fully or partially exempted by co-payment on medical treatments. These variables should then have a negative effect on patient adherence. However, it must also be noted that in the Italian system, patients suffering from diagnosed CVDs never sustain the full cost of the drugs, since they can be totally exempted, due to their health conditions and age. In the empirical analysis we will consider that adherence is exogenously given and will use exemption status by income only to proxy for income and, therefore, for educational status.

parameters for the cohort dummies, indicating that newer cohorts are treated better and this result can be attributed solely to physician behaviour once we accept the principal-agent story.

1.5.2.2 The results

All results are presented in Table 11 and are based on different empirical specification of the model in eq.(7). In what follows we will first discuss the results from our base specification (model 1) and then compare them with the alternative specifications (models from 2 to 5) to test the robustness of our results. All estimates are based on a RE model.[34]

In our base model, our reference patient is a male, below 50 years, living in the Islands, treated with Fluvastatin, non-smoker, not exempted from copayments, without co-morbidities and in primary prevention.

As expected we find that the speed of recovery is positively related with an patient lagged cholesterol level, $LDL_{i,t-2}$. A higher lagged cholesterol level is associated with a quicker reduction of the health gap. This may, on the one hand, be a consequence of the patient's effort, since they are conscious of their health risk and are pursuing their doctor's advice more carefully. At the same time, implicit characteristics of statins make the therapy more effective when the lagged level is higher. When the LDL value is fairly low, the only role for statins is to keep cholesterol constant, without reducing it further.

Looking at the role of technical progress in terms of "product" innovation we have identified five active ingredient-specific dummy variables (Simvastatin, Pravastatin, Fluvastatin, Atorvastatin and Rosuvastatin) with the aim of verifying whether, ceteris paribus, being under treatment with newer drugs provides an actual benefit. According to our results, "second generation" statins (i.e. Atorvastatin and Rosuvastatin) are found to be more effective than the alternatives. A smaller, but still significant role is played by Simvastatin, which belongs to the "first generation" group and still holds an important share of the Italian statin market (around 30% in 2005, the last available year of our data). Pravastatin does not provide any additional effect to our baseline active ingredient in driving the speed of recovery.

The role played by "process" innovation is captured by the cohort dummy parameters that identify the physician learning process. These parameters show an increase in the recovery speed, cohort by cohort. Do cohorts identify the evolution of doctors' know-how, which improves their ability to reduce patient cholesterol levels, or, in a more pessimistic view, just conservativeness in treating longer-lasting patients? We should probably assume that both forces are working, since if only a learning process was acting, then even patients from the first cohort would have benefited from those positive externalities and we would have observed same mean cholesterol levels in all patient cohorts in 2006. This behaviour is confirmed looking at the results presented in Table 12,

[34] The estimation of the empirical model trough a FE approach it is not feasible given our framework, as we could not use some key covariates such as the cohort dummies, which are time-invariant. Furthermore, patient characteristic variables (gender, area of residence and some of the health related information) have a null or reduced variation trough time. Naturally, the adoption of a RE model implies the assumption that individual effects and covariates are independent each other. We are confident that, although a certain degree of correlation between the individual effects and the covariates is certainly possible, this is reduced to a small amount by the wide and heterogeneous set of regressors included.

where we report the share of active ingredient prescriptions by cohort for a reference year. As we can clearly see, in 2005 the cohort of patients that started the treatment in 2002 receives Simvastatin in about 35% of the cases, while the cohort that started the treatment in 2005 receives Simvastatin in only 28% of the cases. If we look at the Rosuvastatin (the most innovative product), the 2002 cohort receives this active ingredient in only 9% of the cases, while the cohort of 2005 receives the same active ingredient in 20% of the cases.

In terms of patient behaviour, we see that persistence in the treatment is found to be influential. This indicator works as a good proxy for the attitude of patients to follow the therapy continuously. Similarly, we find strong positive signs from the secondary prevention dummy, while a milder role is played by hypertension. Patients that already suffered a serious cardiovascular event or are experiencing co-morbidities are, understandably, more careful thus showing a faster speed of recovery.

Individual unhealthy habits are proxied by their attitude to smoke. The related parameter is found to be negative, although not significant. The sign direction is as expected, while the non-significance of the variable might be due to the fact that a vast majority of our sample are non-smokers, implying limited variation in the dummy variable.

When controlling for socio-demographic characteristics, we see that men seem to experience a faster recovery. This may be the consequence of stronger incentives for men to reduce LDL levels due to their higher exposure to cardiovascular risks compared to women.

In terms of age, we observe that elderly patients converge to lower levels of cholesterol more rapidly than younger patients and the speed of recovery increases monotonically with age. This may be the result of elderly patients who benefit more from formal and informal care givers, who help them in a correct and constant observation of therapy rules.

Italian sub-areas are very dissimilar to one another in terms of lifestyle, social habits and degree of urbanization, all of which potentially influence the exit status of the cure. Despite of this, we find a relative small effect played by geographical dummies: in the northern areas, in which lifestyle and diet is somehow more similar to Central Europe, there appear to be a lower recovery speed.

Finally, the dummies for exemption status show that exemption for health conditions (CVDs related) and for age are the only forms of significant exemption. While we don’t have an explanation for the behaviour of age exemption, which acts in the opposite side of the age categories, we find that exempted for a CVD appear to converge even more strongly to their LDL healthy level. None of the other exemptions conditions seem to play a role, in particular exemption by income. This result is in part expected given the universalistic principle of the Italian health care system and the attention provided to access care for serious chronic conditions.

These results have been confirmed with minor changes by the other specifications. In particular, in model 2 we included as a further control the prescribed daily dosage (in terms of milligrams of active ingredient), in order to separate the compound effect from the dosage effect. This aspect is not entirely trivial, as physicians can use an older statin with an increased dosage, potentially getting the same effect of a newer compound at a reduced dosage[35]. The results from this new specification substantially confirm the evidence obtained in the base model, and provides additional evidence of the positive effect that a higher dosage has on the speed of recovery. Also, we must remark that, in both specifications, the hierarchy between the compounds' effectiveness remains unchanged and is fairly stable in terms of proportions.

In model 3 we have added the interaction terms between active ingredient and daily dosage. In this case the daily dosage and the interaction terms lose their significance, while the size of the coefficients is unaltered (with Simvastatin loosing its significance).

In order to check if the unconditional means reported in figure 2 could be replicated even after conditioning on all controls we have included, in model 4 we have added the interactions between the cohort dummies and a linear time trend, with the aim of recovering a set of parameters that could help replicate those patterns. For each cohort, each interaction term can be seen as a differential effect on the speed of cholesterol reduction from the treatment starting year[36]. According to this interpretation, comparable effects sizes to what is found in the years 2005 and 2006 for the 2002 cohort, are already observed in the third and fourth year of 2003 cohort and in the second and third year of 2004 cohort. We can conclude that progressively the rapidity with which the therapy acts increases steeply by cohort. An exception should be made for 2005, whose only time interaction variable is found to be comparatively high but non significant.

Finally, in the last specification we also include the age and the gender of the GP, as it is possible that those characteristics (age in particular) have an influence on the therapy. However, none of those information results to be significant for the speed of recovery.[37]

As a sensitivity check we performed the same estimates for the model 5 in the reduced samples and with both the samples weighted by probability of inclusion, in order to control for attrition. The weights have been estimated using a probit model on the basis of the main descriptive statistics regarding demographic and health related information. According to Table 13a and 13b, age, male gender and being diabetic have a role in favouring the inclusion in the sample, while other health statuses like being secondary prevention or hypertensive promote the exit. Further, it seems that having a younger GP favours the inclusion, while their gender has no role.

In Table 14 we publish the results of model specification 5 for the full sample and for the reduced sample, with and without weighing the estimates for the attrition found. While the weighting has an impact on the results as covariates such as age and gender reduce their effects or looses their significance, we find that the main parameters of the

---

[35] The fact that similar effects could be obtained by mixing active ingredients and dosages is a highly debated issue among physicians and health care administrators. In fact, as older generation statins cost less, health care administrators tend to recommend the prescription of such statins at higher dosages. However, physicians and specialists are in favour of second generation statins as they can get the same results with lower dosage, thus reducing the potential side effects that high dosages can have on patients.

[36] It must be noted that while our estimates are based on quarterly data, our linear time trend has been based on a yearly base to save on interaction parameters (one fourth!). This implies that in the interaction term for the 2001 cohort the base year is an average of 4 quarters of year 2001 and the differential effects are computed starting from year 2002.

[37] As a robustness check, we added to this last specification regional based variables representing the shares of second generation statins prescribed by doctors, in order to evidence whether the doctors' choices could be influenced by regional authorities, which in Italy have the control of health and drug policies. However, the related parameters resulted not to be significant.

model remains unchanged. In particular, we underline how the use of a reduced sample does not affect the significance of the cohort-time interactions, exception made for some of the coefficients belonging to the cohort 2002. In that case, as it is possible to understand by comparing Table 7 and Table 9, the loss of information which derives by the sample reduction is the biggest, as only 53% of the observations from that cohort belong to both samples. The other findings, such as drug effects, cohort dummies and doctors' related characteristics remains unchanged.

1.5.3 Who is responsible for your health: a quantitative assessment

As last step we provide an answer to the initial question: who is responsible for your health. In order to achieve this goal we estimate the impact on the predicted values for the three factors, standardizing to 100 the values of their sum, for each of the five model specifications in Table 11 and for the three additional specifications included in Table 14. The role of patients is determined by the parameters associated with the persistency variable, the role of doctors is captured by the cohort dummy parameters, while medical technology impact is represented by the active ingredient dummy parameters. The results regarding the unweighted estimates are presented in Table 15. In the first specification, where we are not controlling for the drug dosage, technology and medical behaviour share almost the same percentage impact (37%), while the role of patients appears to be minor. When including drug dosage alone, as in specification 2 or interacted with active ingredient dummies, as in the specification 3, technology becomes by far the most important factor. This result is confirmed even if we consider the specification 4 and 5 when significant interactions between the cohorts and the time variable are also added. In fact, technology effect in specifications from 2 to 5 ranges between 54% and 61%. However, while in models 2 and 3 the doctors' role appear to be greater than the role of patients, when moving to specifications 4 and 5 the doctors' effect halves. In Table 16 we try a further sensitivity check by comparing the same effects as Table 15 on the reduced and on the weighted samples. Results are computed only for the model 5 and are consistent with what was found for the full sample in any sample redefinition..

We believe that these results are quite reliable as they have been obtained after controlling for a wide and accurate set of demographic and past health-history variables that should take into account a large fraction of patient and physician heterogeneity, and who have been proven to be highly explicative in our model.

As a further evidence, separate computations of the effects for men and women (not reported here) have shown that the relative importance of the three factors is not significantly affected by gender issues.

In conclusion, we recognize that the only effect which prevails steadily on the others derives from the use of drugs. This is not surprising, as the result of doctors and patients effort is to be thought as conditioned on the available pharmacological tools. However, our estimates clearly prove that technology can at the best explain only 62% of the speed of recovery to a better health status, and its efficacy is mediated by physician and patient behaviours.

1.6 Conclusion

In this work we have analysed the determinants of a better health status (measured in terms of speed of recovery to a better health level), disentangling the roles played by patients, physicians and technology, and how a faster recovery is reflected in terms of reduction in hospitalization rates.

In particular we focused on individuals who suffer from hypercholesterolemia and are treated with statin-based drugs. We explored the patterns of their LDL cholesterol levels, which represent a key piece of information for understanding the effectiveness of the therapy and the consequent probability of incurring a CVD in the immediate future.

Our results show that the speed of recovery is a good predictor of future hospitalization rates: better treated patients experience lower hospitalization rates for CVDs.

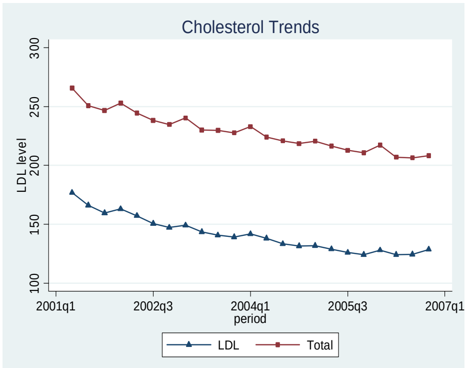
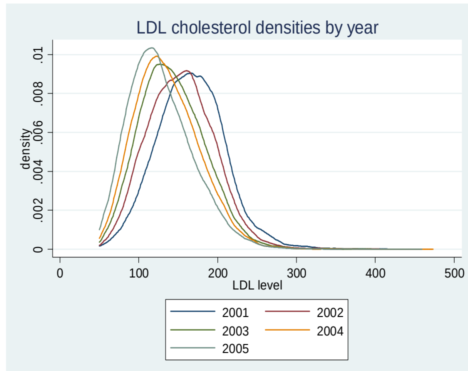
More importantly, we found that treatments with newer drugs, even after controlling for dosage, leads to a faster recovery to better health conditions. However, this effect could be seriously undermined if patients do not adhere to treatment and if GPs refuse to prescribe newer drugs to patients. In fact, we found evidence of a certain degree of conservativeness in GP behaviour, who tended to persist in the use of older statins for the long-term treated group, partially slowing down their recovery process. From a quantitative perspective, we observe that technology, although being the driving factor in increasing the speed of recovery, can explain at the best 62% of the total effect, with patients and physicians who are responsible for the remaining 25% and 13% respectively.

In conclusion, the evidence obtained from this work sheds light on the importance of technical progress (both in terms of product and process innovation) for a full and faster health recovery, which could be even more effective if this technical advancement was made immediately available to all patients. It could be then worthwhile for health care decision makers to promote specific educational campaigns aimed at disseminating the results of studies like EUROASPIRE or REALITY (Ruiz and Ibáñez, 2004) among GPs and specialists, demonstrating the importance of reaching LDL target and trying to resolve potential problems induced by physicians affected by high degree of conservativeness. Overall, as also proven in Atella et al. (2009), such a policy could in the future lead to potential savings for the health care system as a whole.

Acknowledgements: This work would have not been possible without the enthusiastic support provided by people at the Italian College of General Practitioners (SIMG) and in particular Claudio Cricelli, Iacopo Cricelli and Gian Piero Mazzaglia. We are also grateful to Jay Bhattacharya, Decio Coviello, Domenico Depalo, Claudio Jommi, Partha Deb, Randall Ellis, Walter Holland, Marisa Miraldo, Carol Propper, Franco Peracchi, Olmo Silva, Tommaso Valletti and all seminar participants at Boston University, Harvard University, Imperial College London, London School of Economics, MIT, Stanford University, University of Brescia, University of Roma Tor Vergata, and SIE Conference 2009 in Rome for their useful comments and suggestions at different stages of the research. The authors gratefully acknowledge financial support from Fondazione SIMG. All usual disclaimers apply.

Tables and figures

Figure 1
Trends and Distributions of Cholesterol Levels in the Health Search Population

Panel a	***Panel b***
	

Source: Our calculation based on HSD data.

Figure 2
Mean LDL Cholesterol by Cohort

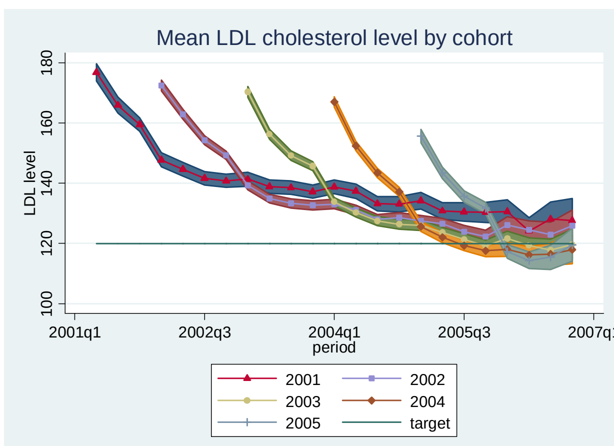


Figure 3
Trends in Market Share of Statin Drugs Sold in Italy

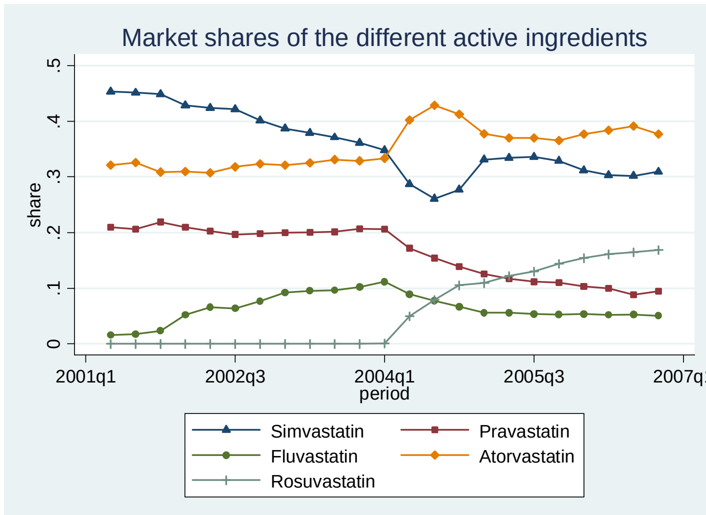


Figure 4
Trends in Market Share by Pack Size

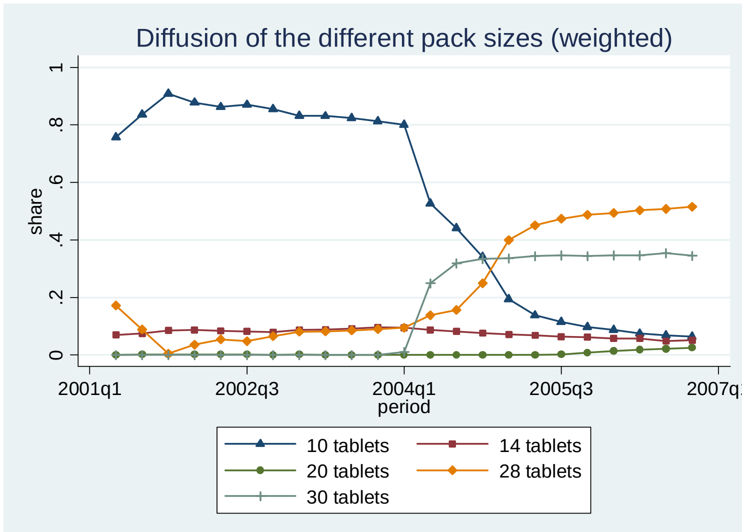


Figure 5
Possible Patterns of Cholesterol Trends

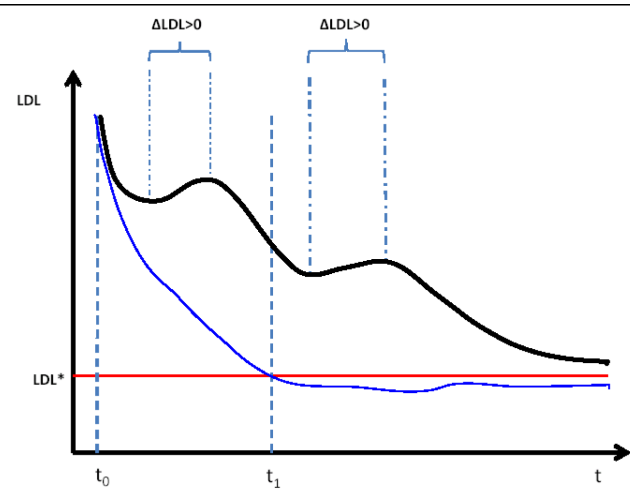


Table 1a Average LDL Levels by Year					
	Cohort				
Year	2001	2002	2003	2004	2005
2001	162.6	-	-	-	-
2002	139.3	153.2	-	-	-
2003	134.9	130.6	148.2	-	-
2004	132.0	126.3	125.4	143.2	-
2005	128.0	123.0	120.4	118.3	136.2
2006	127.0	123.6	118.7	114.4	114.6

Table 1b Variations in Average LDL Levels					
	Cohort				
Year	2001	2002	2003	2004	2005
2001	-	-	-	-	-
2002	-14.3%	-	-	-	-
2003	-3.1%	-14.8%	-	-	-
2004	-2.2%	-3.3%	-15.4%	-	-
2005	-3.0%	-2.6%	-4.0%	-17.4%	-
2006	-0.8%	0.5%	-1.4%	-3.3%	-15.9%

Table 2 Observations Selection		
Selection Criteria	No. Obs.	No. Patients
Initial sample level	1272797	42140
Dropping residual double records	1272607	42140
Merging with adherence information	1823677	42140
Collapsing data to quarters	1011360	42140
Dropping observation with missing LDL values	292940	34474
Dropping observation outliers for λ (dep. Var.) - 1^{st} and 99^{th} percentiles	93042	9510
Dropping observation with missing values for λ	91182	9443
Dropping patients from first quarter 2001	43649	6945
Dropping missing prescriptions in the quarter	23857	4917
Dropping for missing double lagged LDL	21200	4293

Table 3
Descriptive Statistics for the Full Sample

Variable	Obs	Mean	Std. Dev.	Min	Max
λ	21200	0.131	0.655	-4.3	5.6
Log LDL (t-2)	21200	4.972	0.244	4.0	6.1
Male	21200	0.509	0.500	0	1
Age Class 39-50	21200	0.049	0.215	0	1
Age Class 50-60	21200	0.230	0.421	0	1
Age Class 60-70	21200	0.484	0.500	0	1
Age Class 70+	21200	0.237	0.425	0	1
North West	21200	0.247	0.431	0	1
North East	21200	0.260	0.438	0	1
Centre	21200	0.146	0.353	0	1
South	21200	0.235	0.424	0	1
Islands	21200	0.112	0.315	0	1
Simvastatin	21200	0.319	0.466	0	1
Pravastatin	21200	0.164	0.370	0	1
Fluvastatin	21200	0.118	0.323	0	1
Atorvastatin	21200	0.310	0.462	0	1
Rosuvastatin	21200	0.089	0.284	0	1
Hypertensive	21200	0.594	0.491	0	1
Diabetes	21200	0.287	0.452	0	1
Secondary Prevention	21200	0.094	0.291	0	1
Persistent	21200	0.702	0.457	0	1
Cohort 2001	21200	0.174	0.379	0	1
Cohort 2002	21200	0.410	0.492	0	1
Cohort 2003	21200	0.250	0.433	0	1
Cohort 2004	21200	0.135	0.342	0	1
Cohort 2005	21200	0.030	0.171	0	1
Smoker	21200	0.017	0.130	0	1
Drug Milligrams	21200	27.448	20.087	5	80
Exemption: Age	21200	0.279	0.449	0	1
Exemption: CVD	21200	0.100	0.301	0	1
Exemption: Invalidity	21200	0.059	0.235	0	1
Exemption: Income	21200	0.048	0.213	0	1
Doctor is a Female	21200	0.119	0.324	0	1
Doctor's Age	21200	50.712	4.034	35	67
Year 2002	21200	0.084	0.277	0	1
Year 2003	21200	0.229	0.420	0	1
Year 2004	21200	0.309	0.462	0	1
Year 2005	21200	0.284	0.451	0	1
Year 2006	21200	0.093	0.292	0	1

Table 4
Sample Means by Cohort for the Full Sample

Variable	Cohort				
	2001	2002	2003	2004	2005
λ	0.111	0.208	0.221	0.369	0.399
Initial LDL	166.1	170.1	170.5	167.6	161.8
Target LDL	102.0	102.2	101.1	99.1	98.4
LDL (t-1)	149.6	149.3	146.9	145.2	145.7
LDL (t-2)	150.8	152.6	152.2	153.1	156.1
Hospitalized	0.011	0.016	0.011	0.017	0.023
Patients reaching Target	0.062	0.102	0.110	0.162	0.159
Male	0.482	0.468	0.473	0.512	0.509
Age	63.382	62.868	63.051	63.272	62.763
North West	0.254	0.251	0.232	0.234	0.212
North East	0.239	0.239	0.247	0.244	0.232
Centre	0.203	0.128	0.137	0.154	0.138
South	0.180	0.264	0.274	0.255	0.274
Islands	0.125	0.117	0.110	0.113	0.144
Hypertensive	0.494	0.571	0.594	0.638	0.673
Diabetes	0.227	0.232	0.254	0.320	0.399
Secondary Prevention	0.064	0.107	0.079	0.098	0.130
Persistent	0.581	0.564	0.639	0.715	0.632
Good Habits	0.238	0.324	0.343	0.473	0.491
Bad Habits	-0.127	-0.116	-0.121	-0.104	-0.092
No. Patients	473	1396	1148	936	340

Table 5
Descriptive Statistics for the Reduced Sample

Variable	Obs	Mean	Std. Dev.	Min	Max
λ	14595	0.131	0.629	-4.2	4.8
Log LDL (t-2)	14595	4.975	0.241	4.1	6.1
Male	14595	0.506	0.500	0	1
Age Class 39-50	14595	0.047	0.211	0	1
Age Class 50-60	14595	0.228	0.420	0	1
Age Class 60-70	14595	0.486	0.500	0	1
Age Class 70+	14595	0.240	0.427	0	1
North West	14595	0.244	0.430	0	1
North East	14595	0.260	0.439	0	1
Centre	14595	0.157	0.364	0	1
South	14595	0.223	0.416	0	1
Islands	14595	0.116	0.320	0	1
Simvastatin	14595	0.325	0.469	0	1
Pravastatin	14595	0.171	0.377	0	1
Fluvastatin	14595	0.121	0.326	0	1
Atorvastatin	14595	0.304	0.460	0	1
Rosuvastatin	14595	0.078	0.269	0	1
Hypertensive	14595	0.592	0.492	0	1
Diabetes	14595	0.292	0.455	0	1
Secondary Prevention	14595	0.085	0.279	0	1
Persistent	14595	0.753	0.431	0	1
Cohort 2001	14595	0.208	0.406	0	1
Cohort 2002	14595	0.374	0.484	0	1
Cohort 2003	14595	0.240	0.427	0	1
Cohort 2004	14595	0.141	0.348	0	1
Cohort 2005	14595	0.037	0.189	0	1
Smoker	14595	0.016	0.126	0	1
Drug Milligrams	14595	27.692	20.109	5	80
Exemption: Age	14595	0.301	0.459	0	1
Exemption: CVD	14595	0.096	0.295	0	1
Exemption: Invalidity	14595	0.057	0.232	0	1
Exemption: Income	14595	0.040	0.196	0	1
Doctor is a Female	14595	0.120	0.325	0	1
Doctor's Age	14595	50.705	4.010	35	67
Year 2002	14595	0.122	0.327	0	1
Year 2003	14595	0.247	0.431	0	1
Year 2004	14595	0.299	0.458	0	1
Year 2005	14595	0.254	0.435	0	1
Year 2006	14595	0.079	0.269	0	1

Table 6 Sample Means by Cohort for the Reduced Sample					
Variable	Cohort				
	2001	2002	2003	2004	2005
λ	0.115	0.208	0.254	0.379	0.417
Initial LDL	168.3	174.7	174.5	172.1	163.1
Target LDL	101.6	101.3	101.2	98.7	97.9
LDL (t-1)	148.7	147.9	147.2	146.2	145.3
LDL (t-2)	150.3	152.1	154.0	155.7	157.1
Hospitalized	0.008	0.012	0.011	0.021	0.023
Patients reaching Target	0.057	0.096	0.117	0.162	0.169
Male	0.493	0.474	0.454	0.494	0.532
Age	63.026	63.018	63.142	63.302	62.805
North West	0.259	0.244	0.241	0.202	0.198
North East	0.230	0.246	0.270	0.245	0.248
Centre	0.201	0.138	0.144	0.163	0.155
South	0.192	0.256	0.234	0.269	0.263
Islands	0.117	0.116	0.111	0.121	0.137
Hypertensive	0.500	0.563	0.588	0.642	0.679
Diabetes	0.227	0.241	0.249	0.326	0.398
Secondary Prevention	0.067	0.094	0.072	0.077	0.131
Persistent	0.679	0.641	0.708	0.775	0.662
Good Habits	0.228	0.316	0.361	0.467	0.500
Bad Habits	-0.114	-0.108	-0.106	-0.088	-0.083
No. Patients	343	741	667	595	278

Table 7 Attrition by Cohort for the Full Sample					
	Cohort				
Year	2001	2002	2003	2004	2005
2002	343	741	-	-	-
2003	363	1104	667	-	-
2004	351	992	942	595	-
2005	279	761	690	731	278
2006	124	296	282	282	177
Patients	473	1396	1148	936	340

Table 8 Distribution of the First Year of Appearance in the Full Sample					
	Cohort				
Year	2001	2002	2003	2004	2005
2002	343	741	-	-	-
2003	77	475	667	-	-
2004	36	111	398	595	-
2005	15	54	73	306	278
2006	2	15	10	35	62
Patients	473	1396	1148	936	340

Table 9 Attrition by Cohort for the Reduced Sample					
	Cohort				
Year	2001	2002	2003	2004	2005
2002	343	741	-	-	-
2003	286	629	667	-	-
2004	256	519	544	595	-
2005	195	396	371	425	278
2006	81	152	146	160	115

Table 10 Hospitalization Probability		
Variable	Coefficient	Elasticity
λ_{t-1}	-0.226**	-0.079
LDL *t-1*	-0.140	-1.705
Male	0.260**	0.310
Age class 50-60	-0.083	-0.050
Age class 60-70	0.098	0.122
Age class 70+	-0.012	-0.005
North West	-0.306*	-0.181
North East	-0.253	-0.153
Centre	-0.143	-0.053
South	-0.130	-0.079
Hypertensive *t-1*	0.208**	0.280
Diabetes *t-1*	0.082	0.050
Secondary Prevention *t-1*	0.516***	0.102
Cohort 2002	-0.129	-0.106
Cohort 2003	-0.060	-0.039
Cohort 2004	-0.071	-0.032
Cohort 2005	0.105	0.011
Constant	-1.415	
N. Patients	3323	

Table 11 RE Panel Estimates					
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Log LDL (t-2)	0.691***	0.692***	0.691***	0.727***	0.727***
Male	0.045**	0.043**	0.043**	0.043**	0.043**
Age Class 50-60	-0.004	-0.004	-0.003	-0.007	-0.007
Age Class 60-70	0.056	0.057	0.058	0.044	0.044
Age Class 70+	0.103***	0.104***	0.105***	0.079**	0.079**
North West	-0.062**	-0.056*	-0.056*	-0.057*	-0.058*
North East	-0.061*	-0.055*	-0.054*	-0.056*	-0.056*
Centre	-0.049	-0.049	-0.051	-0.053	-0.054
South	-0.018	-0.012	-0.012	-0.011	-0.012
Simvastatin	0.077***	0.214***	0.142	0.139	0.139
Pravastatin	-0.041	0.083**	0.056	0.048	0.049
Atorvastatin	0.137***	0.301***	0.220**	0.210**	0.211**
Rosuvastatin	0.206***	0.386***	0.364***	0.338***	0.339***
Hypertensive	0.031*	0.030*	0.030*	0.022	0.022
Diabetes	0.001	0	0.001	-0.002	-0.002
Secondary Prevention	0.144***	0.140***	0.139***	0.131***	0.131***
Persistent	0.079***	0.079***	0.080***	0.085***	0.085***
Cohort 2002	0.072**	0.071**	0.071**	0.023	0.022
Cohort 2003	0.088***	0.084***	0.084***	0.021	0.021
Cohort 2004	0.157***	0.154***	0.154***	0.084**	0.084**
Cohort 2005	0.174***	0.171***	0.171***	0.147***	0.148***
Smoker	-0.028	-0.027	-0.027	-0.027	-0.027
Exemption: Age	-0.067***	-0.067***	-0.067***	-0.046***	-0.046***
Exemption: CVD	0.053**	0.051**	0.050**	0.046**	0.046**
Exemption: Invalidity	0.004	0.005	0.005	0.007	0.006
Exemption: Income	-0.014	-0.014	-0.014	-0.023	-0.023
Drug Milligrams		0.003***	0.002*	0.002	0.002
Simvastatin x mg			0.001	0.001	0.001
Pravastatin x mg			0	0	0
Atorvastatin x mg			0.002	0.002	0.002
Rosuvastatin x mg			-0.002	-0.003	-0.003
Cohort 02 x Year 03				0.037	0.038*
Cohort 02 x Year 04				0.045*	0.046*
Cohort 02 x year 05				0.085***	0.087***
Cohort 02 x year 06				0.132***	0.135***
Cohort 03 x year 04				0.067***	0.068***
Cohort 03 x year 05				0.090***	0.091***
Cohort 03 x year 06				0.137***	0.139***
Cohort 04 x year 05				0.113***	0.114***
Cohort 04 x year 06				0.102***	0.103***
Cohort 05 x year 06				0.084	0.085
Doctor is Female					0.000
Doctor's Age					-0.001
Constant	-3.516***	-3.727***	-3.673***	-3.829***	-3.797***
No. Observations	21200	21200	21200	21200	21200
No. Patients	4293	4293	4293	4293	4293

Table 12 Percentage Distribution by Cohort of the Active Ingredients Prescribed (considering sample year 2005)				
Active Ingredient	**2002**	**2003**	**2004**	**2005**
Simvastatin	**34.38**	**32.37**	**26.95**	**27.8**
Pravastatin	**16.79**	**18.77**	**16.17**	**12.65**
Fluvastatin	**8.54**	**9.22**	**6.98**	**7.34**
Atorvastatin	**31.21**	**30.55**	**31.88**	**31.75**
Rosuvastatin	**9.08**	**9.08**	**18.02**	**20.47**
N. Obs.	**19,517**	**18,703**	**22,860**	**17,861**

Table 13a Inclusion Probability in the Full Sample (N=21200)						
Variable	**Coef.**	**Std. Err.**	**z**	**P>z**	**L.I.**	**U.I.**
Male	0.137***	0.018	7.440	0.000	0.101	0.173
Age	0.067***	0.016	4.130	0.000	0.035	0.099
Age squared	-0.001***	0.000	-4.170	0.000	-0.001	0.000
North West	0.059*	0.033	1.770	0.076	-0.006	0.124
North East	0.128***	0.032	3.940	0.000	0.064	0.191
Centre	0.048	0.036	1.330	0.184	-0.023	0.119
South	-0.038	0.033	-1.150	0.250	-0.102	0.027
Hyperthensive	-0.027	0.019	-1.440	0.150	-0.064	0.010
Diabetes	0.069***	0.020	3.390	0.001	0.029	0.109
Secondary Prevention	-0.053*	0.031	-1.720	0.086	-0.115	0.008
Doctor is Female	-0.010	0.028	-0.340	0.732	-0.065	0.046
Doctor's Age	-0.072**	0.034	-2.130	0.033	-0.138	-0.006
Doctor's Age squared	0.001	0.000	1.550	0.122	0.000	0.001
Constant	0.726	0.993	0.730	0.464	-1.219	2.672

Table 13b Inclusion Probability in the Reduced Sample (N=14595)						
Variable	**Coef.**	**Std. Err.**	**z**	**P>z**	**L.I.**	**U.I.**
Male	0.127***	0.023	5.620	0.000	0.083	0.171
Age	0.060***	0.020	2.950	0.003	0.020	0.099
Age squared	-0.000***	0.000	-2.990	0.003	-0.001	0.000
North West	0.050	0.040	1.230	0.219	-0.029	0.129
North East	0.109***	0.039	2.770	0.006	0.032	0.186
Centre	0.025	0.043	0.570	0.566	-0.060	0.110
South	-0.055	0.040	-1.380	0.168	-0.134	0.023
Hyperthensive	-0.046**	0.023	-2.010	0.045	-0.091	-0.001
Diabetes	0.061**	0.025	2.440	0.015	0.012	0.109
Secondary Prevention	-0.061	0.040	-1.530	0.125	-0.138	0.017
Doctor is Female	0.022	0.035	0.620	0.533	-0.046	0.090
Doctor's Age	-0.032	0.041	-0.780	0.434	-0.112	0.048
Doctor's Age squared	0.000	0.000	0.280	0.776	-0.001	0.001
Constant	0.052	1.213	0.040	0.966	-2.326	2.429

Table 14
RE Panel Weighted Estimates (Model 5)

Variable	Full Sample	Weighted Full	Reduced Sample	Weighted Reduced
Log LDL (t-2)	0.727***	0.730***	0.709***	0.714***
Male	0.043**	-0.007	0.031	-0.012
Age Class 50-60	-0.007	-0.037	-0.003	-0.027
Age Class 60-70	0.044	0.015	0.035	0.010
Age Class 70+	0.079**	0.066*	0.050	0.039
North West	-0.058*	-0.078**	-0.038	-0.057
North East	-0.056*	-0.103**	-0.033	-0.071
Centre	-0.054	-0.070*	-0.059	-0.070*
South	-0.012	0.007	0.012	0.033
Simvastatin	0.139	0.13	0.203*	0.17
Pravastatin	0.049	0.039	0.109	0.073
Atorvastatin	0.211**	0.197**	0.260**	0.227**
Rosuvastatin	0.339***	0.329***	0.340**	0.295**
Hypertensive	0.022	0.032*	0.037**	0.052**
Diabetes	-0.002	-0.027	-0.029	-0.049
Secondary Prevention	0.131***	0.152***	0.115***	0.137***
Persistent	0.085***	0.085***	0.075***	0.075***
Cohort 2002	0.022	0.021	0.034	0.033
Cohort 2003	0.021	0.019	0.052	0.049
Cohort 2004	0.084**	0.085**	0.103***	0.105***
Cohort 2005	0.148***	0.146***	0.174***	0.174***
Smoker	-0.027	-0.033	-0.029	-0.035
Exemption: Age	-0.046***	-0.055***	-0.030*	-0.034*
Exemption: CVD	0.046**	0.055**	0.026	0.030
Exemption: Invalidity	0.006	0.009	0.028	0.033
Exemption: Income	-0.023	-0.028	-0.027	-0.032
Drug Milligrams	0.002	0.002	0.003*	0.003
Simvastatin x mg	0.001	0.002	0.000	0.000
Pravastatin x mg	0.000	0.000	-0.001	-0.001
Atorvastatin x mg	0.002	0.003	0.003	0.004
Rosuvastatin x mg	-0.003	-0.003	0.000	0.002
Cohort 02 x Year 03	0.038*	0.044	0.068***	0.078***
Cohort 02 x Year 04	0.046*	0.055*	0.008	0.009
Cohort 02 x year 05	0.087***	0.105***	0.061**	0.071**
Cohort 02 x year 06	0.135***	0.166***	0.145***	0.178***
Cohort 03 x year 04	0.068***	0.083***	0.084***	0.102***
Cohort 03 x year 05	0.091***	0.110***	0.090***	0.108***
Cohort 03 x year 06	0.139***	0.171***	0.120***	0.146***
Cohort 04 x year 05	0.114***	0.136***	0.122***	0.144***
Cohort 04 x year 06	0.103***	0.124***	0.069	0.083
Cohort 05 x year 06	0.085	0.103	0.096*	0.115*
Doctor is Female	0.000	0.004	0.000	-0.009
Doctor's Age	-0.001	0.008	0.000	0.009
Constant	-3.797***	-3.905***	-3.818***	-3.939***
No. Observations	21200	21200	14595	14595
No. Patients	4293	4293	2624	2624

Table 15					
Relative Importance of Different Actors in Determining Speed of Recovery					
RE Panel					
Model Specification	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Patient behaviour	**26.2%**	**16.4%**	**19.5%**	**25.2%**	**25.3%**
Technology	**37.0%**	**61.3%**	**54.1%**	**62.0%**	**62.1%**
Medical behaviour	**36.8%**	**22.3%**	**26.4%**	**12.8%**	**12.6%**

Table 16				
Relative Importance of Different Actors in Determining Speed of Recovery (Model 5)				
RE Panel				
Sample Specification	**Full Sample**	**Weighted Full**	**Reduced Sample**	**Weighted Reduced**
Patient behaviour	**25.3%**	**26.3%**	**19.4%**	**21.5%**
Technology	**62.1%**	**60.8%**	**64.9%**	**61.0%**
Medical behaviour	**12.6%**	**12.9%**	**15.7%**	**17.5%**

References

Acemoglu, D. and S. Johnson. 2007. "Disease and development: the effect of life expectancy on economic growth." *Journal of Political Economy,* 115(6): 925-985.

American Heart Association. 2005. *Heart Disease and Stroke Statistics—2005 Update*. Dallas, Tex.

Antonopoulos, S. 2002. "Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report." *Circulation,* 106: 3143-421.

Arnett DK, McGovern PG, Jacobs DR Jr, Shahar E, Duval S, Blackburn H, Luepker RV. 2002**.** "Fifteen-year trends in cardiovascular risk factors (1980–1982 through 1995–1997): the Minnesota Heart Survey." *Am J Epidemiol*., 156:929 –935.

Arnett Donna K., David R. Jacobs, Russell V. Luepker, Henry Blackburn, Christopher Armstrong, Steven A. Claas. 2005. "Twenty-Year Trends in Serum Cholesterol, Hypercholesterolemia, and Cholesterol Medication Use The Minnesota Heart Survey, 1980–1982 to 2000–2002." *Circulation*., 112:3884-3891.

Atella V. and the TECH Investigators. 2003. "The relationship between health policies, medical technology trend, and outcomes: A perspective from the TECH Global Research Network" in *"What is Best and at What Cost: A Disease-Based Approach to Comparing Health Systems"*, OECD, Paris, France.

Atella, V. F. Peracchi, et al. 2006. "Drug Compliance, Co-Payment, and Health Outcomes: Evidence from a Panel of Italian Patients." *Health Economics* 15: 875-892.

Atella V., Belotti F., D'Amico F., Catapano A., Cortese C. and Cricelli C. 2009. "Can better drug treatment improve long-run financial sustainability of NHS in Italy? The case of Statins", *CEIS Research Paper.*

British Heart Foundation Statistics Database. 2005. *Coronary heart disease statistics*, [http://www.heartstats.org] website

Cutler D. M., McClellan M. 2001. "Is Technological Change In Medicine Worth It?" *Health Affairs*, 25, no. 2, pp.34-47.

Cutler, D. M., A. B. Rosen, et al. 2006. "The value of medical spending in the United States, 1960-2000." *The New England Journal of Medicine* 355(9): 920.

Eliasson M., Janlert U., Jansson J.-H., Stegmayr B. 2006. "Time Trends In Population Cholesterol Levels 1986–2004: Influence Of Lipid-Lowering Drugs, Obesity, Smoking And Educational Level. The Northern Sweden MONICA Study", *Journal Of Internal Medicine*, 260: 551–559.

EUROASPIRE II. 2001. "Lifestyle and risk factor management and use of drug therapies in coronary patients from 15 countries. Principal results from EUROASPIRE II Euro Heart Survey Programme", *European Heart Journal*, 22:554-572.

Fabiani, L., M. Scatigna, et al. 2004. "Health Search: istituto di ricerca della società italiana di medicina generale; la realizzazione di un database per la ricerca in medicina generale." *Epidemiol & Prev* 28: 156-62.

Federman, A.D., Alyce, S.A., Ross-Degnan, D., Sourmerai, S.B. and Ayanina, J.Z. 2001. "Supplemental Insurance and Use of Elective Cardiovascular Drugs among Elderly Medicare Beneficiaries with Coronary Heart Disease." *JAMA*,, 286(14), pp. 1732-39.

Ford ES, Mokdad AH, Giles WH, Mensah GA. 2003. "Serum total cholesterol concentrations and awareness, treatment, and control of hypercholesterolemia among US adults: findings from the National Health and Nutrition Examination Survey, 1999 to 2000.", *Circulation*. 2003;107:2185–2189.

Giampaoli S., Palmieri L., Mattiello A., et al. “Definition of high risk individuals to optimise strategies for primary prevention of cardiovascular diseases”, Nutr Metab Cardiovasc Dis 2005; 15: 79-85.

Gowrisankaran, G. and Town R.J. 2004. “Managed Care, Drug Benefit and Mortality: An Analysis of the Elderly.” *NBER Working Papers,* No. 10204.

Grootendorst P., Piérard E., and M. Shim. 2007. “The life expectancy gains from pharmaceutical drugs: a critical appraisal of the literature”, SEDAP Research Paper No. 221.

Grossman, M. 1972. *The demand for health: a theoretical and empirical investigation*, National Bureau of Economic Research; distributed by Columbia University Press.

Gurmankin, A. D., J. Baron, J. C. Hershey, P. A. Ubel. 2002. “The Role of Physicians’ Recommendations in Medical Treatment Decisions” *Medical Decision Making*, Vol. 22, No. 3, 262-271

Hyman DJ, Pavlik VN and Vallbona C., 2000. “Physician Role in Lack of Awareness and Control of Hypertension”, *J Clin Hypertens* (Greenwich). Oct;2(5):324-330.

Johansson, E., Diwana, V. K. , Huongb, N. D. and Ahlberga, B. M. 1996. **“**Staff and patient attitudes to tuberculosis and compliance with treatment: An exploratory study in a district in Vietnam”, *Tubercle and Lung Disease,* Volume 77, Issue 2, April 1996, Pages 178-183

Jones, Charles I. 2005. “More Life vs. More Goods: Explaining Rising Health Expenditures”, *FRBSF Economic Letter* 2005-10 (May 27).

Lichtenberg F. R. 1996. “Do (More and Better) Drugs Keep People Out of Hospitals?”, *The American Economic Review*, Vol. 86, n. 2, pp. 384-388

Lichtenberg F. R. 2000a. “Sources of U.S. longevity increase, 1960-1997”, *CESifo Working Paper* n. 405.

Lichtenberg F. R. 2001. “Are the benefits of newer drugs worth their cost? Evidence from the 1996 MEPS.” *Health Aff (Millwood)*, 20 (5), pp. 241-51.

Lichtenberg F. R. 2006. “The effect of using newer drugs on admissions of elderly Americans to hospitals and nursing homes: state-level evidence from 1997-2003”, *Pharmacoeconomics* 24 Suppl 3:5-25.

Lichtenberg, F. R. 2006a. "On New Cardiovascular Drugs: Pattern of Use and Association with Non-Drug Health Expenditures'". *Inquiry*. Spring; 43(1): 80-2.

Lichtenberg, F. R. 2007. "Pharmaceutical Innovation and U.S. Cancer Survival, 1992-2003: Evidence from Linked SEER-MEDSTAT Data." *Forum for Health Economics & Policy*: Vol. 10: Iss. 1.

Lichtenberg F. R. 2008. “Have newer cardiovascular drugs reduced hospitalization? Evidence from longitudinal country-level data on 20 OECD countries 1995-2003”, *NBER Working Papers*, No 14008.

Long G., Cutler D. M., Berndt E. R. 2007. “The value of antihypertensive drugs : a perspective on the value of improved blood pressure control in the USA”, *European Heart Journal Supplements*, Vol. 9, Suppl. B.

McClellan M.B., Kessler D.P. 2002, *Technological change in health care. A global analysis of heart attack*, The university of Michigan Press, Ann Arbor.

McClellan M. and Kessler D. 1999. “A Global Analysis Of Technological Change In Health Care: The Case Of Heart Attacks”, *Health Affairs*, Vol. 18, No. 3, pp. 250-255.

McCombs J., Thiebaud, P., McLaughlin-Miley, C. and Shi., J. 2004 “Compliance with drug therapies for the treatment and prevention of osteoporosis”, *Maturitas*, Volume 48, Issue 3, Pages 271-287

Murphy, K. M. and R. H. Topel 2006. "The value of health and longevity." *Journal of Political Economy* 114(5): 871-904.

Muurinen, Jaana-Marja 1982. “A Generalised Grossman Model”**,** *Journal of Health Economics*, Vol 1, p. 5-28.

Pearson TA et al. 2000. “The lipid treatment assessment project (L-TAP): a multicenter survey to evaluate the percentages of dyslipidemic patients receiving lipid-lowering therapy and achieving low-density lipoprotein cholesterol goals.”, *Arch Intern Med.*, 28;160(4):459-67.

Ruiz, F. J. G., A. M. Ibáñez, et al. 2004. "Current Lipid Management and Low Cholesterol Goal Attainment in Common Daily Practice in Spain: The REALITY Study." *Pharmacoeconomics* 22: 1.

Sedgley M. and T. Hockley 2006. “A Comprehensive Review of the existing literature on Cholesterol Management”, in *Cholesterol: The Public Policy Implications of Not Doing Enough*, Stockholm Network.

Sessa, E., F. Samani, et al. 2004. "La creazione di un campione validato di Medici di Medicina Generale nel database di Health Search." *SIMG* 3: 10-4.

Soumerai, S.B., Ross-Degnan, D., Avorn, J., McLaughlin, T.J. and Choodnovskiy, I. 1991. “Effects of Medicaid Drug-Payment Limits on Admission to Hospitals and Nursing Homes.” *New England Journal of Medicine*, 1991, 325(1072-1077).

Soumerai, S.B., McLaughlin, T.J., Ross-Degnan, D., Casteris, C.S. and Bollini, P. 1991. “Effects of Limiting Medicaid Drug-Reimbursement Benefits on the Use of Psychotropic Agents and Acute Mental Health Services by Patients with Schizophrenia.” *New England Journal of Medicine*, 331(650-655).

Tamblyn, R., Laprise, R., Hanley, J.A., Abrahamowicz, M., Scott, S., Mayor, N., Hurley, J., Grad, R., Latimer, E., Perreault, R., et al. 2001. “Adverse Events Associated with Prescription Drug Cost-Sharing among Poor and Elderly Persons.” *JAMA*, 285(4), pp. 421-29.

Unal B., Critchley J.A., Capewell S. 2004. “Explaining the Decline in Coronary Heart Disease Mortality in England and Wales Between 1981 and 2000”, *Circulation*;109:1101-1107.

Valletta R. 2007. “The Costs and Value of New Medical Technologies: Symposium Summary”, *FRBSF Economic Letter*, Number 2007-18, July 6, 2007.

Yang, Z., Gilleskie, D. B., and Norton, E. C. 2009. “Health Insurance, Medical Care, and Health Outcomes: A Model of Elderly Health Dynamics.” *Journal of Human Resources, 44(1)*, 47-114.

Weisfeldt, ML, and SJ Zieman 2007,. “Advances in the prevention and treatment of cardiovascular disease” *Health Affairs* 26(1), Jan-Feb, 25-37.

Chapter 2

Stochastic Frontiers and Technical Efficiency: evidences from a panel of Italian hospitals [38]

2.1 Introduction

The main objective of this paper is to evaluate how the productive structure and level of specialization of a hospital affect its technical efficiency. Here, we define productive structure as the degree of capitalization of the hospital, while the degree of specialization refers to the number of different types of cases treated within the organization. To this end, we report an economic analysis measuring the evolution of technical efficiency in hospitals located in the Italian region of Lazio in the 2000-2005 period. Subsequently, we assess the robustness of the hospitals' optimizing behavior about ten years after the introduction of the Diagnosis Related Group (DRG) system, and explore differences in efficiency as related to the hospitals' ownership structure. Finally, we offer some insights as to how hospital scale of activity and productive conditions are determined by the institutional framework. In Italy, the introduction of the DRG system in 1992 served both as a means of prospective payment and as an instrument for efficient allocation of resources. This should have made it possible to properly classify, measure, and assess hospitals' performance by using industrial organization and management methods. Moreover the system applied to all organizations, independent of whether they were privately or publicly owned. As a consequence of the reform, hospitals were made responsible for their own outcomes, partially assuming the burden of financial risk. Further, budget constraints were made more binding for financing institutions, and this affected incentives to curb health consumption and production. As a result, two patterns emerged: on one hand, producers were encouraged to optimize their productive processes given the available inputs, and on the other hand, resource availability was reduced.
The theoretical industrial organization literature refers to several factors which might affect the level of technical efficiency reached by different ownership structures. Alas, there is no consensus regarding the net direction and size of these effects. Neoclassical theory advocates that nonprofits (both public and private) have a propensity to opt for administrators more interested in providing high-quality service than in producing profits. This type of hospitals might use more input to produce the same output as for-profit hospitals. Further, the presence of residual claimants should represent a powerful incentive to efficient production among for-profit hospitals (Alchian and Demsetz, 1972). In addition to these neoclassical arguments, developments on a strand of economic

38 This chapter is based on “Technical Efficiency, Specialization and Ownership Form: evidences from a Pooling of Italian Hospitals” by F. D’Amico and S. Daidone, published in *Journal of Productivity Analysis*, 32:203-216 (2009).

literature that concentrates on information asymmetries have been used to provide a rationale for the existing differences in productive efficiency. While the lack of a residual claimant *a-la* Alchian-Demsetz, might reduce managerial efforts and consequently having a negative impact on efficiency, the non distribution constraint could represent an influential tool for controlling information asymmetries among diverse stakeholders (Hansmann, 1996), increasing their efficiency by augmenting demand. Nonetheless, the impact of these characteristics on efficiency is still unclear.

The empirical literature on hospital efficiency in Italy consists of a substantial number of studies with a variety of methodologies, scopes, and results. Unexploited economies of scale are a recurrent theme (see Grassetti et al., 2005). Public hospital trusts (Aziende Ospedaliere, hereforeward referred to as AO) appear more efficient than other hospital types, such as acute-care and rehabilitation hospitals directly managed by local health authorities (Presidi Ospedalieri, ASLH)[39], but such a crystal clear difference is not observed in a comparison of public and private hospitals. Finally, technical efficiency seems to have decreased in the second half of the nineties, after the introduction of prospective payments under the DRG (see Barbetta et al., 2007).

Any analysis of hospital efficiency must take into account the so-called "Newhouse criticism" (Newhouse, 1994). Model specifications are, in fact, generally restrictive: they omit some relevant inputs and outputs without taking into account the quality characteristics of the services provided. Further, hospital outputs are extremely heterogeneous. The number of discharged patients gives a rough measure of overall hospital production, if we do not take into account other aspects of treatment, such as the type and the severity of illness, the presence of other illnesses, the overall characteristics of the patient, and the like. We address these critiques by: 1) using hospital discharge records that allow to construct precise measures of output and controls of hospital case-mix; 2) representing technology with distance functions, which are more adequate than simple production functions in a multi-input multi-output setting. However, we acknowledge that this approach is still limited, since distance functions may be still mis-specificied. In absence of good measures of hospital quality, we decided not include the available indicators in our analysis and assessed that this does not invalidate our result.

The paper is organized as follows: in the next section, we describe the econometric technique used for the estimation of efficiency scores. Subsequently, we present our data and provide some summary statistics, and in section 2.4.1 we report the results. We conclude by commenting on our findings.

2.2 Stochastic distance functions

The notion of technical efficiency refers to producers' choices to allocate the resources at their disposal to obtain the maximum possible output from given inputs, or to use the minimum possible inputs in the production of a given level of outputs. Therefore, the analysis of technical efficiency may be defined as either output-oriented or input-

39 For a good and quick review of the different nature of these structures, see European Observatory of Health Care Systems (2001).

oriented. When multiple inputs are used to produce multiple outputs, Shephard's distance functions (Shephard, 1970) provide a characterization of the structure of production technology. Concept and properties of distance functions are well described in Kumbhakar and Lovell (2000}, to whom we refer the reader. In what follows we have considered the number of discharges corrected for its case-mix complexity as final goods of the productive process. More than technical efficiency, we are measuring what has been defined by Zweifel and Breyer (1997) “internal medical efficiency”. However we are totally ignoring “external medical efficiency”, or hospitals' effectiveness, defined as the ability of a hospital to improve the health status of a patient, for a given number of discharges or a given length of stay. A definition of output including external medical efficiency would be conceptually more accurate than the definition we use, but this would be empirically far more difficult to analyze, due to the lack of data reporting indicators of patients' health status following post-discharge recovery.

A general cross-sectional multiple-output stochastic frontier model can be written as

$$D^O(x_i, y_i, z_i; \beta) = \exp(v_i - u_i) \tag{1}$$

where $D^O(x_i, y_i, z_i; \beta)$ is the output distance function used to represent the distance from the frontier, which allows technological interactions across and among inputs and outputs. y_i represents the output vector of the *i*-th hospital, x_i is a vector of inputs and z_i is a vector of hospital specific characteristics other than inputs. β is the parameter vector which describes the structure of the technology.

The error component is divided in two parts: v_i is an idiosyncratic normal term with zero mean, whereas u_i is a one-sided asymmetric negatively skewed distribution which captures inefficiency among observations. An analogous substitution relationship between factors and outputs showing deviations from the frontier is given by the input distance function

$$D^I(x_i, y_i, z_i; \beta) = \exp(v_i - u_i) \tag{2}$$

Equations (1) and (2) can be rewritten as stochastic distance function models

$$1 = D^O(x_i, y_i, z_i; \beta) = \exp(v_i - u_i) \tag{3}$$

$$1 = D^I(x_i, y_i, z_i; \beta) = \exp(v_i - u_i) \tag{4}$$

The dependent variable in (3) and (4) is a constant, which has zero variance. Therefore in order to empirically estimate both equations we have to convert them into estimable models. This task can be accomplished by exploiting the fact that D^O is linear homogenous in outputs, while D^I in inputs, i.e.

$$D^O(x_i, \omega y_i, z_i; \beta) = \omega D^O(x_i, y_i, z_i; \beta) \quad \forall\ \omega > 0 \tag{5}$$

$$D^I(\omega x_i, y_i, z_i; \beta) = \omega D^I(x_i, y_i, z_i; \beta) \quad \forall\ \omega > 0 \tag{6}$$

As it has been pointed out by Lovell et al. (1994), one way of imposing such restriction is to normalize D^O and D^I respectively by one of the outputs and one of the inputs (so called ``ratio" model), i.e.

(7) $D^O(x_i, y_i, z_i; \beta)/y_{1i} = D^O(x_i, y_i^*, z_i; \beta)$ where $y_i^* = \dfrac{y_i}{y_{1i}}$

(8) $D^I(x_i, y_i, z_i; \beta)/x_{1i} = D^O(x_i^*, y_i, z_i; \beta)$ where $x_i^* = \dfrac{x_i}{x_{1i}}$

which lead to

(9) $D^O(x_i, y_i, z_i; \beta) = y_{1i} D^O\left(x_i, \dfrac{y_i}{y_{1i}}, z_i; \beta\right)\exp(u_i - v_i)$

(10) $D^I(x_i, y_i, z_i; \beta) = x_{1i} D^O\left(\dfrac{x_i}{x_{1i}}, y_i, z_i; \beta\right)\exp(u_i - v_i)$

Substituting equalities (9) and (10) into equations (3) and (4) and dividing both sides respectively by y_{1i} and x_{1i} generates the following estimable composed error models

(11) $(y_{1i})^{-1} = D^O\left(x_i, \dfrac{y_i}{y_{1i}}, z_i; \beta\right)\exp(u_i - v_i)$

(12) $(x_{1i})^{-1} = x_{1i} D^O\left(\dfrac{x_i}{x_{1i}}, y_i, z_i; \beta\right)\exp(u_i - v_i)$

In the multi-output version of the model, the dependent variable is the reciprocal of the normalizing output, and the regressors are the inputs and the normalized outputs. Finally u_i provides the basis for a reciprocal measure of output-oriented technical efficiency. Similar considerations apply for the multi-input case.

Some authors like Kumbhakar and Lovell (2000) have proposed the Euclidean norm of the outputs (or inputs) in order to respect the linear homogeneity restriction (so called “norm” model). In both the “norm” and “ratio” model there is an issue of endogeneity bias, since normalized outputs and inputs appearing as regressors may not be exogenous. Some authors like Coelli and Perelman (2000) or Morrison and Johnston (2000) argue the normalization $\dfrac{y_i}{y_{mi}}$ creates an output mix vector that is more likely to be exogenous than either y_i or $\dfrac{y_i}{|y_i|}$ It is not the objective of the paper to enter into this debate. For practical purposes we preferred the "ratio" model, since normalizing by the norm of the outputs increases considerably the degree of multicollinearity, due to the introduction of exact linear dependencies.

Note that while modelling multiple-input, multiple-output technologies we have to restrictively assume that the disturbance terms affect the output vector y multiplicatively, i.e. all outputs are assumed to be proportionally affected by the same disturbance.

Three major drawbacks affect cross-sectional stochastic frontier models:

1. Estimations strongly rely on distributional assumptions on each error component.
2. Technical (cost) inefficiency might be correlated with the regressors.
3. Technical and cost efficiency cannot be consistently estimated, because the variance of the conditional mean or mode does not go to zero as the sample size increase.

2.3 Data and summary statistics

We analyze data provided by the public health agency (PHA) of the Lazio region consisting of hospital discharge records (HDRs). Hospitals' administrators have to fill out HDR according to law as informative tool over which to base patients' admission financing. In each HDR, data concerning the patient are recorded. Beyond vital statistics, we observe health-related information collected during the patient's internment including date of admission and discharge, date of surgery, surgery type, transfers to other hospitals, and the like. Further, we observe the DRG related to each patient and its weight, i.e. the level of complexity.

Generally, the number of discharges is strongly correlated with the sum of DRG weights, and this is supported by our dataset.

However when we focus on groups of more complex cases, this correlation may be lower. This, for example, is what typically is observed in hospitals with emergency rooms. Generally studies of hospital efficiency do not have such disaggregated data at disposal and are forced to use all discharges or admissions adjusted by a measure of complexity. Therefore, the observation of the DRG weight is extremely important since it not only makes it possible to capture the precise number of treated cases, but also provides a measure of case-mix control for each output.

The aggregation of the DRGs followed a classification system which is commonly used at an international level, and which has already been applied to Italian data (see Fabbri 2003). Within this classification, hospital activity may be summarized in twenty-eight production categories. These are made up of DRG groups consistent with a standard of productive homogeneity[40]. In order to make the model empirically more manageable, we have further aggregated these 28 lines into the following groups: complex surgery, emergency room treatments, cancers and HIV, general medicine, and general surgery.

---

40 A hospital is basically viewed as a human service enterprise whose primary function is the provision of diagnostic and therapeutic medical services. Production lines are specific sets of services provided to individual patients and largely coincide with an appropriate definition of treatments within each ward typology.

As far as the measurement of output is concerned, frontier techniques seem to work best when the product is homogeneous and one-dimensional such as, for instance, kilowatt-hours in the electricity industry. This is not the case for hospital care, which exhibits wide variation in the quality of the product and its dimensionality, both on the input and the output sides. Therefore, it would be possible that productive units being assessed might not use the same technologies. In order to control at least partially for these types of differences, we have restricted our focus to acute patients. Acute care refers to the necessary treatment of a disease for a short period of time in which a patient is treated for a brief but severe episode of illness. The goal of the hospital is to discharge the patient as soon as he or she is deemed sufficiently healthy and stable following the critical period. Acute care differs from long-term care and rehabilitation care, which are characterized by a combination of treatments provided once the acute phase of the disease has been overcome. These treatments aim to stabilize the disease towards two possible outcomes: recovery or management of a chronic condition.

Hospitals devoting their activity exclusively to long-term and/or rehabilitation care, therefore, have very different production functions from acute-care hospitals, which would make the two groups impossible to compare. For this reason, we did not include such hospitals in the sample. For those hospitals dedicated only partially to post-acute care, we did not include the DRG weights from post-acute care activities in any of the output aggregates. Relatedly, when hospital activity is limited to the treatment of acute patients, this simplifies the debate as to whether the variable measuring output should be the number of discharges or the number of inpatient days. The latter variable may more heavily reflect the assistance component of hospital production, which is unique to structures dedicated to long-term care and rehabilitation, and hence it may reflect a productive choice of the hospital rather than the hospital's efficiency level[41].

We obtained data concerning inputs from the Italian Ministry of Health. These included: number of beds as a rough measure of capital and number of physicians, nurses and other personnel (teaching plus ancillary staff) as a measure of labor. By classifying labor into different categories we recognized differing skill requirements. We are aware that number of hours worked may be a better indicator of the labor factor, since it reveals more about the use of the workforce, but unfortunately such information is not available[42].

As mentioned above in discussing the measure of outputs, some hospitals undertake both acute and post-acute care activity. Therefore, for consistency, we must eliminate this element from the inputs. For the bed variable, this exclusion was straightforward, since we received data on beds for rehabilitation and long-term care. However, hospital staff numbers were reported with no information on the type of care they provided. Therefore we constructed a simple measure of utilization, dividing the number of inpatient days for acute patients by the total number of inpatient days. This ratio was multiplied by each category of workers and to provide the measure of inputs used in the estimation.

---

41 See for instance Rosko and Broyles (1988).

42 We attempted to use an index of hospital machineries as a more refined measure of capital, attaching their average costs as a weight. However the set of known costs is incomplete, and even if this missing data concerns only a couple of machines, we believe this might bias the index value. This may affect parameter estimates, as happened in most of the adopted model specifications.

Both public and private hospitals providing health care services are present in the sample. Public hospitals are financed by public funds, while both for-profit and not-for-profit hospitals rely on a mix of public and private funds. In order to compare productive units and make them as homogeneous as possible, we must control for differences in the source of funding. For this reason, in private hospitals, we consider only those services covered by public funds. In this way we are excluding those hospitals which did not sign any type of agreement with the National Health System (NHS) and are exclusively devoted to a ``pure'' for-profit activity. For the sake of consistency, for private hospitals we included in our sample only the number of beds accredited with the NHS and a proportional fraction of their personnel (i.e. the number of workers in each category multiplied by the share of beds in agreement with the NHS divided by the total number of beds).

This gives us a final sample of a weakly balanced panel of 108 hospitals of the Lazio region observed during the years 2000/2005, which sums 625 observations. In table 1 we show the distribution of the hospitals and acute patients per year, and with respect to ownership structure. From the combined reading of input and output statistics (table 2), it appears private hospitals are smaller with respect to all size-related measures: fewer beds and personnel, and a lower sum of DRG weights. Further, they are specialized and concentrated on less complex cases, as can be seen, for instance, by the low number of emergency room treatments. This appearance is confirmed by the hospital specialization index we have computed. Formally, for the *h*-th hospital, the Gini ratio is equal to

$$(13)\quad gini_h = \frac{\sum_{i=1}^{N-1}(p_i - q_i)}{\sum_{i=1}^{N-1} p_i}, \quad i = 1, 2, ..., N$$

where N is equal to 20, the total number of existing Major Diagnostic Categories (MDC), q_i is the fraction of the total discharges treated by the first *i* MDCs, while p_i is the ratio of the number of the *i*-th MDC over the total number of MDCs. The index varies between 0 and 1: it is equal to zero in case of perfect equidistribution (polispecialistic medical center), since all the differences $p_i - q_i$ are null. Meanwhile it is equal to one in case of maximum specialization, since equation (13) boils down to quantity $\sum p_i \big/ \sum p_i = 1$.
The Gini ratio may be computed by using DRGs instead of MDCs and in this case N of equation 13 would equal 489, the total number of existing DRGs. Counting the number of DRGs for this measure of specialization would be more appropriate, since it better captures the range of all hospital production. However due to the fact that hospitals tend to focus on more remunerative and less costly practices, the array of DRGs chosen by hospitals tends to be very narrow. Therefore the index of specialization is very high for all hospitals and on average it is similar for all ownership categories. In order to get a variable with a greater and more significative range of variation, we use as MDCs, where diagnoses correspond to a single organ system or etiology and are associated with a particular medical specialty[43].

43 Farley and Hogan (1990) made an in-depth analysis of hospital specialization measures. They proposed an Information Theory Index, where specialization is given by caseload deviation from that of the typical hospital. The main pitfall of this Information Theory Index is therefore its inability to distinguish between

2.4 Empirical implementation

2.4.1 The models

The functional form of $f(\cdot)$ may take different aspects. With a Cobb-Douglas specification we can interpret coefficients as output elasticities, since the covariates are all expressed in logs. In this application however we cannot estimate this log-linear function, since in a distance function context, Cobb-Douglas has the wrong curvature in the $\frac{y_i}{y_{1i}}$ and $\frac{x_i}{x_{1i}}$ spaces. Therefore we have estimated a translog function, where the presence of squared and interaction terms gives a high degree of flexibility.

For the multi-output multi-input stochastic frontier consider the two following models

(14)

$$-\ln y_{1i} = \beta_0 + \sum_{m=1}^{M} \alpha_m \ln y^*_{mi} + \sum_{k=1}^{K} \beta_k \ln x_{ki} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N} \alpha_{mn} \ln y^*_{mi} \ln x^*_{ni} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N} \beta_{kj} \ln x_{ki} \ln x_{ji} + \sum_{m=1}^{M}\sum_{n=1}^{N} \beta_k \alpha_m \ln x^*_{ki} \ln y_{mi} + \sum_{h=1}^{H} z_h + v_i - u_u$$

(15)

$$-\ln y_{1i} = \beta_0 + \sum_{m=1}^{M} \alpha_m \ln y^*_{mi} + \sum_{k=1}^{K} \beta_k \ln x_{ki} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N} \alpha_{mn} \ln y^*_{mi} \ln x^*_{ni} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N} \beta_{kj} \ln x_{ki} \ln x_{ji} + \sum_{m=1}^{M}\sum_{n=1}^{N} \beta_k \alpha_m \ln x^*_{ki} \ln y_{mi} + \sum_{h=1}^{H} z_h + v_i - u_u$$

where in both models *i* denotes hospital, *k,j* denote inputs, *m,n* denote outputs, and *h* denotes shifting factors. y_{mi} is the *m*-th output variable for hospital *i* and a symmetry constraint has been imposed on the interaction terms,
i.e. $\alpha_{mn} = a_{nm}$. Further x_{ik} is the *k*-th input variable for hospital *i* and a symmetry constraint has been imposed on the interaction terms, i.e. $\beta_{kj} = \beta_{jk}$.[44] In model (14) y_{1i} is the output used for normalization, which is given by the sum of the DRG weights of

---

hospitals that treat either a very narrow or a very broad range of cases, since both will tend to have relatively high index values. Hence, we decided to make use only of the Gini ratio.

44 Note that each input and output variable has been standardized with its median.

the discharges in complex surgery. On the other hand in model (15) the number of beds, x_{1i}, is the input used for normalization. Since we treat the z_h factors as fixed effects, this translog function is not fully flexible. In addition to input variables we included:

- Time dummies for each year of the sample, using 2000 as the base year. Using a set of time dummy variables is the same as running time fixed effects without considering panel effects.

- Two variables concerning ownership: one dummy if the hospital is private and another one if the hospitals is NFP, i.e. if it has been assimilated to a public structure. ”Fully” public hospitals represent the base category.[45]

- Geographical dummies for each ASL. For the county of Rome we distinguished between hospitals directly managed by Local Health Authorities (**Rome-asl**) and self-managed hospitals (**Rome-self**).

Before getting into the details of the model and commenting the results we would like to remark that our estimations were based exclusively on a pooling of data. That is, we treated the observations as part of a single cross-section. The reasons are explained through a simple decomposition of the total sum of squares (*SST*). For a generic variable z_{it} observed for hospital *i* at time *t*, *SST* is equal to the sum of the between hospitals sum of squares, *SSB(i)*, and the within hospitals sum of squares, *SSW(i)*, i.e.

$$SST = SSB(i) + SSW(i)$$

$$SST = \sum_i \sum_t \left(z_{it} - z_{..}\right)^2$$

$$SSB(i) = \sum_i \left(z_{i.} - z_{..}\right)^2$$

$$SSW(i) = \sum_i \sum_t \left(z_{it} - z_{i.}\right)^2$$

where $z_{..}$ is the global mean and $z_{i.}$ denotes the average of z_{it} over *t*. To standardize the results we divide *SSB(i)* and *SSW(i)* by *SST* in order to calculate the percentage of both components. Since most of the variation in the input and output variables is between rather than within the hospitals it seems there is very little panel data variation in the sample, which is similar to a cross-section.

2.4.2 Estimates' results

45 Public hospitals include hospitals directly managed by ASL and AOs. Hospitals assimilated to public structures are the following: 1) "Istituto qualificato presidio della USL"; 2) "Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS)"; 3) "Ospedali classificati o assimilati ai sensi dell'art.1 u.c. L.132/68"; 4) "Policlinici Universitari".

As a first step of our analysis, we ran a pooled OLS regression in order to provide a simple test for the presence of technical inefficiency in the data. If there were no technical inefficiency, the error term would be symmetric (being $u_i = 0$, then $\varepsilon_i = v_i$) and the data would not support a technical inefficiency story. A skewness/kurtosis test for normality (results not shown) rejected the null hypothesis of normal residuals. Hence, it seems there is evidence of the presence of inefficiency in the data.
The second stage consisted of a general-to-specific estimation and test approach. In order to estimate technology parameters and technical efficiency we added the following set of assumptions:

(16) $v_i \sim N\left(0, \delta_{vi}^2\right)$
(17) $u_i \sim N^+\left(0, \delta_{vi}^2\right)$
(18) $\mu_{it} = q_{it}\phi$
(19) $\delta_{vi}^2 = \exp(w_i \delta)$
(20) $\delta_{ui}^2 = \exp(t_i \gamma)$

Equation (16) and (17) tell us the variance of the idiosyncratic error term and the variability of the inefficiency term are not constant. These have been modelled in equations (19) and (20) respectively. Further by using the truncated normal distribution for the inefficiency term u_i we can parameterize its mean μ. We must note that seeking to address the problem of heteroscedasticity by parameterizing the variance of the inefficiency error term and parameterizing the mean of the truncated normal distribution (thus changing its shape) can be seen as another approach to study the exogenous effects on inefficiency, as has been pointed out by Kumbhakar and Lovell (2000). Degree of competitive pressures, managerial characteristics or even ownership form may influence the structure of the technology by which conventional inputs are converted to outputs or may influence the efficiency with which inputs are converted to outputs. Moreover this particular combination of distributional assumptions allows us to accomodate non-monotonic efficiency effects. This implies that variables having such relationship can be positively related in part of the parameter space while negatively related in the rest (see Wang 2002 for more details). Last but not least, Wang and Schmidt (2002) have demonstrated that this model specification allows one-step estimation of the parameters, avoiding two-step procedures which give biased results if the model estimated at the first step is misspecified. Further they note that the vector of variables affecting the frontier may overlap the vector of variables affecting technical efficiency.
In this paper we are intended to study the impact of specialization and capitalization on hospital efficiency.
Therefore we have included for both models in equation (18) the Gini ratio (**Gini**) as proxy for specialization and the nurses per bed ratio as proxy for capitalization. This latter variable has been already used as determinant of cost efficiency in previous studies. Farsi and Filippini (2006) showed that a higher nurses per bed ratio decreases efficiency, indicating that quality of care is costly. For both models, we have subsequently added another common inefficiency determinant representing hospital dimensions: the logarithm of beds (**Beds**) in the output-oriented model and the logarithm of the sum of

weights for acute patients (**Weighted Patients**) in the input-oriented model. Further, in the input distance function model, we achieved a better fit by adding the mean age of the patients (**Age**) to explain mean inefficiency. We assumed the sources of idiosyncratic noise vary in both models with the hospital size, with **Beds** and **Weighted Patients** again serving as a proxy for this variable. Finally, in the output model, we assumed hospital dimensions affected the variability of the inefficiency term. We implemented some general likelihood ratio (LR) tests, while posing restrictions on the unrestricted translog model, in order to get to a "preferred" model. The LR test is given by $\lambda = 2(\ln L_1 - \ln L_0)$, where $\ln L_0$ and $\ln L_1$ denote the maximum log-likelihood value under the null hypothesis H_0 and the alternative H_1, respectively. The LR tests conducted are presented in Table 3. For both models, the first two null hypothesis assessed the appropriateness of using the half-normal distribution $\mu_i = 0$, with and without modeling heteroscedasticity. We then tested the importance of modelling the variance of the error terms $(w_i = 0, t_i = 0)$. Finally we tested whether exogenous inefficiency variables, as a group, have a significant impact on technical inefficiency. All these hypothesis were strongly rejected. More generally we can say that we cannot do without unmodeled heteroscedasticity in the error components and nor without modelling the mean of the inefficiency error term.

In table 4 we show scale and input/output elasticities, which are more meaningful than the simple technology parameters in a translog function context. Elasticities have been estimated at the means of the variables in the data. Standard errors were computed by applying the delta method to linearize the elasticity functions around the estimated parameter values and then using standard formulas for the variances and covariances of linear functions of random variables. The results given in table 4 suggest significant increasing returns to scale for both models, since scale elasticities $\varepsilon_{Y,X} > 1$ and $\varepsilon_{X,Y} < 1$ and the null hypothesis of constant returns to scale is rejected. From an administrative point of view, this is the typical situation where there is an incentive to centralize operations.

The individual input and output elasticities underlying the scale elasticities are also provided. In the output distance model, elasticity for nurses and beds is high while the marginal product of other staff labor is slightly negative, even if not statistically significant. Therefore it seems there is some excess in the size of the group of workers with administrative and technical duties. Moreover from the input-oriented model we can see that all marginal costs, except from ER treatments, are positive and that not surprisingly general medicine (*Y*5) and general surgery (*Y4*) are the outputs primarily contributing to input use.

Focusing on existing differences among ownership structures, which we do not report here, scale elasticity varies a great deal between hospitals, and is greater for private hospitals and lower for NFP structures in both models. However this is not surprising, and may simply reflect the existence of ceilings on fee-for-service financing, which have been introduced in Lazio region. The main difference between public and NFP hospitals and private structures is that the latter cannot surpass fixed volume limits, while the former are at least reimbursed at a reduced rate once having reached the ceiling, or may even be fully reimbursed because of the political necessity of avoiding hospital failures.

Therefore, a possible effect of this different treatment is input-minimizing behavior in private structures, which are forced to work at a reduced scale. Further, private units are slightly over-capitalized with respect to nursing staff, while public and more heavily NFP hospitals have a slight excess of administrative and technical staff, suggesting the opportunity for a re-allocation of resources within these structures.

In table 5 we show maximum likelihood estimates of the remaining parameters of the hybrid translog production functions, other than technology parameters. We also show the equations of the two heteroscedasticity terms and the mean of the inefficiency term. With respect to ownership, the coefficient for NFP hospitals is always significant (at least at a 10% level) with a positive sign.[46] The role played by private hospitals is, in contrast, less straightforward.

They positively and significantly contribute to an upward shift of the frontier in the input-oriented model but the sign is reversed in the output model, even though it is not statistically significant. Time dummies show a positive trend, with ever-increasing coefficient values, and are not significant only for the first two years of the output distance model. It might be interpreted that we are capturing a linear time trend or disembodied technical change. As far as geographical dummies are concerned, the base category is represented by the hospitals in the county of Rome, but not within this municipality. It appears that hospitals in all the other counties, except those self-managed hospitals located in Rome, contribute to an upward shift of the frontier.

Regarding the determinants of the mean of the inefficiency error term, for the input model, mean age is positively correlated with inefficiency, while for both models size measures, capitalization and specialization are instead negatively correlated.

Particularly, nurses per bed and gini are 1% significant in the input model, but only at 5% and 10% in the output distance model respectively. The result obtained for specialization is strengthened if we look at the sign of cross-output terms (see table 6, that allow us to evaluate input and output complementarities. From the input specification we observe that only three cross-output terms over ten are negative, two of them being significant.[47] The absence of output jointness is not a proof of pro-efficiency specialization, but a fact that is consistent with this hypothesis.

Before commenting on efficiency results we would like to remark that in our models we have attempted to control for quality. This can be broadly divided into outcome and structural indicators. With respect to the outcome indicators, we have computed readmission and mortality rates as quality adjustments or scaling factors for output measures. As noted by Milne and Clarke (1990), both indicators have big drawbacks. Further we did not feel confident with using readmission rates, since we could not distinguish between patients with planned vs. unplanned readmission.

Hence the computed rate depends subjectively only on the number of days used as threshold. Mortality rates, in contrast, were easily measured, but their impact on the

---

46 Remember that for a correct interpretation, the sign in the distance functions must be reversed. This implies that in the output distance model, the upward shift of the frontier is given by a negative sign. While for an input distance function by a positive sign.

47 We remind that signs and magnitudes of the cross-effects represent input/output jointness. In the input specification negative cross-output terms represent output jointness, while in the output specification positive cross-input effects imply input complementarities.

estimation is negligible, likely due to the fact that they are very rare across inpatient specialties.
For structural measures of quality, we tried including the teaching status of the hospitals in the frontier of both models, using the number of teaching staff as a proxy, and an attraction index, proxied by the rate of discharges of people coming from a different ASLH. Signs were not significant, and efficiency estimates with and without these variables were highly correlated.
Hence, we decided not to include them in the models. It is therefore possible that without being able to properly account for quality our estimates may suffer from omitted variable bias.
However as Rosko and Mutter (2008) pointed out, these omissions may not be as serious as commonly thought.

2.4.3 Technical efficiency: trends and transitions

In figure 1 we show estimated sample mean efficiencies for each hospital ownership category. In the output distance model, public and NFP hospitals are on average more efficient than the accredited private ones. We observe a decline in technical efficiency over time for NFP hospitals, and a smaller reduction for private ones, while public hospitals seem to remain at a roughly constant efficiency level throughout the period. In order to explain such differences, as we mentioned in the discussion of scale elasticities, private hospitals have a legal limit to the maximum output they can produce for the NHS.
This fact, coupled with the different and penalizing reimbursement rate among hospital types, not only induced distortions in the production scale, but also in the efficiency level.
Looking again at figure 1, we may note that, in the input distance model, the three ownership categories are much closer to each other and all show an increase in efficiency patterns over time. Public hospitals at the end of the period show a slight reduction, converging to NFP averages, while private structures are now at the same level of the other categories. These results do not contradict the output distance model, since we are estimating two different frontiers under two different sets of theoretical assumptions. With the input distance function we are assuming a cost-minimizing behavior of the hospital, which chooses an input vector, evaluated at exogenously determined input prices, that minimizes the cost of producing a given output vector. This assumption clearly is valid if the hospital acquires inputs in competitive markets. The behavioral assumption we are making through the output distance function is that the firm chooses an output vector that maximizes revenue at given output prices for a given input vector. The plausibility of this assumption is linked to the sale of outputs in competitive markets.
In this study, neither assumption is strictly respected. In fact, none of these models is perfectly congruent with the present incentive framework within the NHS. With a simplifying assumption, we may believe private and NFP hospitals try to minimize costs and hence aim at reducing the use of inputs. In such a context, where we have to take into account mechanisms of control over output volumes, an input-oriented model is preferable. Contrastingly, public hospitals seem to have more discretion with respect to

their level of production. Therefore an output-oriented model cannot be discarded a priori, since it shifts the focus from costs to revenues.
In figure 2 the horizontal and the vertical axis represent efficiency levels for all the hospitals respectively in 2000 and 2005. For the output distance model there is more mobility in technical efficiency over time. This does not happen in the input distance case, but this may be due to greater stability in input endowments. In each of the two models, at the maximum values of the distribution, distances from the 45º line are not very large, implying stability over time for the estimated efficiencies. In the input distance case, apart from a few observations, this happens for the full range of the efficiency distribution. Furthermore, if we look carefully at both graphs, we can see that the majority of public hospitals passes the 45º line, indicating an increase in technical efficiency. This confirms what emerged from the graphs of the efficiency trends.
In tables 7a and 7b we show transition matrices from 2000 and 2005 for our model specifications. These tables confirm that a significant percentage of the hospitals belonging to the extreme quintiles of the technical efficiency distribution in 2000 still belongs to the same quintile in 2005. The diagonal elements of the transition matrix show more mobility. As expected, those diagonal elements are ``fatter'' when considering shifts year-by-year (data not shown). The fact that individual hospital units tend to remain in the same or close to the same quintile of the efficiency distribution is not a surprise. In fact, we do not expect hospitals to change dramatically their operational behaviour in a so short a time period, or year by year.

2.4.4 Capital-labor distribution

In our estimates, we have found a wide distribution of technical efficiencies. We attempted to ascertain whether productive structures are related to efficiency values, in order to make our analysis more precise in distinguishing among heterogeneous units.
In fact, since our dataset contains the entire population of Lazio's hospitals, we have very different units in terms of size and kind of activity. Although we have partially purged heterogeneity by using control variables for ownership type, time and location, using only observations related to acute patients, in a flexible model context that allows also for heteroscedasticity, these measures seem not to have eliminated all the possible sources of diversity from the dataset. Our idea was to draw a curve corresponding to the efficiency frontier. We made a scatter-plot of dots corresponding to the position of each hospital. In particular, the estimated efficiency determined the distance between the unit and the frontier in the graph. The radial-position is fixed depending on the capital-labor ratio value. The polar coordinates are given by the couple:

$$\left[\cos\left(\arctan\left(\frac{K_i}{L_i}\right)\right)\times TE_i, \sin\left(\arctan\left(\frac{K_i}{L_i}\right)\right)\times TE_i\right]$$

where the technical efficiency level of the *i*-th hospital $TE_i \leq 1$ if it lies below or on the frontier curve. This kind of figure has the advantage of pointing out how estimated efficiencies vary among different capital-labor ratios and gives the possibility to distinguish particular typologies of units.

Specifically we differentiated hospitals among ownership structures for both models (figure 3). At first glance, we may notice that private and public structures seem to sort into two different “cones” of capital-labor ratio, while NFP hospitals appear to be more evenly spread. Cones exclude outliers for both categories, i.e. they do not include hospitals below the tenth percentile or above the ninetieth percentile of the capital-labor ratio. Further there appears to be an inverse relationship between the capital-labor ratio and technical efficiency. This evidence has been already captured in the frontier estimation, where the nurse-per-bed ratio is negatively correlated with inefficiency. The more capitalized the hospital, as is the case for private hospitals, the less efficient it appears to be.

It seems reasonable that units operating with different productivity structures cannot be directly compared. The capital-labor ratio is indeed quite variable among the different structures. We tried to define the labor factor in different ways, in order to verify whether the definition made a difference. In particular, we used three typologies of labor aggregation: the first one was the simple sum of all workers operating in the structure; the second was the simple sum of physicians and nurses operating in the hospital; and the last corresponded to a sum of five categories of workers, weighted by their estimated wage. For estimating wages we gleaned data from AOs' balance sheets, taking the total amount of salaries divided by the number of workers for each of three categories (physicians, nurses and other staff). The only labor definition which widens the spread of hospitals in the graph is the sum of physicians and nurses. However, in each of the three cases, the relative proportions are nearly the same.

Another question arises: what are the causes of relatively lower labor usage in private hospitals and higher usage in public ones? We believe that public hospitals operate in a more rigid institutional framework, inclined to keep a high number of workers, while private ones do not. Given that the latter face greater economic risk, it is reasonable to assume that they decide to turn this risk toward workers, instead of toward capital (meaning medical equipment with related costs). Another explanation from the public finance literature is that one role of public institutions is to provide employment. This phenomenon causes a high value for the capital-labor ratio in private structures, but according to our evidence, seems to penalize technical efficiency.

2.5 Conclusions

In this paper, we studied the productive process of the hospitals of the Lazio region by means of an economic analysis. We used data from the hospital discharge records provided by the Public Health Agency of Lazio in order to derive measures of output, such as the number of discharged patients. Further, from the Ministry of Health we obtained data concerning labor and capital used as inputs. Our study has been limited to the pooled cross-section case, since a simple sum of squares analysis showed very little panel variation in the sample.

OLS residuals analysis showed the presence of inefficiency in the data. Subsequently, we implemented a stochastic frontier analysis to assess the level of technical efficiency achieved by the hospitals. When deriving technology parameters, we took into account case-mix complexity. With this correction of the left hand-side of the frontier equation, we were able to address one of the main points of Newhouse's critique, i.e. the impossibility of assessing hospital production by means of efficiency analysis tools, because of excessive simplification of the productive process, especially with respect to the quality of services.

Inefficiency is negatively associated with specialization and positively with capitalization. Capitalization is typical of private hospital structures which, on average, make a less efficient use of resources when compared to public and not-for-profit hospitals. As far as the productive structure of the hospital is concerned, there seem to be increasing returns to scale, suggesting centralization of operations. Private units work in slightly over-staffed conditions for medical staff, while public and more heavily NFP hospitals do the same for technical and administrative staff, suggesting the opportunity for a re-allocation of resources within these structures.

Our efficiency estimates are strictly related to the choice of a specific model, which depends on the theoretical and empirical assumptions one is willing to make. The input distance function model is based on a cost-minimization hypothesis, and empirically is more appropriate in studying private and NFP hospital behavior.

Meanwhile, the output distance function is based on a revenue-maximization hypothesis and is more appropriate to the study of public hospital patterns. Being aware of the limitations of the model, we tried to paint a picture from the combined reading of both.

The results suggest that public and NFP hospitals make a more efficient use of resources. In fact the level of the estimated mean technical efficiency appears to be significantly higher than that of private structures in the output-oriented model, and slightly higher in the input-oriented model. This large gap in the former model can be attributed to the different reimbursement rates among ownership categories.

This turned out to be the cause of under-sizing even in private structures. The minor differences in efficiency levels reported in the input-oriented model, on the other hand, are more easily explainable as a result of the cost-minimizing assumption upon which the model is based.

Acknowledgements: We would like to thank: Vincenzo Atella and Federico Belotti for continuous and stimulating discussion; two anonymous referees; Giorgia Marini; the participants of the Sixteenth European Workshop on Econometrics and Health Economics and the Masterclass in applied Health Economics held in Bergen; the participants of the Econometrics PhD Students' seminars, held at University of Rome Tor Vergata; the participants of the Technical Efficiency Parallel Session at the Third Italian Congress of Econometrics and Empirical Economics, held at University of Ancona; and

Hung-Jen Wang for the assistance with his new STATA package and for his suggestions concerning convergence of the models. We also thank Cristina Tamburini from the Ministry of Health for the data used throughout the paper.

Tables and figures

Table 1
Number of hospitals and acute patients by year and ownership type

Year	Public	NFP	Private	Total
(a) Hospitals				
2000	49	16	40	105
2001	49	16	38	103
2002	50	17	38	105
2003	52	17	39	108
2004	49	17	39	105
2005	45	18	36	99
Total	294	101	230	625
(b) Acute patients				
2000	502'960	356'543	134'291	993'794
2001	507'771	378'632	138'035	1'024'438
2002	524'445	411'879	140'125	1'076'449
2003	540'208	427'752	152'219	1'120'179
2004	536'407	452'790	163'419	1'152'616
2005	520'596	487'316	175'097	1'183'009
Total	3'132'387	2'514'912	903'186	6'550'485

Table 2
Average input and output levels by year and ownership type (Standard deviation in parentheses)

Year	X1		X2		X3		X4		Y1		Y2		Y3		Y4		Y5	
(a) Public																		
2000	10.18	(363.51)	260.47	(525.64)	283.96	(790.62)	251.18	(723.68)	831.99	(4,619.60)	320.34	(307.44)	859.57	(6,066.48)	3,061.86	(6,894.97)	5,705.47	(12,730.93)
2001	123.94	(364.49)	292.41	(550.88)	215.90	(812.05)	248.82	(589.02)	931.45	(4,679.89)	301.22	(284.89)	882.19	(6,297.75)	3,277.69	(7,378.87)	5,658.58	(13,706.07)
2002	129.38	(345.76)	290.06	(609.48)	209.38	(791.27)	234.92	(559.86)	1,044.07	(4,795.72)	276.05	(296.18)	910.75	(5,368.57)	3,416.62	(7,154.56)	5,696.95	(14,199.20)
2003	134.48	(317.14)	293.48	(556.32)	248.42	(701.58)	215.75	(502.01)	1,045.99	(4,679.07)	249.30	(275.42)	946.01	(5,142.27)	3,383.72	(6,964.78)	5,706.03	(14,781.81)
2004	128.53	(309.94)	281.65	(593.69)	228.47	(745.73)	219.86	(509.28)	1,179.58	(5,001.92)	235.20	(280.82)	960.47	(5,022.41)	3,607.16	(7,205.18)	5,882.19	(15,466.18)
2005	141.76	(306.74)	316.93	(465.72)	305.60	(976.94)	233.38	(490.17)	1,377.72	(4,759.24)	261.45	(309.55)	1,076.66	(4,960.56)	3,726.86	(7,552.57)	6,294.96	(16,476.76)
(b) NFP																		
2000	239.56	(363.51)	446.00	(525.64)	528.19	(790.62)	524.31	(723.68)	2,309.48	(4,619.60)	251.37	(307.44)	3,835.24	(6,066.48)	6,199.18	(6,894.97)	11,237.66	(12,730.93)
2001	246.06	(364.49)	465.56	(550.88)	543.06	(812.05)	478.06	(589.02)	2,435.05	(4,679.89)	238.47	(284.89)	4,133.73	(6,297.75)	6,601.94	(7,378.87)	11,917.19	(13,706.07)
2002	244.76	(345.76)	477.71	(609.48)	525.65	(791.27)	452.53	(559.86)	2,560.05	(4,795.72)	264.87	(296.18)	4,032.31	(5,368.57)	6,970.77	(7,154.56)	12,161.03	(14,199.20)
2003	238.41	(317.14)	462.18	(556.32)	493.18	(701.58)	432.41	(502.01)	2,693.21	(4,679.07)	255.55	(275.42)	4,108.43	(5,142.27)	7,222.96	(6,964.78)	12,483.86	(14,781.81)
2004	247.71	(309.94)	473.35	(593.69)	512.12	(745.73)	430.94	(509.28)	3,077.25	(5,001.92)	262.53	(280.82)	4,177.61	(5,022.41)	7,645.00	(7,205.18)	12,941.68	(15,466.18)
2005	231.94	(306.74)	359.44	(465.72)	562.28	(976.94)	408.33	(490.17)	2,917.56	(4,759.24)	256.40	(309.55)	4,014.03	(4,960.56)	7,895.34	(7,552.57)	13,130.73	(16,476.76)
(c) Private																		
2000	28.98	(24.66)	35.97	(35.72)	47.63	(39.78)	90.32	(57.03)	71.18	(146.70)	47.75	(111.11)	146.94	(161.95)	1,608.93	(2,322.44)	1,449.47	(1,167.74)
2001	31.37	(28.26)	37.87	(37.74)	48.87	(41.13)	88.82	(60.32)	84.81	(190.20)	42.57	(94.26)	164.57	(178.16)	1,818.65	(2,565.21)	1,532.32	(1,162.38)
2002	30.50	(25.66)	38.00	(37.23)	52.82	(41.78)	89.39	(58.75)	98.44	(191.00)	44.00	(94.20)	177.96	(177.53)	1,759.58	(1,694.00)	1,615.98	(1,226.36)
2003	28.13	(25.42)	34.00	(34.85)	50.38	(40.69)	86.08	(59.26)	104.45	(220.28)	41.06	(84.06)	190.00	(188.48)	1,949.14	(1,873.98)	1,651.48	(1,170.07)
2004	31.10	(29.71)	37.49	(37.53)	49.92	(42.69)	86.36	(61.70)	141.22	(317.65)	51.12	(85.64)	205.32	(226.94)	2,200.36	(2,122.48)	1,650.19	(1,191.93)
2005	40.22	(30.37)	47.81	(54.06)	57.33	(43.48)	84.28	(60.94)	312.51	(915.72)	82.91	(134.86)	220.61	(240.61)	2,682.13	(2,959.23)	1,856.20	(1,185.41)

Notes: X1: Physicians, X2: Nurses, X3: Other staff, X4: Beds, Y1: Complex Surgery, Y2: ER treatments, Y3: HIV and tumors, Y4: General surgery, Y5: General medicine

Table 3
LR tests on functional form restrictions

Null hypothesis	Log-likelihood	λ	Decision
(a) Output distance function: Log-likelihood= -119.6972			
$H_0 : \mu_i = 0$	**-197.367**	**155.34**	**Reject**
$H_0 : \mu_i = w_i = t_i = 0$	**-267.731**	**296.07**	**Reject**
$H_0 : w_i = t_i = 0$	**-153.164**	**66.93**	**Reject**
$H_0 : q_i = 0$	**-195.325**	**151.26**	**Reject**
(b) Input distance function: Log-likelihood= -133.1165			
$H_0 : \mu_i = 0$	**80.0164**	**106.2**	**Reject**
$H_0 : \mu_i = w_i = t_i = 0$	**-85.0366**	**436.61**	**Reject**
$H_0 : w_i = t_i = 0$	**78.6063**	**109.02**	**Reject**
$H_0 : q_i = 0$	**84.9903**	**96.25**	**Reject**

Note: H1: translog model with ‘‘full’’ heteroscedasticity and mean inefficiency

Table 4
Scale and output/input elasticities evaluated at the average values for the entire sample

Output Distance			Input Distance		
Parameter	**Est.**	**Std. Err.**	**Parameter**	**Est.**	**Std. Err.**
$\varepsilon_{Y,X}$	**1.257****	**0.257**	$\varepsilon_{X,Y}$	**0.703****	**0.029**
□ $\varepsilon_{Y,X1}$	**0.173***	**0.101**	□ ε_{X,X_1^*}	**0.071**	**0.047**
□ $\varepsilon_{Y,X2}$	**0.646*****	**0.181**	□ ε_{X,X_2^*}	**-0.157*****	**0.062**
□ $\varepsilon_{Y,X3}$	**-0.074**	**0.089**	□ ε_{X,X_3^*}	**0.112*****	**0.035**
□ $\varepsilon_{Y,X4}$	**0.512*****	**0.153**	□ ε_{X,X_4^*}	**0.974*****	**0.057**
□ ε_{Y,Y_1^*}	**0.007**	**0.027**	□ ε_{X,Y_1}	**0.013**	**0.012**
□ ε_{Y,Y_2^*}	**0.089*****	**0.034**	□ ε_{X,Y_2}	**-0.007**	**0.02**
□ ε_{Y,Y_3^*}	**0.177*****	**0.055**	□ ε_{X,Y_3}	**0.017**	**0.017**
□ ε_{Y,Y_4^*}	**0.190*****	**0.056**	□ ε_{X,Y_4}	**0.248*****	**0.029**
□ ε_{Y,Y_5^*}	**0.537*****	**0.059**	□ ε_{X,Y_5}	**0.432*****	**0.028**

Table 5 Shifting factors parameters of the stochastic frontier models and ancillary equations.		
Prod. frontier	**Output distance**	**Input distance**
Private	**0.008**	**0.178****
Not-for-profit	**-0.204*****	**0.108***
Year 2001	**-0.018**	**0.056*****
Year 2002	**-0.061**	**0.101*****
Year 2003	**-0.101****	**0.162*****
Year 2004	**-0.117****	**0.167*****
Year 2005	**-0.144*****	**0.206*****
Viterbo	**-0.223****	**0.196*****
Latina	**-0.082*****	**0.02**
Rieti	**-0.211**	**0.188*****
Frosinone	**-0.158****	**0.135****
Rome-asl	**-0.057**	**0.011**
Rome-self	**0.021**	**-0.082**
Intercept	**-0.154***	**0.407*****
Mean of u		
Weighted patients	**–**	**-0.032*****
Beds	**-2.356*****	**–**
Gini	**-9.022***	**-2.229**
Nurses/Bed	**-12.923****	**-0.404*****
Age	**–**	**0.006*****
Intercept	**9.442****	**1.704*****
Variance of u		
Beds	**-0.069**	**–**
Intercept	**0.386**	**-5.321*****
Variance of v		
Weighted patients	**–**	**-0.871*****
Beds	**-1.087*****	**–**
Intercept	**-3.491*****	**-3.135*****
N. obs.	**625**	**625**
Likelihood	**119.7**	**133.12**

Note: The dependent variable in the Output Distance model is the inverse of complex surgery weights. In the Input Distance model the dependent is the inverse of the number of beds

Table 6
Second-order effects of the input distance model

	Complex surgery	ER treatments	HIV and tumors	General surgery	General medecine
Complex surgery	**-0.012**				
ER treatments	**0.010**	**-0.001**			
HIV and tumors	**-0.006**	**0.010**	**-0.082*****		
General surgery	**-0.018****	**-0.032*****	**0.034*****	**-0.032****	
General medecine	**0.010**	**0.025**	**0.061*****	**0.010**	**-0.123*****

Table 7
Transition matrix: quintile of technical efficiency (TE) from 2000 to 2005

	Quintile of TE at 2005					
Quintile of TE 2000	***(a) Output distance model***					
	1	**2**	**3**	**4**	**5**	**.**
1	**11**	**7**	**0**	**0**	**0**	**3**
	52.38	**33.33**	**0**	**0**	**0**	**14.29**
2	**6**	**6**	**3**	**4**	**1**	**1**
	28.57	**28.57**	**14.29**	**19.05**	**4.76**	**4.76**
3	**1**	**1**	**8**	**4**	**2**	**5**
	4.76	**4.76**	**38.1**	**19.05**	**9.52**	**23.81**
4	**0**	**1**	**4**	**8**	**7**	**1**
	0	**4.76**	**19.05**	**38.1**	**33.33**	**4.76**
5	**0**	**3**	**4**	**4**	**8**	**2**
	0	**14.29**	**19.05**	**19.05**	**38.1**	**9.52**
	(a) Input distance model					
1	**9**	**9**	**1**	**1**	**0**	**1**
	42.86	**42.86**	**4.76**	**4.76**	**0**	**4.76**
2	**10**	**3**	**3**	**3**	**1**	**1**
	47.62	**14.29**	**14.29**	**14.29**	**4.76**	**4.76**
3	**1**	**5**	**7**	**5**	**0**	**3**
	4.76	**23.81**	**33.33**	**23.81**	**0**	**14.29**
4	**0**	**2**	**6**	**6**	**3**	**4**
	0	**9.52**	**28.57**	**28.57**	**14.29**	**19.05**
5	**0**	**0**	**2**	**4**	**12**	**3**
	0	**0**	**9.52**	**19.05**	**57.14**	**14.29**

Note: The last column represents those observations that were part of the sample in 2000 but missing in 2005

Figure 1
Trends in technical Input Distance efficiency (2000–2005)

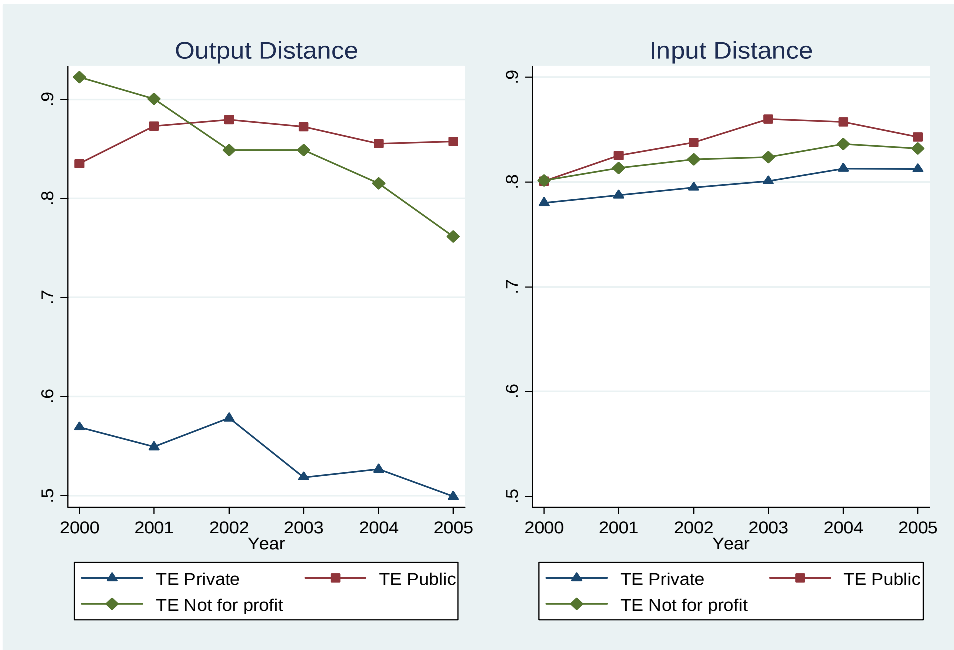


Figure 2
Transitions in technical Input distance efficiency (2000–2005)

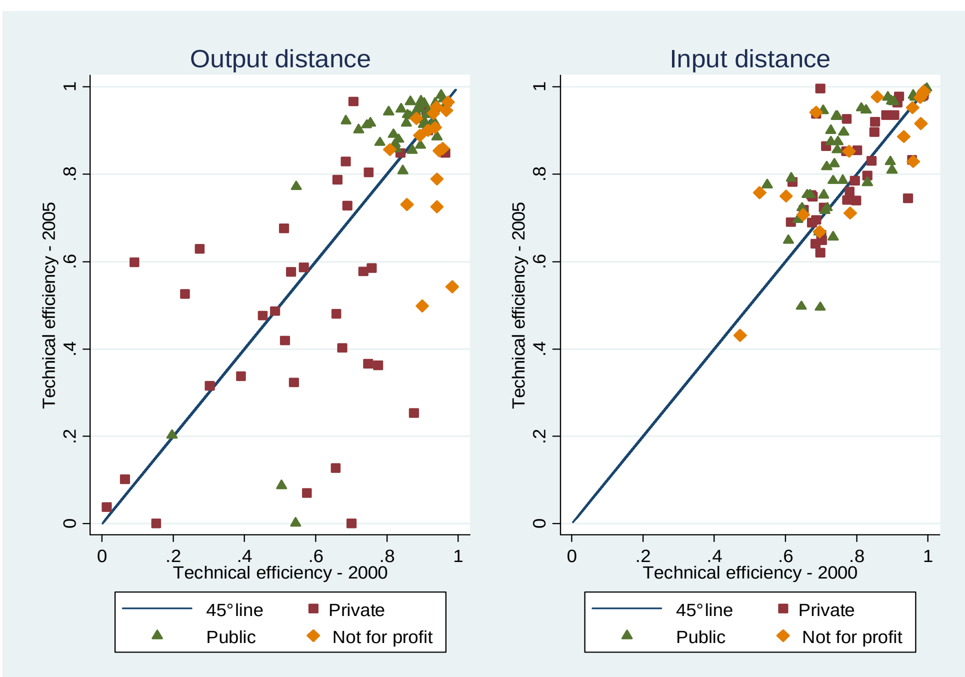
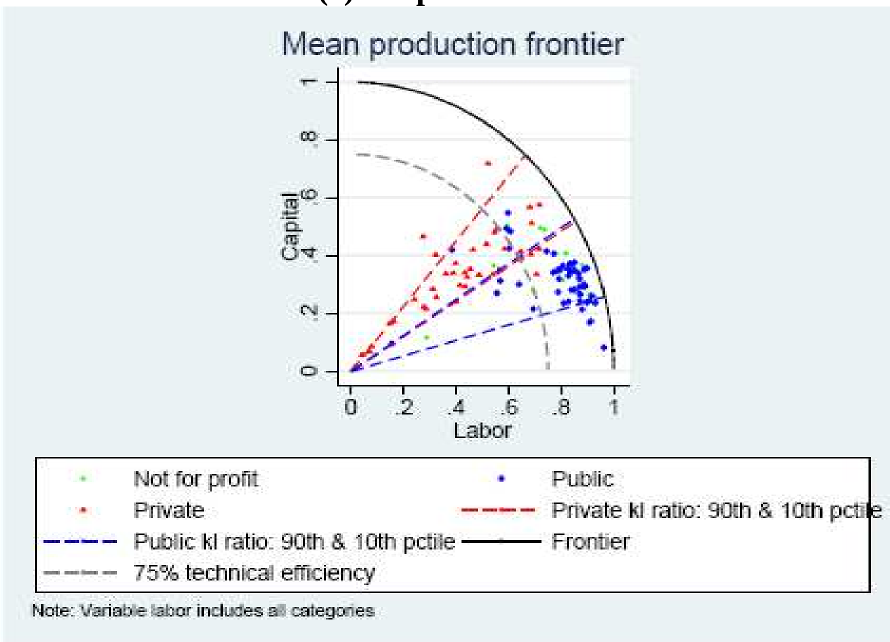
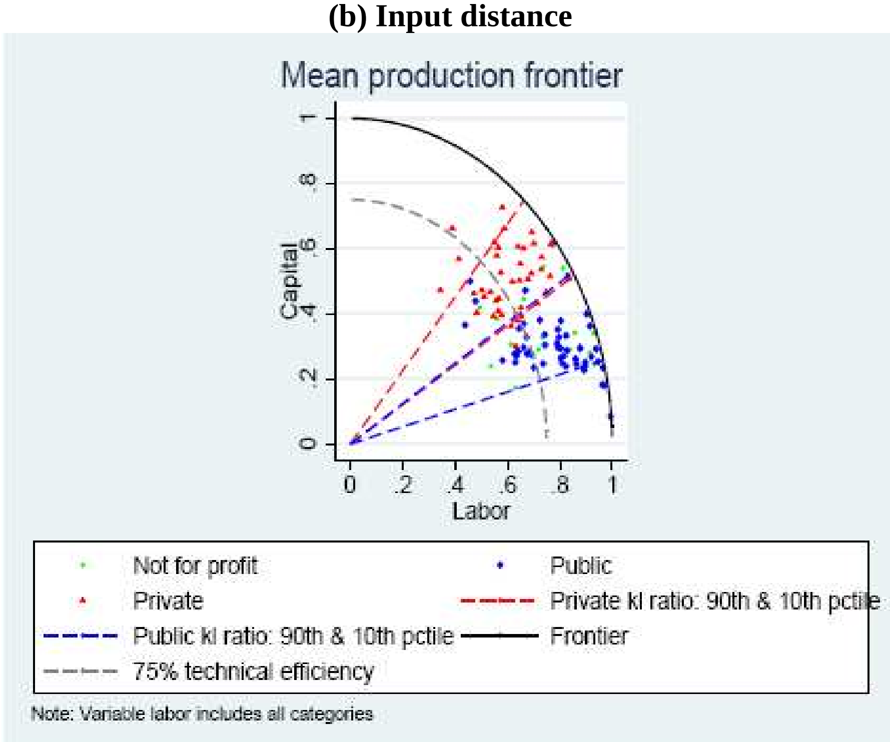


Figure 3
Capital–labor ratio, production frontier, and technical efficiency

(a) Output distance



(b) Input distance



References

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

European Observatory of Health Care Systems. 2001. Health Care Systems in Transition - Italy, http://www.euro.who.int/document/e73096.pdf.

Fabbri D. 2003. "L'efficienza tecnica e di scala degli ospedali pubblici in Italia", in *L'efficienza dei servizi pubblici*, Banca d'Italia.

Farley, D.E. and C. Hogan 1990. "Case-mix Specialization in the Market for Hospital Services" in *Health Services Research*, 25(5): 757-783.

Farsi, M. and M. Filippini 2006. "An analysis of efficiency and productivity in Swiss hospitals", *Swiss Journal of Economics and Statistics*, 142(1): 1-37.

Grassetti, L., E. Gori and S. Minotti. 2005. "Multilevel flexible specification of the production function in health economics", *IMA Journal of Management Mathematics*, 16(4): 383-398.

Kumbhakar S.C. and C.A. Lovell. 2000. *Stochastic Frontier Analysis*, Cambridge University Press.

Lovell, C.A.K., S. Richardson, P. Travers and L.L.Wood 1994. "Resources and functioning: a new view of inequality in Australia" in *Models and measurement of welfare and inequality*, eds. W. Eichhorn, Springer, Berlin.

Milne, R. and A. Clarke. 1990. "Can readmission rates be used as an outcome indicator?", *British Medical Journal*, 301: 1139-1140.

Newhouse, J.P. 1994. "Frontier estimation: How useful a tool for health economics?", J. Health Econ., 13: 317-322.

Rosko, M. and R. Broyles. 1988. *The economics of healthcare: A reference handbook*, Westport, CT: Greenwood Press, Inc.

Rosko, M.D. and R.L. Mutter. 2008. "Stochastic frontier analysis of hospital inefficiency. A review of empirical issues and an assessment of robustness", *Medical Care Research and Review*, 165:2: 131-166.

Shephard, R.W. 1970. *The theory of cost and production functions*, Princeton University Press.

Wang, H. 2002. "Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model", *Journal of Productivity Analysis*, 18: 241-253.

Wang, H. and P. Schmidt 2002. "One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels", *Journal of Productivity Analysis*, 18: 129-144.

Zweifel, P. and F. Breyer. 1997. *Health Economics*, Oxford University Press.

Chapter 3

A stochastic frontier approach to assess the efficiency of English councils with social services responsibility[48]

3.1 Introduction

Recent decades have been characterized by a substantial increase in average life-expectancy in OECD countries, due to improved living standards and to a more generalized provision of health care. Whilst enjoying the benefit of a longer average lifespan, most Western countries are facing an increase in the health and social care demand which threatens their financial sustainability, just as the so called 'baby-boomers' enter old age.

According to Gurria[49] (2009), while Long Term Care (LTC) expenditure accounted for just 1% of the GDP in OCED countries in 2005, in 2050 such spending will reach between 2% and 4%. EUROSTAT official figures predict an increase in LTC expenditure as a proportion of GDP of between 106% in the United Kingdom and 173% in Italy by 2050 (EPC, 2001). Similarly, Costa-i-Font et al. (2008) forecast a more reduced but still very significant expenditure growth, with a predicted increase ranging between 95% and 120% in the base scenario. According to population projections for the United States, by the end of 2050, long-term care expenditure by the Medicaid programme will nearly quadruple from the levels of 2000, to $379 billion dollars (United States General Accounting Office, 2002).

These predicted increases in demand for LTC services have spurred widespread reforms in the provision of social care. In Europe for example, countries such as Sweden, the Netherlands, the United Kingdom, France and Germany introduced policy changes in their care provision systems in the 90s and 2000s. The four main types of reforms were the introduction of private-market incentives, the promotion of cash-payments in substitution of the direct provision of the services, greater funding for informal carers and the personalization of the assistance (Pavolini and Ranci, 2008)..

---

48 This chapter is based on “Measuring inefficiency in social services commissioning: evidences from English local authorities’ activity” by F. D’Amico and J.L. Fernandez.

49 The exact statement from the OECD Secretary-General is “Population ageing will increase spending further, mainly through long-term care needs of the frail elderly. In 2005, long-term care expenditure accounted for just over 1% of GDP across OECD countries. We project that such spending will reach between 2% and 4% of GDP by 2050. Improving the efficiency of health care systems is thus imperative to accommodate future pressures” and has been pronounced during the OECD Industry Partners Session “Strategic options to finance pensions and healthcare in a rapidly ageing world” meeting in Davos, 30 January 2009.

In England, the first attempts to address structural changes in long-term care provision go back to 1997, when the Department of Health (DoH, 1997; DoH, 1998) defined a new vision for social care, giving a primary role to the promotion of independence of the services' clients through an extended use of direct payments and raising quality standards, through the use of specific incentive programmes. This policy drive was reaffirmed in the Green Paper "Independence, well-being and choice: our vision for the future of social care for adults in England" (DoH, 2005) and White Paper "Our health, our care, our say: a new direction for community services" (DoH, 2006). The role of non-residential care (i.e. community care) was further promoted, as a means to increase users' independence and reductions in expenditure. Community care includes long term services delivered either directly in the person's home or in the community, such as day-care centres. Naturally, community care is not always suitable, particularly for some individuals with the highest needs, as the cost of the home assistance would be prohibitive. In such cases, residential assistance is still the preferred option.

In the year 2000 Direct Payments were extended to older people in England. These cash payments can be used by people in need to buy social services in the private market or to buy respite care for informal carers allowing them to have short breaks from their caring duties. From a strategic point of view, the use of cash-refunds also represents an exit of the public social care system from the direct provision of the services. Further, as cash-refunds give to the individuals the freedom to choose their preferred provider, they also introduce private market style incentives.[50]

Social services in England are commissioned by local authorities[51], which have legal duties to arrange their provision. Commissioning is a form of organization which implies the possibility for the authorities to provide directly, *in-house*, or to buy from external providers (private or from the third sector), *contracting-out* the social services which are required. Contracting-out is an option that opens the social care market to competition and is one that English councils tend largely to use.

Using a six-year panel (2002-2007) containing council level aggregate data, we analyze the performances of the English local authorities in the provision of social care services. In particular, through the use of a stochastic frontier cost model, we estimate inefficiency scores for 148 English councils, controlling for the output produced and for local factor prices. The study of the expenditure inefficiency patterns can help the policy-maker better understand which measure a further optimization of the resources by local authorities can help reduce the financial pressure over the public balance.

In the model specification we adjust for local competition effects which are expected to increase with a wider presence of external providers. In particular, we control for the proportions of private and voluntary suppliers when estimating inefficiency scores. In order to gauge the different effects of competition in residential and community care, we compute separate rates for the different types of social care providers. The estimation of efficiency trends, cost-output elasticities and scale economies gives new

---

50 For official documentation on Direct Payments, see also http://www.dh.gov.uk/en/SocialCare/Socialcarereform/Personalisation/Directpayments/DH_080273 on the Department of Health official website.

51 We will use interchangeably throughout the paper the terms "local authority", "council" or the abbreviated form "LA".

evidence on the performances of the councils and quantifies the distortions in expenditure levels from the best performance.

The paper has the following structure: in section 2 we give more detailed information about the functioning of the English Social care system, in section 3 we explain the method of analysis used, in section 4 we present the data, in section 5 we present the model specification, in section 6 we show the results and in the last section we make our final conclusions.

3.2 Long-term care provision in the English system

Long-term care can be defined as the "range of medical and social services for persons who are dependent on help with basic activities of daily living caused by chronic condition of physical or medical disability" (European Commission, 2009). The aim of long-term care is to help individuals maintain their level of autonomy. Individuals might lose their autonomy when they are not able to manage one or more daily activities, as a consequence of a disease or as a consequence of aging. Such tasks are also known as Personal Activities of Daily Living (PADLs), for instance maintaining personal hygiene, dressing, eating, moving about and Instrumental Activities of Daily Living (IADLs), such as preparing a meal or doing light housework. Some demonstration projects in the US, Canada, Italy and UK have evaluated the preventative effect that social care programmes have in terms of health outcomes (Johri et al., 2003). Further, it has been found that variations in social care services can explain local variations in health care performance, such as acute hospitalization rates (Fernandez and Forder, 2008).

Social care systems can be distinguished in two main categories: systems which focus their activity on the direct production of social care services (service-led models) and systems which use predominantly monetary allowances to individuals in needs to buy services on the private market or to sustain informal carers (informal care-led model) (Pavolini and Ranci, 2008). In Europe, the first modality is predominant in the Nordic countries, while the second type is more typical of countries like France or Italy. Other social care systems, such as German and English, are in an intermediate position. The English social care system offers either direct services, such as assistance in care homes, hours of home-help services, meals on wheels and similar, but also it provides direct payments to individuals. The complexity of the English system makes it a difficult target of analysis and other complications derive from the availability of fragmentary data. Organization and provision of public services (including social care) in England is decentralized to 150 local authorities that have some autonomy in assessing eligibility for services and for commissioning such services. National standards exist for quality, minimum salaries for social workers and care-home charging rules. The guidance document from the Department of Health, "Fair access to care services" (DoH, 2003) establishes which categories of people should be covered by public assistance. According to this document, eligibility for services depends on the level of risk to an individual's independence in four areas:

- Health and safety including freedom from harm, abuse and neglect;
- Autonomy and freedom to make choices;
- Ability to manage personal and other daily routines;
- Involvement in work, family and wider community life.

Four bands of needs are recognized: critical, substantial, moderate and low (DoH, 2010). Most of the councils have decided to concentrate their assistance only on the highest needs bands. While such autonomy of the councils could represent an advantage for the residents, as they can influence the provision of those services, for example through the electoral process, it can certainly create an equity issue, as individuals residing in different councils are likely to receive different types of services, in terms of coverage rates and quality (this issue is also know as the “post-code lottery”).

Councils define a specific budget devoted to social care services. The major part of social services is financed by a public grant from the central government (the Revenue Support Grant). In order to determine the size of the grant, central government adopts a formula that takes into consideration the local level of needs. The remaining part of the expenditure is funded by the local tax-payer (council tax and business rates). Further, an increased proportion of the gross expenditure is covered by charges to the final users.

In general, councils can commission social care services from public, private or voluntary/charity providers. According to our elaborations on Care Quality Commission data, only 9% of the registered social care providers are publicly owned (by the local authority itself or by the NHS), while the majority is private (72%) or voluntary (19%). Seven typologies of providers can be distinguished, on the basis of their core-service: residential-type providers which include care homes, nursing homes, non-medical care homes and community-oriented providers like home-care agencies, houses for the placement of adults (in the context of the “adult placement scheme”) and nursing agencies. The presence of private and voluntary providers represents a stimulus for the system in saving costs, as they increase competition and may also be able to produce services at lower unit costs, due to their internal organizations. The providers delivering care are selected by the councils among those who have been registered with the public system. Competitive tendering allows local authorities to select the more convenient packages suppliers, respectful of the minimum quality standards.

As a general rule, care in a residential or nursing home is provided to individuals after means testing: in particular, only individuals whose assets level lie below the limit of £23,000 can receive public institutional care, with a variable charging level. The consequence of such a tight asset threshold is that individuals owning their own homes do not receive publicly funded long-term residential care.[52] The concentration of the English system over the most economically deprived individuals is theme of debate, as there is the risk that such mechanism could disincentivise savings, especially in the latter years of life.

As has been discussed, a series of policy innovations have been addressed by the DoH in recent years, especially with the 2005 Green Paper and the 2006 White Paper[53],

52 Nursing care is not means tested and the costs of any nursing care are met by the state.
53 “Independence, Well-being and Choice: Our Vision for the Future of Social Care for Adults in England”, DoH, 2005 and “Our health, our care, our say: a new direction for community services”, DoH, 2006.

focusing on promoting independence and choice for social care users. Councils can promote the independence of older people through more intense use of community care, avoiding admission to nursing or residential homes when possible. Community care consists of services such as home care, assistance in day-care centres, delivery of hot meals, and the provision of adaptive equipment. These policies pursue an improvement in the cost efficiency while providing social care services: community care allows a reduction of the expenditure by avoiding all the "hotel" costs of residential care.

Opportunities for users to choose a provider has increased through the usage of Direct Payments, "cash payments made to individuals who have been assessed as needing services, in lieu of social service provisions".[54] Few studies have found that local and social group variations exist in the uptake of such cash payments (see for instance D. Leece and J. Leece, 2006).

The recent 2009 Green Paper, *Shaping the Future of Care Together* (DoH, 2009) discussed different options for future financing of social care, comparing scenarios where some long-term care is fully funded by the government with options considering different levels of cost-sharing. Estimations in the Green Paper forecast that by 2026, the number of people aged over 85 will have doubled while the number of people aged over 100 will have quadrupled. While currently there are around four people under 65 for every person aged over 65, by 2029 this number is expected to reduce to three. So far, a big impact on social care spending has already occurred, as it has increased by 46% between 2000/01 and 2007/08 (Audit Commission, 2010).

Some major issues about the current English system need to be addressed: means testing, charging fees, concentration on the most deprived users and post-code lottery problems. At least the first three issues listed could be addressed if an improvement in efficiency is attained, as such a change would allow councils to assist more individuals in need. Further, it is important to address the issue of meeting unmet need, particularly in those older people who live alone. In this respect, among British older people who live alone, between 16% and 37% (depending on the type of impairment considered) receive care from an outside source (Walker et al., 2002), numbers that are more than double that for individuals who do not live alone (6% to 15%). An improved efficiency in provision could help reducing the problem of unmet needs, as it would make available new financial resources for the services: there is evidence that currently 85% of the older people do not use council care services (Audit Commission, 2010).

Where an aging population is creating an increased demand for social care services and central government is seeking a reduction in expenditure, the dilemma for a local authority can be thought of as a minimization problem, since the more efficient authorities will be able to spend less, controlling for the amount of services supplied and for the local factor prices. While most models consider efficiency in social care at the provider level, we find that the same question can be posed at the commissioner level, given demand and supply factors. Demand mainly depends on local long-term care needs and on economic deprivation levels, while supply depends on the local social care market, in terms of factor prices and in terms of providers.

54 http://www.dh.gov.uk/en/SocialCare/Socialcarereform/Personalisation/Directpayments/DH_080273

3.3 Methods

In this section we present a review of the existing literature on social care cost and production relations[55] and discuss the econometric approach adopted in the development of our analysis.

From the pioneering study of Hughes (1988), who adopted a cost stochastic frontier approach for the estimation of cost efficiency scores of residential child care in England, many attempts, using different techniques, have been made in order to evaluate cost and production relations in social care. Among the more recent works in the field, Farsi and Filippini (2004) estimated cost efficiency for nursing homes in Switzerland, while Laine and colleagues (2005) computed productive efficiency levels for hospitals and residential houses in Finland. Similar analyses, focusing in general on nursing or residential homes, are those from Kooreman (1994), Fillippini (2001), Nyman and Bricker (1989), Vitaliano and Toren (1994) and Crivelli et al. (2002). All of these papers analyze provider level data, while there are fewer examples in the literature analyzing efficiency levels of local authorities. In general, the latter approach is adopted depending on the specific national framework. Some examples are available in Denmark and Norway, where the provision of social care is responsibility of the municipalities. Hougaard and colleagues (2004) and Borge and Haraldsvik (2009) estimated DEA (Data Envelopment Analysis) efficiency scores for the local government levels. In England, Jimenez and colleagues (2003), using a non-parametric Malmquist Index approach, evaluated productivity changes among 39 English councils in the period between 1992 and 1995. Bebbington and Kelly (1995) in a previous work analyzed variations in services unit costs and volumes by local authority.

In our analysis we choose to adopt a cost stochastic frontier approach in order to determine the cost-efficiency scores of the local authorities using data on their actual expenditure levels, conditioning on the services supplied and input prices. Stochastic frontier models have been introduced in separate works by Aigner, Lovell, and Schmidt (1977) and Meuser and Van den Broeck (1977). The model assumes a cost or a production function where the total error component (ε_t) is split in two parts: one corresponding to the inefficiency (u_i) and the other corresponding to the idiosyncratic error (v_t). The two error components follow different distributions: inefficiency is positively skewed (e.g. distributed as a half-normal or as a truncated normal), while the measurement error is normally distributed. The two parts are distributed independently of each other and of the regressors.

The basic assumptions for a (cross-sectional) cost model, under a half-normal model distribution hypothesis for the inefficiency term, are synthesized by the following equations:

$$(1)\ C_i = C(y_i, w_i, z_i; \beta)\exp(\varepsilon_i)$$

55 For a theoretical and practical review of issues related to social care production see for instance Knapp (1984).

$$(2)\ \varepsilon_i = u_i + v_i$$
$$(3)\ u_i \sim N^+\left(0, \sigma_u^2\right)$$
$$(4)\ v_i \sim N\left(0, \sigma_v^2\right)$$

Costs (C_i) are a function of the quantity of output (y_i) and of the input prices (w_i). A vector of control variables (z_i) can be added, in order to capture heterogeneity between the units observed[56]. Stochastic frontiers represent the major parametric approach to the inefficiency estimation problem. In literature, as we illustrated, similar problems have also been faced with non-parametric frameworks, such as DEA. The advantage of using a parametric framework is that it takes into account the existence of random deviations from the frontier while DEA, leaving out any statistical assumption about the residual, computes any deviation as a part of the inefficiency. Non-parametric methods on the other hand have the advantage of not imposing any distributional assumption a-priori. Both methods present advantages and drawbacks: we decide to adopt a parametric approach also in order to have a superior level of information from the cost function specification.

The model (1-4) can be further extended by relaxing the assumption of homoskedasticity over the variances of the two error components. In fact, such an assumption might not be correct when a set of heterogeneous units is being analyzed. The consequences of ignoring heteroskedasticity, when it is actually present in the data, are particularly serious in the stochastic frontier context. While in the linear regression model, ignoring heteroskedasticity leads to a non-efficient but still unbiased and consistent estimation, in stochastic frontiers it can affect the parameters of the model as well as the inefficiency estimates, especially in case the term affected by heteroskedasticity is u_i (Kumbhakar and Lovell, 2000).

Adopting the definition included in Wang (2002), we adopt a CFCFGH[57] specification which allows the parameterization of the variance terms of both the error components:

$$(5)\ \sigma_{u,i} = \exp\left(t_i \xi\right)$$

$$(6)\ \sigma_{v,i} = \exp\left(w_i \zeta\right)$$

Within this framework, through the use of the vectors t_i and w_i, it is possible to specify one or more variables which have an impact on the variance. Typically the magnitude of the variance is thought to be correlated with size-related characteristics of the observations. The hypotheses made on variance specification are empirically testable, as well as the distributional assumptions.

Finally, we want to remark that three major drawbacks affect in particular cross-sectional stochastic frontier models: estimations strongly rely on distributional

56 In this framework, economic inefficiency u_i can be due either to technical or to allocative inefficiency reasons. Those two components can be only separated using information on input quantities.

57 The model CFCFGH assumes heteroskedasticity on both of the error components, combining the papers of Caudill and Ford (1993), Caudill, Ford and Gropper (1995), and Hadri (1999).

assumptions on each error component, cost inefficiency might be correlated with the regressors and that technical and cost efficiency cannot be consistently estimated, because the variance of the conditional mean or mode does not go to zero as the sample size increases. In order to control if our cross-sectional estimates are biased for some of these issues, we perform random effect and true-random effect estimations, although excluding hypotheses on heteroskedasticity.

3.4 Data sources

All of our data derive from publicly available sources and cover local authorities in the period included between 2002 and 2007. Due to their widely recognised uncharacteristic nature, City of London and Isles of Scilly were dropped out from the dataset, which then includes 148 units and 888 observations, making the panel balanced. In order to perform a cost analysis, we collected information about gross expenditure, quantity of outputs, price of the inputs (included in the form of proxy information) plus a vector of control variables.

Expenditure information is collected by the Department of Health, which provides detailed data about Personal Social Services costs for people aged 65 and over. Outputs from social care consist of heterogeneous services, which include very different typologies, for quality and for production characteristics. Following the data made available by the NHS Information Centre[58], we have at our disposal yearly information regarding the number of weeks of residential and nursing care and the number of home-help hours which have been provided in any local authority to older people. For other community care services, the best approximation offered by official sources is the number of accesses in a single day of the year (namely the 31st of March). For those services, no annual quantification is available. However, considering that social care services are provided to individuals on a daily basis (individuals in need may require assistance even on Bank Holidays), we approximate their annual score by simply multiplying the daily information by 365.

We make the assumption that the main input factors used to produce social care are the social workforce (L) and the building stock (H). We lack the ideal information, represented by registered providers' balance sheets, in order to estimate input prices. The difficulty in obtaining the data from local authorities or from providers sources constrain us to adopt a second-best strategy, selecting the best available proxy measures in order to provide a geographically variable measure of input prices.

Regarding social workforce salaries, some information is included in two surveys, the Labour Force Survey (LFS) and the Survey of Personal Incomes (SPI). Those surveys however give this information aggregated only at regional level[59] and the relatively small

58 Data are for older people services are publicly available at the URL: http://www.ic.nhs.uk/statistics-and-data-collections/social-care/older-people.

59 In England there are 9 regions, which is a small amount with respect to the 150 councils providing social

number of observations does not guarantee consistent estimations of the workforce salary by region. For this reason, we adopt a different strategy deciding to use a more general but largely available source of information such as the median salary of the female employees, as the major part of the social care workforce is composed of women. This information, included in the ASHE (Annual Survey of Hours and Earnings), has the advantage of being able to capture local authority differentials in salaries and, until a better indicator is made available, to represent a suitable approximation under the assumption that the salaries of the social workforce move together in space and time with the salaries whole set of employees.

A similar problem exists for the rental cost of buildings, which are not obtainable from official sources. In England various sources of house price evaluations are available, for example Hometrack, Nationwide Building Society, Halifax and related data are also produced by a public agency such as the Land Registry. Despite the abundance in house price indicators, less information is available on house rental costs. We are aware that the inclusion of house prices as a proxy for P_H, other than being incorrect from a methodological point of view, would introduce a bias in the estimates as house prices and rental costs do not move together. Our strategy is to firstly estimate a price-rent indicator by region, using the data included in the Survey of English Housing (SEH)[60], which provides information on the level of rents in the nine regions in England. The price-rent indicator is obtained by combining the Land Registry house-price regional averages (at the numerator) and the SEH values (at the denominator). In formula:

$$(7)\ \text{price-rent}_{jt} = \frac{\text{house price}_{jt}}{\text{house rent}_{jt}}$$

where the subscript j represents the nine English regions and t is the year of observation. The following step is to obtain a local authority variable measure of house rental costs by dividing the Land Registry council information for the index estimated in (7), which in formula is:

$$(8)\ \text{house rental cost}_{ijt} = \frac{\text{house price}_{ijt}}{\text{price-rent}_{jt}}$$

where the subscript i indicates the 148 English local authorities included in the final sample.

Some caveats exists regarding the use of this proxy: one limitation is that the house price value is related only to the houses which have been sold on the market. However, this problem represents a general issue for house-price indicators, which are all computed on such a selected sample. The price-rent indicator, as it has been estimated on our data, shows an average of 30.3 (cfr. Table 2) which appears to be coherent with what is available from some commercial sources.[61] According to the descriptive statistics included in Table 1, the estimated rental cost of housing is smaller than expected,

---

services.

60 http://www.communities.gov.uk/housing/housingresearch/housingsurveys/surveyofenglishhousing/

61 An extract from the Global Property Guide reports an average of 28 for London as it is shown at http://www.globalpropertyguide.com/Europe/United-Kingdom/price-rent-ratio.

probably due to the underlying house size composition. However, as median standardized values are used in the cost function, differences in sizes are not relevant for the final parameter estimation.

In order to be able to differentiate the units on the basis of the characteristics of their social care provider market, we use Care Quality Commission data about the number and the typology of the registered social care providers operating in the authorities[62.] Data are provided on a yearly basis and report information over the typology and the ownership of the provider and, in case it is a care home, the number of registered places available.

Demographic statistics and deprivation figures such as the standardized mortality ratio (SMR) are provided by ONS (Office for National Statistics). The standardized mortality ratio is a health-risk measure which is defined as the ratio of actual deaths (D_{it}) over the expected number of deaths, standardized on the population age structure. In formula, according to the ONS methodology is:

$$(9)\quad SMR_{it} = 100 \cdot \frac{D_{it}}{\sum_{k=1}^{K} P_{it,k} M_k}$$

where P_k and M_k are the population number and the age-specific death rate respectively both associated with the *k-th* group, in the *i-th* authority. The denominator represents then an estimation of the expected number of deaths. The K age groups are defined as follows: 0-4, 5-14, 15-24, … , 65-74, 75 and over.

Descriptive statistics by year of all of the variables utilized in the model specification are included in Table 1 and Table 2. As the panel is balanced, the number of observations for all of the variable correspond to the full sample. In Table 1 we include the variables related to expenditure, output quantities and estimated input prices, while in Table 2 we include the other sources of information including the ad-hoc variables inserted in the cost function.

From Table 1 we can observe that average gross expenditure for social care services, in real terms, has increased until 2006, with a small decrease in the last year of the panel. Wages show an increasing trend while our proxy for average house rental cost decreases slightly until 2004 and then starts to increase, following the peaks of the real estate bubble. Output level patterns respond to the reforms implemented by Central Government: we observe a decreasing trend in the number of residential and nursing care-weeks provided, with a corresponding increase of home-help hours, a clear signal of a substitution phenomenon. The other community services, which in this table appear in their estimated annual value, show different patterns. For instance, the use of Direct Payments is increasing steeply, in line with central government policy guidelines, which stated the strategic need for an increase in their uptake rates. Also, other typologies of services like the provision of equipment and adaptations and professional support are

---

62 The more recent set of data is available at http://www.cqc.org.uk/guidanceforprofessionals/socialcare/careproviders/statisticsonregisteredproviders.cfm

increasing their numbers, while day-care accesses and meals-on-wheels see a reduction in their public supply.

According to Table 2, total population and its older component are increasing (values reported represent averages for the local authorities). In particular, the number of individuals aged 85 and over are growing at a greater rate, in keeping with the overall aging population story. The standardized mortality ratio remainins fairly constant, not showing any clear trend. The descriptive statistics of registered providers show a reduction in the number of residential and nursing homes. This phenomenon of "care-home closure" is a consequence of the competetive market mechanisms which are operating in long-term services (see also Netten at al., 2005). On the other side, the presence of home care and nursing agencies is growing, accordingly to the increasing importance placed on community care support. In terms of the ownership of registered providers, we observe that while the numbers of public and of voluntary providers are quite stable throughout the sample, the number of privately owned registered providers is increasing.

3.5 Model specification

Local authorities' efficiency estimation is modelled through the use of a multi-output translog cost function, implicitly including the hypothesis of cost minimization. Such an assumption may not be appropriate in this case: even if local authorities face political and financial incentives from central government to reduce their expenditure levels, their behaviour cannot be explained by a completely unconstrained cost-minimization strategy. In fact, the public nature of the services and also the role of regulation in the quality standards and in the labour markets may affect the optimization process: for this reason we interpret our model in the framework of the behavioural cost functions (Bös, 1986).

It must be noted that council LAs' expenditures do not correspond to the production costs but to the amount that they reimburse to providers, the actual producers of the services. Local authority social care managers aim to reduce expenditure levels conditioned on the output but have no direct control over the providers' production process, tending to select the cheapest suppliers of services, *ceteris paribus,* through the tendering process. Given that within every local authority there will be different proportions of public, private and voluntary providers, the level and the type of inner competition is likely to differ, influencing the average expenditure for the services. For these reasons we include in the model variables regarding the ownership proportions of the providers, distinguished by the type of service offered. Naturally, non-public providers will tend to be located in areas where the demand for services is higher, while public providers will tend to cover such market gaps.

As previously introduced, we adopt a translogarithmic specification since we are analyzing a set of units which are providing a plurality of outputs. In fact a Cobb-Douglas cost function, rather than being a less flexible approach, "cannot accommodate multiple

outputs without violating the requisite curvature properties in outputs space" (Kumbhakar and Lovell, 2000) and for this reason has not been considered.
The dependent variable is represented by the gross expenditure for social services related for older people (C), where the adjective gross refers to the fact that such expenditure also includes the contribution charged to users. A possible alternative measure would have been the use of the net expenditure, which excludes charges. However, a model using net expenditure would answer to a slightly different research question and further, the use of this variable in an inefficiency study would seriously bias the results as the ability to pay charging fees by the resident population differs by authority and it is difficult to control.

The first of our outputs is the number of weeks supplied to older people in a residential or a nursing care home (y_1). This variable represents a more precise approximation of output information required than a simple count of care home residents, as individuals have different lengths-of-stay in the sheltered homes. Further, it corresponds to the weekly counterpart of the residential days. Note also that residential and nursing care are aggregated: even if a disaggregation of the two components would have been more appropriate, as nursing services are more complex and costly than general care, such disaggregation is not supported by the currently available data.

The second output is the number of home-help hours (y_2) provided to older residents. This variable is important from a policy perspective, as a comparative analysis can add information about the current trade-off in expenditure between home-help and residential and nursing care.

The third output contains the number of all accesses to "other community care services" (y_3), i.e. day-care, meals on wheels, short-term breaks, direct payments, professional support, equipment and adaptations and 'others'[63]. The main drawback in aggregating heterogeneous outputs is that the collapse in a single variable of different services in terms of typology and in terms of quality might lead to a non-precise final output measure. The issue of output (or input) aggregation is well-known in the literature and typically occurs in frameworks where multi-output producers (possibly using a multi-input production process) are being analyzed.[64] However, a certain degree of output aggregation is necessary in order to contain the number of covariates in the empirical model, also considering that some of the services present very small numbers in certain authorities.

Social care services are labour-intensive outputs and care does not require the use of complex medical technologies or drugs. In this framework we are assuming two production factors for social care services: the labour workforce and the housing stock. The labour factor is the entire social workforce that, according to the definition of the Department of Health, is composed of those "who work in public services that are provided, directly or commissioned, by local councils to discharge their personal social

63 This taxonomy, including the residual `others' category, blindly follows the labels included in the RAP (Referrals, Assessments and Packages) data, available in the NHS Information Centre database at http://www.ic.nhs.uk.
64 To use a related example, hospital cost-production literature supplies a wide number of analyses where a plurality of services (e.g. the number of discharges under different DRG types) is aggregated into aggregated variables. For a theoretical approach to the problem see also Brown et al. (1979) and Hall (1973).

services (PSS) responsibilities"[65]. In the housing stock, we include all buildings and houses in which social care services are provided or where related administrative tasks are performed. While all social services can be generally defined as labour-intensive[66], some of them make a relatively greater use of house stock. In particular, this is the case of residential and nursing care, as this form of care requires institutionalization. Furthermore, some types of community-care such as assistance in day-care centres require a certain amount of house stock.

The complexity of social care services is in general standardized within a service typology for any specific level of needs (critical, substantial, moderate or low). In fact, the services provided as social care do not require medical treatments, but are a simple help to perform some daily tasks. However, particular local authorities where a higher fraction of individuals in need is resident will eventually be forced to concentrate their activity on a very impaired fraction of older people, presenting multiple limitations in terms of PADLs or IADLs. These particular situations can be the source of an increase in the unit costs of the services provided. For this reason we need to include in the model an indicator for the "population complexity". For example, local authorities with a higher presence of older residents are more likely to face multiple needs. However, when including in the model as a covariate the ratio of older residents (people aged 85 and over) we did not find a significant effect. The reason is to be found in the fact that councils mostly support the poorest older individuals. For this reason, we created an interaction between the rate of people older than 85 and the local standardized mortality ratio (*SMR*). The standardized mortality ratio represents an age-standardized indicator which is able to capture economic deprivation, based on the assumption that the component of mortality not dependent on age is mainly related to low income. Distinguishing the "population complexity" variable from a simple older person rate is crucial also considering that some of the authorities in England are characterized by a higher fraction of wealthy older people, which moved there in the retirement age, biasing the original age distribution of the areas.

While we are controlling for population complexity, we are not including in the model a pure-quality indicator. Surveys regarding the satisfaction of the clients have been performed over the years in a non-continuous way by the CSCI (Commission for Social Care Inspection)[67] in the PAF (Performance Assessment Framework) quality framework. The lack of continuity does not make these data appealing for our analyses.

A more continuous quality indicator, collected on a yearly basis, is named CPA index (Comprehensive Performance Assessment). Its measure is based on a "star-rating" appraisal, from zero to four stars, which is proportional to the quality of their services (Audit Commission, 2006). The CPA quality index has been recently used as a dependent variable in literature for the evaluation of the local government performances (Revelli, 2010). After the inclusion of this variable in the model as a covariate, we found empirically that it does not have a significant impact on the expenditure and consequently

---

65 No differentiation criteria is provided upon the level of experience or of the tasks performed in this definition.

66 The other factors which are involved in social services production are vehicles, equipments and adaptations. We do not include their prices in the model as they contribute to a marginal part of the costs.

67 A discussion regarding the quality issue and some of the related available indicators in English long-term care can be found in Clarkson and Challis (2006).

decided not to include it in the model. One problem could be that CPA does not by construction have a great enough variation by authority, as it composed of only five categories. However, the same results apply when dummies for the single-rating star categories are adopted. In this regard, in the results section we will try to analyze, through some descriptive analysis, why the variable seems to be unrelated to costs.

In order to control, at the *i* authority level, for the ownership distribution of the registered care-suppliers, we include in the model different variables regarding the proportions of private (PR_i) and voluntary (VR_i) providers supplying residential care and for the proportions of private (PC_i) and voluntary (VC_i) providers operating in community care. Such proportions are simply defined as the care suppliers' numbers for each specific ownership type as a fraction of the total number and captures effects deriving from the presence intensity of non-public suppliers in each authority. We expect that the greater the non-public presence, the bigger the competitive effects will be. As reference categories (omitted in the model specification) we consider the proportions of public providers. Further, since we want to separate the "typology of care" effect from ownership, we also include in the model the proportion of providers community-care oriented ($COMM_i$).

The estimated model is specified as follows:

(10)

$$\ln\frac{C_i}{P_{H,i}} = \alpha_0 + \sum_{m=1}^{M}\alpha_m \ln y_{mi} + \beta_L \ln\frac{P_{L,i}}{P_{H,i}} + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{mn}\ln y_{mi}\ln y_{ni} + \frac{1}{2}\beta_{LL}\left(\ln\frac{P_{L,i}}{P_{H,i}}\right)^2$$

$$+\sum_{m=1}^{M}\gamma_{mL}\ln y_{mi}\ln\frac{P_{L\,i}}{P_{K\,i}} + \delta_{COMPL}COMPL_i + \delta_{PR}PR_i + \delta_{VR}VR_i + \delta_{PC}PC_i +$$

$$+\delta_{VC}VC_i + \delta_{COMM}COMM_i + v_i + u_i$$

where the subscript *i* represents the council.

As we are dealing with a set of heterogeneous councils, we also control for variance heteroskedasticity, according to the model explained in the methodological section. In both the vectors t_i and w_i, from the equations (5) and (6) we include a single size variable which is the local authority population.

We impose linear homogeneity in input prices dividing C_i and $P_{L,i}$ by the house stock price $P_{H,i}$. Further, before the log-transformation of the model variables, we standardized all the costs, output and input price variables by their median. Monetary values have been deflated using the index provided by the NHS Information Centre.[68]

68 See for example the Personal Social Cervices Report at the URL http://www.ic.nhs.uk/webfiles/publications/PSSEX10708/Personal%20Social%20Services%20Expenditure%20and%20Units%20Cost%2C%20England%202007-08_1.1.pdf

3.6 Results

In this section we present the results of our model estimations. Main estimates are performed through STATA[69] software and are based on a pooled cross-sectional analysis. Standard errors are estimated using a robust variance clustered by council. A cross-sectional study is preferable to a panel regression as most of the variation in the costs and output variables has been found to occur mostly between the units rather than within them, and also due to the fact that our data are aggregated. We will also compare our cross-sectional results with a random effects (RE) and a true random effects (TRE) model (Greene, 2005) in order to assess the stability of our estimates.

The SFMODEL command allows us to make assumptions about the mean and the variance of the inefficiency term, giving us the opportunity to test different model specifications. In our context we parameterize only heteroskedasticity, testing empirically the assumptions through a LR test.[70] Heteroskedasticity of both error components is modelled over the population resident in the authority (in logarithmic values) and over a constant term. On this respect, we are following the general design that assumes heteroskedasticity as a function of size variables. According to the results of the LR tests included in Table 6, homoskedasticity of both the error components is rejected in all the combinations tested.[71]

In Table 3 we show coefficients and the related significances obtained from the cost model while in Table 4 we present the variance-related coefficients and the ancillary parameters for the stochastic frontier. Theoretical constraints of price homogeneity and symmetry over the parameters have been imposed a priori, while homotheticity and concavity are checked after the estimation. We firstly notice that, despite theoretical requirements need a translog functional form to be used in a multiple output context, just a few of the second order and interaction parameters among are significant. However, given the theoretical requirements and the major flexibility allowed by a translog, we do not consider the use of a Cobb-Douglas.[72]

The principal parameters of the cost model behave consistently with expectations in that costs are positively related with input prices and outputs. Labour input has a major impact on expenditure as the first order parameter is estimated at 0.768 and is highly significant. All of the outputs show positive and strongly significant coefficients and their size is in accordance with policy experience. In particular, residential and nursing care represents the more costly form of care as it implies a more intense using of the housing factor. Coherently, the related parameter is far bigger than those of the other two outputs. Home help and the other forms of community care, which are more cost-saving forms of assistance, have a smaller effect on the expenditure as expected.

---

69 In order to perform our estimates it has bee used the SFMODEL Stata® command by Wang (2002)

70 At this stage of the work, only half-normal distributed inefficiency term has been included in the estimates due to a difficult convergence for truncated-normal based models.

71 Decisions on the tests are made on the basis of a 95% confidence level.

72 Just for the sake of curiosity we performed a LR test over the Cobb-Douglas specification which would be rejected with a 95%, but not with a 99% confidence level.

Due to the presence of interactions and squared effects, first-order parameters are not interpretable directly as elasticities. Estimations of cost-output elasticities are included in Table 5, as well as their standard errors, computed using the delta method. Residential and nursing care cost elasticity is around 6.5 times bigger than home-help elasticity (0.749 vs. 0.115). This means that a percentage increase in the current provision of residential and nursing care would have the same impact on expenditure as increasing the quantity of home-help provided by the 6.5%. Such a gap is even bigger when we consider the trade-off between residential and other community care, as the ratio between the two parameters is around 15 (0.749 vs. 0.051). As a consequence, home-help services have an impact on the expenditure which is double that of other types of community care (0.115 vs. 0.051). Such a hierarchy represents a relevant finding for the English policy maker when redefining the characteristics of long-term care supply.

After the estimation of the three cost-output elasticities, we are able to compute the returns to scale parameter according to the following formula which considers the multi-output framework:

$$(11)\quad RTS = 1 \Bigg/ \left(\sum_{m=1}^{M} \frac{\partial \ln C}{\partial \ln y_m} \right)$$

where the subscript m represents an indicator for the M different outputs included in the model. Our estimate of returns to scale is 1.09, significantly different from the threshold value of 1 at 95% level. This results shows that local authorities are providing services under the optimum. This is not surprising as public agents in a regulated context are not expected to optimize the scale of provision, as happens with private firms. Further, the existence of unexplored scale economies is also coherent with the existing constraints in the expenditure levels, which cannot be easily increased in order to get scale optimality.

Our “population complexity” indicator, which is the interaction between the rate of people older than 85 years and the local standardized mortality ratio affects expenditure with a positive sign as expected, but is significant only at the 10% level. Nevertheless, such an indicator should at least partially control for the average complexity of the services provided.

As regards providers’ market indicators, we have evidence suggesting that higher proportions of private and voluntary providers operating in residential and nursing care are able to curb social care spending. The related parameters are significant at the 95% and 99% level of confidence respectively. Given that these proportion variables have different means, we also computed elasticity values in order to have comparable parameters. The expenditure elasticity value for the private residential providers is estimated to be -0.23, while the same effect for voluntary providers is -0.12. From these numbers we can conclude that the private providers produce a saving effect in a measure which is nearly double that of voluntary ones. Cost-elasticities of private and voluntary providers operating in community care are -0.09 and -0.01 respectively, but only the first of those is significant, at the 10% level. Competitive tendering promoted by the presence of external providers seems to have a milder effect on the provision of community care than those existing for residential and nursing care.

The existence of a time-trend in the cost framework is usually interpreted as an indicator of a Hicks-neutral technological change. In our model all the time-dummies have a positive effect, except for 2003. No major technological changes have affected the sector, apart from the innovation of telecare[73] which is a recent development and in the period of analysis was adopted only in a small number of authorities. In a similar context, more changes occurred in the production and organization of providers, as several mergers and closures occurred. Two arguments could arise from this natural selection process: either such evolution of the market has reduced unit costs, since bigger providers were more able to optimize their production processes, or it has contributed to their increase, giving them greater contractual power. The unit costs of services are estimated by the NHS by dividing the amount of related expenditure by the quantity of services provided. In Figure 1 we plot the deflated trends of these estimates for residential care and home care. We observe that the unit cost of a residential care week has increased constantly, with the exception of a fall in 2003. In part this phenomenon could have been driven by a tendency of local authorities to provide residential care to the more complex cases. The unit cost of one hour of home-help has instead increased until 2005, at which point it started to fall steeply. According to the joint analysis of the time dummies and the descriptive evidence, there seems to be a prevalence of factors with increasing costs. Is difficult however to ascertain which part of the increase is due to the increase of some providers' market power and which part is due to the increased complexity of assistance. Further, unit costs might have been enhanced by wage regulations for social care workers (Machin, S. and Wilson, J. 2004).

An alternative explanation for the behaviour of the time-dummies is that it is a consequence of a change in the quality of services. While the CPA index presents a too small intra-unit variation when analyzed at the local authority level, the observation of its trend in a descriptive analysis could provide more valuable evidence. Plotting the average trend of the "star-rating" CPA quality index, as in Figure 2, we observe a steady increase in the average number of stars attained by the councils. While this could be easily interpreted as a signal of a steady quality increase, we also found that the CPA index plotted has a surprisingly high correlation (0.98) with a simple time-trend. This result is quite suspicious and could induce us to think that the CPA quality indicator presents a certain "political bias". Whether or not this is the case, any evidence obtained through this variable should be interpreted cautiously.

In table 4 we observe that the council population size has a positive impact on the inefficiency heteroskedasticity and a negative impact on the heteroskedasticity of the idiosyncratic component. In fact, as has been shown in Wang (2002), the variables included in the inefficiency variance specification give a positive contribution to the inefficiency mean marginal effect. We could then conclude that bigger authorities (in terms of population size) are more likely to be inefficient. This conclusion is somehow expected, as the greater the population to assist, the more serious the resource optimization problem would be.

After commenting on the coefficients of the cost model, we focus our attention on the inefficiency score estimates, which have been calculated as $Ineff_i = E(\exp(u_i) \mid \varepsilon_i)$

73 Telecare can be defined as "the use of information and communications technology (ICT) to support care in the wider community" (Barlow et al., 2005),

according to the Battese and Coelli (1988) formulation. The γ parameter, which represents the fraction of the total error variance due to inefficiency, results are significant but lower than expected, with a value of 0.37. There are two possible reasons for such a small value: either differences in inefficiency among councils exist but policy instruments are able to compress their amounts or, in a more pessimistic view, the model is not able to capture the entire inefficiency component. In order to verify the validity of cross sectional estimates, we also performed two different panel estimates, with the random effects and with the true random effects specification. These models however do not take into account the heteroskedasticity problem. Results published in Table 7 show that cross sectional results are substantially robust: there are not big changes in the main parameters, apart from β_L coefficient, which is lowered to 0.51. We tested the stability of our cross sectional inefficiency score by using the Spearman's rank correlation in order to verify the correspondence with the two panel-based inefficiency scores: the resulting value was 0.79 with the random effects and 0.62 with the true random effects estimate. Comparative descriptive statistics with the cross-sectional inefficiency score are available in Table 9.

The inefficiency trend curve (cfr. Table 8 and Figure 3) shows a slight negative slope: while in the first year the expenditure distortion from the frontier has been on average 8%, in the last year of the panel it has been reduced to 7.5%. However, since confidence intervals on the trend are not narrow enough for supporting an inefficiency reduction story, we must conclude that no significant changes in inefficiency have occurred. The inefficiency distribution presented in Figure 4, averaged by council, shows that most of the authorities are included in an inefficiency range between 1 and 1.10, meaning that the distortion in expenditure is contained to 10%. Fewer authorities appear in the range 1.10-1.20 while only a small number of cases are included between 1.20 and 1.35. These numbers provide a realistic estimate of the potential saving impact occurring whether the commissioning process would be further optimised, once the sources of inefficiency are identified.

In order to assess the stability of the results and at the same time to provide evidence about the mobility of authorities in terms of inefficiency rank, we provide in Table 10 information about variation in the quintile membership for authorities between the first and the last year of the panel. At first glance, what looks most impressive is the stickiness of the last quintile: 62% of the authorities which were found in the last quintile at the beginning remain in that same part of the inefficiency distribution. In the other quintiles the persistence rate is more limited: in particular, the first and the fourth quintiles show persistency rates of the 40% or about, while in the second and in the third quintiles more movement is observed.

In Table 11 we present regional descriptive statistics of the inefficiency results. On average, the worst performances are found in the East, where the mean inefficiency score reaches 1.132. On the other side, North-East seems to include more virtuous authorities as its mean inefficiency is contained to 1.056. The general ranking is confirmed when looking at the median inefficiencies, showing that these results are robust to outliers.

In Table 12 we show returns to scale and cost-output elasticities displayed by region. The estimated returns to scale are higher in North-Western councils (1.113) and are lower in the Greater London (1.073). As regards the geographical patterns of cost-output elasticities, we observe that North-West has the highest values for residential care (0.786) and the lowest for home-help (0.070) and other community care (0.043). At the other end of the scale, the cost elasticity for home help and other community care are highest in South-West (0.152 and 0.060 respectively), while the cost elasticity for residential care is the lowest (0.708). As residential and community care services are substitute services, is not surprising that their elasticities show inverted patterns. The interpretation of such variations is that where elasticities are found to be particularly high, either an excessive amount of care is delivered or it is produced at particularly high costs.

3.7 Conclusions

This paper proposes an empirical model in order to estimate inefficiency scores for English local authorities in their provision of long-term care provision. Available data allowed for the analysis of 148 local authorities providing social services for older people in a period between 2002 and 2007, when diverse policy inputs might have influenced inefficiency patterns.

Complications in efficiency studies for social care provision in England arise because councils may choose to produce long-term services internally (in-house) or to commissioning them to an external provider (contracting-out). In this respect, we approach the inefficiency estimation problem by taking into account the role of private and voluntary providers in the social care provision market, accounting for their diffusion in the authorities. Empirically, the unavailability of balance sheet data from care providers forces us to adopt some second-best solutions such as the use of proxy-variables when determining the unit price of the production factors.

Parameters estimated from the cost-function model quantify the different impacts that different types of care have on the total social care expenditure. In particular, the cost-output elasticity for residential care is found to be around 7 times bigger than that of home help and about 15 times bigger than that of the other types of community care. On the basis of these elasticities, we are able to evaluate the existence of returns to scale in councils' activity. We found that local authorities are providing services on average under the optimal scale, as the return to scale parameter is found to be 1.09. This evidence is not surprising, as in a regulated context where public services are being provided, the policy maker does not follow optimal scale considerations.

The analysis of inefficiency trends in our specific time-interval gives evidence of a period characterized by important policy innovations such as the endorsement of community care services and the promotion of cash refunds as a form of support. We found a smooth decrease in inefficiency, as the related score decreased from 1.080 in 2002 to 1.076 in 2007. As confidence intervals for the trend are wide, we conclude that no significant changes in efficiency levels occurred.

Councils seem to benefit from a positive competition effect when the social care provider market is more complex and includes a greater proportion of private and voluntary providers. In the residential and nursing care market, a greater proportion of private structures present in the area is associated with cost saving effects which, measured in terms of elasticity, are double the size (-0.23) of those effects relating to voluntary providers (-0.12). In the community care market, such saving effects are milder (-0.09 and -0.01) and less statistically significant.

Some major issues exist for implementing a more precise efficiency estimation. Apart from the unavailability of input prices, the absence of precise indicators controlling the quality levels may still be biasing the estimates, which can only control for the local complexity of care. In this sense, some political bias seems to be affecting the CPA quality index. Irrespective of this issue, more precise and articulated indicators are required in order to distinguish quality by single type of service and by category of service recipient.

Further studies could arise from this research: we discussed the effect that providers' ownership has on expenditure levels. On a similar note, a topic not addressed in this article is the investigation of the effects that concentrations of care provider markets have on expenditure levels. When only a small number of providers is present, competitive tendering effects could be reduced or nullified. Also, since providers' activity could be diffuse in the neighbouring authorities, it would be interesting to investigate whether geographical patterns exist with the use of spatial econometric techniques.

As a final remark we want to underline that the estimation of inefficiency scores leaves out any consideration about the level of provision. Some authorities could be very efficient in the care levels they provide, but still unmet needs could be present. While there is not necessarily a direct relation between efficiency in provision and the level in which needs are met, a systematic improvement in the efficiency could make room for increasing the number of individuals that are supported.

Acknowledgements: We would like to thank Vincenzo Atella, Silvio Daidone, Domenico Depalo, Juliette Malley and Tom Snell for their comments. A special thank to Catherine Henderson for her comments and the stimulating discussions, to Hung-Jen Wang for the assistance with his STATA package, for his suggestions concerning convergence of the models and for the provision of more updated routines. We also thank Care Quality Commission (CQC) and Dave James for the provision of data about the distribution of the providers in the local authorities. Finally, we remark that all of the possible errors in this work are of our exclusive responsibility.

Tables and figures

Table 1
Descriptive Statistics (Part A)

Variable		EXP	W	HRC	RES	HH	DC	M	ST	DP	PS	E&A	O
2002	**Mean**	**49954.5**	**13623.0**	**6227.5**	**78252.1**	**1046957.0**	**296000.0**	**344264.5**	**56665.0**	**3911.4**	**215060.0**	**336162.5**	**179718.1**
	Std. Dev.	***34726.3***	***3532.0***	***3537.0***	***60728.7***	***781839.4***	***313132.9***	***364719.0***	***104062.8***	***6916.4***	***409220.5***	***429625.7***	***286957.1***
2003	**Mean**	**50832.5**	**14478.7**	**6167.3**	**79944.2**	**1091815.0**	**280556.8**	**301472.7**	**63343.6**	**7985.6**	**271653.7**	**370980.6**	**170551.2**
	Std. Dev.	***36523.8***	***3617.8***	***3325.6***	***61943.1***	***866488.0***	***299345.6***	***323047.4***	***115106.0***	***13943.4***	***685291.0***	***404784.3***	***265422.8***
2004	**Mean**	**53731.6**	**14104.3**	**5957.8**	**77545.4**	**1154557.0**	**238950.0**	**270916.9**	**77900.0**	**15371.9**	**233875.0**	**386944.7**	**157337.2**
	Std. Dev.	***38916.6***	***3498.8***	***2868.9***	***59948.7***	***936364.6***	***203405.1***	***305973.1***	***145970.9***	***19669.4***	***346831.2***	***507789.5***	***243087.3***
2005	**Mean**	**55788.2**	**14316.2**	**6012.5**	**75996.4**	**1231404.0**	**237600.0**	**246966.0**	**75733.8**	**24908.8**	**269550.0**	**428624.2**	**118097.2**
	Std. Dev.	***40516.9***	***3451.4***	***2937.2***	***58905.0***	***1027486.0***	***207054.3***	***286267.2***	***109098.2***	***31128.1***	***397012.9***	***538599.3***	***230247.2***
2006	**Mean**	**56025.5**	**14571.5**	**6156.2**	**73657.8**	**1287432.0**	**222575.0**	**212112.2**	**66737.5**	**33861.2**	**319715.0**	**439978.9**	**114454.6**
	Std. Dev.	***40006.9***	***3422.3***	***3144.8***	***57075.9***	***1058735.0***	***192358.4***	***235456.1***	***121383.8***	***38702.6***	***511771.8***	***492095.9***	***236726.2***
2007	**Mean**	**55006.3**	**14636.9**	**6323.3**	**66411.1**	**1339767.0**	**211800.0**	**188106.6**	**37502.5**	**51889.2**	**326145.0**	**506888.2**	**129121.2**
	Std. Dev.	***39165.8***	***3475.5***	***3315.3***	***52779.2***	***1182624.0***	***178499.9***	***216901.5***	***64403.6***	***70821.6***	***476828.6***	***530881.0***	***246109.8***
Total	**Mean**	**53556.4**	**14288.4**	**6140.8**	**75297.6**	**1191989.0**	**247988.0**	**260686.1**	**62980.8**	**22988.0**	**272664.1**	**411550.5**	**144879.9**
	Std. Dev.	***38328.4***	***3507.4***	***3190.0***	***58641.8***	***987114.6***	***239720.1***	***296989.8***	***113157.6***	***40149.6***	***484662.4***	***488185.1***	***252745.4***

Legend: EXP = Gross Expenditure (in thousand £), W = Wage (in £), HRC = House Renting Cost (in £), RES = Care-Home Weeks, HH = Home-help Hours, DC = Day-Care Accesses, M = Meals On Wheels, ST = Short-Term Breaks, DP = Direct Payments, PS=Professional Support, E&A= Equipment and Adaptations, O= Other Services.

Table2
Descriptive Statistics (Part B)

Variable		P-R	POP	POP85	SMR	RH	NH	HCA	NA	AP	PUB	PRI	VOL
2002	**Mean**	**24.4**	**335.4**	**6.5**	**100.4**	**159.3**	**16.4**	**0.8**	**3.0**	**0.0**	**2.9**	**107.3**	**26.9**
	Std. Dev.	**2.9**	***251.0***	***5.7***	***10.9***	***154.6***	***16.2***	***5.1***	***4.0***	***0.0***	***5.9***	***110.4***	***24.2***
2003	**Mean**	**27.7**	**336.8**	**6.3**	**102.2**	**150.6**	**14.9**	**12.7**	**6.3**	**0.0**	**12.5**	**111.3**	**26.6**
	Std. Dev.	**3.1**	***252.5***	***5.6***	***10.7***	***143.8***	***14.8***	***15.1***	***6.0***	***0.0***	***12.7***	***111.4***	***23.7***
2004	**Mean**	**31.3**	**338.4**	**6.4**	**101.7**	**135.2**	**13.6**	**27.6**	**6.2**	**0.3**	**15.6**	**119.3**	**27.7**
	Std. Dev.	**2.4**	***253.8***	***5.6***	***11.7***	***131.4***	***13.6***	***24.1***	***5.8***	***0.5***	***14.3***	***115.9***	***24.5***
2005	**Mean**	**32.2**	**340.7**	**6.7**	**101.6**	**131.3**	**13.4**	**31.2**	**5.8**	**0.8**	**16.4**	**120.6**	**27.4**
	Std. Dev.	**1.7**	***255.5***	***6.0***	***11.5***	***129.2***	***13.2***	***26.3***	***5.2***	***0.6***	***14.9***	***116.3***	***24.0***
2006	**Mean**	**32.9**	**342.9**	**7.1**	**101.7**	**129.9**	**13.5**	**31.8**	**5.2**	**0.9**	**15.5**	**120.9**	**26.4**
	Std. Dev.	**1.9**	***258.4***	***6.4***	***12.4***	***128.8***	***13.3***	***26.0***	***4.6***	***0.7***	***14.0***	***116.3***	***23.5***
2007	**Mean**	**33.5**	**345.2**	**7.4**	**101.7**	**127.9**	**13.7**	**32.8**	**4.8**	**0.9**	**14.5**	**122.3**	**25.7**
	Std. Dev.	**2.3**	***266.4***	***6.8***	***12.2***	***127.7***	***13.4***	***26.8***	***4.4***	***0.6***	***13.3***	***117.8***	***23.0***
Total	**Mean**	**30.3**	**339.9**	**6.7**	**101.6**	**139.0**	**14.3**	**22.8**	**5.2**	**0.5**	**12.9**	**116.9**	**26.8**
	Std. Dev.	**4.1**	***255.6***	***6.0***	***11.6***	***136.4***	***14.1***	***25.1***	***5.2***	***0.7***	***13.6***	***114.5***	***23.8***

Legend: P-R = Price-Rent Ratio, POP = Total Population (in thousands), POP85 = Population aged 85 and over (in thousands), SMR = Standardized Mortality Ratio, RH= Residential Homes, NH = Nursing Homes, HCA = Home-Care Agencies, NA = Nursing Agencies, AP = Adult Placement Schemes, PUB = Public Sector Providers, PRI = Private Sector Providers, VOL = Third Sector Providers

Table 3
Cost-frontier Estimates

Parameter	Coefficient	SE
β_L	0.768***	0.033
α_1	0.754***	0.043
α_2	0.108***	0.034
α_3	0.051***	0.015
γ_{1L}	0.116	0.091
γ_{2L}	-0.179**	0.080
γ_{3L}	-0.013	0.060
α_{12}	0.089	0.067
α_{23}	0.008	0.027
α_{13}	-0.027	0.030
β_{LL}	0.015	0.197
α_{11}	-0.088	0.087
α_{22}	-0.063	0.063
α_{33}	0.013	0.011
δ_{COMPL}	0.044*	0.026
δ_{PR}	-0.320**	0.148
δ_{PC}	-0.114*	0.068
δ_{VR}	-0.564***	0.159
δ_{VC}	-0.124	0.136
δ_{COMM}	-0.172	0.117
Year 2003	-0.069***	0.021
Year 2004	0.036	0.028
Year 2005	0.074**	0.029
Year 2006	0.071**	0.030
Year 2007	0.121***	0.031
Constant	0.253	0.160
N	888	
Log-likelihood	508.91	

Table 4
Variance effects and ancillary parameters

Variable	Coefficient	SE
Variance of u		
Population (in logs)	**0.998****	**0.453**
Constant	**-10.492*****	**3.188**
Variance of v		
Population (in logs)	**-1.108*****	**0.200**
Constant	**1.949***	**1.029**
Ancillary Parameters		
γ	**0.373*****	**0.005**
σ^2	**0.026*****	**0.01**

Table 5
Cost-Output Elasticity Estimates and Returns to Scale

Parameter	Estimate	SE	Lower C.I.	Upper C.I.
$\varepsilon_{C,y1}$	**0.749*****	**0.062**	**0.628**	**0.870**
$\varepsilon_{C,y2}$	**0.115****	**0.053**	**0.011**	**0.218**
$\varepsilon_{C,y3}$	**0.051***	**0.027**	**-0.002**	**0.104**
RTS	**1.094****	**0.048**	**1.001**	**1.187**

Table 6
LR test on the Model Specification

Null hypothesis	*Log-L*	λ	*p-value*	*AIC*	*BIC*	*d.o.f.*	*Decision*	*Convergence*
$H_0 : \mu = 0$	**508.91**	-	-	**-957.81**	**-814.14**	**1**	**Chosen**	**Yes**
$H_0 : \xi = \zeta = \mu = 0$	**460.49**	**96.84**	**0.000**	**-864.97**	**-730.88**	**2**	**Rejected**	**Yes**
$H_0 : \xi = \mu = 0$	**501.49**	**14.84**	**0.000**	**-944.97**	**-806.09**	**1**	**Rejected**	**Yes**
$H_0 : \zeta = \mu = 0$	**461.42**	**94.97**	**0.000**	**-864.83**	**-725.95**	**1**	**Rejected**	**Yes**

Figure 1
Unit Costs Trends

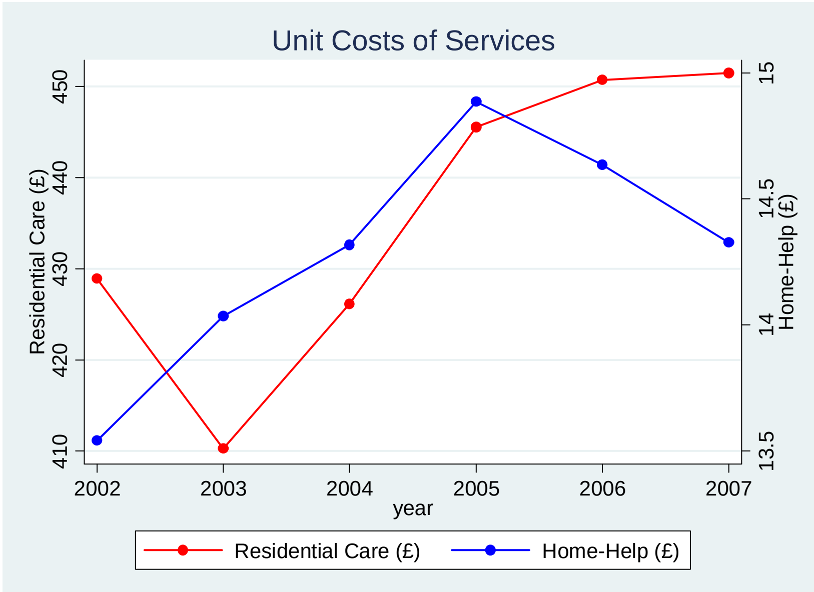


Figure 2
Trends in CPA Star Rating Index by Region

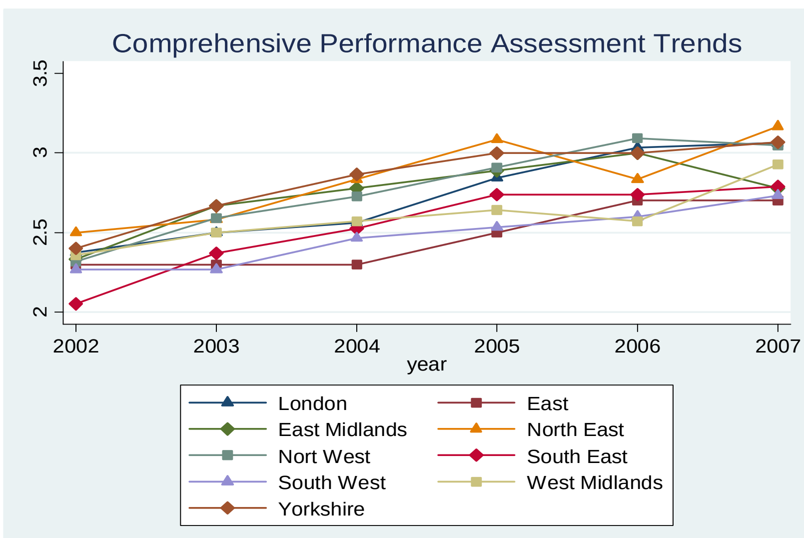


Table 7
RE and TRE Panel Estimates

Parameter	RE	TRE
β_L	0.517***	0.513***
α_1	0.698***	0.747***
α_2	0.136***	0.153***
α_3	0.030***	0.035***
γ_{1L}	0.092*	-0.024
γ_{2L}	-0.119**	-0.093**
γ_{3L}	0.036	0.053*
α_{12}	-0.067*	-0.088***
α_{23}	0.006	0.000
α_{13}	-0.020	-0.020
β_{LL}	0.321***	0.095
α_{11}	0.122**	0.016
α_{22}	0.066*	0.038
α_{33}	0.000	-0.005
δ_{COMPL}	0.037	0.053**
δ_{PR}	-0.023	-0.006
δ_{PC}	-0.074**	-0.042
δ_{VR}	-0.422***	-0.461***
δ_{VC}	-0.094	-0.141*
δ_{COMM}	-0.090	-0.118*
Year 2003	-0.041***	-0.040**
Year 2004	0.047***	0.053***
Year 2005	0.084***	0.089***
Year 2006	0.078***	0.081***
Year 2007	0.111***	0.116***
Constant	0.517***	-0.236**

Figure 3
Trends in Inefficiency

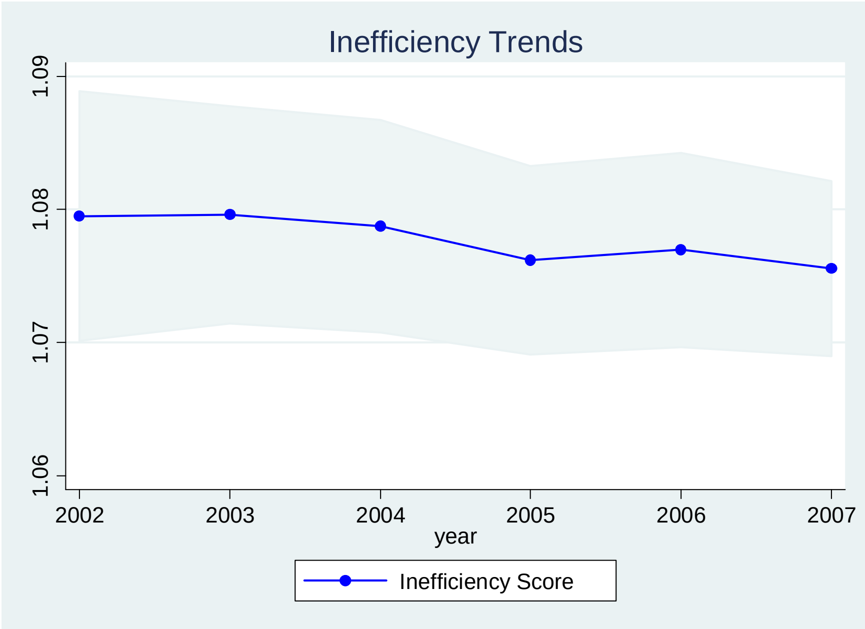


Figure 4
Inefficiency distribution

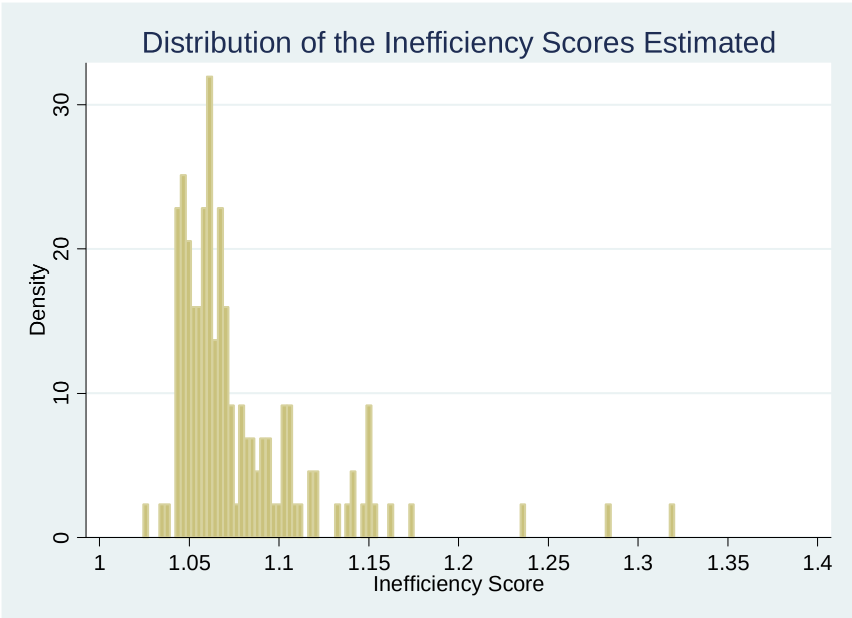


Table 8
Descriptive Statistics for Inefficiency by Year

Year	Mean	Std. Deviation	Median	Min	Max
2002	1.080	0.058	1.062	1.020	1.445
2003	1.080	0.051	1.061	1.024	1.351
2004	1.079	0.050	1.062	1.024	1.302
2005	1.076	0.044	1.062	1.025	1.367
2006	1.077	0.045	1.064	1.024	1.335
2007	1.076	0.041	1.062	1.024	1.283
Total	1.078	0.048	1.062	1.020	1.445

Table 9
Inefficiency Scores Descriptive Statistics from the Alternative Models

Model	Mean	Std. Deviation	Median	Min	Max
Cross Sectional	1.078	0.048	1.062	1.020	1.445
Random Effects	2.675	0.461	2.666	1.000	3.878
True Random Effects	1.279	0.180	1.260	1.010	1.823

Table 10
Transition matrix: Quintiles of Cost Inefficiency from 2002 to 2007

	Inefficiency Quintile Year 2007					
Inefficiency Quintile Year 2002	1	2	3	4	5	1
1	13 43.33	7 23.33	6 20.00	3 10.00	1 3.33	30 100
2	10 33.33	11 36.67	6 20.00	1 3.33	2 6.67	30 100
3	3 10.34	8 27.59	7 24.14	9 31.03	2 6.90	29 100
4	3 10.00	2 6.67	7 23.33	12 40.00	6 20.00	30 100
5	1 3.45	2 6.90	3 10.34	5 17.24	18 62.07	29 100
Total	30 20.27	30 20.27	29 19.59	30 20.27	29 19.59	148 100

Table 11
Inefficiency Descriptive Statistics by Region

Region	Mean	Std. Dev.	Median	Min	Max	N
East	1.132	0.102	1.090	1.041	1.445	60
East Midlands	1.079	0.058	1.061	1.020	1.275	54
London	1.066	0.030	1.061	1.027	1.346	192
North East	1.056	0.014	1.054	1.035	1.101	72
North West	1.066	0.032	1.057	1.031	1.193	132
South East	1.073	0.041	1.060	1.021	1.302	114
South West	1.081	0.038	1.065	1.038	1.199	90
West Midlands	1.098	0.052	1.085	1.041	1.311	84
Yorkshire	1.082	0.037	1.074	1.033	1.260	90
Total	1.078	0.048	1.062	1.020	1.445	888

Table 12
Elasticity Estimates and Returns to Scale by Region

Region	$\varepsilon_{C,y1}$	$\varepsilon_{C,y2}$	$\varepsilon_{C,y3}$	RTS
East	0.752***	0.123**	0.044**	1.088*
East Midlands	0.747***	0.109**	0.052**	1.101**
London	0.746***	0.134**	0.052*	1.073
North East	0.752***	0.096**	0.053**	1.109**
North West	0.786***	0.070	0.043	1.113**
South East	0.732***	0.141**	0.052*	1.082
South West	0.708***	0.152***	0.060**	1.086**
West Midlands	0.759***	0.096**	0.050**	1.105**
Yorkshire	0.749***	0.096**	0.055**	1.111**
Total	0.749***	0.115**	0.051*	1.094**

References

Aigner, D.J., C.A.K Lovell and P. Schmidt. 1977. “Formulation and Estimation of Stochastic Frontier Production Function Models”, *Journal of Econometrics*, 6, 21-37.

Audit Commission. 2010. *Under Pressure: Tackling the financial challenge for councils of an ageing population*, Local Government Studies, Audit Commission, London.

Audit Commission. 2006. *Briefing on the Audit Commission’s Comprehensive Performance Assessment frameworks*, Local Government Studies, Audit Commission, London.

Barlow, J., Bayer, S. and Curry, R. 2005 “Flexible homes, flexible care, inflexible organisations? The role of telecare in supporting independence”, *Housing Studies*, 20, 3, 441-456.

Battese, G. E. and Coelli, T. J., 1988. "Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data," *Journal of Econometrics*, 38(3), 387-399.

Bebbington, A. and Kelly, A. 1995. “Expenditure planning in the personal social services: unit costs in the 1980s”, *Journal of Social Policy*, 24, 3, 385-411.

Borge, L.-E. and Haraldsvik, M. 2009. “Efficiency potential and determinants of efficiency: an analysis of the care for the elderly sector in Norway”, *International Tax and Public Finance*, 16: 468–486

Bös D. 1986. *Public Enterprise Economics*, North-Holland, Amsterdam.

Brown, R. S., D. W. Caves and L. R. Christensen, "Modelling the Structure of Cost and Production for Multiproduct Firms." *Southern Economic Journal*, 46, 256-73.

Caudill, S. B. and J. M. Ford. 1993. “Biases in Frontier Estimation Due to Heteroscedasticity.”, *Economics Letters*, 41, 17–20.

Caudill, S. B., J. M. Ford and D. M. Gropper. 1995. “Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity.”, *Journal of Business and Economic Statistics*, 13, 105–111

Challis D, Clarkson P, Warburton R. 2006. *Performance Indicators in Social Care for Older People*. Ashgate: Aldershot.

Costa-i-Font, J., Wittenberg, R., Patxot, C., Comas-Herrera, A., Gori, C., Di Maio, A., Pickard, L., Pozzi, A., and Rothgang, H. 2008. “Projecting long-term care expenditure in four European Union member states: the influence of demographic scenarios”, *Social indicators research*, 86 (2).303-321.

Crivelli, L., Fillipini, M., & Lunati, D. 2002. “Regulation, ownership and efficiency in the Swiss nursing home industry”, *International Journal of Health Care Finance and Economics*, 2, 79–97.

Department of Health. 1997. *Social services achievement and challenge*, White Paper, Department of Health, London.

Department of Health. 1998. *Modernising social services promoting independence, improving protection, raising standards*, White Paper, Department of Health, London.

Department of Health. 2003. *Fair access to care services - guidance on eligibility criteria for adult social care*, Guidance, Department of Health, London.

Department of Health. 2005. *Independence, Well-being and Choice: Our Vision for the Future of Social Care for Adults in England*, Green Paper, Department of Health, London.

Department of Health. 2006. *Our health, our care, our say: a new direction for community services*, Green Paper, Department of Health, London.

Department of Health. 2010. *Prioritising need in the context of Putting People First: a whole system approach to eligibility for social care - guidance on eligibility criteria for adult social care*, Guidance, Department of Health, London.
Economic Policy Committee. 2006. *Impact of ageing populations on public spending, on pensions, health and long-term care, education and unemployment benefits for the elderly.* Brussels, 6 February 2006.
Farsi, M. and M. Filippini. 2004. "An Empirical Analysis of Cost Efficiency in Non-profit and Public Nursing Homes", *Annals of Public and Cooperative Economics*, vol. 75(3), pages 339-365, 09.
Fernandez, Jose-Luis and Forder, J. 2007. "Consequences of local variations in social care on the performance of the acute health care sector', *Applied Economics*, 1 - 16
Fillippini, M. 2001. "Economies of scale in the Swiss nursing home industry", *Applied Economic Letters*, 8, 43–46.
Hadri, K. 1999. "Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function." *Journal of Business and Economic Statistics,* 17, 359–363.
Hall, R. E., "The Specification of Technologies with Several Kinds of Outputs." Journal of Political Economy, 81, 878-92
Hougaard, J. L., Kronborg, D., & Overgård, C. 2004. "Improvement potential in the Danish elderly care", Health Care Management Science, 7, 225–235.
Hughes, M.D. 1988, "A Stochastic Frontier Cost Function for Residential Child Care Provision", *Journal of Applied Econometrics*, vol. 3(3), pages 203-14, July-Sept.
Knapp, M. 1984. *The Economics of Social Care*, Macmillan, London.
Kooreman, P. 1994. "Nursing home care in the Netherlands: a non-parametric efficiency analysis", *Journal of Health Economics*, 13, 301–316.
Kumbhakar S.C., Lovell C.A. 2000. *Stochastic frontier analysis.* Cambridge University Press
Jiménez, S.J., Chaparro, F.P. and Smith, P.C. 2003. "Evaluating the introduction of a quasimarket in community care", *Socio-Economic Planning Sciences*, 37, 1-13.
Johri, M., Beland, F. & Bergman, H. 2003. "International experiments in integrated care for the elderly: a synthesis of the evidence". *Journal of Geriatric Psychiatry*, 18, 222-35.
Laine, J., Linna, M., Häkkinen, U., & Noro, A. 2005. "Measuring the productive efficiency and clinical quality of institutional long-term care for the elderly", **Health Economics**, 14, 245–256.
Leece, D. and J. Leece. 2006. "Direct Payments: Creating a Two-Tiered System in Social Care?", *British Journal of Social Work*, 36(8):1379-1393.
Machin, S. and Wilson, J. 2004. 'Minimum Wages in a Low-wage Labour Market: Care Homes in the UK', *Economic Journal*, 114(494).
Meeusen, W. and J. van den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions With Composed Error", *International Economic Review*, 18, 435-444.
Netten, A., Williams, J. and Darton, R. 2005. "Care-Home Closures in England: Causes and Implications", *Ageing and Society*, 25: 319–38.
Nyman, J. A., & Bricker, D. L. 1989. "Profit incentives and technical efficiency the production of nursing home care", *Review of Economics and Statistics*, 56, 586–594.
Pavolini, E. and Ranci, C. 2008. "Restructuring the welfare state: reforms in long-term care in Western European countries", *Journal of European Social Policy*, Vol. 18, No. 3, 246-259
Revelli, F. 2010. "Spend more, get more? An inquiry into English local government performance", *Oxford Economic Papers*, 62, 185-207.

United States. General Accounting Office. 2002. "Long term care: aging baby boom generation will increase demand and burden on federal and state budgets", *Testimony; GAO-02-544 T*, Washington, D.C.

Vitaliano, D. F., & Toren, M. 1994. "Cost and efficiency in nursing homes: a stochastic frontier approach", *Journal of Health Economics*, 13, 281–300.

Walker, A., O'Brien, M., Traynor, J., Fox, K., Goddard, E. & Foster, K. 2002. *Living in Britain: Results from the 2001 General Household Survey*, London, HMSO.

Wang H 2002. "Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model", *Journal of Productivity Analysis*, 18:241–253.