Production Factors, Productivity Dynamics and Quality Gains as Determinants of Healthcare Spending Growth in U.S. Hospitals[1]

Juan Contreras
jutirado@gmail.com
Banco de México

Elena Patel
elena.s.patel@gmail.com
U.S. Department of
the Treasury

Ignez Tristao
ignezt@iadb.org
Interamerican
Development Bank

October 30, 2013

Abstract

We analyze the contribution of production factors to revenue growth in almost the complete universe of U.S. hospitals, accounting for quality and productivity. Production factors (capital, labor, energy, materials and drugs) contributed 70% (drugs alone contributed 52%), better health outcomes (higher quality) contributed 5%, and better use of resources (productivity) contributed 25%. We find increasing returns to scale, a markup of between 15% and 36% and a much larger productivity dispersion in the hospital sector than the one found in manufacturing, with gains coming mainly from within-hospital productivity growth and almost zero coming from net entry.

---

1 Introduction

Per-capita health care spending has grown, on average, about 5 percent per year in real terms over the past four decades in the US, which is about 2 percentage points above GDP growth in the same period.[2] This growth differential has lead to a permanent increase in the share of health care spending in the federal budget, becoming a main concern for policymakers because of its implications for fiscal sustainability and public policy.

The consensus among health economists is that new medical technologies are the main driver of the growth in health care spending; estimated contributions typically range from 38 percent to 65 percent.[3] In the aggregate, technology growth is commonly measured in health care as the residual of aggregate cost growth not explained by income growth, population aging, prices growth, administrative cost growth, changes in third party payments or defensive medicine and supplier-induced demand growth.[4]

In the health cost growth literature, however, the definition of technology is ambiguous. Besides the elements that allow a firm to produce more output with fewer resources, the term technology used in this literature also includes capital, demand shocks (embedded in the prices), new drugs and increases in health care quality. When applied to health care, not all new treatments and machines increase productivity or represent technology in an economic sense.

While technology is commonly associated with productivity, there are differences between them: Technology comprises all of the elements that allow a firm to produce more output with fewer resources (capital, labor, etc.), perhaps with a higher quality level, while productivity measures how well firms use the available technology. Consider for example the acquisition of an MRI machine by a hospital. In much the health cost growth

---

[2]While health care spending growth from 1960 to 1990 was about 5.5 percent, it slowed dramatically from 1994 to 1998, never exceeding 2.8 percent. In contrast, health care spending growth rocketed to 12.5 percent in 1999 and to 14.5 percent in 2004. Hospital spending is the largest among all other categories of total heath care spending in the U.S.: it was 31 percent in 2011, a share that has been roughly constant in the last decade, although it represented 40 percent in 1982 and slowly fell to reach 30 percent in 2003.

[3]See Newhouse (1992), Cutler (1996) and CBO (2007).

[4]The specific measures vary, but several studies attribute 2 percent to aging, between 10 and 13 percent to changes in third-party payments, up to 23 percent to personal income growth, between 11 to 22 percent to prices in the health care sector and between 3 and 13 percent to administrative costs. For more details, see "Technological Change and the Growth of Health Care Spending", CBO Paper, 2007.

literature, such machine would be counted as technology but in the macroeconomic or industrial organization literature, it would be counted as capital. The technology component of the MRI machine, if any, would be the possibility of delivering a more precise diagnosis with fewer resources (i.e. fewer physical exams). More precise diagnosis would represent the gains in quality from the new technology. The productivity component would be the way that a hospital is able to use the potential gains of the new machine to deliver those diagnosis with fewer resources.[5]

In this paper, we estimate the productivity at the hospital level of the universe of U.S. Medicare-certified institutional providers, which correspond to 95% of U.S. hospitals. Specifically, our measure of productivity is the residual of a hospital production function estimated at the micro level, with medical services as the measure of output. We decompose this term in true productivity and quality, controlling for prices.

We measure the productivity of hospitals in terms of the medical services they deliver and not in terms of health outcomes, which are considered in our quality measures. Triplett (2013) characterizes the measurement of productivity in health systems in two ways, depending on how output is defined. The first measures output as the medical services that depend on inputs like capital, labor, energy and drugs. In this case, productivity change is given by the growth of medical services when the production inputs change. The second measures health as the outcome. In this case, a *partial* measure of productivity can be calculated as the change in health given by a change in medical care, or, alternatively, as the change in health given by the change in the inputs used to produce medical care. This measure is the "productivity of the medical care resources used to improve health". This is a partial productivity measure because health depends on many other factors different from medical services. In order to measure the productivity of medical services to produce health, we would have to account for all the determinants of health to isolate the specific spending on medical services, this is, genetics, exercise, diet, lifestyle, etc. McKeown (1976) documents that most long term improvements in health can not be accounted for by medical interventions, and Ford (2007) shows that only 46.6 percent of the decline in US deaths from coronary disease between 1980 and

---

[5]We will call this term "True Productivity" in the rest of the paper.

2000 is accounted for medical/surgical treatments. Explanations of the lack of cross sectional correlation between spending and outcomes in the U.S. health care system consider the production function of health as starting point. Instead, we estimate the production function of medical services, allowing for the possibility of an efficient system in delivering medical services. In this way we can be agnostic about the productive efficiency of the delivery of health.

To illustrate our output measure, and the analysis of the production factors, productivity and quality involved in the hospital sector, consider again the example of the MRI machine. Suppose that a given hospital buys this machine to be used in the process of breast cancer diagnosis. In this example, the medical services that the hospital provides are the MRI exam itself, the diagnosis, any additional necessary exams and the different treatments that doctors apply to the patient, like radiation and chemotherapy. All those would be our measured outputs. Our inputs would be the hours spent by doctors and nurses in this example, the drugs administered to the patient, the capital involved in the diagnosis and treatment, which includes the MRI machine, the energy spent by the hospital and the medical supplies needed for the diagnosis and treatment. Our quality measure would account for the amount of days the person survives after the treatment is administered, and our productivity measure would account for the amount of services the hospital was able to provide given a certain amount of resources. Our measure of quality (survival days after treatment) could also be used also as a measure of output. However, in this case we would have to account for all the determinants of health to isolate the specific spending on medical services, this is, genetics, exercise, diet, lifestyle, etc. and the productivity concept would be different, as Triplet (2013) discussed.

To separate the effect of hospital-level prices from the productivity growth, we use the method proposed by Klette and Griliches (1996), which introduces a demand system to account for unobservable prices at the individual level. Estimating productivity of hospitals also requires a careful treatment of the prices that are charged to consumers, since an increase in medical expenditure can reflect an increase in price, an increase in health output or an increase in the amount of services provided by the hospital, the last two being true indicators of productivity.

To control for the endogeneity of productivity and input choices, we use the method proposed by Olley and Pakes (1996), which exploits the fact that investment choices are the result of a dynamic decision wherein productivity is a state variable, such that the unobserved productivity can be recovered as an inverse function of the observed investment choices.[6] We also are able to separate quality from true productivity by using direct measures of health care quality; in particular, we use data on mortality rates from the Dartmouth Atlas of Health Care.

We use data from the Health Care Provider Cost Reports, which are a set of facility level data files on the universe of Medicare-certified institutional providers from 1996 to 2009. This dataset is ideal for our study for two main reasons. First, it comprises 95 percent of all the hospitals in the US and has information on every state, county and referral region, making it representative of the entire hospital system in the US.[7] Second, this data set has information on every input each hospital uses in day to day activities, including high-quality information on capital and depreciation, labor (medical and non medical), materials (including drugs), and energy. In order to receive medicare reimbursements, providers are required to submit an annual cost report, containing information on facility characteristics, utilization data, cost, charges by cost center (in total and for Medicare), Medicare settlement data, and financial statement data. The data are available for hospitals, skilled nursing facilities, renal facilities, home health agencies, and hospices. To our knowledge, no other data set contains this level and quality of information, which allows us to estimate a hospital production function in the US.

Our results show that (i) the contribution of true productivity growth to revenue growth (whose correlation with cost growth is 0.9) is about 25 percent on average between 1996 and 2009, (ii) the contribution of better health outcomes (higher quality) is about 5

---

[6]Although productivity measurement has a long tradition in the economic literature, two problems plague these estimates. The first is the endogeneity of productivity to the input factor choice. This problem arises because the choice of production factors is not independent of the productivity level, introducing biases in the estimated parameters. The second problem is the lack of data on individual quantities and prices: when these elements are not observed, productivity estimates cannot distinguish between true productivity and demand or price variation.See for example Klette and Griliches (1996) and more recently Foster, Haltiwanger and Syverson (2008).De Loecker (2011) uses this method in the context of international trade.

[7]We define later in the paper Health Referral Regions and illustrate their importance in our identification of the effects of quality separate from productivity

percent and (iii) the contribution of the production factors is 70 percent, with capital contributing 6 percent, labor contributing 8 percent, materials contributing 4 percent and drugs contributing 52 percent.

If we define technology as it is commonly defined in the health cost growth literature, that is, by including capital investment and drugs in addition to the true productivity term, the contribution is very high, of about 82 percent. This is even larger than estimates suggested by previous studies. However, if we limit the definition of technology to true productivity, our estimate is much lower at 25 percent.

There is a temporal dimension to these gains as well. Our results show that the contribution of drugs to revenue growth represented the majority of the share until 2000, but between 2005 and 2007 this contribution has come to represent only 20 percent. The contribution of capital has been stable through the period of analysis, but the contribution of labor has come to represent almost 20 percent of revenue growth in the last years of the sample. The contribution of energy has been almost neutral with large fluctuations through the years.

Within the same empirical procedure, we also estimate a large price-cost mark-up for the hospital sector in the U.S., which is consistent with earlier findings of a highly concentrated hospital market. This suggests that the large gains in efficiency we find have not been translated into lower prices for patients, but rather have been due to factors other than the production of health, like the payment systems.

Our analysis of the dynamics of productivity in the hospital sector shows that most productivity growth comes from within hospitals, with almost no contribution from net entry or reallocation of market share towards more productive hospitals. At the same time, true productivity shows a much larger dispersion than that of other sectors in the U.S. The relationship between our measure of true productivity and quality is positive, and true productivity is also positively related to spending. This complements our analysis and gives additional support to our interpretation of the data: hospitals are very efficient in delivering more medical services that are not necessarily related to quality.

Taken together, our results are consistent with the view that technology is the main

driver of health cost growth in the US, but we go further this assertion and precisely qualify and quantify this statement. In a time series, increases in productivity and drugs are the main drivers of revenue growth, and they are part of a broad definition of technology. At this point, it is important to emphasize that, according to our definition of output, gains in true productivity are gains in efficiency in delivering medical services, but they do not necessarily imply gains in health outcomes. In fact, our proxy measure for quality (adjusted mortality) grows slowly compared to true productivity. In a cross section, our results are also consistent with a health system in which true productivity, measured as the amount of medical services delivered given a determined amount of inputs, can vary greatly across regions but is not reflected in gains in quality (survival).

This paper makes several contributions to the measurement and understanding of the role of technology and productivity in the health care sector. First, to our knowledge we are the first to measure productivity separately from technology, as is commonly defined, for almost the complete universe of U.S. hospitals (as opposed to the measure in a single medical procedure or in a specific region). Second, we are the first to measure the role of different input factors like capital, labor, medical supplies, drugs and energy in the hospital production function at the micro level and within a common framework. Having good measures of capital and materials, afforded to us by the HCRIS data, is crucial to understanding how much of "unexplained growth" in health costs is truly unexplained, rather than coming from growth in the costs of drugs or machines (like MRIs). Third, we can separately identify the contribution of factor growth, productivity growth and quality growth to the rise of health costs in U.S. hospitals. Fourth, we estimate the returns to scale for the US hospitals, a key input when analyzing the impact of mergers and prices. Fifth, we are able to estimate within this framework the elasticity of demand for hospital care and therefore the Lerner index in the US hospital industry. Finally, we further document the large dispersion in productivity across U.S. hospitals and provide evidence that productivity growth comes mainly from within US hospitals and not from a reallocation process.

This paper is related to several strands of the literature. First, it contributes to the literature that investigates the role of technology in health care cost growth. Newhouse

(1992) is the first to attempt a systematic decomposition of health care growth, and Cutler (1995) updates his estimates. Other studies that investigate the relationship between health care and productivity are Cutler (1995, 2007), Cutler and McClellan (2001), Wiesbrod (1991), Gaynor (2006), Gaynor, Kleiner and Voigt (2008), Skinner and Staiger (2009), Chandra and Skinner (2008), and Lee, McCullogh and Town (2012). Our study uses a unique data set that is representative of all U.S. hospitals and that contains detailed information about production factors. By taking advantage of this data set, we add to this literature by estimating the contribution of each production factor and defining precisely the role of technology as studied in other fields in economics. In addition, we estimate a high degree of market power in the sector by using the same econometric framework, suggesting that other factors different from technology allow hospitals to keep health care prices high.

Our approach complements previous studies that try to explain the apparent lack of cross-sectional correlation between outcomes and spending using measures of productivity of health. Chandra and Staiger (2007) argue that productivity spillovers in the adoption of technology can explain this pattern. Skinner, Staiger and Fisher (2006) offer an explanation based on heterogeneity in production functions across regions. Skinner and Staiger (2009) offer a related explanation based on the heterogeneity in technology diffusion and adoption. We try to explain aggregate patterns based not on a single treatment, but in the production of medical services at the hospital level.

Our results are also consistent with Chandra and Skinner (2012), who pose that a mix of moral hazard and principal agent is the cause of rapid cost growth in the U.S. health system. That is, because insurance is the primary payor, more services are offered and there is an incentive to increase the use of new technologies, which may or may not be worth the price. They propose a model in which health care productivity, measured in terms of health output, depends on the heterogeneity of treatment effects across patients, the shape of the health production function and the cost structure of different procedures. Although our measure of productivity depends on medical services and not on health outputs, our results fit their basic story, which also fits the description by Horwitz and Nichols (2009), who provide evidence that at least some non-profit hospitals maximize

output, while all others maximize profits.

This paper also contributes to the literature that estimates production functions in the health care sector. Some authors have estimated production functions for physicians (Reinhardt (1972)) or cost functions for hospitals (Vita (1990), Gaynor, Kleiner and Voigt, 2008, Keeler and Ying(1996), Bilodeau, Cremieux and Pierre Ouellette (2000), Hughes and McGuire (2003), and Preyra and Pink (2006)) but it has been hard to find good measures of capital stock,[8] or they have used data that is not representative of the sector. Other authors have estimated efficiency in the health care sector using stochastic frontier analysis, facing similar problems.[9] Our contribution is to estimate a production function that controls for endogeneity, that is representative of the US hospitals, and that has very good measures of all production factors, especially capital.

2   Data and Measurement

2.1   HCRIS hospital cost reports

We use annual data from 1996 to 2009 from the Health Care Provider Cost Reports (HCPCR) to estimate the hospital production function for medical services. Those reports are a set of facility level data files on the universe of Medicare-certified institutional providers from 1996 to 2009. The Centers for Medicare & Medicaid Services (CMS) collect data for all Medicare-certified institutional providers (a sample that includes 95 percent of all U.S. hospitals), which are required to submit an annual cost report after the end of each fiscal year to a fiscal intermediary. The fiscal intermediary, in turn, reports the data to the Healthcare Cost Reporting Information System (HCRIS) of the CMS. The cost reports contain provider information such as facility characteristics, utilization data, cost and charges by cost center (in total and for Medicare), Medicare settlement data, and financial statement data. The HCRIS includes subsystems for the Hospital

---

[8]Reinhardt, for example, approximates the stock of capital with measures of physician annual depreciation on furniture and equipment and his annual cost of renting or owning office. Others have used the number of beds space, for example.

[9]See, for example Hollingsworth (2003) and his references for detail, and Newhouse (1994) for a critique of this method.

Cost Reports (HCR), Skilled Nursing Facility Cost Report, Home Health Agency Cost Report, Renal Facility Cost Report and Hospice Cost Report. We use the subsystem for the the Hospital Cost Report.

The Medicare Cost Reports (MCR) have been traditionally used to determine Medicare's share of allowable costs and to provide a basis for calculating Medicare payments to providers. It has been used also by the Prospective Payment Assessment Commission to make recommendations to Congress for Prospective Payments System (PPS) changes. Our measures for output, employment, capital, materials and energy come from the MCR. We deflate all measures by the corresponding BEA price index, taking 2000 as the base year.

2.2 Output and Input Measures

Triplett (2013) describes two parallel ways of measuring output in health care. The first measure of output is medical care, which is a function of production factors (capital, labor, energy, medical supplies and drugs) and productivity, which accounts for all inputs that had not been accounted for or not fully measured. The alternative measure of output is health itself, which is a function of medical care, time, R&D and individual behaviors. The problem with this measure is that health depends on many other factors different from medical services. In order to measure the productivity of medical services to produce health we would have to account for all the determinants of health to isolate the specific spending on medical services (genetics, exercise, diet, lifestyle, etc). In addition, Triplett argues that although measuring medical care output by health outcomes is appealing, it has the potential of mixing production and productivity measurement in the health care sector with the measurement of the determinants of health.[10]

We measure output as medical services. We think that revenues reflect variation in the level of medical services produced by a hospital. Specifically, our measure of hospital

---

[10]Triplett (2001) makes the analogy of the health production process with the production function of car repairs. In the national accounts, output in the car repair sector is given by the quantity of repairs, which in theory can be adjusted for the quality of repairs. In the health care sector, however, this quality plays a much more important role in the measurement, and the price paid for medical procedures do not necessarily translate in better outcomes.

revenues is hospital charges excluding contractual discounts (i.e. list prices). Because prices are included in the revenue measures, we must add a procedure to control for those prices and to determine the associated markup. We will come back to this point in the next section.[11]

Labor is measured as the number of full time equivalent employees. Energy is measured by the reported annual costs dedicated to plant operations, which include the maintenance and service of utility systems such as heat, light, water, air conditioning and air treatment. Materials are measured as the sum of charges to patients and non-patients for medical and surgical supplies. Drugs are measured as the charges to patients and non-patients for medical drugs.

One of the most important advantages of our data set is that it allows us to construct very precise measures of capital stock by each individual hospital. Capital is constructed directly from the cost reports using a perpetual inventory method. Hospitals report capital balances at the beginning and at the end of the fiscal year, and also report acquisitions, disposals and retirements and depreciation. To start the series, in the initial period we have:

$$K_{it0} = \frac{K_{it0} + K_{it0-1}}{0.5P_{INt0} + 0.5P_{INt0-1}} \tag{1}$$

where K_{it0} and K_{it0-1} are the book values of capital for each hospital i at the beginning and at the end of the period calculated as the sum of beginning balances, purchases and donations less disposals, fully depreciated, and P_{INt} is the implicit deflator for capital formation from the BEA. The series for capital is created as follows:

$$K_{it} = (1 - \delta_{it})K_{it-1} + \frac{I_{it}}{P_{INt}} \tag{2}$$

Here, I_{it} represents investment and is calculated based on the sum of purchases, donations, and retirement of new and old capital, as defined in the HCRIS cost reports, and δ_{it} is the rate of depreciation for hospital i, calculated from the depreciation and capital stock book values such that $\delta_{it} = \frac{\text{Depreciation}_{it}}{K_{it}}$.

---

[11]Output activity includes the sum of total inpatient routine care services, ancillary services, outpatient services, home health agency, ambulance, outpatient rehabilitation providers, ASC, and hospice capital.

	Median	Mean	Std
Revenue	53,402	175,734	317,700
Capital	17,479	52,445	96,962
Labor	308	791	7,801
Energy	1,010	2,586	4,382
Medical supplies	998	4,735	9,970
Drugs	1,737	5,128	9,654
Investment	66,344	460,886	1,756,629
Hospital beds	80	137	881
Patients per year	2,346	6,015	43925
Medicare discharge share	0.45	0.46	0.19

Revenue, capital, energy, medical supplies and drugs are measured in thousands of 2000 US\$. Labor is measured as full time equivalent employees. The total number of observations (hospital-year)is 64681. $\rho(total revenues, total costs) = .9$

The data set includes roughly 66000 observations from 1996 to 2009 from about 6000 hospitals. We drop observations with negative or zero inputs (565 observations) and observations with input grow of more or less than 1000 percent from one year to the next.[12] Table 1 provides the main descriptive statistics.

The average hospital has 137 hospital beds serviced by 791 fulltime equivalent employees (FTEs). These hospitals care for 6,015 patients per year on average, 46% of which are paid for by Medicare or Medicaid. The average hospital earns \$176 million in total patient revenues.[13] This revenue measure reflects a "charge," or the list-price of a service performed. The average hospital spends per year \$2.5 million in energy, \$4.7 million in medical supplies and \$5.1. million in drugs. As in other industries, investment is highly skewed, and almost 13% of the hospital-year observations present zero investment (those are valid data points and are not dropped). The statistics show the lumpiness of investment

We use Age-Sex-Race (ASR) adjusted mortality rates by Hospital Referral Region (HRR) from the Darmouth Atlas on Health Care to construct a proxy measure for qual-

---

[12]We manually checked many of the dropped observations and most of the dropped outliers were typos, or wrong or incomplete records. In the capital series, for example, for many of the outliers the initial capital is very small because of wrong data entries.

[13]Total Patient Revenues can be found on Worksheet G-3 Line 1.

ity. HRRs are geographical regions that group together several hospitals based on a common network of hospitals accepting difficult cases. Specifically, the Dartmouth Atlas on Health Care defines a HHR by documenting where patients were referred for major cardiovascular surgical procedures and for neurosurgery. There are 306 referral regions. Geweke, Gowrisankaran and Town (2003) argue that mortality rates are not good measures of hospital quality because "hospital admission is not random and some hospitals may attract patients with greater unobserved severity of illness than others". In our case, the aggregation of mortality by referral region homogenizes the measures of quality across regions such that the admissions in the referral regions depend on the geographical location in a wide sense and the severity of the patients is the same across regions. This, however, limits our analysis to a higher level of aggregation than the hospital level, which, in turn, poses some additional challenges in terms of aggregation of revenues, outputs and inputs.[14]

3   The Hospital Production Function

Rather than estimating a hospital production function, health economists have focused on the estimation of cost functions to, for example, analyze the efficiency gains after a hospital merger. However, some studies have estimated production functions for physicians and other hospitals, while others have estimated at the national level production functions using stochastic frontier analysis.

The studies that estimate health cost functions face unique challenges, such as defining outputs, inputs, prices and fixed production factors. Moreover, these researchers must grapple with the endogeneity of outputs and inputs and the availability of data. For example, studies that estimate health or hospital production functions at the national level use production factors that depart from the traditional economic production factors, like the spending per capita or the level of education.

---

[14]Another option is to use other, more comprehensive, measures of quality from the Hospital Compare Database, created by the Centers for Medicare & Medicaid Services (CMS) and the Hospital Quality Alliance. This database is at the hospital level, includes mortality ratings and has a more comprehensive system of scoring hospitals on a variety of quality measures. However, data are available just from 2005, which prevents us to use this type of data for a long enough time series.

We depart from previous analysis in our approach by directly measuring the production function of hospitals. To our knowledge, we are the first to estimate a production function for the universe of the U.S. hospitals using more traditionally defined economic outputs and inputs. This has some important advantages. First, by using revenues we capture a measure of activity that reflects the quantity of medical services delivered to the patients. This is possible in part because our estimation procedure captures variation in activity across hospitals that is independent of variation in prices and market power. Second, we have very detailed measures of economic inputs. In particular, the HCRIS data contains a precise measure of capital (equipment and structures) by hospital, something not common in earlier studies estimating production functions. This capital measure includes the detailed information that each hospital reports about investment and depreciation adjustments as explained in the previous section.

In estimating a hospital production function, we must handle several common problems including simultaneity and selection issues and biases due to the lack of data on individual prices. More specifically, simultaneity problems arise because the choice of production factors is correlated with unobserved productivity, present in the residual term, and selection issues are generated by the entry and exit of hospitals. Moreover, while individual prices are very rarely observed, this is even more true in the health sector. When the econometrician does not control for unobserved individual prices, a bias arises, leading to a confusion between productivity and quality. We will explain this point in more detail later in the same lines as Griliches and Kettle (1996) and Foster et al. (2008), among others. Finally, when quality is unobserved, as is the case in the hospital sector, this further complicates the identification of true productivity.

We use the method proposed by Olley and Pakes (1996) to control for simultaneity and selection problems, the method proposed by Klette and Griliches (1996) to address unobserved individual prices, and we use proxy measures of quality to separately identify the true productivity term.

Consider hospital i that earns revenue, R_{it}, in period t as the result of the production of some amount of medical services Q_{it}. Hospital i uses a vector of factors $X_{it} = K_{it}, L_{it}, E_{it}, M_{it}, D_{it}$ in the production of medical services. Here K_{it} represents

capital, L_{it} represents labor, E_{it} represents energy, M_{it} represents materials, and D_{it} represents drugs. Moreover hospital i has a true productivity level of ω_{it}. We use a Cobb-Douglas production function to express this relationship such that:[15]

$$Q_{it} = K_{it}^{\alpha_k} L_{it}^{\alpha_l} E_{it}^{\alpha_e} M_{it}^{\alpha_m} D_{it}^{\alpha_d} \exp^{\{\omega_{it} + u_{it}\}} \tag{3}$$

In this equation, u_{it} is an iid disturbance representing measurement error and idiosyncratic shocks to production. In logs, with lower case letters representing the logs of the variables:

$$q_{it} = \alpha_0 + \alpha_k k_{it} + \alpha_l l_{it} + \alpha_e e_{it} + \alpha_m m_{it} + \omega_{it} + u_{it} \tag{4}$$

Estimates of this equation are biased because of the simultaneity between true productivity and input choices. For example, a hospital's choice of the number of nurses and doctors depends on how well those nurses or doctors are able to attend all patients at a given time. If they are very efficient (productive) they would use fewer nurses and doctors. Olley and Pakes (1996) propose a solution to this problem noting that if the function that links measured productivity to investment is invertible, the econometrician can use the observed investment to recover the measured productivity shocks. We use this by assuming that investment has a monotonically increasing relationship with true productivity.

Formally, investment is the solution to a dynamic programming problem in which true productivity is a state variable. The policy function is given by:

$$i_{it} = h(k_{it}, \omega_{it}) \tag{5}$$

If investment is a monotonically increasing function of true productivity, we can invert this policy function to get true productivity as a function of observable variables on the

---

[15]Olley and Pakes (1996) and De Loecker (2009) use the same production function and note that results do not differ, and that the identification of the parameters do not depend on this particular functional form.

part of the econometrician,

$$\omega_{it} = \phi(k_{it}, i_{it}) \tag{6}$$

This relationship allows us to correct for the simultaneity bias between production inputs and productivity. Further, the Olley and Pakes methodology corrects for the selection bias that occurs if the entry or exit decision of a hospital is based on a productivity shock that depends upon its size. This is, larger firms are less likely to exit in response to a small productivity shock when compared to smaller firms.

In addition, unobserved prices at the hospital level generate other biases. Kettle and Griliches (1996) and Foster et al. (2008) illustrate this point. Econometricians do not typically directly observe quantities. We are only able to observe an overall price level (measured by the BEA) and individual hospital revenue. We do not observe individual prices directly, and this is a source of bias. If we were able to observe individual prices P_{it} and revenue R_{it}, we could calculate the quantity of medical services as $Q_{it} = \frac{R_{it}}{P_{it}}$. The log transformation of the Cobb-Douglas production function that we assumed before more clearly reveals the bias induced by the unobserved individual prices by using Eq. (4) to substitute for q_{it} :

$$
\begin{aligned}
r_{it} &= q_{it} + p_{it} \\
r_{it} &= \underbrace{\alpha_0 + \alpha_k k_{it} + \alpha_l l_{it} + \alpha_e e_{it} + \alpha_m m_{it} + \alpha_d d_{it} + \omega_{it} + u_{it}}_{q_{it}} + p_{it}
\end{aligned} \tag{7}
$$

Instead of the individual price level, p_{it}, we can only observe the aggregate price index of the industry P_{It}, leading to the following estimating equation:

$$
\begin{aligned}
\tilde{r}_{it} &= r_{it} - p_{It} \\
&= \alpha_0 + \alpha_k k_{it} + \alpha_l l_{it} + \alpha_e e_{it} + \alpha_m m_{it} + \alpha_d d_{it} + \underbrace{p_{it} - p_{It}}_{\text{bias source}} + \omega_{it} + u_{it}
\end{aligned} \tag{8}
$$

Following Klette and Griliches (1996) and De Loecker (2011), we use a standard hori-

zontally differentiated product demand system to control for the lack of price information at the individual level:

$$Q_{it} = Q_{It} \left(\frac{p_{it}}{p_{It}} \right)^{\eta} \exp\left(\chi_{it} + v_{it}\right) \tag{9}$$

where η is the constant elasticity of demand, χ_{it} is hospital quality, and v_{it} represents iid demand shocks. In addition, we make the assumption that the hospital market is monopolistically competitive. This assumption is both due to convenience and to the fact that it describes "tolerably well", as Dranove and Satterthwite (2000) argue, the market for most health services. As in Klette and Griliches (1996) and De Loecker (2011), these two assumptions imply a constant mark-up over marginal cost, or Lerner Index $\left(\frac{1}{|\eta|} \right)$, for the hospital industry that we will be able to estimate.

Solving for the individual price level in the demand equation and substituting it in the production function in Eq. (4) results in an equation that can be estimated using observable variables:

$$
\begin{aligned}
\tilde{r}_{it} &= \frac{\eta + 1}{\eta} \alpha X_{it} - \frac{1}{\eta} q_{It} + \frac{\eta + 1}{\eta} (\omega_{it} + u_{it}) - \frac{1}{\eta} (\chi_{it} + v_{it}) \\
&= \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_e e_{it} + \beta_m m_{it} + \beta_d d_{it} + \beta_\eta q_{It} + \tilde{\omega}_{it} + \varepsilon_{it}
\end{aligned} \tag{10}
$$

where $\beta_j = \frac{\eta + 1}{\eta} \alpha_j$, η is the average price elasticity of demand across hospitals, $\beta_\eta = \frac{1}{|\eta|}$ $\varepsilon_{it} = \frac{\eta + 1}{\eta} u_{it} - \frac{1}{\eta} v_{it}$. X_{it} represents the group of factor to consider, which in this case is $X = k, l, e, m, d$ -capital labor, energy, materials and drugs-. Finally $\tilde{\omega}_{it} = \frac{\eta + 1}{\eta} \omega_{it} + \beta_\eta \chi_{it}$ represents measured productivity, a term that combines both productivity and quality shocks. We additionally assume that investment has a monotonically increasing relationship with both true productivity and quality, and that they can be combined in a single term in the estimation.[16] Because of this, we do not directly estimate the quality component in the Olley and Pakes procedure. Instead, we decompose the measured productivity into true productivity and quality from the estimated residual using a separate estimator.

This estimating equation highlights the biases introduced when the econometrician cannot observe individual prices. In particular, the marginal products of productive

---

[16]This assumption is confirmed later in the paper where we observe that their correlation is positive.

inputs must be rescaled by the elasticity of demand, the logged industry output, q_{It}, must be included, and measured productivity term, $\bar{\omega}_{it}$, now includes hospital quality. Fortunately, the only change in the Olley and Pakes procedure is the inclusion of industry quantity, which we measure using aggregate deflated revenues as a proxy for industry output.

4   Estimates of the Hospital Production Function

Table 2 shows the results of several specifications of the production function estimated at the hospital level. Columns (1) through (3) show the results for a production function that includes only capital and labor (KL). Columns (4) through (6) show the results for a production function including capital, labor, energy, medical supplies and drugs (KLEMD) production function. Each production function specification is analyzed using three estimators: OLS, an estimator as in Olley-Pakes and an estimator that augments the Olley-Pakes procedure by incorporating a demand specification as illustrated in the last section. We estimate a KL specification to compare the results with previous studies in other sectors, and we estimate a KLEMD specification to relax the restrictions that the KL form imposes in terms of the elasticity of output to energy and materials. In addition, the inclusion of energy can help account for different levels of capital utilization that might be correlated with energy use. The inclusion of drugs in the production function for health can illustrate the importance of the introduction of new drugs to health treatments.

Our estimates may have three main sources of bias. First, input choices are positively correlated with the unobserved productivity: a highly efficient hospital would like to choose more input factors in order to increase production and therefore earn more revenue. This is especially true for the factors that are easy to adjust, like labor. Second, larger firms are more likely to survive to lower productivity realizations. If we would take the balanced panel alone, we would have a negative correlation between capital and the unobserved disturbance term due to the selection of the surviving firms. This bias is corrected in the OP estimator by accounting for age of the hospitals in the panel. And the third source of bias is an omitted price bias when deflating by an industry-wide price

Table 2: Production Function Estimates of the Hospital Sector in the US.

	(1) OLS	(2) OP	(3) Demand‡	(4) OLS	(5) OP	(6) Demand‡
Capital	0.121	0.153	0.225	0.041	0.032	0.038
	(0.003)	(0.021)	(0.069)	(0.002)	(0.010)	(0.020)
Labor	0.844	0.822	1.144	0.302	0.284	0.352
	(0.003)	(0.028)	(0.037)	(0.003)	(0.024)	(0.025)
Energy				0.201	0.111	0.143
				(0.003)	(0.013)	(0.014)
Medical Supplies				0.060	0.084	0.099
(excluding drugs)				(0.001)	(0.006)	(0.006)
Drugs				0.429	0.531	0.585
				(0.003)	(0.014)	(0.021)
Age		0.033	0.025		0.005	-0.001
		(0.234)	(0.018)		(0.002)	(1.433)
Average quantity †			0.264			0.133
			(0.016)			(0.010)
Elasticity of demand			-3.789			-7.527
Lerner Index			1.359			1.153
# Obs	63495	50347	50347	56406	50347	50347
Scale	0.965	0.975	1.368	1.033	1.041	1.217

This table shows the estimates of a production function of hospitals using Ordinary Least Squares (OLS, columns 1 and 4), The Olley & Pakes methodology (OP, columns 2 and 5) and the Olley & Pakes methodology controlling for prices through demand (Demand, columns 3 and 6). Standard errors are in parenthesis. † The coefficient of average quantity represents the Lerner Index $L = \frac{1}{|\eta|}$, where η is the elasticity of demand. ‡ We calculate the coefficients for the OP demand as $\alpha = \beta \left(\frac{\eta}{\eta+1} \right)$ where β are the coefficients obtained directly from the estimation. See section II for more details.

index. Klette and Griliches show that using an industry wide deflator tends to create a downward bias in the scale estimate and this bias decreases with an increase in the elasticity of demand. Because the omitted price bias goes against the bias created by the endogeneity of inputs and productivity, the final result is more an empirical question than a theoretical one.

The KL specification in columns (1) to (3) shows clearly the biases that each estimator attempts to correct. By correcting for endogeneity, the OP estimator in column (2) produces a lower labor coefficient and a higher capital coefficient as is predicted by theory. By correcting for the price bias, the estimator in column (3) produces much higher coefficients for both labor and capital. This large increase in both coefficients suggest that in the hospital sector the price bias is much more important than the endogeneity bias. In

addition, not accounting for prices would produce decreasing returns to scale, when the scale estimate is considerable higher than one and equal to 1.368. The implausible coefficient for labor when correcting for endogeneity and price bias suggest that estimating a production function using only those two factors can produce misleading results.

The problem with the KL specification in terms of estimating total factor productivity is that it imposes implausible restrictions in terms of the elasticity of output to energy, materials, and in this case, also drugs. For this reason, we estimate a KLEMD specification that includes capital, labor, energy, materials and drugs as inputs (as mentioned before, the inclusion of energy can help account for different levels of capital utilization, under the assumption that the latter is positively correlated with the use of energy).

Columns (4) to (6) show the results for the KLEMD specification. A specification that uses only capital and labor overestimates the contribution of those factors to the production of health. As expected, factor elasticities for capital and labor become smaller, as energy and materials are allowed to impact output in a nonlinear fashion. When correcting for endogeneity, the OP estimator in column (5) shows smaller coefficients for capital, labor and energy, but larger coefficients for medical supplies and drugs. Once we correct for endogeneity and price bias, the results in column (6) show higher coefficients and higher returns to scale, suggesting elasticities for capital, labor, energy, medical supplies and drugs of 0.038, 0.352, 0.143, 0.099, and 0.585 respectively and a scale factor of 1.217.

Our period of analysis, 1996-2009, overlaps with a decade dominated by consolidation among US hospitals. During the 1990s, over 900 mergers and acquisitions were undertaken in the hospital industry. To this end, Gaynor and Town (2012) show that the mean Herfindahl-Herschmann Index (HHI), a metric often used by antitrust authorities to asses the level of concentration within a market, grew from 2,340 in 1987 to 3,261 by 2006. The US Department of Justice classifies a market as "highly concentrated" when the HHI exceeds 2,500.

We are able to identify the elasticity of hospital demand and the associated Lerner Index by including a CES demand system under monopolistic competition within the Olley-Pakes estimation framework to control for unobserved prices. Columns (3) and (6)

show that the elasticity of demand is between -3.8 and -7.5, which translates to a constant mark-up, or Lerner Index, of between 36 percent and 15 percent. These results provide further evidence of the considerable market power held by hospitals.

Previous estimates of the elasticity of demand for hospital care or health care have produced varied results, steadily increasing as the techniques for estimating elasticities have evolved. Regardless, our estimates of the elasticity of demand for hospital care are at the higher end of previous studies. Ringel et al. (2002) review earlier analyses attempting to estimate the elasticity of demand for health care services and relying primarily on reduced form regression analysis. They report a range of estimates between -0.04 (Cherkin et al.(1989) and Phelps and Newhouse (1974)) and -0.75 (Eichner (1998)). Feldstein (1973) estimates a price elasticity of -0.5 using data for hospitals. More recent estimates of the price elasticity in health care markets employ structural estimation techniques and suggest a much higher elasticity. Gaynor and Vogt (2003) for example, estimate a structural model of demand in the style of Berry et al. (1995) tailored to the nonprofit hospital organizational form, and they find an elasticity for hospitals of -4.85. Because their estimation, which relies on detailed micro data from California hospitals in 1995, focuses specifically on the hospital industry, rather than health care demand broadly, we believe that their estimated elasticity is most closely related to our own results. Finally, Duarte (2012) estimate elasticities of expenditures across health care services up to -2.08 in the case of psychologist visits.

A possible explanation for the high elasticity that we estimate can be found in Dranove and Satterthwite (2002). They differentiate two types of demand in the health care markets under monopolistic competition: One demand curve determines the change in quantities demanded from an individual hospital i as it varies the price from an initial price p_0, holding other hospital prices as fixed. The other demand curve determines the quantities demanded from this hospital when it moves prices in tandem with other hospitals, which describes the industry demand curve. They show that the individual demand curve is more elastic than the aggregate demand curve. We interpret our results as an estimate of the average individual demand curve, which have a much higher elasticity than the estimates of the industry demand curve present in previous studies.

We also estimate the production function at the HRR level for several reasons. First, in order to differentiate true productivity from quality, we need to use proxy measures of quality that are only available aggregated at the HRR level. Therefore, we estimate the contribution of true productivity and quality using two different methods for aggregation, one of which requires coefficients of the production function that are estimated at the HRR level of aggregation. And second, we explore the possibility that the high estimates of the elasticity of hospital demand when measured at hospital level do not hold when estimated at the HRR level because it is harder for patients to move from one region to another.

Before estimating the production function at the HRR level, we need to aggregate output and inputs because different factors may have different productivity. We measure aggregate output and inputs as the weighted sum of individual variables, using revenue shares as the weights. At the HRR level, we average the age of the firms and the number of exits per period by HRR to include in the Olley and Pakes procedure.

Estimates of the production function at the HRR level differ in several ways to the estimates at the hospital level (Table 3). First, returns to scale are lower than the estimates at the hospital level, although they are still increasing when correcting for the price bias and endogeneity. Second, capital has a higher contribution to production in this case, while labor has a lower contribution, and the contribution of drugs is roughly stable. Finally, the estimate of the elasticity of hospital demand is lower, but nonetheless is still high relative to previous estimates. The estimated lower elasticity across regions suggests that markets show a certain degree of segmentation. Interestingly, all the age coefficients are positive and statistically significant, showing an important effect of age in the production function of hospitals at the HRR level. This may come from the individual effect of entry and exit decisions.

From the production function estimates for US hospitals, we can draw several important conclusions. First, capital and labor alone cannot determine the production function of US hospitals. Not considering other factors would allocate a disproportionate contribution of those factors. In particular, drugs seem to be very important in the production for health, with a share of more than 50 percent. Second, our results suggest increasing returns

Table 3: Aggregate Production Function Estimates of the Hospital Sector in the US

	(1) OLS	(2) OP	(3) Demand‡	(4) OLS	(5) OP	(6) Demand∗
Capital	0.491	0.357	0.492	0.181	0.140	0.171
	(0.011)	(0.072)	(0.100)	(0.009)	(0.030)	(0.043)
Labor	0.377	0.485	0.704	0.128	0.110	0.158
	(0.011)	(0.080)	(0.096)	(0.009)	(0.025)	(0.037)
Energy				0.041	0.057	0.081
				(0.006)	(0.033)	(0.029)
Medical Supplies				0.040	0.050	0.055
(excluding drugs)				(0.005)	(0.021)	(0.018)
Drugs				0.534	0.553	0.634
				(0.011)	(0.042)	(0.029)
Age		0.027	0.020		0.018	0.012
		(0.005)	(0.007)		(0.003)	(0.004)
Average quantity †			0.286			0.166
			(0.080)			(0.044)
Elasticity of demand			-3.492			-6.007
Markup			1.401			1.200
# Obs	4264	4041	4041	4264	4041	4041
scale	0.868	0.842	1.196	0.925	0.911	1.099

This table shows the estimates of a production function calculated aggregating the variables across hospitals in each Health Referral Region (HRR). We use Ordinary Least Squares (OLS, columns 1 and 4), the Olley & Pakes methodology (OP, columns 2 and 5) and the Olley & Pakes methodology controlling for prices through demand (Demand, columns 3 and 6). Standard errors are in parenthesis. † The coefficient of average quantity represents the Lerner Index $L = \frac{1}{|\eta|}$, where η is the elasticity of demand. ‡ We calculate the coefficients for the OP demand as $\alpha = \beta \left(\frac{\eta}{\eta + 1} \right)$ where β are the coefficients obtained directly from the estimation. See section II for more details.

to scale in US hospitals. Third, price bias is very important, and not accounting for individual prices in the production function estimation would significantly underestimate the coefficients in the production function. And fourth, our estimated Lerner Index of between 15 and 36 percent is consistent with other evidence of considerable market power held by US hospitals.

4.1 Production Factors' Contribution to Hospital Revenue Growth

Our goal is to estimate the contribution of each factor to total revenue growth. In order to do this, we use a growth accounting methodology following Solow (1957). This well

Table 4: Average Production Factors' Contribution to Revenue Growth in US Hospitals (Productivity Accounting)

	Capital $\alpha_k \frac{\Delta K}{K}$	Labor $\alpha_l \frac{\Delta L}{L}$	Energy $\alpha_e \frac{\Delta E}{E}$	Medical Supplies $\alpha_m \frac{\Delta M}{M}$	Drugs $\alpha_d \frac{\Delta D}{D}$	Productivity $\frac{\Delta \tilde{\omega}}{\tilde{\omega}}$	Revenue $\frac{\Delta R}{R}$
1997	0.5	0.6	0.2	0.2	2.2	-0.1	3.6
1998	0.4	0.4	1.8	0.4	4.8	-3.1	4.7
1999	0.4	-0.1	-0.3	0.3	4.4	-0.8	3.9
2000	0.3	-0.5	-2.0	-0.2	1.1	5.1	3.8
2001	0.3	0.4	0.5	0.2	3.9	-0.4	4.9
2002	0.1	0.7	1.5	0.4	5.1	-0.3	7.6
2003	0.0	0.5	-1.3	0.1	3.0	1.9	4.3
2004	0.2	0.2	-0.6	-0.2	0.4	3.7	3.7
2005	0.3	0.8	-1.2	0.1	1.5	4.0	5.5
2006	0.3	2.1	0.4	0.5	2.0	2.7	8.0
2007	0.3	0.8	-0.2	0.6	1.1	2.2	4.8
2008	0.3	0.6	-1.6	-0.1	-1.3	7.0	4.9
2009	0.2	-1.7	2.7	0	1.6	-4.8	-2.0
Average	0.3	0.4	0.0	0.2	2.3	1.3	4.5
Contribution to revenue growth (%)							
	6.2	8.2	0.0	4.0	52.2	29.5	

This table shows the contribution of each production factor to revenue in U.S. Hospitals. The decomposition uses the growth accounting method of Solow (1957). It takes the production function coefficients estimated previously to weight the growth of each factor in their contribution to revenue.

known equation is:

$$\frac{\Delta R_{it}}{R_{it}} = \alpha_k \frac{\Delta K_{it}}{K_{it}} + \alpha_l \frac{\Delta L_{it}}{L_{it}} + \alpha_e \frac{\Delta E_{it}}{E_{it}} + \alpha_m \frac{\Delta M_{it}}{M_{it}} + \alpha_d \frac{\Delta D_{it}}{D_{it}} + \frac{\Delta \tilde{\omega}_{it}}{\tilde{\omega}_{it}} \qquad (11)$$

In this equation, the contribution of factor X to revenue growth is given by $\alpha_x \frac{\Delta X}{X}$, and the contribution of productivity to revenue growth is $\frac{\Delta \tilde{\omega}}{\tilde{\omega}}$. We can calculate this contribution for each hospital by using the estimated coefficients and the growth rates of all factors. Table 4 shows the average contribution of each factor to hospital revenue growth by year during the period of analysis (1997 to 2009).

The first thing to notice from Table 4 is that the average revenue growth between 1997 and 2009 is 4.5 percent, 2.1 percentage points above GDP growth during the same

period. This is consistent with the excess cost growth literature for the US. The only period of negative revenue growth was during 2009, and here only labor and productivity contributed to the fall in revenue. This suggest that hospitals may have adjusted revenue in response to the recession through labor, and especially through productivity. The decrease in productivity during the recession year is consistent with the literature on procyclical productivity. As Basu and Fernald (2001) suggest, this may reflect variable utilization and resource allocation, meaning that hospitals reacted to decreased demand by providing fewer services with the same resources and shifting resources to other activities.

Overall, capital accounts for 6.2 percent of revenue growth, labor accounts for 8.2 percent, energy does not contribute on average, and medical supplies account for 4.0 percent. Moreover, as Table 4 shows, the average contribution of capital, labor and medical supplies is relatively stable during the period of analysis. That the capital contribution does not vary much is expected because capital adjustments can take considerable time. Labor is relatively more volatile, but its contribution is also roughly stable. The same can be said about medical supplies. Energy contribution to revenue growth is highly variable, and the fall in energy usage might reflect a move towards using more energy efficient processes. However, on average it does not have any impact on revenue growth.

Drugs are the main contributor to revenue growth in the US during this period. On average, across time they account for more than half of total revenue growth in US hospitals. However, Table 4 also shows the fading importance of drugs. They accounted for most of revenue growth during the 90's, had a contribution between 20 percent to 30 percent between 2005 and 2007, and a negative contribution in 2008.

Productivity, measured as the Solow residual, accounts on average for 29.5 percent of total revenue growth, the second most important factor. Although productivity decreased between 1997 and 2002 (except in 2000), it grew at an average of 3.6 percent between 2003 and 2008. Considering the entire period of analysis, productivity increased by 1.3 percent. This result, however, may come from either gains in quality or gains in productive efficiency. We will discriminate those two sources next.

4.2 Decomposing Measured Productivity into Quality and True Productivity

We estimate the contribution of true productivity and quality to revenue growth by using a proxy variable for quality and the relationship derived in (10) between measured productivity $\tilde{\omega}$, true productivity ω and quality χ. We interpret our estimates of w as the true productivity measures, which indicates how better resources are used given constant quality of care.

The proxy we use for quality χ_{it} is a measure that considers a transformation of the adjusted mortality rates in hospitals (by age, sex and race). Two points are worth emphasizing. First, as mentioned before, mortality is not a good measure of quality because some hospitals may attract patients with more difficult cases. If this is true, hospital admission is not random. Aggregating hospitals by Health Referral Regions (HRR) addresses this problem by homogenizing the severity of the treated cases by geographical location, but we have to face the issue of aggregation. And second, we estimate the contribution of quality outside of the Olley and Pakes procedure. As mentioned before, the Olley and Pakes procedure corrects the biases caused by the endogeneity between production factors and the unobserved productivity, which contains both the true productivity and the quality measures. Estimating the contribution of quality inside the Olley and Pakes procedure would not affect our results.

We use first a simple mean of the measured productivity across hospitals by Health Referral Region to estimate the relationship between the measured productivity and the true productivity and quality. Recall that given our assumptions about the demand curve and the production function we have from the relationship derived in (10) that $\tilde{\omega}_{it} = \frac{\eta+1}{\eta}\omega_{it} + \beta_\eta\chi_{it}$. In principle, we could use this expression directly to estimate the contribution of quality to measured productivity. We do not observe, however, direct measures of quality nor we have measures of true productivity. With our measured productivity, we disentangle the contribution of each term by regressing measured productivity on our proxy measure for quality. This is, we estimate the following relationship

between measured productivity, true productivity and quality:[17]

$$\tilde{\omega}_{ht}\left(\frac{\eta}{\eta+1}\right) = \gamma_\chi \chi_{ht} + \gamma_o + \underbrace{\sum_{t=2}^{9} \gamma_t + \kappa_t}_{\omega_{ht}} \tag{12}$$

For each HRR h in period t, the true productivity component of Eqn.(12) is given by the sum of the time dummies, γ_t, and the error term, κ_t. The contribution of quality is given by $\gamma_\chi \chi_{ht}$. Our measure for quality, χ_{ht}, is the inverse of the ASR mortality rate in HRR h at time t taken from the Dartmouth Atlas on Health Care. We choose the inverse to obtain a straightforward interpretation of the coefficient: the lower the mortality rate, the higher the quality. Data availability prevents us from estimating this equation for all periods for which we have information on hospitals. The Dartmouth Atlas only provides mortality rates from 1999 to 2007. For this reason, our estimate of the contribution of quality to revenue growth applies only to this period.[18]

The estimated coefficient for the inverse of the ASR adjusted mortality rate is 1.987 (std. 0.256). This value is close to the theoretical one of 1.15, which is equal to $\frac{\eta}{\eta+1}$. The difference may contain the measurement error in previous estimations.

We next incorporate this estimate of the contribution of quality and true productivity into the productivity accounting. Table 5 expands the productivity accounting equation to incorporate those terms, and Figure 1 shows the contribution of true productivity and quality to revenue growth by year. As Table 5 shows, quality accounts for 5 percent of total revenue growth. Much of what is driving the movement in the measured productivity comes from true productivity growth, accounting for 25 percent of total revenue growth.

Quality growth has contributed relatively little to revenue growth. Although our proxy for quality is a broad measure, our results suggest that revenue growth is not a reflection of large increases in the quality of health. Instead, revenue growth is a reflection of increases in efficiency, which has likely not been passed along to patients given the relative

---

[17]We tried several specifications, including a single constant and a quadratic trend but results were very similar.

[18]This period represents about 65 percent of our sample.

Table 5: Productivity accounting including mortality (As percentage of revenue growth)

	Capital $\alpha_k \frac{\Delta K}{K}$	Labor $\alpha_l \frac{\Delta L}{L}$	Energy $\alpha_e \frac{\Delta E}{E}$	Medical Supplies $\alpha_m \frac{\Delta M}{M}$	Drugs $\alpha_d \frac{\Delta D}{D}$	Quality $\frac{\Delta \chi}{\chi}$	True Productivity $\frac{\Delta \omega}{\omega}$	Revenue $\frac{\Delta R}{R}$
Average (hospital)	6.2	8.2	0.0	4.0	52.2	4.8[†]	25.8[†]	100
Weighted Average (HRR)	15.9	6.9	2.9	5.0	55	0.7[†]	14.6[†]	100

[†] The sum of the contribution of pure productivity and quality is higher than the one estimated for the measured productivity for the period 1997 to 2009 because data availability only allows us to estimate it for the period 1999-2007 using about 60 percent of the complete dataset. The differences, however, are lower than 1%.

market power of hospitals in the industry. True productivity, however, has increased at an important rate especially after 2002, contributing almost as much as the growth in spending on drugs.

Those results are consistent with the observed pattern in the U.S. that health care spending is uncorrelated with outcomes across regions. At least in hospitals, our evidence suggest that spending has gone to more medical services that do no necessarily lead to better outcomes reflected in our measure of quality.

Using these estimates, we can define several measures of the contribution of technology to revenue growth. If we use the traditional economic definition, in which technology is associated with productivity, the contribution is 26 percent. However, we can include new machines and new drugs in the definition of technology. Using this definition, the contribution of technology growth jumps to represent 84 percent of revenue growth, with new drugs accounting for more than half this estimate. Compared with previous studies, this range is wider, but it is precisely defined in terms of production factors as productive efficiency, drugs and capital.

We also aggregate the measures of output,inputs and productivity by HRR as a robustness test and to acknowledge that the proxy measure for quality is at the HRR level.
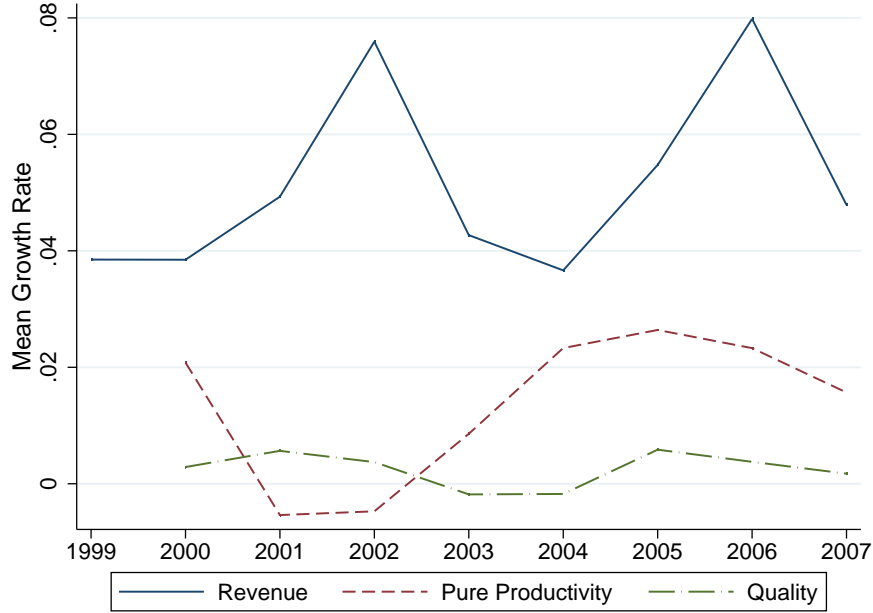
Figure 1: Contribution of True Productivity and Quality to Revenue Growth

Results are also shown in Table 5. To estimate the contribution and productivity by HRR, we use the coefficients of the production function estimated at the HRR level and the measures of inputs and outputs calculated at the HRR level. Those aggregated measures are the weighted sum of each output/input. The weights we use are revenue shares.

Consistent with the production function estimates at the HRR level, capital has a much higher contribution to revenue growth. The contribution of drugs, energy and medical supplies is also higher while the labor contribution is lower. True productivity and quality have a much lower contribution to overall revenue growth. One interpretation is that bigger hospitals have lower productivity, something that is confirmed in the data.[19]

---

[19]This correlation is positive for the US, even after controlling for HRR fixed effects.

5 Implications for Productivity, Quality and Spending

5.1 Productivity Reallocation

Which hospitals contribute more to aggregate productivity growth? We follow the literature on firm productivity and reallocation to investigate how quality and productivity have evolved during the past 15 years in the US hospitals and what have been the sources of this growth. The literature on productivity reallocation has mostly analyzed the manufacturing and retail sectors, finding that aggregate productivity growth comes mainly from within plant growth in the manufacturing sector, although net entry also plays an important role, while in the retail sector net entry accounts for most of the aggregate productivity growth (Foster, Haltiwanger and Krizian, 2006). This type of analysis has not been done in the health care sector before and very little is known about the relationship between productivity reallocation and health care cost growth.

We analyze the relationship between productivity reallocation and health care cost growth by investigating the contribution of the growth of productivity within hospitals, the contribution of net entry and the contribution of reallocation of production between hospitals. With this analysis, we can infer the contribution of each component to health care cost growth using the estimates we have obtained about the contribution of total productivity to health care cost growth.

We use two methods to analyze productivity reallocation in US hospitals. The first method is the one proposed by Foster, Haltiwanger and Krizian (2001), whose decomposition is given by:

$$\Delta\Omega_t = \sum_{e\in C} s_{et-1}\Delta\omega_{et} + \sum_{e\in C}(\omega_{et}-\Omega_{t-1})\Delta s_{et} + \sum_{e\in C}\Delta\omega_{et}\Delta s_{et}$$
$$+ \sum_{e\in N} s_{et}(\omega_{et}-\Omega_{t-1}) - \sum_{e\in X} s_{et-1}(\omega_{et-1}-\Omega_{t-1}) \tag{13}$$

where s_{et} is the share of the hospital e in the hospital industry in period t, C are continuing hospitals, N are entering hospitals and X are exiting hospitals. As a measure of market share, we use revenue shares. Ω denotes industry-wide productivity, measured as the revenue-weighted sum of individual hospital productivities ω. The first term of the decomposition represents the change of productivity within hospitals not considering changes in market shares; the second term represent the change of productivity between hospitals, this is, the change in the market share multiplied by the deviation of its productivity from the average productivity in the industry, not accounting for changes in efficiency. The third term is a covariance-type term and equal to the change in productivity multiplied by the change in market share. The other terms represent the contribution of entering and exiting hospitals to total productivity by multiplying the difference of its productivity with the average productivity by its market share. In sum, the change in aggregate productivity is decomposed in changes due to productivity enhancements within hospitals, in changes due to reallocation across hospitals, in changes due to the interaction of changes in productivity and market share and in changes due to the differentiated productivity of entering and exiting hospitals.

The second method we use is the one proposed by Olley and Pakes (1996), which is given by:

$$\Delta\Omega_t = \bar{\omega}_t - \sum_e (s_{et} - \bar{s})(\omega_{et} - \bar{\omega}) \tag{14}$$

In this decomposition, $\bar{\omega}$ and $\bar{s}$ represent the cross-sectional unweighted mean of productivity and shares. The second term reflects whether production has gone to higher productivity firms in time, this is, if there is reallocation to more (or less) productive hospitals. This decomposition is less sensitive than the previous ones to measurement error and is less sensitive to the measurement of entry and exit.

According to the results shown in table 5, productivity growth within hospitals is the main contributor to aggregate productivity growth, as is the case in other sectors of the economy. Net entry, on the other hand, contributes almost nothing to productivity growth. Moreover, aggregate productivity growth is not driven by the allocation of activity

Table 6: Sources of Productivity Reallocation in Hospitals, 1997-2008.

	total	within	between	covariance	entry	exit	Total (OP)	Average(OP) Productivity
FHK	0.013	0.011	-0.007	0.009	0.001	0.003		
OP				0.167			1.774	1.769

This table uses the method of Foster Haltiwanger and Krizian (2002) (FHK, row 1) and Olley and Pakes (1996)(OP, row2) to decompose the sources of aggregate productivity growth in the hospital sector in U.S.

to more productive hospitals, as it has been the case in other sectors of the economy. This explains why the results from the previous section about the productivity and quality growth in the health care sector do not change much with the method of aggregation, since productivity growth comes mainly from within hospitals and the reallocation term is small in comparison.

Our result that within productivity growth explains most of aggregate productivity growth in the hospital sector is consistent with the prevalent view about health care markets, in which the productivity differences across hospitals are given by idiosyncratic components.[20] In this view, health care markets are not easily affected by competitive forces and there is little scope for reallocation. A recent paper by Chandra et al. (2013), however, argues for the opposite: productivity dispersion in the health care markets is not different from that of other sectors in the economy and it is subject to competitive forces that allocate market shares based on productivity. Although we obtain similar results when we follow their strategy,[21] they are not fully comparable to their results because we use a different measure of health care output, which implies they measure productivity of heart attacks, as opposed to the productivity of hospitals or the healthcare sector. We use a comprehensive dataset that includes nearly all hospitals in the U.S. and our study is not limited to one particular health procedure.

---

[20]See, for example, Cutler (2010) and Skinner (2011)

[21]We regress market share, probabilities of exit and growth in market share on measures of productivity. Our results suggest that productivity is associated with higher market share, lower probability of exit and positive growth in market share. Calculations are available from authors upon request

5.2 Dispersion in Hospital Productivity, Health Quality and Health Spending

Our previous analysis has implications for the relationship between hospital true productivity[22], quality and health spending across U.S. regions. Given the large spending dispersion present in the health care sector as documented by the Dartmouth Atlas of Healthcare, the natural question that follows is about its determinants. We explore in this section the correlation between health care spending, true productivity, and quality. In this analysis, we take advantage of the fact that our measure of true productivity is not derived directly from the ratio of spending to outcomes in the health sector. As a measure of health spending in U.S. regions, we use total medicare reimbursements per enrolle (Parts A and B) from the Dartmouth Atlas of Health for years 2003 to 2010 (in prices of 2003).

We start by analyzing the dispersion in our measure of true productivity and quality. Figure 2 shows the normalized distribution of those variables. We observe that the dispersion of true productivity in the hospital sector is much larger than in other sectors of the economy. Syverson (2004) finds that the difference in the log-productivity between the 90th and 10th percentile plants within four-digit SIC industries in the US is 0,651 and the range's standard deviation is 0.173. We find that those numbers for hospitals in the US are 2.21 and 0.926, 3 and 5 times bigger than the numbers Syverson calculates for the US. This productivity dispersion is comparable with the dispersion in quality, and we find they are positively correlated, although the association is not as strong as the correlation with spending (Figure 3). Moreover, a regional analysis shows they do not correspond to the same Health Referral Region, a point we analyze next.

A geographical analysis (Figure 4) shows that different regions have different typologies in terms of the relationship between true productivity, spending, and quality of health. The difference of our analysis with previous research is that we are able to correlate spending and health outcomes with measures of productivity obtained independently from the simple ratio of spending to outcomes. Under that simple measure, a high productivity

---

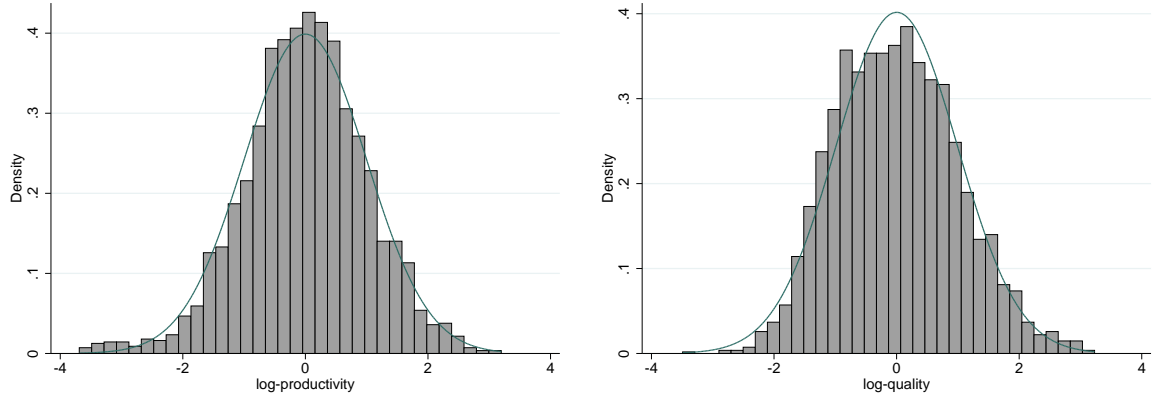[22]In this section we refer to true productivity always.

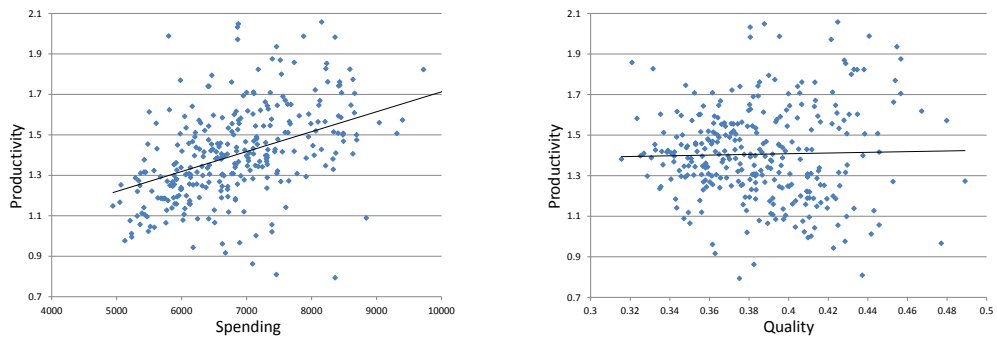Figure 2: True productivity and quality dispersion (normalized)

Figure 3: Relationship between true productivity, spending and quality

region would have low spending, and high quality. In our case, there may be regions with low spending, high quality and low productivity in the delivery of medical services. The best regions in terms of efficiency would be the ones with low spending, high quality, and high productivity, and the worst would be the ones with high spending, low quality, and low productivity. For example, across the U.S., the South Atlantic region presents low quality, high spending and average productivity; The midwest presents high quality, low spending but low productivity. Some areas of the Pacific West present high quality, high productivity but at the same time high spending

Those typologies of spending, productivity and quality complement the alternative views about the relationship between spending and outcomes. On the one hand, the "flat of the curve" explanation poses that medical interventions are done until the marginal return is zero (see, for example, McClellan, McNeil and Newhouse, 1994), suggesting that medical improvements may be realized by cutting spending in high-use regions. Other explanations emphasize that health care productivity depends on heterogeneity of treatments, the shape of the health production function and the cost structure of procedures (Chandra and Skinner, 2012). Those explanations are based in measures of productivity of health outcomes. Our explanation uses measures of productivity of medical services for the hospital sector, allowing the possibility of effectively cutting medical services that are not essential for the main treatment.

6   Conclusions

We have measured the contribution of capital, labor, energy, medical supplies, drugs, productivity and quality to revenue growth in the US hospital sector. Within a consistent framework, we were able to measure simultaneously the returns to scale and the market power of the hospital sector. As a byproduct, we also analyzed the sources of productivity growth and documented the large productivity dispersion across US hospitals. We based our estimation in a panel of 95 percent of the US hospitals spaning 19 years from 1996 to 2009, using a measure of medical services as output instead of health outcomes, which allows us to measure the productivity of hospitals in delivering medical services.
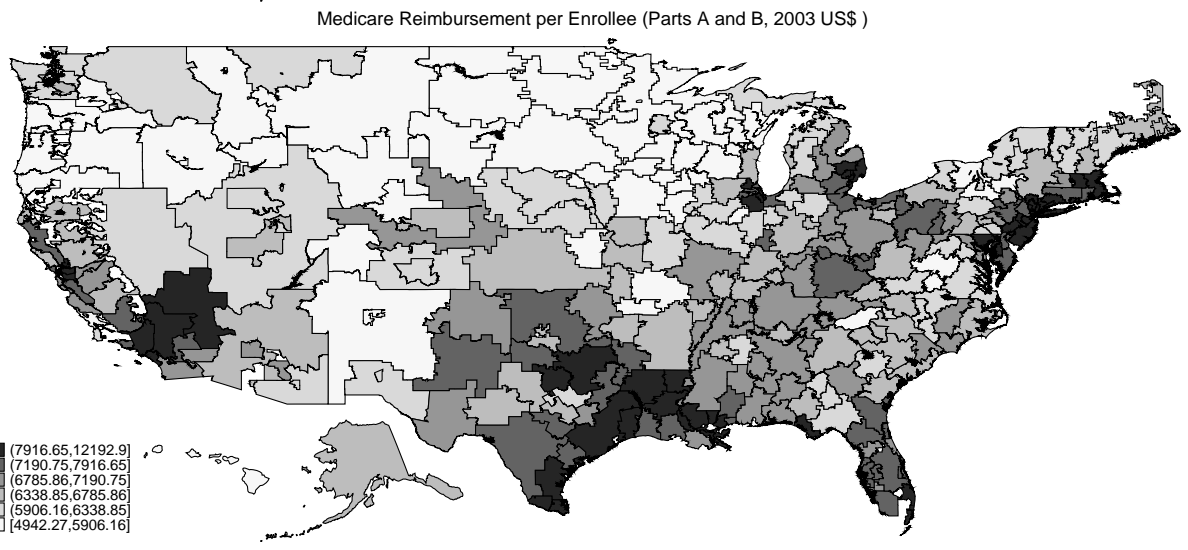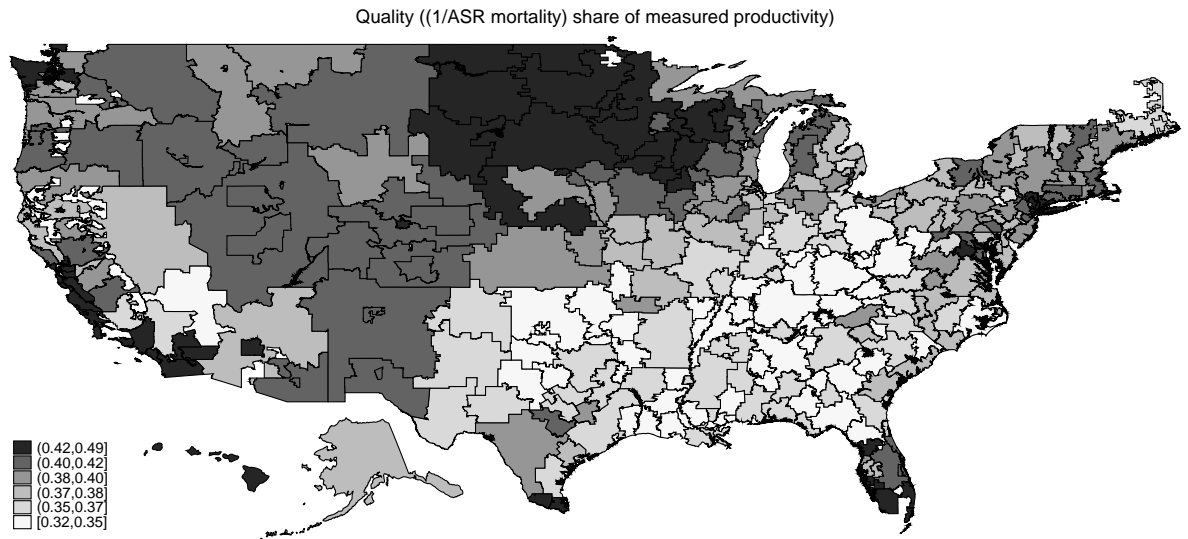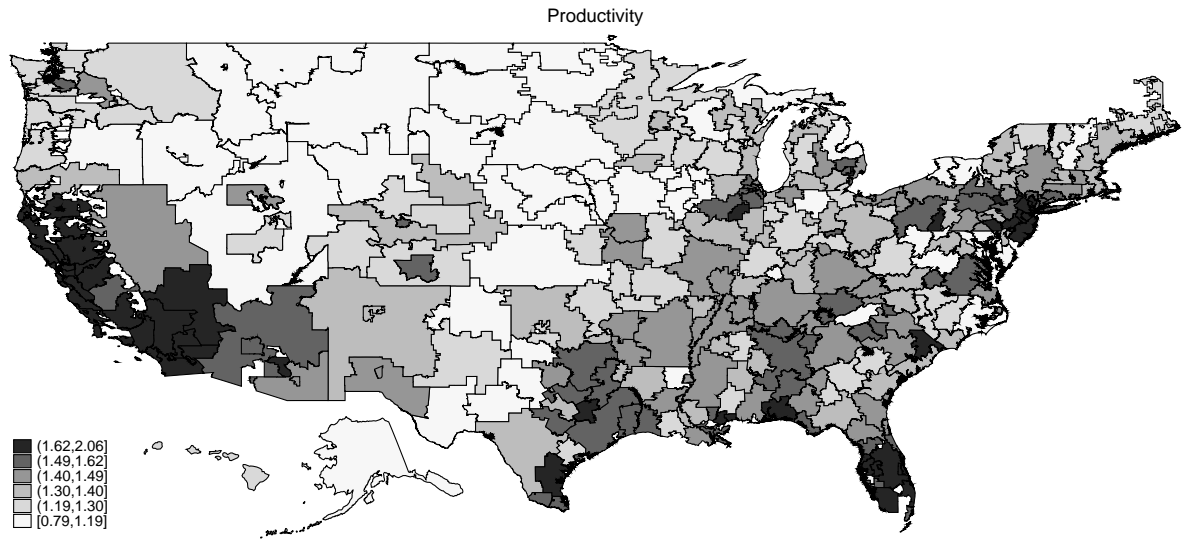
Figure 4: Geographical Distribution of True Productivity, Quality and Spending (average during 1997-2007 and 2003-2010 respectively)

Previous work suggests that technology growth accounts for between 38 percent and 65 percent of health care cost increases. Our assessment of this assertion depends on what is consider technology: If technology is limited to productivity, or the unexplained component of the production function, its contribution is 25 percent. However, if we define technology as including capital, drugs and productivity, its contribution is 82 percent. Drugs alone contribute with 52 percent of the revenue growth, while other factors contribute much less. In particular, health quality improvements account for little of the medical services growth.

Our results are consistent with the notion that the hospital sector has a considerable market power and has had large increases in efficiency in the last decade. This suggests that the increase in health care cost growth comes from factors unrelated to the production of health, like the payments system.

We observe increasing returns to scale in the US hospital sector. By controlling for individual prices using a demand system, we are able to account for the downward bias that this omission generates. In fact, without considering this fact, we observe constant or slightly decreasing returns to scale. This has important economic implications, particularly in merger analysis.

Our measured productivity growth comes mainly from within hospital reallocation. Contrary to other sectors, net entry contributes little to aggregate productivity growth. Reallocation between hospitals and reallocation of market shares to hospitals with higher productivity is not important. At the same time, our results show that productivity dispersion is much larger in the hospital sector than in other industries.

Finally, our analysis shows different geographical typologies of spending, productivity and quality that complement the alternative views about the relationship between spending and outcomes. In particular, by using measures of productivity for medical services in hospitals that are calculated independently of spending and outcomes, we can infer that allowing for the possibility of effectively cutting medical services that are not essential for the main treatment could lead to lower costs without lowering the quality of the treatments.

References

[1] Basu, Susanto, John Fernald and Matthew Shapiro, 2001. "Productivity growth in the 1990s: technology, utilization, or adjustment?," Carnegie-Rochester Conference Series on Public Policy, Elsevier, vol. 55(1), pp. 117-165.

[2] Berry, Steven, James Levinson and Ariel Pakes, 1995. "Automobile Prices in Equilibrium" Econometrica, vol. 63(4), pp. 841-890

[3] Bilodeau, Daniel, Pierre-Yves Cremieux and Pierre Ouellette, 2000. "Hospital Cost Function In A Non-Market Health Care System," The Review of Economics and Statistics, MIT Press, vol. 82(3), pp. 489-498.

[5] Chandra, Amitabh, Amy Finkelstein, Adam Sacarny and Chad Syverson, 2013. "Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector" NBER Working Paper 19200.

[5] Chandra, Amitabh and Jonathan Skinner, 2012. "Technology Growth and Expenditure Growth in Health Care" Journal of Economic Literature, American Economic Association, vol. 50(3), pp. 645-80.

[6] Chandra, Amitabh and Douglas Staiger, 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks" Journal of Political Economy, vol. 115(1), pp. 103-140.

[7] Congressional Budget Office, 2007. "Technological Change and the Growth of Health Care Spending", CBO Paper.

[8] Cutler, David, 1995. "Technology, Health Costs, and the NIH". National Institutes of Health Economics Roundtable on Biomedical Research, September.

[9] Cutler, David, 2007. "The Lifetime Costs and Benefits of Medical Technology" Journal of Health Economics, Elsevier, vol. 26(6), pp. 1081-1100.

[10] Cutler, David and Mark McClellan, 2001. "Is Technological Change In Medicine Worth It?" Health Affairs, vol. 20(5), pp. 11-29.

[11] De Loecker, Jan, 2011. "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," Econometrica, vol. 79(5), pp. 1407-1451.

[12] Dranove, David and Mark Satterthwite, 2000. "The Industrial Organization of Health Care Narkets," Handbook of Health Economics, Vol 1. Ed by A.J. Culyer and J.P Newhouse, Elsevier.

[13] Duarte, Fabian, 2012. "Price Elasticity of Expenditure Across Health Care Services," Journal of Health Economics, 31 824 - 841.

[14] Eichner, Matthew, 1998. "The Demand for Medical Care: What People pay Does Matter," The American Economic Review, May. vol. 88(2), pp. 117 - 121.

[15] Feldstein, Martin, 1973. "The Welfare Loss of Excessive Health Insurance," The Journal of Political Economy vol. 81(1), pp. 251 - 280.

[16] Foster, Lucia, John Haltiwanger and C. J. Krizan, 2001. "Aggregate Productivity Growth. Lessons from Microeconomic Evidence," in: New Developments in Productivity Analysis, pp. 303-372, National Bureau of Economic Research.

[17] Foster, Lucia, John Haltiwanger and Chad Syverson, 2008. Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability? American Economic Review, 98, pp. 394 425.

[18] Garber, Alan and Jonathan Skinner, 2008."Is American Health Care Uniquely Inefficient?," Journal of Economic Perspectives, American Economic Association, vol. 22(4), pp. 27-50.

[19] Gaynor, Martin, 2006. "What Do We Know About Competition and Quality in Health Care Markets?," Foundations and Trends in Microeconomics, December.

[20] Gaynor, Martin and William B. Vogt, 2003. "Competition Among Hospitals," RAND Journal of Economics, vol. 34(4), pp. 764 - 785.

[21] Gaynor, Martin, Samuel Kleiner and William B. Vogt, 2008. "Analysis of Hospital Production: An Output Index Approach," working paper.

[22] Geweke, John, Gautam Gowrisankaran and Robert J. Town, 2003. "Bayesian Inference For Hospital Quality in a Selection Model," Econometrica 71, 1215-1238.

[23] Hollingsworth, B., 2003. "Non-Parametric and Parametric Applications Measuring Efficiency in Health Care," Management Science, vol 6. (4), pp. 203-218(16).

[24] Horwitz, Jill and Austin Nichols, 2009. "Hospital Ownership and Medical Services: Market Mix, Spillover Effects and Nonprofit Objectives," Journal of health economics 28, pp. 924 - 937.

[25] Hughes, David and Alistair McGuire, 2003. "Stochastic Demand, Production Responses and Hospital Costs." Journal of health economics, 22 (6), pp. 999 - 1010.

[26] Keeler, Theodore and John Ying, 1996. "Hospital costs and excess bed capacity: a statistical analysis" Review of Economics and Statistics, 78, pp. 470 481.

[27] Klette, T.J. and Z. Griliches, 1996. "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," Journal of Applied Econometrics, vol 11, pp. 343 361.

[28] Jinhyung Lee, Jeffery McCullough and Robert Town, 2010. "The Impact of Health Information Technology on Hospital Productivity," NBER Working Paper No. 18025.

[29] Newhouse, Joseph, 1992. "Medical Care Costs: How Much Welfare Loss?," Journal of Economic Perspectives, American Economic Association, vol. 6(3), pp. 3 - 21.

[30] Newhouse, Joseph, 1994. "Frontier Estimation: How Useful a Tool for Health Economics?" Journal of Health Economics, vol. 13(3), pp. 317-322.

[31] Olley, Steven and Ariel Pakes, 1996 "The Dynamics of Productivity in the Telecommunications Equipment Industry," Econometrica, 64, pp. 1263 - 1298.

[32] Phelps, C.E. and J.P. Newhouse, 1974. "Coinsurance, the Price of Time and the demand for Medical Care Services," Review of Economic and Statistics 56, pp. 334 - 342.

[33] Preyra, Colin and George Pink, 2006. "Scale and scope efficiencies through hospital consolidations" Journal of Health Economics, vol. 25(6), pp. 1049 1068.

[34] Reinhardt, Uwe, 1972. "A production function for physician services. The Review of Economics and Statistics," vol. 54(1), pp. 55 65.

[35] Ringel, Jeanne, Susan Hosek, Ben Vollaard and Serjei Mohnovski, 2002. "The Elasticity of Demand for Healthcare: A Review of the Literature and Its Application to the Military Health System," RAND Monoghraph Report Series.

[36] Skinner, Jonathan, and Douglas Staiger, 2009. "Technology Diffusion and Productivity Growth in Health Care," NBER Working Paper 14865.

[37] Skinner, Jonathan, Douglas Staiger and Elliott Fisher, 2006. "Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction," Health Affairs, vol. 25(2), pp. 34-47.

[38] Syverson, Chad, 2004. "Product Substitutability and Productivity Dispersion," Review of Economics and Statistics, vol. 86(2), pp. 534 - 550.

[39] Solow, Robert, 1957. "Technical Change and the Aggregate Production Function," Review of Economics and Statistics, vol. 39(3), pp. 312 - 320.

[40] Triplett, Jack, 2001. "What's different about health? Human repair and car repair in national accounts and national health accounts" in David Cutler and Ernst Berndt (Eds) Medical Care Output and Productivity, NBER, University of Chicago Press.

[41] Triplett, Jack, 2013. "Health System Productivity" in Sherry Glied and Peter Smith (Eds) Oxford Handbook of Health Economics, Oxford University Press.

[42] Vita, Michael, 1990. "Exploring hospital production relationships with flexible functional forms" Journal of Health Economics, 6, pp. 305 318.

[43] Wiesbrod, Burton, 1991. "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment," Journal of Economic Literature, June, Vol. XXIX, pp. 523 - 552.