# Healthcare Exceptionalism?
# Productivity and Allocation in the U.S. Healthcare Sector*

Amitabh Chandra, Harvard and NBER
Amy Finkelstein, MIT and NBER
Adam Sacarny, MIT
Chad Syverson, Chicago Booth and NBER

February 2013

*Abstract:* The conventional wisdom in health economics is that large differences in average productivity across hospitals are the result of idiosyncratic, institutional features of the healthcare sector which dull the role of market forces that exists in other sectors. Strikingly, however, productivity dispersion across hospitals is, if anything, smaller than in narrowly defined manufacturing industries such as concrete. While this fact admits multiple interpretations, we also find evidence against the conventional wisdom that the healthcare sector does not operate like an industry subject to standard market forces. In particular, we find that more productive hospitals have higher market shares at a point in time and are more likely to expand over time. For example, a 10 percent increase in hospital productivity today is associated with about 4 percent more patients in 5 years. Taken together, these facts suggest that the healthcare may have more in common with "traditional" sectors than is often assumed.

*Preliminary and incomplete. Do not cite or circulate without permission. The research in this paper was conducted while Syverson was a Special Sworn Status researcher of the U.S. Census Bureau at the Chicago Census Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau. This paper has been screened to insure that no confidential data are revealed. We are grateful to Daron Acemoglu, Iain Cockburn, Mark Duggan, Joe Doyle, Matthew Gentzkow, Michael Greenstone,*

## 1. Introduction

A central observation about the U.S. healthcare sector is the existence of substantial differences in productivity across regions and across hospitals. For example, annual Medicare spending per capita ranges from $6,264 to $15,571 across geographic areas (Skinner, Gottlieb, and Carmichael 2011), yet health outcomes do not positively covary with these spending differentials (e.g. Fisher et al 2003a, 2003b; Baicker and Chandra, 2004; Chandra, Staiger, and Skinner, 2010; Skinner 2011). Similar patterns have been documented across hospitals within geographic markets (e.g., Yasaitis et al 2009). These facts have in turn generated substantial academic interest in understanding the root causes of the underlying productivity dispersion, and what can increase productivity at under-performing hospitals (Skinner, Staiger and Fisher, 2006; Chandra and Staiger, 2007; Staiger and Skinner, 2009). Outside of academia, these "Dartmouth Atlas" facts have attracted consider popular attention (see e.g Gawande's (2009) *New Yorker* article) and were heavily cited by the Obama administration during the discussions leading up to the 2010 Affordable Care Act (see e.g. Pear's (2009) *New York Times* article or the Office of Management and Budget (2009).

The conventional wisdom in health economics is that the driving force behind these large average productivity differences is various idiosyncratic, institutional features of the healthcare sector that effectively reduce competitive pressures on providers. Oft-cited culprits include: uninformed consumers who lack knowledge of the quality differences across providers, generous health insurance that insulates consumers from the direct financial consequences of their healthcare consumption decisions, and public sector reimbursement that provides little incentive for productive efficiency by providers. These are widely believed to dull the basic disciplining

force of demand-side competition that exists in most other sectors. Echoing and advancing this view, Cutler (2010) notes:

> "There are two fundamental barriers to organizational innovation in healthcare. The first is the lack of good information on quality. Within a market, it is difficult to tell which providers are high quality and which are low quality… Difficulty measuring quality also *makes expansion of high-quality firms more difficult* [emphasis added]… The second barrier is the stagnant compensation system of public insurance plans."

In a similar vein, Skinner (2011) states in his overview article on regional variations in healthcare:

> "[low productivity producers are]…unlikely to be shaken out by normal competitive forces, given the patchwork of providers, consumers and third-party payers each of which faces inadequate incentives to improve quality or lower costs…"

This "healthcare exceptionalism" view stands in contrast to a large empirical literature outside of the healthcare sector that has documented extensively – almost without exception – enormous differences in average productivity across producers within narrowly defined industries (see Bartelsman and Doms 2000, Syverson 2011 and references therein). For example, on average within narrow US manufacturing (4-digit SIC) industries, the 90$^{th}$ productivity percentile firm creates twice as much output as the 10$^{th}$ percentile firm, given the same inputs (Syverson 2004a). This dispersion exists both within and across geographic markets (e.g. Syverson 2004a,b).

We estimate that productivity dispersion across hospitals in treating heart attacks is about the same order of magnitude as productivity dispersion within narrowly defined manufacturing industries. Figure 1 (whose construction we describe in much more detail later in the paper) shows, for example, that productivity dispersion across hospitals for heart attack treatment is slightly lower than productivity dispersion across ready-mixed concrete plants. Ready-mixed

concrete is, like healthcare, a spatially differentiated good (in that it is produced and consumed locally), but one in which the product is less differentiated, insurance does not dampen price sensitivity, and prices aren't set administratively. More generally, looking across 450 different narrowly defined (4-digit SIC code) manufacturing industries in the US, average within-industry productivity dispersion in manufacturing is quite similar to our estimates across hospitals for heart attack treatment (Syverson 2004a).

This finding is striking and, we believe, surprising. But, it admits multiple possible explanations. Productivity dispersion has been shown, both theoretically and empirically, to shrink with greater competition within and across industries (e.g. Syverson, 2004a,b, Martin 2008, Balasubramanian and Sivadasan 2009). However, we would not be comfortable drawing any direct inferences about the role of competition in these two very different sectors from comparisons of their productivity dispersion.

Rather, these facts serve as a point of departure that motivates us to re-examine productivity and allocation in the healthcare sector using the analytical insights from this broader productivity literature. In particular, we draw on a long tradition of theoretical and empirical work in manufacturing examining whether higher productivity producers are systematically allocated greater market shares; in healthcare, the prevailing wisdom captured by the Cutler (2010) and Skinner (2011) quotations above – is that these re-allocation forces are weak or non-existent.

Our findings suggest otherwise. Figures 2a and 2b (again discussed in more detail later in the paper) give a qualitative flavor for our results. They show that within a market-year, higher productivity hospitals tend to have greater market share (i.e., patients) at a point in time (Figure 2a) and experience more growth over time (Figure 2b). Quantitatively, we find, for example, that

a 10 percent increase in hospital productivity is associated with about a 25 percent higher market share at a point in time and 4 percent more growth over the next 5 years. These findings suggest that market forces indeed have some self-correcting influence on productivity in health care.

Similar static and dynamic allocation metrics have been examined in a variety of different manufacturing industries. A finding that the market allocates more market share to more productive firms at a point in time and over time is a robust characteristic of US manufacturing industries (Syverson 2011 provides a recent review) but is noticeable absent from manufacturing in less competitive settings such as Central and Eastern European countries at the beginning of their transition to a market economy (Bartelsman, Haltiwanger, and Scarpetta 2009), Chile prior to trade reforms (Pavcnik 2002), or the US steel industry in the 1960s (Collard-Wexler and de Loecker 2012). As a result, these allocation metrics are often interpreted as "signposts of competition."

As in this previous work in manufacturing, we do not establish a causal link between competition and the signs of competition in the data. It could be that competitive market forces re-allocate market share to higher productivity hospitals, or that higher productivity hospitals have other features – such as nice lobbies or good managers – which separately increase demand. But whatever the driving force behind them, some force or forces in the healthcare sector lead it to evolve in a manner favorable to higher productivity producers. This finding puts US healthcare on a very different part of the map than, say, Romanian or Slovenian manufacturing in the early 1990s, where there appears to have been little (or even negative) correlation between a firm's productivity and its market share (Bartelsman et al, 2009). The results are particularly noteworthy given the context of heart attack treatments, where the acute nature of the condition might be expected to generate a smaller role for market forces in allocating patients to the more

productive hospitals than for less time-sensitive conditions such as cancer treatment, the management of chronic conditions, or elective procedures.

Taken together our results suggest that healthcare may have more in common with "traditional" sectors than is commonly recognized in popular discussion and academic research. Continued efforts to uncover what may improve productivity in the US healthcare sector may therefore benefit from greater attention to the theoretical and empirical insights from the manufacturing productivity literature. Naturally, the converse applies as well.

The rest of the paper proceeds as follows. Section 2 describes the analytical framework. Section 3 discusses our estimation of hospital productivity; this is the key empirical input to all our analyses. Section 4 presents our main results on hospital productivity and resource allocation. Section 5 discusses some questions of interpretation, including possible mechanisms behind the findings and various gauges of their magnitude. Section 6 shows that our main findings are robust to a variety of alternative specifications. A concluding section follows.

**2. Analytical approach: static and dynamic allocation**

Our primary empirical exercise examines the correlation between producer (i.e. hospital) productivity and market share at a point in time, and the correlation between producer productivity and growth in market share over time. These relationships have been examined in a variety of different industries and countries as a proxy for the role of competition in these settings (e.g., Olley and Pakes 1996; Pavcnik 2002; Escribano and Guasch 2005; Bartelsman, Haltiwanger, and Scarpetta 2009; Collard-Wexler and De Loecker 2012). Intuitively, competitive forces exert pressure on low productivity firms, causing them to either become more efficient, shrink, or exit.

Models of such reallocation mechanisms among heterogeneous-productivity producers have found applications in a number of fields, including industrial organization, trade, and macroeconomics.[1] While these models differ considerably in their specifics, they share a common intuition: greater competition – as reflected in greater consumer willingness or ability to substitute to alternate producers – makes it more difficult for higher-cost (lower-productivity) firms to earn positive profits, since demand is more responsive to their cost and price differential relative to other firms in the industry. As substitutability increases, purchases are reallocated to more productive firms, raising the correlation between productivity and market share at a point in time (*"static allocation")* and causing more productive firms to experience higher growth over time (*"dynamic allocation").* Appendix A describes this archetypical mechanism slightly more formally.

For the static allocation analysis, we will use the following regression framework:

$$\log\left(N_{h,t}\right) = \beta_0 + \beta_1 a_{h,t} + \gamma_{Mt} + \varepsilon_{ht} \quad (1)$$

where $N_{ht}$ is a measure of the market size of hospital *h* in year *t*, $\gamma_{Mt}$ are market-year fixed effects, and $a_{ht}$ is our estimate of logged TFP of hospital *h* in year *t*; we discuss in detail below how we estimate $a_{ht}$. Thus $\beta_1$ reflects the static relationship between a hospital's log TFP and its market share, within a hospital market-year. If it is positive, as has been found in many U.S. industries (e.g., Olley and Pakes 1996; Hortaçsu and Syverson 2007; Bartelsman, Haltiwanger and Scarpetta 2009), it indicates that higher productivity producers have a greater share of activity. If $\beta_1$ is zero or negative, as has been found for example in some former Soviet-bloc countries in the early 1990s (Bartelsman, Haltiwanger and Scarpetta 2009), in Chile prior to trade reforms (Pavcnik 2002), and in the U.S. Steel industry circa 1960-70 (Collard-Wexler and

[1] See, for example, Ericson and Pakes (1995), Melitz (2003), and Asplund and Nocke (2006).

De Loecker 2012), it indicates that less productive industry producers are the same size or larger than their high productivity counterparts and suggests that forces beyond standard competition are driving the allocation of market activity.

The static allocation analysis in equation (1) can reflect the market's ability to reallocate activity from less productive hospitals to more productive ones. But it shows the outcome of this process rather than the process itself. To measure the actual dynamics of the market's selection and reallocation mechanisms, we employ two additional metrics.

Our first dynamic allocation metric examines the relationship between hospital TFP and its probability of closing. We will estimate:

$$I\left[exit_{h,t+1}\right] = \beta_0 + \beta_1 a_{h,t} + \gamma_{Mt} + \varepsilon_{ht} \quad (2)$$

where $I[exit_{h,t+1}]$ is an indicator equal to one if hospital $h$ exits at time $t+1$, and the right hand side variables are defined as in equation (1). Thus $\beta_1$ reflects the relationship between a hospital's TFP and its probability of exit, controlling for any changes in aggregate exit probabilities across market-years. A negative relationship between TFP and hospital exit is one of the most robust findings in the productivity literature (See Bartelsman and Doms 2000 and Syverson 2011 for surveys). It is indicative of a Darwinian selection process at work: less productive producers find it more difficult to survive.

Our second dynamic measure is the relationship between hospital TFP and future hospital growth. We will estimate:

$$\Delta_{h,t,t+1} = \beta_0 + \beta_1 a_{h,t} + \gamma_{Mt} + \varepsilon_{ht} \quad (3)$$

where $\Delta_{h,t,t+1}$ is a measure of the hospital's growth rate (in terms of number of heart attack patients treated) between year $t$ and $t+1$. A positive correlation between TFP and growth indicates that more productive hospitals see larger gains in patient traffic, and points to the

operation of a selection and reallocation process. While not as robust as the negative TFP-exit relationship, there is widespread evidence in developed country manufacturing and retail that higher TFP producers experience growth in market shares (e.g, Scarpetta, Hemmings, Tressel, and Woo 2002; Disney, Haskel, and Heden 2003; Foster, Haltiwanger, and Krizan 2006).

Regression equations (1) through (3) form the heart of our empirical analysis. They describe the associations between a hospital's productivity and market share and indicate whether forces exist that are favorable to the expansion of higher productivity producers. Although motivated by models in which competitive forces create these re-allocation pressures, the correlations are naturally not causal evidence of the impact of competition in the healthcare sector. After presenting our results, we discuss possible uses and interpretations in light of other forces that may mimic the effects of competition.

## 3. Estimation of the hospital production function.

The key empirical input for estimation of our analytical equations (1) through (3) is a measure of a producer's (i.e. hospital's) total factor productivity, or TFP. We estimate hospital TFP in the specific context of hospital treatment of heart attacks, analyzing the treatment and outcomes of about 3.5 million heart attack patients from 1993 through 2007. TFP is the amount of output a supplier can produce per unit input. In our setting, variation in TFP across hospitals reflects differences in patient survival (output) conditional on treatments (inputs) the patient receives. We describe the data and approach we use to estimate hospital TFP, and discuss key estimation challenges.

### *3.1 Setting: Heart Attack Treatments in US Hospitals*

For a number of reasons, heart attacks present an excellent setting for studying hospital productivity. First, cardiovascular disease, of which heart attacks (AMIs) are the primary manifestation, is the leading cause of death in the United States. Second, the high post-AMI mortality (survival rates at one year are less than 70 percent in our Medicare population) provides an accurately measured outcome with a great deal of variation across hospitals. There is broad agreement that for AMIs, survival is the most important endpoint both clinically and in terms of patient preferences, and therefore a key measure of output, particularly in an elderly population.[2] Third, the emergency nature of heart attacks provides a setting in which the sorting of patients across providers is likely to be more limited than in many other healthcare settings, reducing empirical concerns arising from patients selecting into hospitals on the basis of their underlying health. At the same time, the reduced scope for sorting also makes the null hypothesis that higher productivity hospitals do not attract greater market share a particularly plausible one in this context. Finally, inputs are well measured and there exist rich data on the relevant health characteristics of the patients (called risk-adjusters) which can be used in the estimation. Not surprisingly, therefore, heart attacks have been the subject of considerable study in the medical and economics literature on the value of medical technology and the returns to medical spending (e.g. Cutler, McClellan, Newhouse and Remler, 1998; Cutler and McClellan, 2001; Skinner, Staiger and Fisher, 2006; Chandra and Staiger, 2007).

*3.2 The Hospital Production Function for AMI Patients*

We posit a patient-level health production function of the following form:

$$y_p = A_{ht}\left(\prod_k R_{p,k}^{\alpha_k} x_p\right)^{\mu} e^{\varepsilon_p}, \quad (4)$$

[2] Clinical trials for heart-attack therapies compare treatments by focusing on survival as the key outcome (see for example, Anderson et al., 2003), but this is not true for trials of treatments for more elective coronary conditions such as stable coronary disease where quality of life concerns make it difficult to measure output. A review of over twenty-three trials for heart-attack treatments is provided by Keeley, Boura and Grines (2003).

where $y_p$ is the number of post-AMI survival days of patient *p* treated at hospital *h* in year *t*, and $x_p$ is a measure of hospital inputs used to treat this patient. All production functions relate outputs to inputs; our particular function uses patient survival days as a measure of output and a single index of (dollar-weighted) resources spent on the patient as inputs.[3] Because patients are inherently heterogeneous, survival may also depend on characteristics of the patient, which could potentially also be correlated with input choices. In addition, the marginal effect of inputs on survival may vary with patient characteristics. To capture both of these effects, we follow the literature and adjust inputs for a vector of observable patient-level risk factors, $R_{p,k}$, where *k* indexes the factors. The parameters $\alpha_k$ capture the influence of these risk factors on health. Thus the expression in the parentheses reflects risk-adjusted inputs on the patient. The parameter $\mu$ is the elasticity of survival days with respect to risk-adjusted inputs. Finally, the expression $e^{\varepsilon_p}$ is a patient-level error term that accounts for random variations in health outcomes.

The key input into all of our analyses described in Section 2 is $A_{ht}$. $A_{ht}$ measures the total factor productivity (TFP) of hospital *h* in year *t*. It is common across all (risk-adjusted) patients in that hospital in that year.[4] Holding risk-adjusted inputs constant, differences in $A_{ht}$ across hospitals produces systematic differences in survival length.

The hospital production function model in (4) allows variation across providers in the marginal health product of inputs (i.e. $A_{ht}\mu$ varies across hospital-years) but constrains them to have the same elasticity of output with respect to input (i.e. $\mu$ is common across hospitals). Our

[3] This sort of single-input production function is unusual but convenient; one could reasonably interpret the single input as an index of the use of multiple inputs that go into producing health. If, conditional on a total spending level, the hospital chooses this input mix optimally, there is no loss from collapsing these inputs to a single (dollar-weighted) index. In Appendix C we show the results are robust to the use of a multi-input production function instead.

[4] We allow hospital productivity to vary across years because it allows us to capture intertemporal variation in hospitals' efficiencies and because it is consistent with standard practice in the broader productivity literature outside the healthcare sector.

empirical specification therefore allows for the possibility that the "marginal return to inputs" curve to vary across hospitals, as suggested by Chandra and Staiger (2007) and Garber and Skinner (2008). Figure 3 provides a stylized illustration of our production function specification.

Taking logs, we have our main estimating equation for the hospital production function:

$$ln(y_p) = ln(A_{ht}) + \mu \sum_k \alpha_k ln(R_{p,k}) + \mu ln(x_p) + \varepsilon_p \quad (5)$$

To estimate equation (5) we regress the log of patient survival days on a vector of risk factors $R_{p,k}$, the inputs applied to each patient $x_p$, and a set of hospital-year fixed effects. These hospital-year fixed effects are in turn our log-TFP estimates ($a_{ht} \equiv \ln(A_{ht})$) which we then use as inputs to estimate our main analytical equations (1) through (3).

*3.3 Data and Measurement of Key Variables*

Our primary dataset consists of all Medicare Part A (i.e. inpatient hospital) claims for all heart attacks (AMI) in individuals age 65 and over in the United States from 1993 through 2007. We limit the sample to AMIs in patients who have not had an admission for an AMI in the prior year. We have information on mortality through 2008, so we can observe at least one year of post-AMI survival. In order to have enough data to estimate annual hospital productivity, we eliminate any hospital-year with fewer than 5 heart attack patients that year. This restriction eliminates less than 1 percent of patients, but about 10 percent of hospital-years and 6 percent of hospitals; naturally the dropped hospitals are disproportionately small.

Table 1 presents some basic summary statistics for patients, hospitals, and markets in our sample; we will reference them as we discuss the various components of our estimation approach below. Our final sample consists of about 3.5 million heart attacks in 55,540 hospital-years and 5,346 unique hospitals. The average hospital-year has about 65 patients, but the median hospital-year has only 39 patients**.** We follow the literature in defining a hospital market (*M*) for an AMI

as a Hospital Referral Region (HRR) (see e.g. Chandra and Staiger 2007).[5] Our sample includes 304 HRRs, and on average they have about 12 hospitals in them. The Medicare claims data also include information on patient demographics (age, race and sex) and detailed information on co-morbidities (i.e. admissions for other conditions) during the prior year. We use this information as a basis of our risk factors $R_{p,k}$.

Our baseline output (survival) measure ($y_p$) is the number of days that the patient survives after receiving initial treatment, up through the first year. Survival includes the first day of treatment itself, so $y_p$ is bounded from below at 1 and above at 367 days. As shown in Table 1, average survival through 1 year, censoring anyone who survives more than 1 year at 367 days of survival, is 268 days; about two-thirds of our sample survives past one year. We show below that our core results are robust to alternative time horizons for measuring output (i.e. 30 day or 5 year survival windows).

Our baseline input measure defines hospital factor inputs for a patient as the (dollar-converted) sum of diagnostic-related group (or DRG) weights during the first 30 days following a heart attack. These DRG weights reflect the Centers for Medicare and Medicaid Services' (CMS's) assessment of the resources necessary to treat a patient as a function of the patient's comorbidities and procedures received. This approach is standard in the literature and ensures that we measure real services rendered to patients, purged of reimbursement (price) variation across geographic areas or hospitals (see e.g. Skinner and Staiger 2009, Gottlieb et al 2010). Appendix B gives a detailed description of our baseline input measure and the sources of

[5] The *Dartmouth Atlas of Healthcare* divides the United States into HRRs which are determined at the zip code level through an algorithm that reflects both commuting patterns and the location of major referral hospitals. HRRs may cross state and county borders. A complete list of HRRs can be found at http://www.dartmouthatlas.org/.

variation that contribute to it.[6] About 15 percent of the variation is explained by indicator variables for whether the patient received one of two surgical procedures, bypass or stent.

Table 1 shows that on average about $16,000 worth of hospital inputs are used on one of our patients in the 30 days following a heart attack, with a standard deviation of about $12,000. As is typical in healthcare, inputs are right skewed; the median is about $12,000 and the 90$^{th}$ percentile is nearly $32,000. We show below that our core results are generally robust across a wide range of alternative input measures, as well as across alternative time horizons for measuring inputs.

### *3.4 Estimation challenges*

Estimating productivity in any setting is conceptually straightforward but practically involves a number of measurement challenges (Syverson 2011). In addition to the measurement of output and inputs discussed above, we describe our approach to handling three other challenges to estimating the hospital production function: endogeneity of inputs, differences across hospitals in patient characteristics related to survival, and estimation error.

#### *Endogeneity of inputs*

A general econometric concern that pervades production function estimation is the potential endogeneity of inputs. In a typical setting, productivity is the residual in a firm-level regression of outputs on inputs; therefore, the coefficient on inputs ($\mu$ in our setting) may be biased by a correlation between input choice and the residual (productivity). In our setting, however, because we observe production at the unit (patient) level, we can include hospital-year

[6] As described in Appendix B, we make an adjustment to the prior literature's approach to account for the fact that some of CMS's DRGs are defined partly based on subsequent survival status. We purge our measure of this outcome-based variation in input measurement by assigning the relevant patients the average weight across the DRGs which distinguish otherwise similar treatments based on survival. We also discuss some of the challenges in measuring inputs in other settings (such as the handling of intermediate inputs or different qualities across workers) that we avoid here, as well as shared challenges such as the appropriate weighting of different inputs.

fixed effects, estimating $\mu$ solely from within-hospital variation in observables. By identifying the coefficients on inputs only from variation within hospitals, we control for any tendency for hospitals with different productivity to use different amounts of inputs on average. Of course, any unobserved inputs that do not vary within the hospital (such as, for example, whether the hospital requires its staff to use checklists) will load onto our estimate of hospital productivity. This is not a problem per se; as in the productivity literature more broadly, we think of productivity as the component of output that cannot be explained by observed inputs.

However, our estimates will be biased if, within-hospital-year, hospitals choose different observable input levels for patients who differ unobservably in their latent survival, or if their choice of unobservable inputs is correlated with observed inputs at the patient level. The sign of the bias of the estimate of $\mu$ is not obvious. Moreover, our focus is not on estimating $\mu$. Our primary concern is what impact any bias in $\mu$ will have on our analysis of the relationship between estimated productivity and market share, which are the ultimate objects of interest for the analysis. We therefore evaluate below the robustness of our main results to imposing, rather than estimating, various values for the scale parameter $\mu$. This method amounts to following the index number, or Solow residual, approach to measuring productivity in which factor elasticities are taken from auxiliary data such as factor cost shares. We are re-assured that our main results are quite insensitive to the choice of $\mu$. We discuss below the economic interpretation of this insensitivity.

*Differences across hospitals in patient characteristics*

A related issue is the concern of patient selection of hospitals. Even if $\mu$ is known and imposed based on auxiliary information, if patients at different hospitals differ on average in their unobserved survival probabilities, this variation will cause us to mis-estimate hospital

productivity. As noted earlier, one of the reasons for the focus on heart attacks in the empirical literature is the belief that such patient sorting across hospitals may be less of an issue in an emergency setting. Yet the imperfect ability of patients to choose their hospitals cannot eliminate the concern entirely; indeed, were there no mechanisms by which patients (or their surrogates) actively selected hospitals for AMI treatment, it would be hard to understand any findings that suggested that competitive market pressures were operative in this setting.

Therefore, to try to minimize the impact of any unobserved patient-health differences across hospitals, we follow the standard practice in the literature and include various risk adjusters ($R_{p,k}$) to control for patient factors that are related to health. In particular, our baseline specification controls for a full set of interactions between age (in five-year groupings), gender, and whether the patient is white, as well as various co-morbidities. Each co-morbidity is included as an indicator for whether the patient has been to the hospital for a specific condition in the year prior to the AMI admission. Table 1 shows that on average our patients are 78 years old (recall our sample is for the Medicare population), about half are female, and about 90 percent are white; it also presents the means for the 17 co-morbidities we include in our baseline specification. We show below that our main results are quite insensitive to using fewer risk adjustors.

*Estimation error in TFP measures*

In our sample, the median hospital-year has less than 40 patients, and for 20 percent of our hospital-years we observe less than 15 patients. The consequence of a relatively small number of patients in some hospital-years, together with the stochastic nature of our outcome (survival), means that our key object of interest and input into all of our productivity metrics – hospital log TFP $a_{ht}$ – may be estimated with error. Such estimation error will cause attenuation

bias in our analysis of the relationship between market share and hospital productivity in equations (1) through (3). [7]

We therefore apply the standard shrinkage or "smoothing" techniques of the empirical Bayes literature (e.g. Morris, 1983) to adjust for estimation error in our estimates of hospital productivity.[8] Appendix D provides a detailed description of this procedure. The intuition behind it is that when a hospital's productivity is estimated to be far above (below) average, it is likely to be suffering from positive (negative) estimation error. Therefore, the expected level of productivity, given the estimated productivity, is a convex combination of the estimate and the mean of the underlying productivity process. The relative weight that the estimate gets in this convex combination varies inversely with the noise of the estimate (which is based on the standard error of the hospital-year fixed effect). In practice, as we show in Appendix D, our core finding that hospitals with higher estimated productivity get allocated more market share at a point in time and over time remains statistically significant without the empirical Bayes adjustment, although naturally the magnitude is attenuated.

*3.5 Estimates of the Hospital Production Function*

Table 2 presents our estimates of the "returns to scale" parameter ($\mu$) from estimating equation (5). Column 1 presents our baseline estimates, which use our full set of risk-adjusters. We estimate a coefficient on log patient inputs ($\mu$) of 0.446 (standard error = 0.005), which suggests that every 1 percent increase in inputs per patient is associated with a 0.45 percent

[7] This small-sample problem is probably much less of an issue in more traditional settings for estimating productivity, since the number of units of output produced (the statistical analog of patients in our context) is much larger. Increasingly, however, the productivity literature is also trying to adjust for other sources of measurement error in output (e.g. Collard-Wexler, 2011, Dobbelaere and Mairesse, 2013).

[8] McClellan and Staiger (1999) introduced this approach into the healthcare literature when estimating quality differences across hospitals, and it has since been widely applied in the education literature for estimating and analyzing teacher or school value added measures (e.g. Kane and Staiger 2001, Jacob and Lefgren 2007).

increase in survival days. A comparison of columns 1 through 3 indicates that our estimate of $\mu$ increases from 0.45 to 0.59 as we reduce the set of risk adjusters to just age, race and sex (column 2) or to nothing (column 3), with the age-race-sex risk adjustment accounting for most of the change. Our estimates of $\mu$ are roughly in the middle of the (very wide) range of estimates that papers in this literature have produced.[9]

The key input into our productivity metrics is not our estimate of $\mu$ but rather our estimates of log TFP, $a_{ht}$. These objects are the hospital-year fixed effects from equation (5) and are the key right hand side variables in our estimating equations (1) through (3). As a validity check on whether these estimates are picking up differences in hospital productivity, we verify that our measures correlate positively with observable measures of hospital quality. This exercise is in the spirit of Bloom and Van Reenen (2007), who perform the reverse validation exercise: validating an observable measure of management quality by correlating it with estimates of firm level productivity. The quality measures that we use were first collected by the Center for Medicare and Medicaid Services (CMS) in 2003; they have been publicly reported by the agency's "hospital compare" website (www.hospitalcompare.hhs.gov) since 2005. They are calculated by hospitals and submitted to CMS independently of the data that we use.

These measures, which CMS refers to as scores, are presented as the fractions of patients appropriate for treatments who receive that treatment. We focus on the hospital's score for beta blockers, which are inexpensive drugs that reduce the demands on the heart and are long-

[9] Skinner and Staiger (2009) note that various papers have used different right hand side specifications or sample periods to produce estimates of the "return to spending" in a within-hospital linear probability model of one year survival on one year inputs; their own estimates range from -0.015 to 0.122 from regressions of one year survival on one year resources. In our data the comparable estimates are 0.072 to 0.100. Within-hospital estimates of the return to input use tend to produce a positive relationship between inputs and survival, in contrast to the cross-region comparisons that tend to find no or negative association between inputs and health-related outcomes. One parsimonious explanation for this difference would be if input use were a substitute for TFP.

established as having important benefits for AMI patients after discharge. We also look at a combined score that sums across the number of patients who are given each of eight treatments and divides by the sum of patients appropriate for each of these treatments.[10] All of these measures have been studied in the literature and are considered indicative of good quality care (e.g. Higashi et al., 2007, Skinner and Staiger 2009, Jha et al. 2005, and cites therein).

We find that our productivity measures positively covary with these observable quality measures. Table 3 reports the results from regressing a z-score of a hospital's observable measures of quality of care for AMI patients on our estimate of the hospital's log productivity. We report results both for the first year the data were available and for the hospital-years 2003-2007.[11] Reassuringly, the results indicate a positive and statistically significant correlation between these "external" measures of the quality of AMI care and our estimates of hospital productivity; on average, we find that a 10 percent increase in estimated productivity is associated with about a 0.12 to 0.19 standard deviation increase in the quality score.

## 4. Main results: Static and dynamic resource allocation

Table 4 presents our central results on the static and dynamic allocation of patients across hospitals. In our discussion, we focus on column 1, which presents our baseline estimates based

[10] The eight measures are 1) given aspirin at arrival, 2) given aspirin at discharge, 3) given ACE inhibitor for left ventricular systolic dysfunction (LVSD), 4) given smoking cessation advice/counseling, 5) given beta blockers at arrival, 6) given beta blockers at discharge, 7) given fibrinolytic medication within 30 minutes of arrival, and 8) given percutaneous coronary intervention (PCI) within 90 minutes of arrival.

[11] In both cases, the estimates of hospital-year log productivity come from our full sample estimates of equation (2). We separate the first year of data from later years because it is possible that once the scores were reported and easily accessible, hospitals had more of an incentive to improve them (despite no direct financial incentives to improve), and therefore their predictive power for other aspects of hospital quality may attenuate. Consistent with this concern, the distribution of scores is more compressed in later years than in the first reported year.

on the full set of risk adjusters (i.e. the same specification as shown in Table 2, column 1); the results are not sensitive to the choice of risk adjustors (columns 2 and 3).

The first row shows our static allocation analysis based on estimation of equation (1), examining the correlation between a hospital-year's log TFP, $a_{ht}$, and the number of heart attack patients it treats, $\log(N_{ht})$. Because we include market-year (HRR-year) fixed effects, this estimate is within market-year, relating a hospital's market share of heart attack patients to its TFP level relative to other hospitals in its market. Our right-hand side measure of $a_{ht}$ is the estimate of log TFP from estimation of the hospital production function in equation (5). We bootstrap the standard errors, clustering at the market level.

The results show a statistically significant positive relationship between productivity and market share, suggesting that within markets, more market share (patients) tends to be allocated to more productive hospitals at a point in time. In particular, our baseline estimate suggests that a 10 percent increase in a hospital's productivity is associated with about a 25 percent increase in market share.[12] A visual presentation of the results is given in Figure 2(a).

The second row shows our analysis of the TFP-exit relationship based on estimation of equation (2), which examines the within market-year relationship between a hospital's log TFP $a_{ht}$ and an indicator variable for whether the hospital "exits" next year. The regression's right-hand side and standard errors are calculated as in the static allocation analysis. We define the dependent variable $I[exit_{h,t+1}]$ equal to one if hospital $h$ has less than 5 heart attack patients in

[12] Our sample is limited to hospital-years with at least 5 patients, raising potential concerns about selection on the dependent variable in the static analysis. (This is not a concern for the subsequent dynamic analysis). We explored the sensitivity of our static allocation results to an alternative, Tobit-style truncated regression and found that the static allocation results were slightly strengthened by this adjustment.

each year from year *t+1* to *t+5*.[13] We measure exit as the lack of more than 5 patients in each of five subsequent years to try to ensure that we've captured a "permanent" reduction in volume, as opposed to measurement error stemming from idiosyncratic fluctuations in the number of patients that a hospital receives.

We find a statistically significant negative relationship between hospital productivity and subsequent exit. The baseline results suggest that a 10 percent increase in hospital productivity within a market-year is associated with a statistically significant decline in the probability of exit next year of about 0.3 percentage points (about an 8 percent decline relative to the baseline exit rate of 4.4 percent).

The bottom row of Table 4 shows our analysis of the TFP-growth relationship based on estimation of equation (3), which examines the within market-year relationship between a hospital's log TFP ($a_{ht}$) and its subsequent one-year growth. The right-hand side and standard errors are calculated as in the prior analyses. For our left-hand-side measure of the hospital's one-year growth rate $\Delta_{h,t,t+1}$ we define

$$\Delta_{h,t,t+1} = \frac{N_{h,t+1} - N_{h,t}}{\frac{1}{2}\left(N_{h,t+1} + N_{h,t}\right)}$$

where $N_{h,t}$ is one again the number of heart attack patients treated by hospital *h* in year *t.* Our measure of the hospital's one-year growth rate thus divides the change in the number of patients between this year and next year by the average number of patients across these two years.[14]

[13] There are a non-trivial number of hospital mergers over our time period. If hospital A merges with hospital B and physically shuts down, hospital A is coded as having 0 patients in subsequent years. If however, hospital A and B both continue to exist physically and admit their own patients (e.g. Beth Israel and Deaconess), they continue to be coded as separate hospitals with each still assigned the AMI patients whom they admit.

[14] This monotonic transformation of the standard percentage growth rate metric bounds growth between -2 (exit) and +2 (growth from an initial level of 0). An attraction of this transformation is that it reduces the chance that the results are skewed by a few fast-growing but initially small hospitals that would have very large percentage growth

Once again, the estimates are statistically significantly different from zero. The baseline results suggest that a 10 percent increase in hospital productivity within a market-year is associated with over a 1 percent increase in the number of patients the hospital treats in the next year.[15] Figure 2b gives a visual presentation of this relationship between hospital productivity and growth.

## 5. Interpretation and Discussion

### *5.1 Mechanisms*

The above findings indicate that more productive hospitals have statistically significantly higher market share at a point in time, and are more likely to increase that market share over time. These findings contrast with the conventional wisdom – summarized in the introductory quotations – that there is little in the healthcare sector to encourage the growth of higher productivity providers or weed out lower productivity ones. And they place US healthcare, at least qualitatively, in the same part of the spectrum as US manufacturing, and distinct from many less competitive manufacturing settings where these relationships have been found to not exist or even to have the opposite sign.

What mechanisms might act to allocate more patients to higher productivity hospitals in an emergency setting like heart attacks? A definitive answer is beyond the scope of this paper. However, we present some suggestive, initial evidence by examining whether the positive relationship between productivity and market share is primarily driven by a positive output-market share relationship or a negative input-market share relationship. Figures 4a and 4b show

rates. This growth rate transformation has been used in other contexts to avoid unnecessary skewness in the growth rate measure; see, for example, Davis, Haltiwanger, and Schuh (1996).

[15] Table 4 reports negative average annual growth; this is primarily due to the fact that our measure conditions on the hospital initially being in the market.

the within market-year correlation between, respectively, risk-adjusted survival and market share (conditional on risk adjusted inputs) and risk-adjusted inputs and market share (conditional on risk adjusted survival).[16] The results suggest that the productivity-market share relationship is primarily driven by the relationship between risk-adjusted survival and market share. The correlation between risk-adjusted survival and market share (Figure 4a) is virtually the same as that between risk-adjusted productivity and market share in Figure 2a. The correlation between risk-adjusted inputs and market share (Figure 4b) is less than half the size. In other words, patients and their surrogates appear to seek out hospitals that achieve higher (risk-adjusted) survival (conditional on risk adjusted inputs) rather than ones that use fewer (risk-adjusted) inputs (conditional on risk-adjusted survival). In practice, we find that risk adjusted survival and risk adjusted productivity are extremely highly correlated; consistent with this, as can be seen from Figures 4a and 4b, the residual variation in risk-adjusted inputs (conditional on risk-adjusted survival) is much smaller than the residual variation in risk-adjusted survival (conditional on risk adjusted inputs).

It is not immediately obvious how patients know which hospitals offer longer survival. This ambiguity is not unique to our study. Indeed, a long-standing question in the field – dating back at least to Arrow (1963) – is how patients can acquire information on provider quality. One possibility is some form of market-learning; hospitals acquire a reputation for good outcomes and this reputation spreads through physicians' professional networks and patients' social networks and influences patients, family members, physicians, and ambulance drivers to request treatment at hospitals that are better at producing survival. In a related setting, Johnson (2011)

[16] As with our productivity estimates, we use an Empirical Bayes correction to adjust our estimates of risk-adjusted survival and of risk-adjusted inputs for measurement error; our procedure accounts for the correlation in measurement error between these two objects.

finds that cardiac specialists who have higher risk-adjusted survival rates for their patients are less likely to stop practicing; she interprets this and related evidence as consistent with a model of market learning by the referring physician.

Alternatively, the correlation between productivity and market share could reflect omitted factors that independently drive demand and correlate with productivity. For example, higher productivity hospitals might also have nicer lobbies, which in turn influence hospital demand, or better managers might improve both the production process and separately increase demand for the hospital.

For many economic and policy questions, the mechanism by which market share is allocated to higher productivity firms is quite important. In our setting as well as in the prior work in manufacturing, more work is needed to establish to what extent the observed signs of competition are the direct result of competition or the result of other factors that are correlated with both productivity and demand.

*5.2 Magnitudes*

But the exact mechanism is less important for forecasting whether and to what extent the market is evolving in a manner that favors higher productivity firms. Here, the quantitative importance of the productivity-market share relationships we estimate becomes important. In the remainder of this section, we provide a variety of ways to shed some light on these magnitudes.

To begin, we investigate how a hospital's productivity correlates with its within-market growth and exit over longer horizons than the one-year horizon examined in Table 4. Specifically, we re-estimate equations (2) and (3) replacing the dependent variables $I[exit_{h,t+1}]$ and $\Delta_{h,t,t+1}$ with $I[exit_{h,t+k}]$ and $\Delta_{h,t,t+k}$, respectively.

Table 5 shows the results. The first row shows results one year out (i.e. the results from Table 4, where $k$=1), and the subsequent rows show results up to 10 years out ($k$=10). The relationship between productivity and growth or exit strengthens (in absolute value) over time. For example, a 10% increase in hospital productivity is associated with about 1 percent more patients next year, 4 percent more patients in 5 years, and almost 6 percent more patients in ten years.[17]

Another way to provide a sense of magnitude is to report the results in terms of the market re-allocation associated with a standard deviation change of productivity. Table 6 reports our estimates of the standard deviation of productivity dispersion, as well as several other dispersion measures. Appendix D provides more details on how these statistics were computed. The results are quantitatively stable across alternative sets of risk adjustors.

Our baseline estimate of the national standard deviation of hospital log productivity is 0.17. Thus a hospital that has one standard deviation higher log productivity has about 40 percent higher market share at a point in time, and grows about 6 percent more over the next five years. On the other hand, variation in hospital productivity accounts for little of the variation in market share. We estimate a partial $R^2$ on log productivity in the static allocation regression (1) of about 5 percent, and in the growth regression (3) of about 0.06 percent.

Another way to provide a sense of the magnitudes of these relationships is by comparing them to those in other industries. We therefore produced estimates of the static and dynamic allocation analyses for the ready-mixed concrete sector, a physically homogenous product. Details on the data, estimation and results can be found in Appendix E. Like healthcare, concrete

[17] Because our data on growth and exit ends in 2007, as $k$ rises, a smaller sample of hospital-years is available for these analyses. We verified that the findings that these relationships strengthen over time also holds (with quite similar magnitudes) if we restrict our sample to productivity estimates for hospital-years prior to 1998 (not shown).

is consumed and produced locally, so that spatial differentiation (i.e. physical distance) can be an important barrier to competition. Otherwise, however, concrete lacks many of the features deemed to be important impediments to competition in healthcare; prices are not set administratively, consumers are likely well informed about their choices, and they bear the financial consequences of their decisions.

Across all of our static and dynamic allocation measures, the results indicate a stronger (often an order of magnitude larger) relationship between producer productivity and market allocation for hospitals than for concrete plants. Likewise, Figure 1 showed that national productivity dispersion appears larger for concrete than for hospitals; we estimate a standard deviation of 0.25 in concrete, compared to 0.17 for hospitals.[18]

This comparative finding is not limited to concrete. Productivity dispersion in other U.S. manufacturing industries also tends to be similar to (indeed, somewhat smaller than) our estimates for healthcare. [19]

[18] We follow the tradition of the existing productivity literature and compute productivity dispersion metrics at a nationwide (within-year) level, even though the market for treating heart attacks is (like many of the manufacturing industries studied) plainly local. This standard practice arose in part because manufacturing industries, the focus of the previous literature, are often geographically broad. But the literature has also typically reported nationwide numbers even for those industries that are more locally oriented, such as ready-mix concrete (Syverson 2004b), in part because geographic differentiation is itself one of the possible causes of productivity dispersion within an industry. In practice, we find within-market year dispersion to be only slightly lower (standard deviation about 0.16) than our national dispersion estimate. Put another way, we estimate that about 88 percent of the within-year variation in hospital productivity is within (rather than across) markets. For concrete, we estimate that about 70 percent of the variation in productivity is within market.

[19] Compared to our estimate of a standard deviation of hospital productivity of 0.17, Foster, Haltiwanger and Syverson (2008) estimate an average within-industry standard deviation of productivity of 0.22 across a dozen manufacturing industries in the US selected for having physically homogeneous products (e.g. white pan bread, block ice, raw sugar cane, etc.), and Bartelsman, Haltiwanger and Scarpetta (2009) estimate an average within-industry standard deviation of or the average estimate of 0.38 across a broader range of manufacturing industries. Across 450 different narrowly defined (4-digit SIC code) US manufacturing industries, Syverson (2004a) estimates an average within-industry interquartile range of logged plant productivity of 0.29, compared to our estimate in Table 6 of 0.23 for hospitals. Although most of the work in productivity dispersion has focused on the manufacturing sector, the more limited work on productivity dispersion in service industries suggests that in general it is roughly similar to that found in manufacturing. For example, Fox and Smeets (2011) estimate productivity dispersion in four Danish service industries and four Danish manufacturing industries and find generally comparable estimates. Similarly, looking at 4-digit retail industries, Foster, Haltiwanger and Krizan (2006) estimate an average

We are not the first to perform such cross-industry comparisons in productivity dispersion. For example, looking across narrowly defined manufacturing industries, Syverson (2004a) finds that the extent of within-industry productivity dispersion is negatively correlated with proxies for the amount of substitutability or competition across firms within that industry. We caution, however, against drawing inferences about the extent of competition in such different settings as heart attack treatment and manufacturing from comparisons of productivity dispersion. Basic measurement differences – such as differences in the output definition (survival vs. revenue), how inputs are measured, and estimation error – raise real comparability concerns, albeit without creating a clear direction of bias.[20] Moreover, as noted earlier, the causal force behind reduced dispersion is unclear, and may well not be competitive pressure.

Nonetheless, at a broad level, the comparison may serve as a useful benchmark against which to assess the quantitative relationships we have estimated for productivity and allocation in the US healthcare sector. They also seem inconsistent with the conventional wisdom that the variations in inputs across areas and hospitals without concomitant output gains are unique to healthcare and must therefore result from idiosyncratic features of the sector.

## 6. Robustness

interquartile range for logged labor productivity which is comparable to Syverson (2004a)'s estimate of the interquartile range for logged labor productivity in manufacturing.

[20] To take but one example, the extent of measurement error in output – which would serve to attenuate estimates of the correlation between productivity and market share and to increase estimated dispersion – is likely different in healthcare than in manufacturing, although the sign of the difference is unclear. On the one hand, measurement error may be smaller in our setting since survival using death records is likely measured with less error than revenue in the Census of Manufactures and we observe a direct measure of output (survival) rather than a proxy for it in the form of revenue (P*Q) which has been shown to be problematic in some cases (Foster, Haltiwanger, and Syverson 2008 and 2012). On the other hand, in manufacturing industries output is more-or-less a deterministic function of inputs, while output (survival) in our setting is stochastic, which should work to create more measurement error, especially given the relatively small number of patients in some hospital-years. As discussed, we use the Empirical Bayes "shrinkage" estimator to try to adjust for this stochastic element and relatively small sample size within hospitals.

We explored the robustness of our findings along a number of dimensions and were generally quite reassured by the results. We have already showed that our core results are robust to our choice of risk adjustment. Here we briefly describe some of our remaining robustness analyses concerning the measurement of inputs, the measurement of output, and the potential endogeneity of inputs. Although our focus is on the robustness of the static and dynamic allocation measures, for completeness we also show the robustness of the dispersion measures from Table 6.

We face several key choices with the construction of our input measure. One is how coarsely or finely to measure inputs. There is a tradeoff between our relatively coarse baseline measure of inputs (with its associated measurement error stemming from input variation that we do not capture) and more granular measures which suffer from potential survivorship bias (a patient cannot receive many procedures if she does not survive very long); we experimented with considerably more granular input measures based on the individual procedures received and the length of hospital stay. We also explored using these inputs directly in a multi-input production function rather than aggregating them to a single index as in our baseline approach. Finally, our baseline measure follows standard practice and defines inputs based only on hospital inpatient treatments, thereby excluding physician inputs – which may occur both inside and outside the hospital – and other outpatient inputs. We tried an alternative input measure that incorporates non-hospital inputs. Again there is a trade-off; some non-hospital inputs may be closely linked (or indeed part of) the care received in the hospital, others may be quite distinct. These

alternative input measures are each described in more detail in Appendix B and the general robustness of the results is discussed in Appendix C (particularly Table A3). [21]

Another issue concerns the time horizon over which we measure inputs and outputs. Our baseline measures use a 30 day window for inputs and a 1 year window for output (survival days). We explored the robustness of our results to shorter and longer time horizons – 7 days and 1 year on the input side, and 30 days and 5 years on the output side. Again, there are tradeoffs in the length of time horizon.[22] Appendix C (and particularly Table A4) shows the general robustness of our results to these alternative input and output horizon windows.

Finally, as noted earlier, a pervasive concern in the productivity literature is the potential endogeneity of inputs to producer productivity. This can bias the estimates of the returns to scale parameter $\mu$. There is a wide range of estimates of this parameter in the literature (see e.g. Cutler et al. 1998, Fisher et al. 2003b, and Baicker and Chandra 2004) and uncertainty as to the "right" estimate. We are therefore reassured that our main static and dynamic allocation results are quite robust to imposing (rather than estimating) a range of "reasonable" values of $\mu$ and then calculating productivity under different imposed values; the dispersion estimates are also robust.

[21] Estimation in more traditional settings must also deal with input measurement problems, including issues we do not confront here stemming from differential qualities across types of workers and capital, trying to capture the flow of capital services using measures of capital stocks, and intermediate inputs typically measured by expenditures rather than quantities. Additionally, and more directly to the issue here, these inputs must also be aggregated to a single-dimensional input index by weighting the individual inputs appropriately; the theoretically correct weights are the elasticities of output with respect to the respective inputs. Estimating these elasticities involves its own set of measurement challenges. Our approach in the hospital sector avoids many of these additional issues.

[22] On the input side, a shorter time horizon will miss some of the resources the patient receives, while a longer horizon creates greater scope for survival bias as well as the issue that treatments are increasingly linked to providers other than the original hospital. On the output side, for our baseline measure we chose the relatively standard 1-year horizon since it seemed substantively more of interest than shorter-term (e.g. 30 day) survival. Analysis of a shorter horizon might capture aspects of hospital productivity that reflect only a slight postponement in death, and might not capture aspects that affect outcomes through long-term mechanisms such as the management of complications due to co-morbidities and the quality of the hospital's follow-up care. On the other hand, with a longer output horizon there is greater scope for the impact of non-hospital factors – such as patient compliance in terms of diet, smoking and medication, and the impact of doctor quality regardless of whether the doctor was associated with the initial hospital – on our productivity estimates.

These results are shown in Table 7. We impose a $\mu$ of 0.1, 0.3, and 0.9. The lack of sensitivity of our static and dynamic allocation results to alternative values of $\mu$ is consistent with the results in Figures 4a and 4b that the correlation between market share and estimated productivity is driven primarily by the correlation between market share and risk-adjusted survival.[23]

## 7. Conclusion

This paper has examined the relationship between productivity and market allocation in healthcare, specifically for hospital treatment of Medicare patients' heart attacks. We have done so by drawing on the insights of several decades of theoretical and empirical work in productivity more broadly. Qualitatively, we find that higher productivity hospitals have greater market share at a point in time, and grow more over time. Quantitatively, a hospital with a one standard deviation higher log productivity has about 40 percent higher market share at a point in time, and grows about 6 percent more over the next five years.

These relationships, which are driven primarily by the relationship between risk-adjusted survival and market share, mean that over time the healthcare market evolves in a manner favorable to higher productivity producers. This qualitative pattern is generally viewed by the broader productivity literature as an empirical sign of the workings of competition; it has been consistently found within manufacturing industries in the United States but not in less competitive settings such as post-Soviet Eastern block countries or Chile prior to trade reforms.

[23] Referring back to the basic estimating equation for hospital log productivity (equation (5)), the fact that the market share-productivity covariance is not sensitive to $\mu$ must mean that there is little variance in risk-adjusted inputs and/or a low covariance between risk-adjusted inputs and market share – otherwise, changes in the value of $\mu$, which ties risk-adjusted input variation to our estimate of hospital's productivity levels, would change the correlation between estimated productivity and market share.

Our more speculative quantitative comparisons between healthcare and manufacturing industries in the US suggest that, if anything, these re-allocation results are stronger, and dispersion similar or smaller, in healthcare.

Taken together, our qualitative and quantitative findings suggest that the healthcare sector may not be as idiosyncratic as the conventional wisdom has claimed. In this sense, our results are in the same spirit as Skinner and Staiger's (2007) finding of a common "innovativeness factor" across healthcare and other sectors within a geographic area; they found that areas of the country that were early adopters of hybrid corn in the 1930s and 1940s were also early adopters of beta blockers for heart attacks at the beginning of the current century.

Such findings suggest that, going forward, research on the determinants of productivity in the health care sector may benefit from more attention to the insights, both theoretical and empirical, from research on other industries about productivity and allocation. By the same token, insights from the health care sector may likewise be a useful laboratory for thinking about other industries. A recent series of papers by Bloom, Van Reenen and co-authors have begun to do just this, empirically investigating the role of such factors as management style and labor quality on hospital performance (usually survival rates; see Bloom et al., 2010, Propper and Van Reenen (2010), and Bloom et al., 2012).

Of course, a given amount of re-allocation to higher productivity producers – or a given improvement in this re-allocation process – may be much more valuable in healthcare than in manufacturing, not to mention of greater consequence for public sector budgets. Moreover, in our healthcare setting as in the manufacturing setting more broadly, the estimated re-allocation relationships stop far short of indicating what economic or policy forces could be unleashed to create still greater reallocation to higher productivity producers. We see a great opportunity for

further work that tries to estimate the causal impact of competition – or other factors – on resource allocation in healthcare and in manufacturing settings.

## References

Andersen HR, Nielsen TT, Rasmussen K, et al.. 2003. "A comparison of coronary angioplasty with fibrinolytic therapy in acute myocardial infarction," *New England Journal of Medicine* 349: 733-742.

Arrow, Kenneth. 1963. "Uncertainty and the Welfare Economics of Medicare Care." *American Economic Review*. 53(5). 941-973.

Asplund, Marcus, and Volker Nocke. 2006. "Firm Turnover in Imperfectly Competitive Markets." *Review of Economic Studies*, 73(2): 295–327.

Baicker, Katherine and Amitabh Chandra. 2004. "Medicare Spending, the Physician Workforce, and the Quality of Healthcare Received by Medicare Beneficiaries." *Health Affairs*, April, 184-197.

Balasubramanian, Natarajan and Jagadeesh Sivadasan. 2009. "Capital Resalability, Productivity Dispersion and Market Structure." *Review of Economics and Statistics*, 91(3), 547-557.

Bartelsman, Eric J., and Mark Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature*, 38(3): 569–94.

Bartelsman, Eric J., John Haltiwanger, and Stefano Scarpetta. 2009. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." National Bureau of Economic Research Working Paper 15490

Bloom, Nicholas and John Van Reenen. 2007. "Measuring and Explaining Management Practices Across Firms and Countries." *Quarterly Journal of Economics* CXXII (4): 1351-1408.

Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen. 2012. "Management Practices Across Firms and Countries." NBER Working Paper # 17850.

Bloom, Nicholas, Carol Propper, Stephan Seiler, John Van Reenen. 2010. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." NBER Working Paper #16032.

Chandra, Amitabh and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy*, 115(1). 103-140.

Chandra, Amitabh, Douglas O. Staiger, and Jonathan Skinner. 2010. Saving Money and Lives, in The Healthcare Imperative: Lowering Costs and Improving Outcomes. Institute of Medicine. Available at http://www.ncbi.nlm.nih.gov/books/NBK53920/

Collard-Wexler, Allan. 2011. "Productivity Dispersion and Plant Selection in the Ready-Mix Concrete Industry." Center for Economic Studies Working Paper No. 11-25.

Collard-Wexler, Allan and Jan De Leocker. 2012. "Reallocation and Technology: Evidence from the U.S. Steel Industry." Griswold Center for Economic Policy Studies Working Paper No. 230.

Cutler, David M. and Mark McClellan, 2001. "Is Technological Change in Medicine Worth It?" *Health Affairs* 20, no. 5: 11–29.

Cutler, David M., Mark McClellan, Joseph P. Newhouse and Dahlia Remler. 1998. "Are Medical Prices Declining? Evidence from Heart Attack Treatments." *Quarterly Journal of Economics* 93(4): 991 – 1024.

Cutler, David M. 2010. "Where are the Healthcare Entrepreneurs." *Issues in Science and Technology.* July 27(1): 49-56.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge: MIT Press.

Disney, Richard, Jonathan Haskel, and Ylva Heden. 2003. "Restructuring and Productivity Growth in UK Manufacturing." *Economic Journal*, 113(489), 666-94.

Dobbelaere, Sabien and Jacques Mairesse. 2013. "Panel Data Estimates of the Production Function and Product and Labor Market Imperfections." *Journal of Applied Econometrics*, 28(1). 1-46.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62(1): 53–82.

Escribano, Álvaro and J. Luis Guasch. 2005. "Assessing the Impact of the Investment Climate on Productivity Using Firm-Level Data: Methodology and the Cases of Guatemala, Honduras, and Nicaragua." World Bank Policy Research Working Paper No. 3621.

Fisher, Elliott S., David Wennberg, Theresa Stukel, Daniel Gottlieb, F.L. Lucas, and Etoile L. Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 1:The Content, Quality and Accessibility of Care" *Annals of Internal Medicine* 138(4): February 18, 2003. 273-287

Fisher, Elliott S., David Wennberg, Theresa Stukel, Daniel Gottlieb, F.L. Lucas, and Etoile L. Pinder. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care" *Annals of Internal Medicine* 138(4): February 18, 2003. 288-299

Fox, Jeremy and Valerie Smeets. 2011. "Does Input Quality Drive Measured Differences in Firm Productivity?" NBER Working Paper 16853.

Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2006. "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s." *Review of Economics and Statistics*, 88(4): 748–58.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98(1): 394–425.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2012. "The Slow Growth of New Plants: Learnings about Demand?" NBER Working Paper 17853.

Garber, Alan M. and Jonathan Skinner. 2008. "Is American Health Care Uniquely Inefficient?" *Journal of Economic Perspectives*, 22(4). 27-50.

Gawande, Atul. 2009. "The Cost Conundrum." *The New Yorker*, June 1, 2009. Available at http://www.newyorker.com/reporting/2009/06/01/090601fa_fact_gawande

Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan S. Skinner, and Jason M. Sutherland. 2010. "Prices Don't Drive Regional Medicare Spending Variations." *Health Affairs*, 29(3), 1-7.

Higashi, Takahiro, Neil S. Wenger, John L. Adams, Constance Fung, Martin Roland, Elizabeth A. McGlynn, David Reeves, Steven M. Asch, Eve A. Kerr, and Paul G. Shekelle. 2007. "Relationship between number of medical conditions and quality of care," *New England Journal of Medicine* 356 (24): 2496-2504.

Hortaçsu, Ali and Chad Syverson. 2007. "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices." *Journal of Political Economy*, 115(2) 250-301.

Jacob, Brian A. and Lars Lefgren. (2007). "What Do Parents Value in Education? An Empirical Investigation of Parents' Revealed Preferences for Teachers." *Quarterly Journal of Economics*. 122(4): 1603-1637.

Jha, Ashish K., Zhonghe Li, E. John Orav, and Arnold Epstein. 2005. "Care in U.S. Hospitals – the Hospital Quality Alliance Program." *New England Journal of Medicine* 353(3), 265-274. July 21.

Johnson, Erin. 2011. "Ability, learning, and the career path of cardiac specialists." MIT Mimeo. Available at http://www.mit.edu/~erinmj/files/Draft_042011.pdf

Kane, Thomas J. and Douglas O. Staiger. 2001. Volatility in School Test Scores: Implications for Test-Based Accountability Systems. UCLA Graduate School of Public Policy Studies, Working paper.

Keeley, Ellen C., Judith A. Boura, and Cindy L. Grines. 2003. "Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials," *Lancet* 361:13-20.

Martin, Ralf. 2008. "Productivity Dispersion, Competition and Productivity Measurement." CEP Discussion Paper 0692.

McClellan, Mark and Douglas O. Staiger. 1999. "The quality of healthcare providers." NBER Working Paper 7327.

Melitz, Marc J. 2003. "The Impact of Trade on Intraindustry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Morris, Carl. 1983. "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association* 78(381): 47-55.

Office of Management and Budget. 2009. "McAllen Redux" OMB Blog. June 4. Available at http://www.whitehouse.gov/omb/blog/09/06/04/McAllenRedux/

Olley, Steven G. and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263-97.

Pavcnik, Nina. 2002. "Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants." *Review of Economic Studies*, 69(1), 245-76.

Pear, Robert. 2009. "Health Care Spending Disparities Stir a Fight." *New York Times*, June 8. Available at http://www.nytimes.com/2009/06/09/us/politics/09health.html

Propper, Carol and John Van Reenen. 2010. "Can Pay Regulation kill? The impact of labor markets on hospital productivity" *Journal of Political Economy* (2010), 118(2), 222-273,

Scarpetta, Stefano, Philip Hemmings, Thierry Tressel, and Jaejoon Woo. 2002. "The Role of Policy and Institutions for Productivity and Firm Dynamics: Evidence from Micro and Industry Data." OECD Economics Department Working Papers: 329.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Healthcare" Handbook of Health Economics, Volume 2, Mark Pauly, Thomas McGuire and Pedro Barros (eds). Pages 49-93.

Skinner, Jonathan, Daniel J. Gottlieb, and Donald Carmichael. 2011. "A New Series of Medicare Expenditure Measures by Hospital Referral Region: 2003-2008." *Dartmouth Atlas Project Report*

Skinner, Jonathan and Douglas Staiger. 2007. "Technology Adoption from Hybrid Corn to Beta Blockers" in E. Berndt and C.M Hulten (eds) "Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches. University of Chicago Press and NBER.

Skinner, Jonathan and Douglas Staiger. 2009. "Technology Diffusion and Productivity Growth in Healthcare." NBER Working Paper 14865.

Skinner, Jonathan, Douglas O. Staiger, and Elliot S. Fisher. 2006. "Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction." *Health Affairs*, Web Exclusive. February. http://www.dartmouth.edu/~jskinner/documents/SkinnerIsTechnologicalHF.pdf

Syverson, Chad. 2004a. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics*, 86(2): 534–50.

Syverson, Chad. 2004b. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*, 112(6): 1181–1222.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2). 326-365.

Yasaitis, Laura, Elliott S. Fisher, Jonathan S. Skinner, and Amitabh Chandra. "Hospital Quality and intensity of spending: is there an association?" Health Affairs July/August 2009 28:w566-w572.

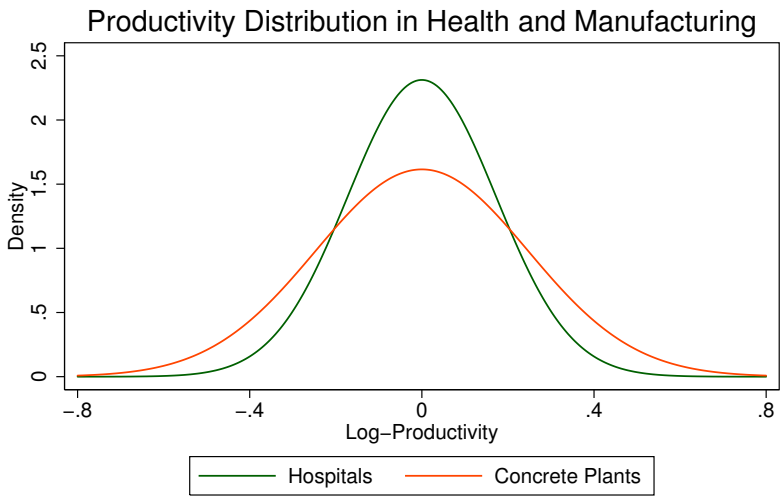
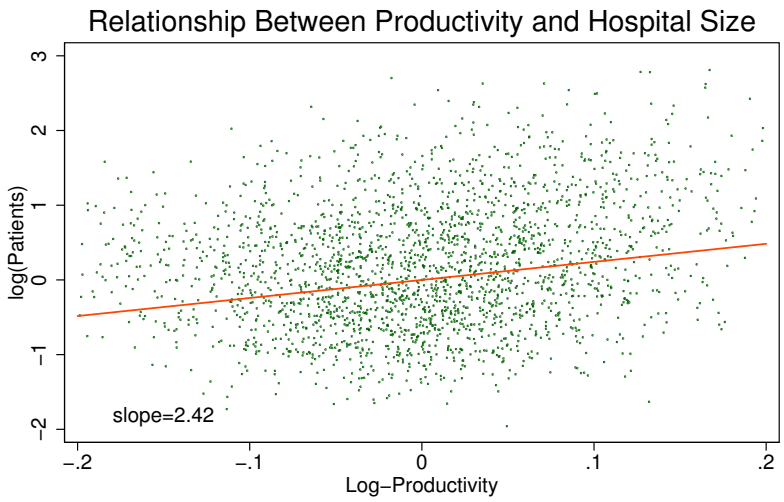
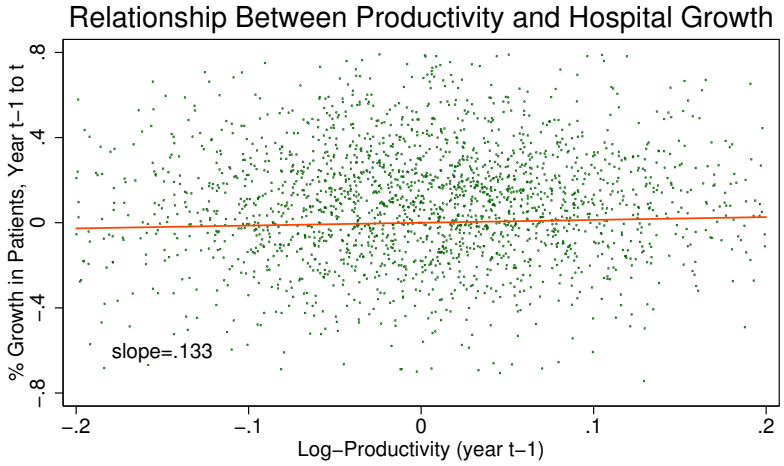

Productivity Distribution in Health and Manufacturing
Density
0
.5
1
1.5
2
2.5
−.8
−.4
0
.4
.8
Log−Productivity
Hospitals
Concrete Plants
Hospital result is fitted normal density and is within−year. Productivity estimates are
from our baseline specification (see e.g. Table 2, Column 1).
Concrete result is fitted normal density derived from Census estimates and is within−year.

Figure 1




Productivity estimates are from our baseline specification (see e.g. Table 2, Column 1). Market−year effects partialed out. Random 5% sample of hospital−years. Hospital−years with less than 11 patients suppressed from scatter.

Figure 2a




Productivity estimates are from our baseline specification (see e.g. Table 2, Column 1). Market−year effects partialed out. Random 5% sample of hospital−years. Y−axis limits to hospitals with residual growth between −.8 and .8, retaining 93.3% of the sample. Linear fit line is calculated on whole sample. Hospital−years with less than 11 patients suppressed from scatter.

Figure 2b

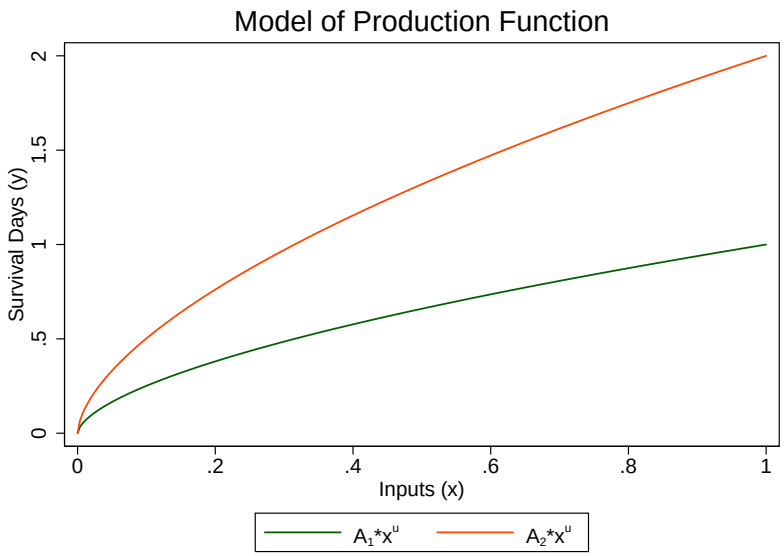

Model of Production Function
Survival Days (y)
0
.5
1
1.5
2
0
.2
.4
.6
.8
1
Inputs (x)
A1*xu
A2*xu

Figure 3

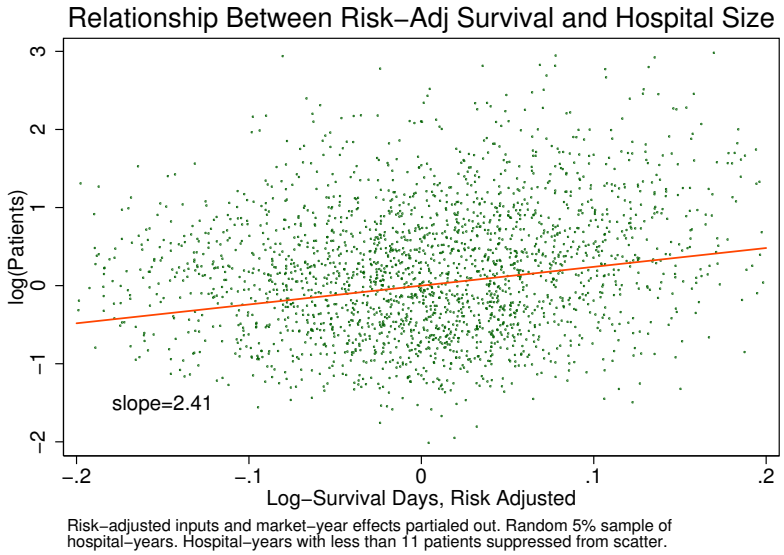



Figure 4a

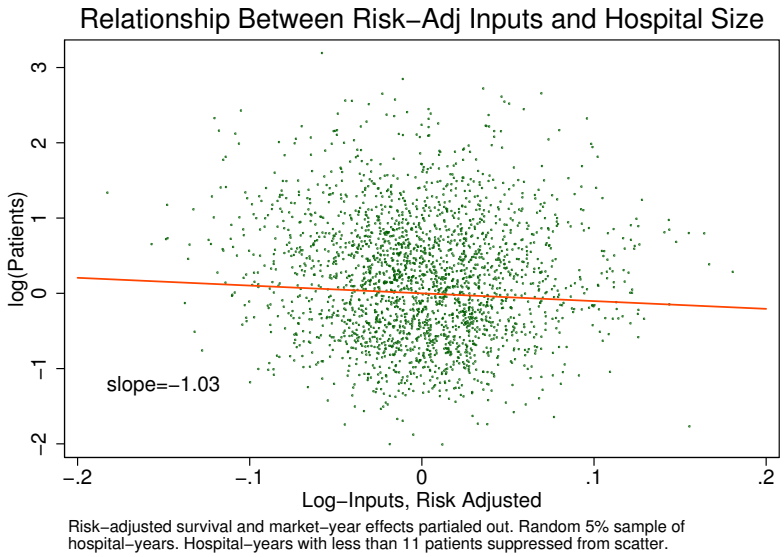



Figure 4b

Table 1a - Hospital and market statistics

| | (1) Mean | (2) SD | (3) Min | (4) Max |
|---|---|---|---|---|
| Hospital-Years (N=55,540) | | | | |
| Patients | 63.57 | 69.63 | 5 | 917 |
| HRR-Years (N=4,560) | | | | |
| Patients | 774.2 | 735.2 | 63 | 5,700 |
| Hospitals | 12.18 | 11.38 | 1 | 97 |

Note: The number of hospitals is 5,346.

Table 1b - Patient Summary Statistics

| | (1) Mean | (2) SD |
|---|---|---|
| Outputs | | |
| Survival (days; censored at 365) | 268.1 | 149.4 |
| Binary: Survival > 365 Days | 0.660 | 0.474 |
| Inputs | | |
| Baseline (30 day) input measure ($) | 15,996 | 12,172 |
| Risk Adjusters | | |
| Age | 78.17 | 7.546 |
| Female | 0.507 | 0.500 |
| White | 0.906 | 0.291 |
| Hypertension | 0.207 | 0.405 |
| Stroke | 0.0232 | 0.150 |
| Cerebovascular Disease | 0.0398 | 0.195 |
| Renal Failure | 0.0521 | 0.222 |
| Dialysis | 0.00670 | 0.0816 |
| COPD | 0.0981 | 0.297 |
| Pneumonia | 0.0592 | 0.236 |
| Diabetes | 0.128 | 0.334 |
| Protein Cal Malnut | 0.0118 | 0.108 |
| Dementia | 0.0412 | 0.199 |
| Paralysis/FD | 0.0256 | 0.158 |
| Periph Vasc Disease | 0.0639 | 0.245 |
| Metastatic Cancer | 0.0117 | 0.107 |
| Trauma | 0.0392 | 0.194 |
| Substance Abuse | 0.0225 | 0.148 |
| Major Psych Disorder | 0.0138 | 0.117 |
| Chronic Liver Disease | 0.00281 | 0.0529 |

Note: The number of observations is 3,530,401.

Table 2 - Production Function Parameter Estimates

| | (1) | (2) | (3) |
|---|---|---|---|
| Risk Adjustment: | Baseline | Age/Race/Sex | None |
| Parameter | | | |
| $\mu$ | 0.446 | 0.481 | 0.589 |
| | (0.00449) | (0.00464) | (0.00505) |

Notes: N = 3,530,401 patients and 55,540 hospital-years. Standard errors are bootstrapped with 50 replications and are clustered at the market level (5,346 hospitals). "Baseline" risk-adjustment includes a full set of interactions between age (in five year groupings), gender and whether the patient is white; it also includes indicators for the various co-morbidities shown in Table 1; column 2 excludes the co-morbidities and column 3 has no risk adjusters.

Table 3 - Beta Blockers and Productivity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable: | Beta-Blockers Score | | Composite Score | |
| Years in Regression: | 2003 | 2003-2007 | 2003 | 2003-2007 |
| $\rho$ | 1.658 | 1.227 | 1.456 | 1.862 |
| | (0.282) | (0.195) | (0.236) | (0.151) |
| Year Fixed Effects | N | Y | N | Y |
| Hospital-Years | 1045 | 8016 | 2183 | 12861 |
| Hospitals | 1045 | 2104 | 2183 | 3164 |

Notes: Dependent variables are z-scores. Right hand side variable is our estimate of hospital-year TFP from our baseline specification (Table 2, column 1), included for the years indicated in the column heading. Standard errors are bootstrapped with 500 replications and are clustered at the market level.

Table 4 - Main Results - Allocation Metrics

| | (1) | (2) | (3) | (A) | (B) |
|---|---|---|---|---|---|
| Risk Adjustment: | All | Age/Race/Sex | None | DV Mean[a] | Observations |
| Static Allocation | 2.418 | 2.496 | 2.618 | 3.641 | 55,540 |
| | (0.0861) | (0.0806) | (0.0695) | | |
| Dynamic Allocation | | | | | |
| Exit Regression | -0.0329 | -0.0353 | -0.0458 | 0.0438 | 40,379 |
| | (0.0118) | (0.0113) | (0.00985) | | |
| Growth Regression | 0.133 | 0.154 | 0.201 | -0.126 | 52,777 |
| | (0.0221) | (0.0213) | (0.0185) | | |

Notes: "Static Allocation" reports the results from estimating the relationship between a hospital's log(patients) and log TFP within a market year given by equation (1). "Exit regression" reports the results from estimating the within-market relationship between a hospital "exit" as defined in the text (over 5 years) and last year's log TFP as given by equation (2). "Growth regression" reports the results from estimating the within-market relationship between a hospital's one-year percent growth and it's base year log TFP as defined in equation (3). Log TFP is estimated based on the corresponding specifications from Table 2. Standard errors are bootstrapped with 50 replications and are clustered at the market level.

[a]"DV mean" reports the mean of the dependent variable for the regressions, which is ln(Patients) for the static allocation regression, 5-year exit for the exit regression, and 1-year growth for the growth regression. See text for more detailed definitions of dependent variables.

Table 5 - Dynamic Allocation Varying Time Horizons

| | Growth from $t$ to $t+k$ | | | Exit in $t+k$ | | |
|---|---|---|---|---|---|---|
| Years ($k$) | Coeff | Mean DV | Obs | Coeff | Mean DV | Obs |
| 1 | 0.133 | -0.126 | 52,777 | -0.033 | 0.044 | 40,379 |
| 2 | 0.207 | -0.224 | 49,954 | -0.056 | 0.077 | 36,864 |
| 3 | 0.270 | -0.314 | 46,961 | -0.085 | 0.108 | 33,163 |
| 4 | 0.345 | -0.392 | 43,742 | -0.122 | 0.137 | 29,338 |
| 5 | 0.365 | -0.462 | 40,379 | -0.147 | 0.166 | 25,359 |
| 6 | 0.397 | -0.530 | 36,864 | -0.165 | 0.195 | 21,320 |
| 7 | 0.477 | -0.598 | 33,163 | -0.203 | 0.226 | 17,226 |
| 8 | 0.526 | -0.666 | 29,338 | -0.224 | 0.255 | 13,050 |
| 9 | 0.573 | -0.735 | 25,359 | -0.242 | 0.284 | 8,761 |
| 10 | 0.587 | -0.807 | 21,320 | -0.212 | 0.313 | 4,412 |

These results report the coefficient from a regression of growth or exit on log-productivity. Each row considers a different time horizon. Longer horizons have smaller samples because data on growth ends in 2007 and data on exit ends in 2003. Standard errors (not shown) are bootstrapped with 50 replications and are clustered at the market level. All coefficients are significant at the 1% level.

The standard deviation of log-productivity is 0.173. Mean DV refers to the mean of the dependent variable (growth or exit) in the sample.

Table 6 - Productivity Dispersion across hospitals.

| | (1) | (2) | (3) |
|---|---|---|---|
| Risk Adjustment: | All | Age/Race/Sex | None |
| 90-10 | 0.442 | 0.469 | 0.521 |
| | (0.0112) | (0.0118) | (0.0130) |
| 75-25 | 0.233 | 0.247 | 0.274 |
| | (0.00588) | (0.00623) | (0.00682) |
| Standard Deviation | 0.173 | 0.183 | 0.203 |
| | (0.00436) | (0.00462) | (0.00506) |

Notes: Log TFP is estimated based on the corresponding specification in Table 2. Dispersion measures in log TFP are constructed nationally each year, and then averaged across years. The top row reports difference in log productivity between the 90th percentile hospital and the 10th percentile hospital; the next row reports the difference in log productivities between the 75th percentile and the 25th percentile hospital; the bottom row reports the estimated standard deviation of the log productivity distribution. Standard errors are bootstrapped with 50 replications and are clustered at the market level.

Table 7 - Sensitivity of Results to $\mu$

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Source of $\mu$: | Estimated | Imposed | | |
| Value of $\mu$: | 0.446 | 0.1 | 0.3 | 0.9 |
| Static Allocation | 2.418 | 2.358 | 2.399 | 2.278 |
| | (0.0861) | (0.0878) | (0.0865) | (0.0787) |
| Dynamic Allocation | | | | |
| Exit Regression | -0.0329 | -0.0361 | -0.0343 | -0.0263 |
| | (0.0118) | (0.0117) | (0.0118) | (0.0111) |
| Growth Regression | 0.133 | 0.144 | 0.138 | 0.107 |
| | (0.0221) | (0.0215) | (0.0218) | (0.0214) |
| Dispersion | | | | |
| 90:10 | 0.442 | 0.449 | 0.445 | 0.457 |
| | (0.0112) | (0.0113) | (0.0112) | (0.0107) |
| 75:25 | 0.233 | 0.237 | 0.234 | 0.241 |
| | (0.00588) | (0.00593) | (0.00590) | (0.00565) |
| Standard Deviation | 0.173 | 0.175 | 0.173 | 0.178 |
| | (0.00436) | (0.00440) | (0.00437) | (0.00419) |

Notes: Column (1) shows results based on estimation of our baseline specification (Table 2, column 1). In the other columns $\mu$ is imposed rather than estimated. Standard errors are bootstrapped with 50 replications and are clustered at the market level.

**Appendix A: Analytical framework**

As mentioned in the text, models of reallocation mechanisms among heterogeneous-productivity producers have found applications in a number of fields, including industrial organization, trade, and macro-economics. While these models differ considerably in their specifics, they share an archetypal mechanism that connects the extent of competition in the market to the shape of the productivity distribution among market producers. We describe this central mechanism here.

Producers (indexed by $i$) earn profits which depend positively on their idiosyncratic productivity levels $A_i$ – more productive firms earn higher profits due to their lower costs – and negatively on the number (or mass, in models with a continuum of firms) of producers in the industry $N$.[24] Hence $\pi_i = \pi(A_i,N)$, with $\partial\pi/\partial A_i > 0$ and $\partial\pi/\partial N < 0$. The monotonic relationship between productivity and profits implies that, for any given $N$, there is a critical cutoff productivity level $A^*(N)$ at which firm profits are zero. Only producers with productivity levels at or above $A^*(N)$ will operate in equilibrium.

The zero-profit cutoff productivity $A^*(N)$ is endogenously determined by a free entry condition, where ex-ante identical potential entrants consider whether to pay a sunk cost $\sigma$ to take an idiosyncratic productivity draw from a known distribution, $G(\cdot)$ with upper bound $\bar{A}$. The expected value of entry, which equals zero by the free entry condition, is:

$$V^e = \int_{A^*(N)}^{\bar{A}} \pi(A,N)g(A)dA - \sigma = 0$$

The expected profits from entry depend upon the equilibrium number of entrants $N$ in two ways. First, an increase in $N$ shifts upward the zero-profit cutoff productivity level $A^*(N)$, reducing the probability that the entrant's productivity draw is high enough to earn nonnegative profits and thus making successful entry less likely. Second, a higher number of firms $N$ also reduces the producer's profits if it does enter. Thus expected profits fall monotonically in $N$. In equilibrium, the number of firms choosing to pay the entry cost yields a number of entrants $N$ that, through these two effects, exactly equates the expected profit from taking a productivity draw to the sunk entry cost.

The endogeneity of $A^*(N)$ means the industry productivity distribution observed in the data is determined in equilibrium. Specifically, it is a truncation of $G(\cdot)$, the underlying productivity distribution from which potential entrants take productivity draws, where the

[24] Standard presentations of these models consider profit-maximizing firms. Although we keep this terminology to be more familiar relative to the existing literature, we note that in the context of hospitals, it might be more appropriate to consider firms as earning (and maximizing) "surplus" rather than "profits". This more general terminology recognizes that many hospitals are legally structured as nonprofits and does not affect the qualitative comparative statics. Nonprofit hospitals are often modeled in the literature as having an objective function that is a convex combination of profits and other objectives; therefore on the margin they should respond qualitatively the same way as for-profit hospitals to factors like competition. Even if a hospital's objective is not profit maximization, it is likely that for any given level of output(s) the hospital produces (in order to meet whatever outcomes are in its objective function), surplus will be larger if the hospital's costs are lower. Finally, in practice, a large empirical literature finds essentially no evidence of differential behavior across for-profit and non-profit hospitals, calling into question whether the non-profit label has any substantive meaning for behavioral responses (see Sloan 2000 for a recent review of this literature).

truncation point is $A^*(N)$. Changes in market primitives that shift the equilibrium location of $A^*(N)$ therefore shift the observed productivity distribution as well.

The primitive that we are interested in here is the extent of competition, as reflected in how easily consumers can (or how willing consumers are to) substitute to alternate producers. The specific mechanism through which primitives map into substitutability may vary, from changes in the differentiation of firms' products, to shifts in openness to trade, to movements in the size of transport costs. The particulars of the mechanism aren't important here; what matters are the effects on the equilibrium.

Higher substitutability has three effects that can be examined empirically. First, it makes it more difficult for higher-cost (lower-productivity) firms to earn positive profits, as demand is now more responsive to their cost and price differential relative to other firms in the industry.[25] In turn, the zero-profit cutoff productivity level $A^*(N)$ rises: the threshold for operation is greater than before. This truncates the equilibrium productivity distribution, reducing observed *productivity dispersion*.[26] Second, higher substitutability means that, among operating firms, market shares are more sensitive to productivity differences. Purchases are reallocated to more productive firms, raising the correlation between productivity and market share at a point in time (*"static allocation")*. Third, over time more productive firms are likely to grow in market share (*"dynamic allocation").*[27]

[25] In the case of hospitals, this demand response can be manifested either directly in patients' choices in response to out-of-pocket costs, or indirectly through insurers' decisions to include the hospital in its covered network.

[26] This dispersion implication requires some additional regularity assumptions on the underlying productivity distribution. Most "standard" distributions exhibit declining second moments as they are truncated from below. The exponential distribution, however, is an example of one that does not. Nevertheless, if we assume the productivity distribution is bounded at the top (i.e., there is some maximum productivity level), as we do here, then all distributions will eventually exhibit decreased dispersion as they are truncated from below.

[27] The model just described is static, so the effects of changes in competition on equilibrium should be thought of as comparing two different markets or the same market across different long-run steady states. However, several of the models in the literature are explicitly dynamic and have similar predictions about the effect of competition on the productivity of entrants and growth of incumbents (e.g., Hopenhayn 1992, Asplund and Nocke 2006).

**Appendix B: Measuring inputs**

Our baseline input measure (as well as many of the alternative measures discussed below) is derived from the formulas used to determine Medicare's Hospital (Part A) reimbursement. Some alternative measures also use information derived from the formulas used to determine Medicare's reimbursement of physicians and outpatient facilities (Part B). It is therefore useful to begin with a very brief overview of the key features of Medicare hospital reimbursement needed to understand the construction and composition of our baseline and alternative input measures. Considerably more detail can be found in CMS (2011).

The amount Medicare reimburses a hospital is determined by the patient's Diagnosis Related Group (DRG), national factors, and hospital-specific factors. A patient's DRG is a function of his principal diagnosis, procedures performed, and secondary complications and comorbidities. Some DRGs also depend on whether the patient died in the hospital.

Each DRG is assigned a (national) weight based on how much it costs to treat the nationwide average patient with that DRG; a national conversion factor is used to convert these DRG weights into dollar payments. The weights and the conversion factor are updated annually. The national rate is then adjusted for hospital-specific considerations. The major adjustments are due to geographic factors (e.g. the local wage rate) and characteristics of the hospital (such as whether it operates a resident training program or has a disproportionate share of patients on Medicare or SSI).

For most stays the hospital will receive payments solely based on the patient's DRG. However, in certain extraordinarily costly cases hospitals receive additional "outlier payments" covering 80 percent of costs beyond a threshold level. To compute costs, the hospital's billed charges are deflated by a hospital-specific cost-to-charge ratio. If a patient is transferred to another hospital, Medicare allocates payments for the patient across the initial and receiving hospital. For our purposes, we assign all inputs for the patient in the time horizon (30 days for our baseline measure) back to the initial hospital.

A1. Baseline input measure: Part A "resources"

Our baseline input measure follows the approach of Gottlieb et al. (2010) and Skinner and Staiger (2009) to purge the "price" variation in the reimbursement formula from the "input" variation. Specifically, our starting point is the DRG weight (multiplied by a national conversion factor to convert it to a dollar metric) plus outlier payments (also in dollars). It does not reflect any variation in reimbursement prices across hospitals due to geographic factors or specific characteristics of the hospital.

According to this measure, the inputs a patient receives equal the sum of his converted DRG weights and outlier payments at all hospital stays in the 30 days following his AMI. Variation across patients in the input measure therefore comes from 3 sources: variation in the patient's DRG(s); whether there are (and the extent of) outlier payments; and the number of hospital stays during the 30 day window. We discuss each in turn.

*Variation across Index Event DRGs*

To give a sense of the nature and variation across DRGs, Table A1 lists the top 20 DRGs for the index event (initial AMI hospital stay), their patient share and their weights in 2000.[28] The top five DRGs account for over 90 percent of the index events, and the top 20 account for virtually 100 percent.

Looking within the top five we see substantial differences in weight based on whether an invasive procedure is performed. There are two separate DRGs for invasive procedures (#107, "Coronary Bypass with Cardiac Catheterization" and #116, "Other Permanent Cardiac Pacemaker Implant or PTCA with Coronary Artery Stent Implant") and they respectively have weights of 5.46 and 2.47. By contrast, the other three DRGs in the top five are medical DRGs (i.e. do not involve invasive procedures) and have

[28] For presentation purposes, we limit Table A1 to one year because DRG weights and classifications change slightly from year to year.

weights ranging from 1.11 to 1.51. For the year 2000, two dummies for these two surgical DRGs (bypass and stent) explain 15 percent of the total variation in our 30 day input measure.

Within the three most common medical DRGs, we see that there is variation for a medically treated AMI based on whether or not the patient died (#123), survived following a stay with major complications (#121) or survived following a stay without major complications (#122). This variation has, to our knowledge, not previously been noted by the large empirical literature on the relationship between inputs for heart attacks and subsequent survival which has used the variation in puts stemming from survival. However, this source of variation in the standard input measure seems suspect: it partly causes in-hospital death – not inputs, per se – to explain survival, an association that must exist trivially.

Therefore, for these three DRGs that refer to the same diagnosis but differ on the basis of patient survival, we eliminate the variation in inputs across DRGs within this group at the hospital-year level. We assign each DRG the patient-weighted average of the different DRG weights. The averaging weights are equal to the share of patients in the DRG in that year. Almost three-quarters of hospital stays were grouped into DRGs that were affected by this fix.[29]

*Variation from Outlier payments*

Approximately 8.2% of our patients trigger outlier payments due to unusually costly cases. These payments are triggered when a hospital's cost of treating a patient exceeds a national threshold. Conditional on receiving an outlier payment, the average outlier payment as a share of DRG reimbursement without outlier payments is 53.9; the standard deviation of outlier payments is 13,154.8. (All statistics calculated for patients in the year 2000.)

*Variation due to number of hospital stays*

Even ignoring outlier payments, the total variation coming from DRGs is in fact larger than that indicated in Table A1 because of the possibility of multiple (and potentially non AMI) hospital stays in the 30 days following the index event (AMI). Our baseline input measure is constructed for the 30 days following the initial AMI, meaning that it includes all hospital stays in these 30 days. On average, an AMI patient has 1.07 stays in this window. Conditional on having multiple stays, the average patient visits the hospital 2.07 times in the month following the AMI.

If a hospital stay straddles the end of the time window (e.g. a patient stays in the hospital for 10 days and is admitted on day 25 days following the heart attack), the inputs attributed to that hospital are reduced; in particular, we multiply our input measure by the share of days in the hospital that were inside the 30 day analysis window. We adjusted all DRGs (not just those associated with index events) to purge variation stemming from mortality in the manner described above.

[29] Note that this "fix" also purges the variation across the three most common medical DRGs in whether the patient had a major complication or not. Although the case in question is the only one where different DRGs are assigned based on patient survival, there are other cases where separate DRGs are assigned based on the presence of complicating conditions (CCs). For example, the 6th-ranked DRG #110, "Major Cardiovascular Procedures with CC" (weight 4.16) and the 18th-ranked DRG #111, "Major Cardiovascular Procedures without CC" (weight 2.23) differ only on this basis. It is a priori unclear to us whether we want to purge variation due to the presence of CCs. On the one hand, conditional on a rich set of patient risk adjustors, the presence of a CC may be a useful measure of the intensity of resources required to treat the condition; on the other hand, with imperfect risk adjustors, it may also capture correlates of mortality (our outcome of interest).
As noted, in practice our approach to purging mortality-based variation across DRGs also purges complications-based variation in the most common DRGs. We experimented with an alternative measure that purged variation due to CCs in all DRGs. The procedure took DRGs that were identical but for the CC requirement and assigned them the same DRG weight within each hospital-year. This DRG weight was a weighted average of the component DRG weights; the averaging weights were the shares of patients in each DRG in the hospital-year. For example, in 2000, DRGS #110 and #111 were assigned the same weight in each hospital-year. This correction affected only a few percent more patients and made no noticeable difference to our findings (results available on request).

Table A2 lists the top 20 DRGs across all stays in the 30 day window following the index event. The index events are included in this table. As expected, there is more variation across these DRGs.

*Empirical variation in baseline input measure*

Figures A1-A3 show the variation in the input measures across patients for one year (2000). Figure A1 shows the variation in the DRG index events (using our "collapsed" DRG measure that purges mortality variation). Figure A2 shows the variation from the DRG index events plus outlier payments in the index event. Figure A3 shows the total 30 day variation, which adds in additional hospital stays (their DRGs and outlier payments) within the 30 days. As would be expected, the input distribution gets less "lumpy'' at each step.

A2. Alternative input measures

We confronted a number of choices in defining our baseline input measure. We therefore constructed several other alternative input measures. This section describes them.

*Alternative measures of hospital inputs*

A central tension in our choice of input measurement is how coarse or detailed we make our input measure. The tradeoff is between the survival bias that can occur with finer input measures—since the longer a patient survives, the more can be done to a patient—and the measurement error which occurs at coarser definitions of inputs. Our baseline measure, following standard practice, is aggregated to a relatively high level, and may therefore measure inputs with a non-trivial amount of error.

We experimented with two alternative hospital-based input measures. One measures Part A spending rather than Part A inputs; it therefore includes variation in reimbursement rates stemming from hospital specific factors like geographic location or type of hospital. As shown in Appendix Figure A4 the distribution of Part A reimbursement is less "lumpy" than our baseline input measure; the correlation between the two is 0.90.

The other measure is designed to be more detailed than our baseline measure to reflect that fact that input use may vary substantially within the relatively coarse DRGs. We used data on the length of hospital stay and the procedures performed during the stay (up to six may be listed). Procedure codes are themselves available at different levels of granularity; there are 3 levels of CCS procedure codes ranging from the least granular level 1 to the most granular level 3; the much larger set of ICD-9 procedure codes is more granular still. The ICD-9 codes account for over 3878 possible procedures that may be performed on patients.

To reduce the dimensionality of the set of procedures, we use the following algorithm. We start with the coarsest set of procedures (level 1 CCS codes, of which there are 16) and move iteratively to the finest set of procedure codes (ICD-9). At each step we aggregate codes that are rare and disaggregate codes that are very common. Thus, beginning with CCS level 1 codes, we include indicators for level 1 procedures that were performed on less than 10% of patients; if the level 1 procedure was performed on 10% or more of patients, we disaggregate it by looking at CCS level 2 components.

In similar fashion, if the CCS level 2 procedures were performed on 1-10 percent of patients, we include an indicator for it. Within a level 1 code, all level 2 codes performed on less than 1 percent of patients are grouped together and included as one indicator. If the level 2 procedure was performed on 10 percent or more of patients, we disaggregate by looking at its level 3 components.

We follow the same process for level 3 components; when we disaggregate these codes we look at the component ICD-9 codes. If the ICD-9 code was performed on at least 1 percent of patients we include an indicator for it. Within a level 3 code, all ICD-9 codes that were performed on less than 1 percent of patients are grouped together and included as one indicator.

This algorithm results in 60 procedure indicators: 18 for ICD-9 codes, 6 for level 3 CCS codes, 22 for level 2 CCS codes and 14 for level 1 CCS codes.

*Incorporating non-hospital inputs*

A limitation of our input measures thus far is that, following standard practice in the heart attack literature, they reflect only inpatient hospital inputs. Notably, they do not include physician inputs, which may occur in an inpatient or outpatient setting. They also do not include outpatient tests and procedures like MRIs.

Many of these inputs are directly related to the treatment of the AMI. For example, the work of physicians who treat the patient surgically or medically in the hospital is obviously an input that may bear on the patient's survival. Likewise, an MRI done in an outpatient facility that is closely affiliated with the hospital will inform treatment decisions and influence mortality.

There are two reasons why we follow most of the literature on heart attacks and do not include inputs by physicians or outpatient facilities in our baseline measure. First, while some of these inputs are closely linked to the care received in the hospital, many of the payments reflect care that is independent of the hospital. In particular, doctor visits and outpatient diagnostic tests at long time horizons from the initial AMI admission may be less dependent on initial treatment decisions. The second reason is practical: data on much of these other input measures are only available for 20 percent of the sample and only since mid-2000, reducing the set of hospital-years in which we can observe at least 5 AMI patients by 70.0%.

Still, we sought to evaluate the sensitivity of our results to including physician and outpatient services. Medicare reimburses physicians based on their assessment of the "Relative Value Units" (RVUs) of the services the physician provided; the RVU of a service is intended to reflect the resources required to provide that service. The RVUs attributed to procedures are constant across geographic areas and practitioners, although Medicare makes further adjustments based on geography and provider type to derive reimbursement rates (see MedPAC [2010a] or Clemens and Gottlieb [2012] for more details). We construct our measure of physician inputs by summing all RVUs associated with the patient in the 30 days following his initial hospital admission. We multiply the RVUs by a national conversion factor to convert them to a dollar metric; the national conversion factor eliminates variation due to Medicare's geographic price adjustments.

Calculating outpatient contributions to the production function is significantly more complicated than calculating physician or inpatient contributions. While physician services and inpatient stays are each reimbursed using a single payment system that is designed to reflect resource utilization, different outpatient services are covered by different types of systems (MedPAC [2010b] provides more details). Some outpatient services are covered prospectively – although the payment groups are so fine that treatment decisions may be reimbursed at the margin. Providers are paid for other services according to a fee schedule that is geographically adjusted. Some services are reimbursed according to local prices.

For the portion of outpatient services covered prospectively, there is a series of classification groups (Ambulatory Payment Classification groups or APCs) which function analogously to DRGs. Each APC is given a weight that is based on its expected resource costs; we translate these weights into a dollar basis using a national conversion factor that is an analogous to the procedure we use to convert DRG weights. For services that are reimbursed on a fee schedule, we mimic the method used for physician inputs by applying the fee schedule prior to geographic adjustments.

These adjustments eliminate much of the variation in outpatient prices that is region- or provider-specific. Still, some payments, like those for certain prescription drugs and new technologies, do not have an associated national fee schedule and are included unadjusted.

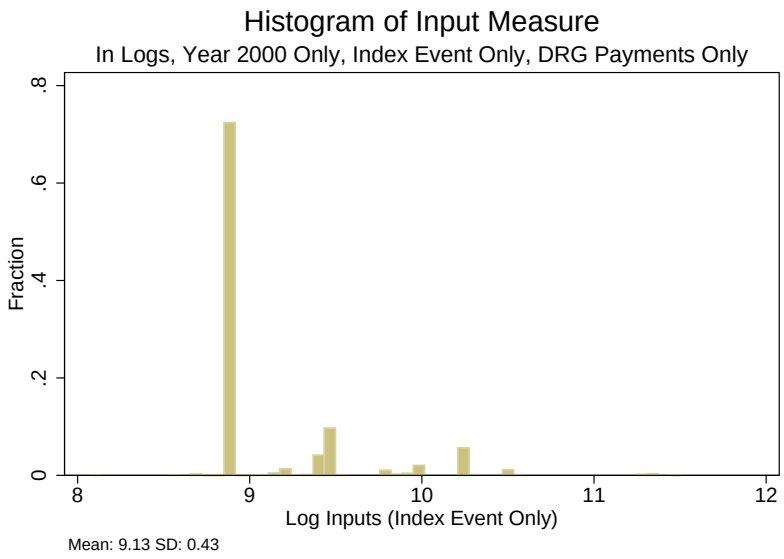



Figure A1

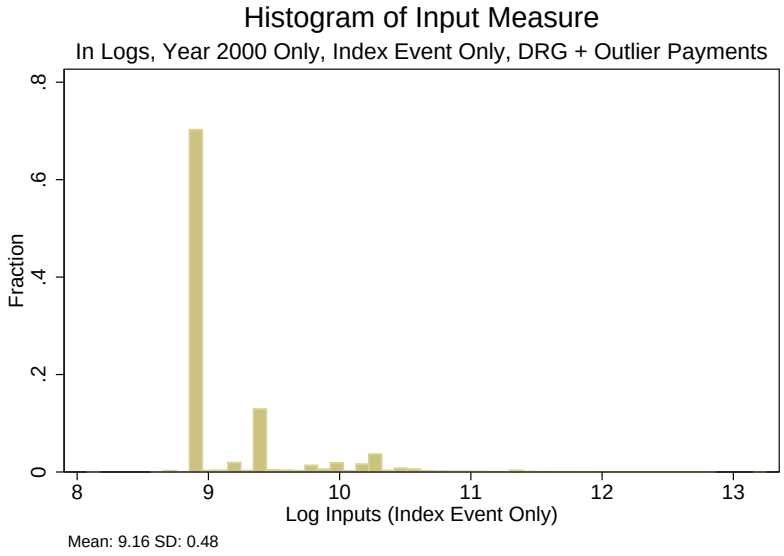



Figure A2

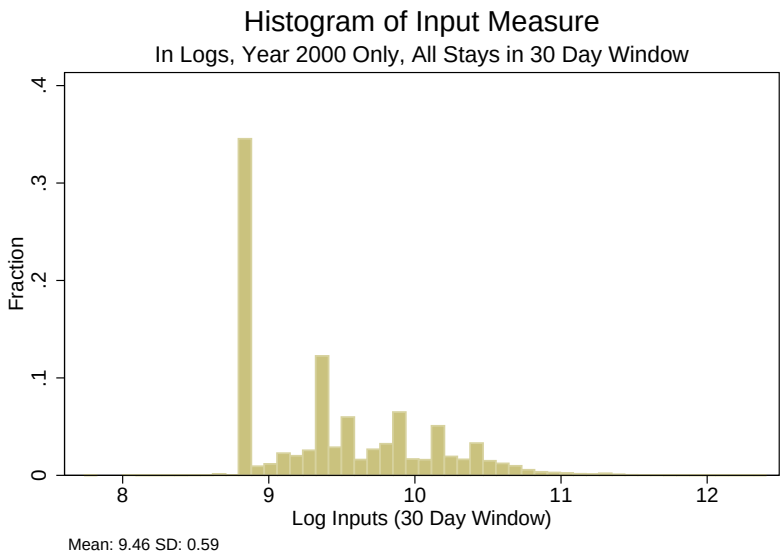



Figure A3

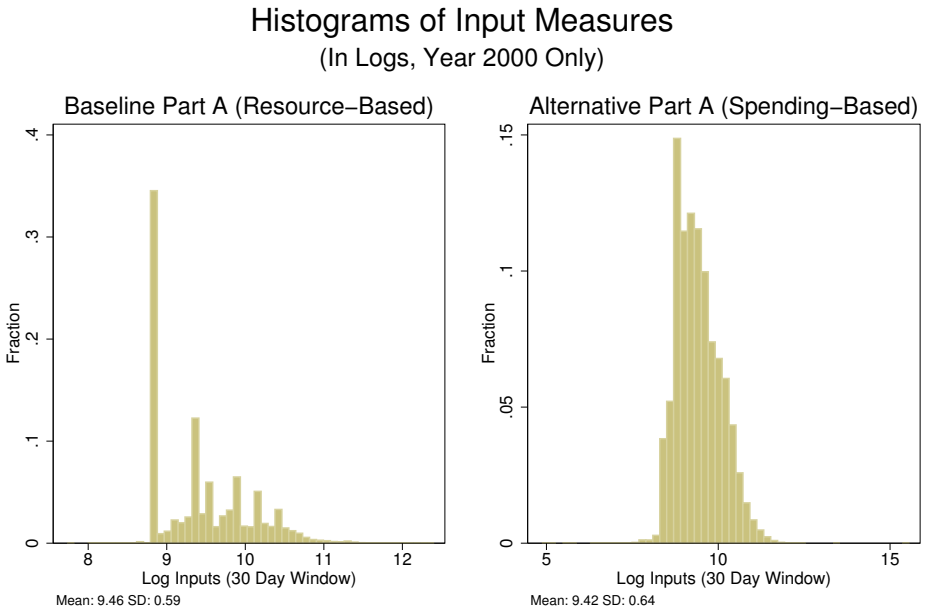



Figure A4

Table A1 - List of Top DRGs for Index Events (Initial Hospital Stays for the AMI Episode) in 2000

| Rank | Number | DRG Name[a] | Weight | Share | Cum. Share |
|---|---|---|---|---|---|
| 1 | 121 | Circulatory Disorders with AMI and Major Complications, Discharged Alive | 1.63 | 41.2% | 41.2% |
| 2 | 122 | Circulatory Disorders with AMI, without Major Complications, Discharged Alive | 1.11 | 20.9% | 62.1% |
| 3 | 116 | Other Permanent Cardiac Pacemaker Implant or PTCA with Coronary Artery Stent Implant | 2.47 | 13.0% | 75.1% |
| 4 | 123 | Circulatory Disorders with AMI, Expired | 1.51 | 10.9% | 86.0% |
| 5 | 107 | Coronary Bypass with Cardiac Catheterization | 5.46 | 5.4% | 91.4% |
| 6 | 110 | Major Cardiovascular Procedures with CC | 4.16 | 2.0% | 93.4% |
| 7 | 112 | Percutaneous Cardiovascular Procedures | 1.92 | 1.6% | 95.0% |
| 8 | 115 | Permanent Cardiac Pacemaker Implant with AMI, Heart Failure or Shock, or AICD Lead or Generator Procedure | 3.47 | 1.0% | 96.0% |
| 9 | 104 | Cardiac Valve and Other Major Cardiothoracic Procedure with Cardiac Catheterization | 7.24 | 0.8% | 96.8% |
| 10 | 483 | Tracheostomy except for Face, Mouth, and Neck Diagnoses | 16.12 | 0.5% | 97.3% |
| 11 | 106 | Coronary Bypass with PTCA | 7.33 | 0.4% | 97.7% |
| 12 | 109 | Coronary Bypass without PTCA or Cardiac Catheterization | 4.04 | 0.4% | 98.1% |
| 13 | 144 | Other Circulatory System Diagnoses with CC | 1.15 | 0.3% | 98.4% |
| 14 | 478 | Other Vascular Procedures with CC | 2.35 | 0.3% | 98.7% |
| 15 | 468 | Extensive OR Procedure Unrelated to Principal Diagnosis | 3.64 | 0.3% | 99.0% |
| 16 | 120 | Other Circulatory System OR Procedures | 2.01 | 0.2% | 99.2% |
| 17 | 108 | Other Cardiothoracic Procedures | 5.77 | 0.2% | 99.4% |
| 18 | 111 | Major Cardiovascular Procedures without CC | 2.23 | 0.1% | 99.5% |
| 19 | 477 | Non-Extensive OR Procedure Unrelated to Principal Diagnosis | 1.77 | 0.1% | 99.6% |
| 20 | 145 | Other Circulatory System Diagnoses without CC | 0.65 | 0.1% | 99.7% |

Notes: "Rank" refers to the share of patients with the DRG; "Number" refers to CMS's assigned number for that DRG; "Weight" is a CMS-assigned value that is designed to be proportional to the average cost of treatment and is used to determine reimbursement - the weights are set by CMS so that the average Medicare patient across all conditions has a weight of 1.

[a]Abbreviations: CC - Complicating Conditions, OR - Operating Room, PTCA - Percutaneous Transluminal Coronary Angioplasty.

Table A2 - List of Top DRGs for All Claims

| Rank | Number | DRG Name[a] | Weight | Share | Cum. Share |
|---|---|---|---|---|---|
| 1 | 121 | Circulatory Disorders with AMI and Major Complications, Discharged Alive | 1.63 | 15.1% | 15.1% |
| 2 | 127 | Heart Failure and Shock | 1.01 | 8.4% | 23.5% |
| 3 | 116 | Other Permanent Cardiac Pacemaker Implant or PTCA with Coronary Artery Stent Implant | 2.47 | 8.0% | 31.5% |
| 4 | 122 | Circulatory Disorders with AMI, without Major Complications, Discharged Alive | 1.11 | 7.3% | 38.8% |
| 5 | 123 | Circulatory Disorders with AMI, Expired | 1.51 | 3.8% | 42.6% |
| 6 | 132 | Atherosclerosis with CC | 0.67 | 2.8% | 45.4% |
| 7 | 107 | Coronary Bypass with Cardiac Catheterization | 5.46 | 2.7% | 48.1% |
| 8 | 462 | Rehabilitation | 1.36 | 2.7% | 50.8% |
| 9 | 89 | Simple Pneumonia and Pleurisy, Age > 17, with CC | 1.09 | 2.5% | 53.3% |
| 10 | 14 | Specific Cerebrovascular Disorders Except TIA | 1.19 | 1.9% | 55.2% |
| 11 | 88 | Chronic Obstructive Pulmonary Disease | 0.94 | 1.8% | 57.0% |
| 12 | 144 | Other Circulatory System Diagnoses with CC | 1.15 | 1.5% | 58.5% |
| 13 | 174 | Gastrointestinal Hemorrhage with CC | 1.00 | 1.2% | 59.7% |
| 14 | 112 | Percutaneous Cardiovascular Procedures | 1.92 | 1.2% | 60.9% |
| 15 | 124 | Circulatory Disorders Except AMI, with Cardiac Cath and Complex Diagnosis | 1.40 | 1.2% | 62.1% |
| 16 | 138 | Cardiac Arrhythmia and Conduction Disorders with CC | 0.82 | 1.2% | 63.3% |
| 17 | 143 | Chest Pain | 0.53 | 1.2% | 64.5% |
| 18 | 296 | Nutritional and Miscelaneous Metabolic Disorders, Age > 17, with CC | 0.86 | 1.2% | 65.7% |
| 19 | 109 | Coronary Bypass without PTCA or Cardiac Catheterization | 4.04 | 1.1% | 66.8% |
| 20 | 182 | Esophagitis, Gastroenteritis, and Miscelaneous Digestive Disorders, Age > 17, with CC | 0.78 | 1.1% | 67.9% |

Notes: "Rank" refers to the share of patients with the DRG; "Number" refers to CMS's assigned number for that DRG; "Weight" is a CMS-assigned value that is designed to be proportional to the average cost of treatment and is used to determine reimbursement - the weights are set by CMS so that the average Medicare patient across all conditions has a weight of 1.

[a]Abbreviations: CC - Complicating Conditions, OR - Operating Room, PTCA - Percutaneous Transluminal Coronary Angioplasty, TIA - Transient Ischemic Attack.

## Appendix C: Robustness of results

*Alternative input measures*

Appendix Table A3 explores the robustness of our results to alternative input measures; more detail on their construction is provided in Appendix B. Column 1 replicates our baseline results. As noted in Section 6, there is a tradeoff between our relatively coarse baseline measure of inputs (with its associated measurement error) and more granular measures which suffer from potential survivorship bias (you cannot have a lot of procedures done if you do not survive very long). Columns 2 and 3 explore the sensitivity of our estimates to more granular measures which use as inputs a series of approximately 60 indicators for whether various procedures were performed as well as a continuous variable measuring the log of the number of days in the hospital during our 30 day window (see Appendix B for more detail).

We incorporate this more granular input measure in two different ways. In column 2 we explore a multi-input production function; specifically, we replace our single index measure with a series of indicators for whether various procedures were performed as well as a continuous variable for log number of hospital days. In column 3 we return to a single-input production function but one that is based on this more granular input measure; we create the single input by regressing log hospital charges on these same procedure indicators and log length of stay variables from column 2, as well as hospital-year fixed effects.[30] We use the coefficients from this regression – ignoring the hospital-year effects – to produce an estimate of predicted large charges for each patient in our data. The correlation between this predicted log charges measure and our baseline log input measure is 0.77 (with actual log charges it would be 0.75). As would be expected from survivorship bias, the returns to scale coefficient $\mu$ in column 3 is substantially higher than that in our baseline column 1.

Yet another alternative approach to inputs is to measure Medicare reimbursement to the hospital for a patient, rather than hospital inputs. Like our baseline approach, this approach is also often used in the literature (e.g. Cutler et al., 1998, Skinner and Staiger 2009). Medicare reimbursement depends not just on the patient's DRGs (our baseline resource measure) but also characteristics of the hospital (such as whether it is a teaching hospital or whether it treats a disproportionate share of low income patients) and its location (MedPAC 2011a). Part A Medicare spending per AMI patient is the standard measure used in the economics literature in studying the relationship between heart attack treatment and outcomes (e.g. Cutler et al. 1998, Skinner and Staiger 2009). The results in column 4 use this Medicare reimbursement measure; the returns to scale parameter $\mu$ is therefore interpreted here as the return to federal expenditures (in the form of post-AMI survival) rather than real inputs. The correlation between our baseline resources measure and the reimbursement measure is 0.90. The main results are all quite robust to this alternative measure.

A final input measure incorporates physician inputs and outpatient hospital inputs for the subsample of hospital years beginning in 2001 (see Appendix B for more details; our sample starts in 2001 because it is the first full year with data). Column 5 shows our baseline results limited to the sample where we can observe these other input measures; this cuts our sample of hospital-years substantially (by about 70 percent). Column 6 shows the results for this same "overlap" sample with our expanded input measure. For the overlap sample, the correlation between our baseline input measure and the expanded measure is 0.98[31].

[30] Hospital "charges" are accounting charges for rooms and procedures and do not reflect transacted prices. They have been used in the literature as convenient, price-weighted summary of treatment, albeit at somewhat artificial prices (Card et al., 2009, Finkelstein et al., 2012). The hospital-year fixed effects in the log charges regression eliminate variation across hospital-years in the charge to cost ratio (i.e. differential hospital markups of list prices above costs).

[31] This high correlation reflects the fact that outpatient resources are, on average, about one-fifth the size of the inpatient resources devoted to one of our patients; in addition there is a high (about two-thirds) correlation between outpatient and inpatient resources devoted to a patient.

Looking across the columns, the basic qualitative findings concerning the role for competition in allocating more market demand to more productive firms both at a point in time and over time are quite robust to alternative input measures. In particular, the static allocation analysis and the growth analysis remain statistically significant in virtually all alternative specifications. The statistical significance of the exit-based regression results is more sensitive to the choice of input measure. Perhaps not surprisingly, the magnitudes of the static and dynamic allocation analyses vary somewhat across the specifications. The dispersion estimates are remarkably robust to alternative input measures.

*Alternative time frames for measuring inputs and outpus*

Appendix Table A4 considers how our metrics are affected by alternative time windows for measuring survival and inputs. Our baseline specification looks at survival over 1 year and at inputs over 30 days. A shorter time horizon for inputs will miss some of the resources provided to the patient. There is also a practical limitation to very short horizons; we observe resources at the level of a hospital stay, not a hospital day or hour; 96% of hospital stays are at most 30 days long, but a measure like 7 day utilization would require arbitrary spreading of resources across the 7 days for the 33% of patients who spend more than 7 days in the hospital. Longer time horizons have their own limitations: issues of survival bias (the longer you live the more that can be done) and as time passes since the first incident, the treatments that are undertaken are increasingly linked to providers outside the original hospital. Columns 2 and 3 show, respectively, that the results are robust to a longer (one year) survival horizon, and a shorter (7 day) survival horizon rather than our baseline 30 day time frame.

In terms of the time horizon for outcomes, we choose a 1-year survival window because it is of more interest than short-term survival which may reflect only a few day postponement in mortality. As a practical matter, censoring is also less prevalent at 1 year than at shorter horizons. Finally, another advantage of our 1-year window is that it will pick up aspects of hospital productivity that affect outcomes through longer term mechanisms such as the management of complications due to co-morbidities like congestive heart failure or diabetes. Longer time windows will also better capture the quality of continuing care like the prescribing of statins and the follow up to make sure the patient is taking these medications. Such inputs are less likely to affect survival at much shorter horizons but can be quite important over longer intervals  On the flip side, the longer measurement horizon introduces greater scope for patient autonomy (e..g in terms of changes in behavior such as diet and smoking, compliance with recommended medications and follow-up visits etc) and for the impact of doctors (regardless of which hospital you went to) or admissions to other hospitals to affect survival and therefore may attenuate differences across hospitals in measured productivity. Our results are robust to moving away from our baseline 1 year survival to 30 day survival (column 4) or to 5 year survival (column 6) which requires that we limit the sample to heart attacks through 2003 so that we observe 5 subsequent years; column 5 shows our baseline 1 year survival measure on this same sample.

Table A3 - Comparison of Input Measures

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Input Measure: | Baseline | Procedures | Fitted Chg | Spending | Baseline | Base+Part B |
| Sample: | Full | Full | Full | Full | With outpatient data | |
| Parameter | | | | | | |
| μ | 0.446 | | 0.714 | 0.395 | 0.369 | 0.399 |
| | (0.00449) | | (0.00521) | (0.00412) | (0.00640) | (0.00677) |
| Static Allocation | 2.418 | 1.497 | 0.972 | 1.749 | 2.347 | 2.252 |
| | (0.0861) | (0.0884) | (0.0991) | (0.0896) | (0.224) | (0.223) |
| Dynamic Allocation | | | | | | |
| Exit Regression | -0.0329 | -0.0199 | -0.00661 | -0.0245 | -0.0331 | -0.0348 |
| | (0.0118) | (0.0117) | (0.0124) | (0.0116) | (0.0546) | (0.0576) |
| Growth Regression | 0.133 | 0.0611 | -0.00515 | 0.0762 | 0.222 | 0.213 |
| | (0.0221) | (0.0241) | (0.0262) | (0.0225) | (0.0705) | (0.0739) |
| Dispersion | | | | | | |
| 90:10 | 0.442 | 0.431 | 0.428 | 0.453 | 0.351 | 0.341 |
| | (0.0112) | (0.00864) | (0.00918) | (0.0101) | (0.0216) | (0.0219) |
| 75:25 | 0.233 | 0.227 | 0.225 | 0.239 | 0.185 | 0.179 |
| | (0.00588) | (0.00455) | (0.00483) | (0.00534) | (0.0114) | (0.0115) |
| Standard Deviation | 0.173 | 0.168 | 0.167 | 0.177 | 0.137 | 0.133 |
| | (0.00436) | (0.00337) | (0.00358) | (0.00395) | (0.00843) | (0.00856) |
| Patients / 1000 | 3,530 | 3,530 | 3,530 | 3,525 | 271.3 | 271.3 |
| Hospital-Years | 55,540 | 55,540 | 55,540 | 55,529 | 15,039 | 15,039 |
| Hospitals | 5,346 | 5,346 | 5,346 | 5,346 | 3,092 | 3,092 |

Notes: Column (1) is baseline specification. All other columns use alternative input measures (described in more detail in Appendix A and B). Column 5 and 6 are limited to the sub-sample of approximately 30 percent of hospital-years for which we observe outpatient data for at least five AMI patients in that hospital-year; in column 6 our baseline input measure (which uses only Part A inputs) is expanded to include Part B inputs; see text for more details. Productivity metrics reflect empirical bayes adjustment. Standard errors are bootstrapped with 50 replications and are clustered at the market level.

Table A4 - Comparison of Results with Varying Survival and Input Horizons

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Survival Horizon: | 1 Year | 1 Year | 1 Year | 30 Days | 1 Year | 5 Years |
| Input Window: | 30 Days | 1 Year | 7 Days | 30 Days | 30 Days | 30 Days |
| Sample Thru: | 2007 | 2007 | 2007 | 2007 | 2003 | 2003 |
| Parameter | | | | | | |
| $\mu$ | 0.446 | 0.790 | 0.172 | 0.292 | 0.451 | 0.585 |
| | (0.00449) | (0.00427) | (0.0103) | (0.00206) | (0.00503) | (0.00754) |
| Static Allocation | 2.418 | 2.694 | 2.421 | 3.992 | 2.347 | 2.047 |
| | (0.0861) | (0.0870) | (0.0907) | (0.136) | (0.0941) | (0.0811) |
| Dynamic Allocation | | | | | | |
| Exit Regression | -0.0329 | -0.0317 | -0.0372 | -0.0660 | -0.0221 | -0.0201 |
| | (0.0118) | (0.0123) | (0.0117) | (0.0212) | (0.0129) | (0.00996) |
| Growth Regression | 0.133 | 0.138 | 0.147 | 0.213 | 0.101 | 0.101 |
| | (0.0221) | (0.0236) | (0.0213) | (0.0397) | (0.0232) | (0.0181) |
| Dispersion | | | | | | |
| 90:10 | 0.442 | 0.422 | 0.450 | 0.224 | 0.446 | 0.583 |
| | (0.0112) | (0.00998) | (0.0113) | (0.00650) | (0.0121) | (0.0144) |
| 75:25 | 0.233 | 0.222 | 0.237 | 0.118 | 0.235 | 0.307 |
| | (0.00588) | (0.00525) | (0.00596) | (0.00342) | (0.00636) | (0.00755) |
| Standard Deviation | 0.173 | 0.164 | 0.175 | 0.0874 | 0.174 | 0.227 |
| | (0.00436) | (0.00389) | (0.00442) | (0.00254) | (0.00472) | (0.00560) |
| Patients / 1000 | 3,530 | 3,530 | 3,530 | 3,530 | 2,702 | 2,702 |
| Hospitals | 5,346 | 5,346 | 5,346 | 5,346 | 5,180 | 5,180 |

Notes: Column (1) is baseline specification. In other columns the time horizon in which we measure survival and/or inputs is modified as indicated in the column headings. Productivity metrics reflect empirical bayes adjustment. Standard errors are bootstrapped with 50 replications and are clustered at the market level.

## Appendix D: Empirical-Bayes Adjustment

### Introduction

In this appendix we describe the Empirical Bayes (EB) procedure we use to adjust our estimates of hospital productivity for measurement error. This procedure is based on Morris (1983). For another example see Jacob and Lefgren (2007).

Definitionally, we call the productivity level of hospital $h$ at time $t$ $A_{ht}$ and call its log-productivity $a_{ht} = \ln(A_{ht})$. We will refer to these objects as the "true" log-productivities and their distribution as the "underlying" distribution of log-productivity. We do not observe these values, but instead estimate them. Call $\hat{a}_{ht}$ the estimated log-productivity. Thus we can write

$$\hat{a}_{ht} = a_{ht} + \eta_{ht}$$

where $\eta_{ht}$ is an error term. The goal of the EB procedure is to adjust our estimated log productivity $\hat{a}_{ht}$ so that the presence of the error term does not introduce bias into our regressions, which use our estimate of log productivity ($\hat{a}_{ht}$) as a key right hand side variable . The procedure adjusts the estimates by shrinking them toward the mean of the true, underlying productivity distribution.

Although true log-productivity is not observable, we show in this appendix that its distribution is estimable. We also show how this shrinkage estimator fixes the attenuation bias that measurement error would otherwise introduce into our regressions.

### Background on Empirical Bayes Procedure

*Statistical Background*

We start with an overview of the EB procedure assuming that all parameters of the distributions are known, and refer to the EB-adjusted estimated log productivity as $a_{ht}^{EB}$. We then describe the feasible EB-adjusted estimate, which we denote $a_{ht}^{EB(f)}$.

Suppose that the estimated log-productivities are independently normally distributed around the true log-productivities with known variance $\pi_{ht}^2$:

$$\hat{a}_{ht}|a_{ht}, \pi_{ht}^2 \sim N\left(a_{ht}, \pi_{ht}^2\right) \text{ independently}$$

One can think of $\pi_{ht}^2$ as the variance of the measurement error of the estimate.

We also assume that the true log-productivities $a_{ht}$ are independently normal with underlying mean $\theta_{ht}$ (known and allowed to differ for each hospital-year) and underlying variance $\sigma_a^2$ (known and common across hospitals within a year). In other words, each hospital's log-productivity is a random variable.

Then we have the prior distribution of the log-productivity $a_{ht}$:

$$a_{ht}|\theta_{ht}, \sigma_a^2 \sim N\left(\theta_{ht}, \sigma_a^2\right) \text{ independently}$$

We call it the prior because it is the distribution we envision before we condition on the estimated log-productivity. $\theta_{ht}$ may be most simply a mean $\mu$, but could also be a linear function of some covariates $x_{ht}\beta$.

If we condition on the estimated log-productivity $\hat{a}_{ht}$, then we get the posterior distribution of $a_{ht}$:

$$a_{ht}|\hat{a}_{ht}, \theta_{ht}, \sigma_a^2, \pi_{ht}^2 \sim N\left(a_{ht}^{EB}, \pi_{ht}^2\left(1 - B_{ht}\right)\right) \tag{1}$$

where $a_{ht}^{EB}$ denotes the Empirical-Bayes adjusted log productivity and

$$\begin{aligned} a_{ht}^{EB} &= \left(1 - B_{ht}\right)\hat{a}_{ht} + B_{ht}\theta_{ht} \\ B_{ht} &= \pi_{ht}^2 / \left(\pi_{ht}^2 + \sigma_a^2\right) \end{aligned}$$

The value $a_{ht}^{EB}$ -- the EB-adjusted estimate of $a_{ht}$ -- is the expected value of $a_{ht}$ conditional on the estimated value $\hat{a}_{ht}$ and the parameters $\theta_{ht}, \sigma_a^2$, and $\pi_{ht}^2$.

In other words, the EB-adjustment amounts to attenuating the estimate $\hat{a}_{ht}$ toward the mean $\theta_{ht}$. As the variance of the measurement error $\pi_{ht}^2$ rises, the EB correction increasingly disregards the value of the estimate and closes in on the mean.

*Feasible Version of Procedure*

The procedure just described assumes all parameters are known. This section describes how we implement the process when parameters must be estimated.

The log-productivity estimate $\hat{a}_{ht}$ is equal to the estimated coefficient on a hospital-year fixed effect from equation (5). The regression that produces the estimated coefficient also yields an estimated standard error. Recall that this standard error is an estimate of the standard deviation of the asymptotic distribution of $\hat{a}_{ht}$. In other words, it is an estimate of $\pi_{ht}$. We estimate $\pi_{ht}^2$ by squaring the estimated standard error of the hospital-year fixed effect. We call this estimate $\hat{\pi}_{ht}^2$.

Still, we lack estimates of $\theta_{ht}$ and $\sigma_a^2$. Morris (1983, section 5) describes how to estimate them, and we reproduce this method here. Note that it includes a degree of freedom correction to account for the estimation of these terms. This correction will usually be trivial in our applications.

Suppose that

$$\theta_{ht} = x_{ht}\beta$$

We will focus on estimating $\beta$. This example can be generalized to the case where $\theta_{ht} = \mu$ by letting $x_{ht} = 1$.

Fix estimates:

$$\begin{aligned} W_{ht} &:= \frac{1}{\hat{\pi}_{ht}^2 + \hat{\sigma}_a^2} \\ \hat{\beta} &:= (X'WX)^{-1} X'WA \\ \hat{\sigma}_a^2 &= \max\left\{0, \frac{\sum_{ht} W_{ht}\left\{\left(\frac{N_{ht}}{N_{ht}-N_x}\right)\left(\hat{a}_{ht} - x_{ht}\hat{\beta}\right)^2 - \hat{\pi}_{ht}^2\right\}}{\sum_{ht} W_{ht}}\right\} \end{aligned}$$

$X$ is the stacked $x_{ht}$, $W$ is a diagonal matrix of the $W_{ht}$, and $A$ is the stacked $\hat{a}_{ht}$. $N_{ht}$ is the number of hospital-years, or equivalently the number of $\hat{a}_{ht}$. $N_x$ is the number of regressors, i.e. the dimensionality of $x_{ht}$.

Looking closely, $\hat{\beta}$ is a WLS regression of the $\hat{a}_{ht}$ on $x_{ht}$. $\hat{\sigma}_a^2$ is the weighted average of the squared deviations of $\hat{a}_{ht}$ from $x_{ht}\hat{\beta}$ less the weighted average of $\hat{\pi}_{ht}^2$. The weights are $W_{ht}$, giving more weight to observations with less measurement error. The max operator is used because in finite samples $\hat{\sigma}_a^2$ could be negative without it.

$\hat{\beta}$ and $\hat{\sigma}_a^2$ are simultaneously determined in these equations, so we must perform an iterative procedure to estimate them. We start with a guess for $\hat{\sigma}_a^2$: the unweighted variance of the $\hat{a}_{ht}$ less the unweighted average of $\hat{\pi}_{ht}^2$. Then we iterate the following procedure to convergence:

1. Using the guess or estimate of $\sigma_a^2$, fix the vector $W_{ht}$
2. Compute $\hat{\beta}$ and then a new estimate $\hat{\sigma}_a^2$
3. If $\hat{\sigma}_a^2$ hasn't converged, return to step 1

Then with these estimates we can make a (feasible) best estimate of the posterior mean $a_{ht}^{EB(f)}$:

$$\begin{aligned} a_{ht}^{EB(f)} &= \left(1 - \hat{B}_{ht}\right)\hat{a}_{ht} + \hat{B}_{ht}x_{ht}\hat{\beta} \\ \hat{B}_{ht} &= \left(\frac{N_{ht} - N_x - 2}{N_{ht} - N_x}\right)\left(\frac{\hat{\pi}_{ht}^2}{\hat{\pi}_{ht}^2 + \hat{\sigma}_a^2}\right) \end{aligned}$$

We can also make an estimate of the variance of log-productivity unconditional on covariates, which we call $\hat{\varsigma}_a^2$. We take the formula for $\hat{\sigma}_a^2$ and replace the conditional mean $x_{ht}\beta$ with the weighted (unconditional) mean $\bar{A}$:

$$\begin{aligned} \hat{\varsigma}_a^2 &= \max\left\{0, \frac{\sum_{ht} W_{ht}\left\{\left(\frac{N_{ht}}{N_{ht}-1}\right)\left(\hat{a}_{ht} - \bar{A}\right) - \hat{\pi}_{ht}^2\right\}}{\sum_{ht} W_{ht}}\right\} \\ \bar{A} &= \frac{\sum_{ht} W_{ht}\hat{a}_{ht}}{\sum_{ht} W_{ht}} \end{aligned}$$

Actual Empirical Bayes Adjustment

We start with the squared standard errors $\hat{\pi}_{ht}^2$ of the estimated log-productivities from equation (5). Next, we assume $\theta_{ht} = \tau_M$, where $M$ is the market (HRR) of hospital $h$. Therefore the underlying mean of the productivities is market-specific.

To match earlier notation, suppose that $\tau_{HRR(h)} = z_{ht}\tau$. Therefore when estimating the underlying distribution of log-productivity, we make $z_{ht}$ a vector of 304 market indicators and $\tau$ a vector of 304 market fixed effects. The underlying log-productivities are distributed normally with underlying market level mean $\tau_M$ and common variance $\sigma_a^2$ independent of any other hospital-level covariates.

We perform the EB procedure separately on each year's estimated log-productivities. Thus the productivity process is estimated to have a market-year-specific underlying mean $\hat{\tau}_{M,t}$ and a year-specific variance $\hat{\sigma}_{a,t}^2$ (conditional on the market-year effects). Running the procedure also produces EB-adjusted log-productivities $a_{ht}^{EB(f)}$ and a year-specific variance $\hat{\varsigma}_{a,t}^2$ (unconditional on the market-year effects) .

Our procedure, which allows the EB adjustment to have a market-year specific mean, ensures that when the EB-adjusted log-productivities are used in our main regressions (equations (1) through (3)) which have market-year fixed effects, all regressors are orthogonal to the measurement error term. Generally, allowing the mean of log-productivity to depend on all other covariates in the regression maintains this orthogonality condition.

Reported productivity metrics

*Standard Deviation*

To estimate the standard deviation of productivity using the EB adjusted values, we rely on the estimates of the yearly underlying unconditional variance of log-productivity $\hat{\varsigma}_{a,t}^2$ that the procedure computes.[1] The root of these estimates is taken, forming $\hat{\varsigma}_{a,t}$. The yearly values are then averaged together.

The EB adjustment produces $\hat{\varsigma}_{a,t}^2$ by taking the weighted empirical variance of the $\hat{a}_{ht}$ and subtracting the weighted average squared standard error $\hat{\pi}_{ht}^2$. Hospital-years with larger standard errors receive lower weights. In effect, this process takes the variance of the noisy productivity estimates and subtracts off the variance due to measurement error.

*90:10 and 75:25*

We define the 90:10 ratio as $F^{-1}(0.9) - F^{-1}(0.1)$ and the 75:25 ratio as $F^{-1}(.75) - F^{-1}(.25)$ where $F^{-1}$ is the inverse CDF of the log-productivity distribution. In other words, the 90:10 is the

[1] While it might seem natural to instead estimate the standard deviation of the EB-adjusted values, this would cause us to erroneously under-estimate dispersion. Underlying log-productivity is composed of a best prediction (the EB-adjusted log-productivity) and the prediction error. These two components are orthogonal. The variance of true log-productivity is thus strictly greater than the variance of EB-adjusted log-productivity (see Jacob and Lefgren 2007).

90th percentile value of the distribution minus the 10th percentile value, and likewise for the 75:25. Since we are working with log-productivity, exponentiating these ratios would produce the 90:10 ratio of the productivity levels distribution (that is, an actual ratio: p90 / p10).

As with the standard deviation, it is not possible to estimate these ratios using the distribution of the $a_{ht}^{EB(f)}$. The EB correction does not produce a variable with the same asymptotic distribution as the underlying process. The procedure is only intended to estimate the parameters of an underlying normal distribution and correct for measurement error in regressions.

To estimate these ratios we consider the underlying normal distribution that the EB procedure uncovers, then compute the 90:10 and 75:25 of that distribution. Note that if $q \sim N(\tau, \kappa^2)$ then fixing $\Phi$ as the standard normal CDF, the 90:10 and 75:25 ratios of $q$ depend only on $\kappa^2$:

$$\begin{aligned} F^{-1}(0.9) - F^{-1}(0.1) &= \kappa\left[\Phi^{-1}(0.9) - \Phi^{-1}(0.1)\right] \\ F^{-1}(0.75) - F^{-1}(0.25) &= \kappa\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right] \end{aligned}$$

So we need only plug in the EB estimate of the standard deviation $\hat{\varsigma}_{a,t}$.

*Allocation Metrics (Patient, Growth, and Exit Regressions)*

Jacob and Lefgren (2007) show that a regression with un-adjusted estimated log productivities on the right hand side will not be consistent, but that with the adjustment the coefficients will be consistently estimated. To see this result, suppose that there is a relationship between growth $g_{ht}$, market-year fixed effects $\gamma_{M,t}$, and log-productivity $a_{ht}$:

$$g_{ht} = \gamma_{M,t} + \delta a_{ht} + \epsilon_{ht}$$

where $\mathbb{E}[\epsilon_{ht}|z_{ht}, a_{ht}] = 0$. Following earlier notation, $z_{ht}$ is a vector of indicators for the market-years. The left-hand side variable could alternatively be the number of patients or an indicator for hospital exit.

Since we do not observe true log-productivity, we use the estimate $\hat{a}_{ht} = a_{ht} + \eta_{ht}$ instead of $a_{ht}$, where $\eta_{ht}$ is measurement error. Then substituting into the equation:

$$g_{ht} = \gamma_{M,t} + \delta\hat{a}_{ht} + (\epsilon_{ht} - \delta\eta_{ht})$$

This shows that if we try to estimate $\delta$ by regressing $g_{ht}$ on market-year effects and $\hat{a}_{ht}$, the error term is $\epsilon_{ht} - \delta\eta_{ht}$. In this case we will get a biased and inconsistent estimate of $\delta$ due to the correlation between $\hat{a}_{ht}$ and $\eta_{ht}$ in the error term. Instead, we must use the EB-adjusted log-productivity $a_{ht}^{EB}$. Equation 1 states that:

$$\mathbb{E}\left[a_{ht}|\hat{a}_{ht}, \theta_{ht}, \sigma_a^2, \pi_{ht}^2\right] = a_{ht}^{EB}$$

Recall that we assume the underlying mean equals a market-year fixed effect, i.e. $\theta_{ht} = z_{ht}\tau = \tau_{M,t}$. We also replace $\sigma_a^2$ with $\sigma_{a,t}^2$ because the EB adjustment procedure is run one year at a time. The expectation becomes:

$$\mathbb{E}\left[a_{ht}|\hat{a}_{ht}, z_{ht}, \sigma^2_{a,t}, \pi^2_{ht}\right] = a^{EB}_{ht}$$

And therefore if we represent the prediction error of the EB procedure as $v_{ht}$:

$$a_{ht} = a^{EB}_{ht} + v_{ht}$$

Then by construction:

$$\mathbb{E}\left[v_{ht}|a^{EB}_{ht}, z_{ht}, \sigma^2_{a,t}, \pi^2_{ht}\right] = 0$$

(Note that we swapped $a^{EB}_{ht}$ for $\hat{a}_{ht}$ because given the parameters, knowing one determines the other)

Notice that the prediction error is orthogonal to $a^{EB}_{ht}$ and any regressor included in $z_{ht}$. Since the $z_{ht}$ are a set of market-year indicators, the prediction error is also orthogonal to market-year effects. Therefore if we regress $g_{ht}$ on market-year effects and $a^{EB}_{ht}$:

$$g_{ht} = \gamma_{M,t} + \delta a^{EB}_{ht} + (\epsilon_{ht} - \delta v_{ht})$$

We see that there is no correlation between any of the regressors and the error term. The consistency of $\delta$ follows as a result.

Comparison of estimates

We run all of our regression analyses with the EB-adjusted log-productivities $a^{EB(f)}_{h,t}$ and calculate our dispersion metrics using the EB-adjusted dispersion estimates as described above. Table A5 explores the impact of the EB-correction on our main results. The first column reproduces the EB-adjusted main results from Tables 2, 4, and 6. The second column shows the results without the EB correction.

To produce the uncorrected allocation metrics, we use the estimates $\hat{a}_{ht}$ rather than $a^{EB(f)}_{h,t}$ in our regressions. Due to measurement error in the estimates, the allocation metrics computed without the EB correction will be attenuated. To create the uncorrected dispersion metrics, we produce statistics that are comparable to the corrected versions, but calculate them with uncorrected estimates of log-productivity. For example, to calculate the standard deviation, the empirical weighted standard deviation of the estimated log-productivities -- $SD\left(\hat{a}_{ht}\right)$ -- is taken year-by-year, then averaged (we use the same weights that were used to calculate $\hat{\varsigma}^2_{h,t}$ so that the statistics are comparable.) Likewise, the 90:10 and 75:25 ratios are calculated by fitting a normal distribution to the estimated, uncorrected log-productivities and reporting the ratios implied by it (the ratios are calculated year-by-year, then averaged). Due to measurement error, the dispersion metrics computed without the EB correction will overstate the true dispersion.

The results show that the EB correction has a substantial effect on our baseline estimates of the expected sign. Comparing our baseline (EB-adjusted) estimates in column 1 with the un-adjusted version in column 2, we see that the allocation results are substantially larger and the dispersion

estimates are substantially lower with the correction. For example, we find that measurement error explains nearly half of the dispersion of the log-productivity estimates; without correcting for measurement error, these estimates have an average yearly standard deviation SD $(\hat{a}_{ht})$ of 0.293, while the EB procedure estimates that the underlying log-productivity process has an average yearly standard deviation $\hat{\varsigma}_{a,t}$ of 0.173.

A quantiatively large impact of the EB correction (i.e. a large amount of measurement error) is not surprising in light of results from other applications. For example, looking at estimates of teacher fixed effects in value added regressions, Jacob and Lefgren (2007) estimate a ratio of the unadjusted standard deviation to the EB-adjusted estimate of the standard deviation of about 1.3 to 1.6. We find ratios of about 1.7.

Table A5 - Sensitivity of Results to EB Adjustment

| | (1) | (2) |
|---|---|---|
| EB Adjustment: | Yes | No |
| Parameter | | |
| $\mu$ | 0.446 | 0.446 |
| | (0.00449) | (0.00449) |
| Static Allocation | 2.418 | 0.440 |
| | (0.0861) | (0.0170) |
| Dynamic Allocation | | |
| Exit Regression | -0.0329 | -0.0138 |
| | (0.0118) | (0.00428) |
| Growth Regression | 0.133 | 0.0373 |
| | (0.0221) | (0.00781) |
| Dispersion | | |
| 90:10 | 0.442 | 0.751 |
| | (0.0112) | (0.0125) |
| 75:25 | 0.233 | 0.395 |
| | (0.00588) | (0.00657) |
| Standard Deviation | 0.173 | 0.293 |
| | (0.00436) | (0.00487) |

Notes: Column (1) is baseline specification. Columns 2 shows results without the Empirical-Bayes adjustment. Standard errors are parametrically bootstrapped with 50 replications and are clustered at the market level.

**Appendix E: Static and dynamic allocation in concrete and healthcare**

We use data on concrete from the Census of Manufactures, which we have for every five years from 1972 – 1997. We observe approximately 2,500 ready-mixed concrete plants per data year; by way of comparison, we have approximately 3,700 hospitals per year. We use these data to estimate plants' physical total factor productivity levels. A plant's physical total factor productivity is the number of cubic yards of concrete it produces per unit input, where inputs are a weighted composite of labor, capital, and intermediates. The weights are the inputs' cost shares. These weights are theoretically correct, equalling the elasticities of output with respect to each input. Our market definition is the Bureau of Economic Analysis' Component Economic Areas, which are approximately 350 mutually exclusive and exhaustive groupings of economically interrelated U.S. counties. (See, e.g., Syverson 2004b for more details on productivity and market measurement in ready-mixed concrete.) To reduce the influence of outliers, we trim the top and bottom 1% of the industry's productivity distribution in each Census of Manufactures.

Table A6 reports the results. Across all of our static and dynamic allocation measures, the results indicate a stronger relationship between market allocation and producer productivity for hospitals than for concrete plants. The first row reports the results for static allocation. We estimate a slight variant of equation (1); as before, the specification regresses output on productivity (both measures are in logarithms) and market-year fixed effects. However, we now use lagged productivity on the right-hand side to facilitate comparisons between hospitals and concrete plants.[32] Strikingly, the correlation between output and lagged productivity is an order of magnitude larger in healthcare than in concrete.

The second row reports our exit analysis, based on equation (4) but modified to account for the fact that in concrete we only have data every five years; therefore, for purposes of comparability, we look at exit five years later for both hospitals and for concrete. However, comparability is limited by the fact that "exit" is defined quite differently in the two data sets.[33]

The final row reports our growth analysis. To make the analysis comparable across the two industries, for both we run the following regression:

$$\frac{N_{h,t+10}-N_{h,t+5}}{\frac{1}{2}(N_{h,t+10}-N_{h,t+5})} = \beta_0 + \beta_1 a_{h,t} + \gamma_{Mt} + \varepsilon_{ht} \qquad \text{(A3)}$$

Here, "size" (N) is defined as the number of patients in hospitals or the amount of physical output for concrete plants.[34]

[32] Due to how productivity is measured for concrete plants, regressing output on contemporaneous productivity would yield spuriously expanded coefficients: for concrete, output is effectively the numerator of the productivity measure. To fix the bias, we use the productivity measure from 5 years earlier on the right-hand side, rather than contemporaneous productivity, The lag is 5 years because data for concrete plants is only available at that frequency.

[33] In the concrete data, exit is directly observed; in the hospital data we infer "exit" based on the hospital having less than 5 patients for five consecutive years. Therefore, for concrete we regress an indicator for whether the firm has exited at year *t+5* on log productivity in year *t* (and market-year fixed effects). For hospitals, we regress an indicator for whether the hospital has less than five patients in every year from year *t+5* to year *t+9* on log productivity in year *t* (and market-year fixed effects).

[34] In order to make the growth analysis comparable for hospitals and for concrete, this regression differs from our baseline growth regression (equation 3) in two ways. First, because the concrete data is only available every five years, it looks at growth between 5 year periods rather than 1 year periods. Second, we lag the log productivity estimate on the right hand side back another time period. As in the static allocation metric, we do this because in manufacturing, our measure of size is output, which also enters the numerator of the productivity estimate; if there is

Table A6 - Allocation Metrics: Concrete vs Hospitals

| | Concrete | | | Hospitals | | |
|---|---|---|---|---|---|---|
| Risk Adjustment: | Estimate | DV Mean | Sample (Approx) | Estimate | DV Mean | Sample |
| Static Allocation | 0.299 | | 5,500 plant-years | 2.166 | 3.585 | 33,155 hospital-years |
| | (0.076) | | | (0.097) | | |
| Dynamic Allocation | | | | | | |
| Exit Regression | -0.066 | 0.20 | 12,400 plant-years | -0.147 | 0.17 | 25,359 hospital-years |
| | (0.018) | | | (0.032) | | |
| Growth Regression | 0.080 | -0.075 | 2,600 plant-years | 0.480 | -0.62 | 18,569 hospital-years |
| | (0.069) | | | (0.081) | | |

Notes: Estimates for concrete are based on data from the quinquennial Census of Manufactures from 1972-1992. Estimates for hospitals are based on Medicare AMI patients from 1993-2007 and use our baseline specification (see Table 2, column 1). See text for further details on metrics and data.

mean reversion in output and we had $a_{h,t+5}$ on the right hand side instead, this would create a negative bias on the $\beta_1$ coefficient.

**References (Additional for Appendices):**

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare Save Lives?" *The Quarterly Journal of Economics*, 124(2). 597-636.

Center for Medicare and Medicaid Services (CMS) 2011. "Hospital Acute Inpatient Services Payment System." http://www.medpac.gov/documents/MedPAC_Payment_Basics_11_hospital.pdf. Last accessed, July 10, 2012.

Clemens, Jeff and Josh Gottlieb. 2012. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" Unpublished mimeo. Available at http://www.people.fas.harvard.edu/~jdgottl/papers/DoctorsIncentives.pdf

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *The Quarterly Journal of Economics*, 127(3). 1057-1106.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5). 1127–50.

MedPAC 2010a. "Physician Services Payment System" http://www.medpac.gov/documents/MedPAC_Payment_basics_10_Physician.pdf Last accessed, July 10, 2012.

MedPAC 2010b. "Outpatient Hospital Services Payment System." http://www.medpac.gov/documents/MedPAC_Payment_Basics_10_OPD.pdf Last accessed, July 10, 2012.

MedPAC 2011a. "Hospital Acute Inpatient Services Payment System." Last accessed February 9, 2012 at http://www.medpac.gov/documents/MedPAC_Payment_Basics_11_hospital.pdf

Sloan, Frank. 2000. "Not-for-profit ownership and hospital behavior". *Handbook of Health Economics*, Volume 1B, Anthony Culyer and Joseph Newhouse (eds).