

# Saving Lives or Saving Money?

## Understanding the Dual Nature of Physician Preferences<sup>†</sup>

Alice Chen\* and Darius Lakdawalla\*\*

November 2015

[PRELIMINARY DRAFT]

Abstract

A longstanding literature has highlighted the tension between the altruism of physicians and their desire for profit. This paper develops new implications for how these forces drive pricing and utilization outcomes in healthcare markets. Altruism dictates that providers reduce utilization in response to higher prices, but profit-maximization does the opposite. Rational physicians will behave more altruistically when treating poorer, more vulnerable patients, and when the financial costs of altruism are lower. These insights help explain the observed heterogeneity in pricing dynamics across different healthcare markets. We empirically test the implications of our model by utilizing two exogenous shocks in Medicare price setting policies. Our results demonstrate that uniform policy changes in reimbursement or patient cost-sharing may lead to unintended consequences.

\* University of Southern California, Sol Price School of Public Policy. [alicejc@price.usc.edu](mailto:alicejc@price.usc.edu).

\*\* University of Southern California, School of Pharmacy and Sol Price School of Public Policy; and NBER. [dlakdawa@usc.edu](mailto:dlakdawa@usc.edu).

<sup>†</sup> We gratefully acknowledge comments from participants at the BU-Harvard-MIT Health Seminar, the USC CESR-Schaeffer Brown Bag, and the Midwest Health Economics Conference. Financial support provided by the National Institute on Aging of the National Institutes of Health under award number P01AG033559.

## I. Introduction

Economists have long emphasized the peculiarities of healthcare markets, compared to other markets for goods and services. Since at least Kenneth Arrow's pioneering paper on the subject, economists have recognized two features in particular: the altruism of healthcare providers towards their patients and the reliance of patients on their physicians for information and guidance (Arrow 1963). Less attention has been paid to the market pricing and utilization implications of these well-known insights into physician behavior.

Altruism encourages physicians to represent the interests of their patients. For example, an altruistic physician will tend to economize on the use of scarce inputs and attempt to maximize the utility of patients subject to their own resource constraints. However, the informational advantage of physicians creates a classic agency problem that physicians might exploit to pursue their own interests instead of their patients' interests. (See, for example, Blomqvist, 1991; Dranove and White 1987; Emanuel and Emanuel, 1992; Mooney and Ryan, 1993; Lu, 1999; and Zweifel and Breyer, 1997). These countervailing incentives produce similarly conflicting implications for pricing dynamics. Self-interested physicians will respond to higher price by doing more. Altruistic physicians, on the other hand, will protect their patients from higher prices by doing less. A full theory of healthcare pricing must thus present a unified framework for analyzing how altruism and agency problems interact.

Figure 1 provides some initial insight into the importance of this issue. The figure depicts the histogram of price elasticities within Medicare services that experienced an approximately 50% increase in annual physician reimbursement rates. Such large and sudden changes are unlikely to reflect changes in demand, but instead more likely to reflect movements along a demand curve. Price increases coincided with increased quantity in half the cases depicted, but with decreased quantity in the other half.<sup>1</sup> This pattern is difficult to explain when relying on either agency problems or altruism alone. Significant policy questions are at stake, since price is often viewed as an important lever for influencing behavior. In some market segments, for example, higher physician

---

<sup>1</sup>The large majority of the procedures depicted in Figure 1 are major or minor procedures. Very few are lab test, imaging, or evaluation and management services.

reimbursements can be expected to curb utilization, while in others, the opposite effect obtains.

In this paper, we study how altruism and agency problems compete to influence price and utilization, and we study the positive and normative implications of this competition. We rely on well-established models of physician behavior but apply these to problems of pricing and utilization that have not been viewed through the lens of physician preferences. From a positive standpoint, we show that exogenous price changes may increase or decrease quantity supplied. When higher prices lower quantity, we say dynamics are primarily “patient-driven,” and when the opposite is true, we say they are primarily “physician driven.” Moreover, the degree to which markets are patient-driven or physician-driven endogenously depends on physician incentives. Specifically, pricing is more likely to be patient-driven when patients are poorer and when healthcare provision is less profitable. In other words, physician altruism is more likely to win out when the value of behaving altruistically is higher and the cost is lower.

From a policy standpoint, patient-driven behavior limits the potential for overuse of healthcare resources, while physician-driven behavior exacerbates it. Thus, we expect less overuse, from the consumer’s perspective, when consumers are poorer, patient cost-sharing is higher, input prices are lower, and profitability is lower. Increases in patient wealth, therefore, are expected to increase “waste” in healthcare, as are expansions in the availability of insurance.

Empirically, we test the conjectures of our model by using two exogenous policy shocks to Medicare payments: the 1997 consolidation of geographic payment regions and the 1999 change in estimation of practice expenses. Our results indicate that the size and sign of the own-price elasticity does vary when there is joint decision making between patients and physicians. We also show that procedures are more likely to follow patient-driven pricing behavior when patient income is lower, patient cost-sharing is higher, and the physician’s price-cost margin is lower. We use these findings illustrate why uniform changes in payment or cost-sharing may not generate the intended responses in quantity.

Our main contribution is offering a theory that unifies theories of physician altruism and agency problems into a single framework for healthcare price theory. The literature thus far has offered piecemeal explanations of the observed heterogeneity in response to

price changes. Some empirical studies observe that higher reimbursements will lead to increased utilization, and the accompanying theory relies on physicians being profit maximizers (Clemens and Gottlieb, 2003; Gruber et al., 1999; and Jacobson et al., 2006). Other empirical studies show that there is a negative relationship between price and quantity (Escarce, 1993; Nguyen and Derrick, 1997; Rice, 1983; and Yip, 1998). Theories used to explain a negative price-quantity relationship include models of physician induced demand and non-fee-for-service reimbursement schemes. For example, Farley (1986) discusses implications of the target-income model. Ellis and McGuire (1986) demonstrate that having a prospective-payment system can lead to too few services being provided if physicians undervalue the benefits of patients relative to hospital profits, and Choné and Ma (2011) and Glied and Zivin (2002) discuss how managed care can restrict quantity. Finally, some studies find a low responsiveness between quantity and price, and they conclude that there is uncertainty in a physician's objective function (Holohan, 1977; Hurley et al., 1990; and Hurley and Labelle, 1995).

We unify these findings by offering a simple modification to the existing theory. In contrast to a number of important prior studies such as Ellis and McGuire (1986), Ellis and McGuire (1990), and Liu and Ma (2013), our model allows physicians to care about patient health *and* patient spending. This latter mechanism generates new insights on when services are likely to be patient-driven versus physician-driven. Our theory highlights that certain demand-side policies may be just as effective as supply-side policies in controlling costs.<sup>2</sup> This work relates to Dickstein (2014), who empirically quantifies the contributions of patient and physician incentives to prescription drug utilization.

Our findings have several notable policy implications. First, patient, physician, and procedure characteristics have systematic and predictable effects on healthcare price elasticities. Second, changes in physician reimbursement rules – for instance, in a public health insurance system – will have systematically different directional effects across different types of patients and procedures. If policymakers know these effects, they can better target reimbursement reforms and thus considerably strengthen their impacts.

---

<sup>2</sup> While policymakers have traditionally focused on controlling Medicare expenditures by altering Medicare payments, demand-side policies, such as changes to patient cost-sharing and supplemental insurance, are currently being debated (Gruber, 2013; National Commission on Fiscal Responsibility and Reform, 2010; and Zuckerman et al., 2010).

Finally, relying simply on aggregated estimates of price elasticities may lead to interventions with unintended consequences for certain patients, procedures, or markets.

The rest of the paper is organized as follows. In Section 2, we propose a theoretical framework for the joint decision-making between patients and physicians, and we derive the normative implications from our model. In section 3, we discuss the empirical approach for testing conjectures derived from our model. In Section 4, we present the empirical results, and Section 5 concludes.

## 2. Theoretical framework

Physician altruism and joint patient-physician decision making create unique relationships among pricing, utilization, and other economic forces. We demonstrate these points in a simple and standard theoretical model that traces back to Becker (1957). The model has been used by a number of health economists to study physician behavior (e.g., Ellis and McGuire 1986; Ellis and McGuire 1990; McGuire 2000; McGuire and Pauly, 1991)

### 2.1. Simple illustration

For pedagogical purposes, we first illustrate in a very simple, perfectly competitive model how physician and patient decisions interact. Initially, we assume that physicians earn zero economic profits, and patients bear the full cost of healthcare.

Suppose health is produced using a good or procedure  $X$ , according to  $F(X)$ , where  $F_{XX} < 0$ . We also suppose that this good is initially health-improving, but eventually health-reducing if overused. In this context, suppose for simplicity that a fully informed representative patient maximizes the value of health net of the cost of production. This results in the following household production function for health:

$$\max_X vF(X) - p_X X$$

It is straightforward to show in this context that the derived demand for  $X$  is falling in price

$$p_X, \text{ as in } \frac{\partial x}{\partial p_x} = \frac{1}{vF_{xx}} < 0.$$

Now, however, suppose that the representative patient is not fully informed but instead receives care from a physician, who bears cost  $c(X)$ , where  $c_{XX} \geq 0$ . The physician maximizes a weighted average of patient well-being and her own income, as in:

$$\max_X (1 - \alpha)[p_X X - c(X)] + \alpha[vF(X) - p_X X]$$

The parameter  $\alpha$  is an index of altruism. With relatively minor modifications, it can also be thought of equivalently as the patient's relative bargaining leverage in a Nash-bargaining problem between patients and physicians.

Observe in this framework that the physician's objective function can be rewritten as:

$$\max_X \alpha vF(X) - (1 - \alpha)c(X) + (1 - 2\alpha)[p_X X]$$

This has the following first-order condition:

$$\alpha vF_X - (1 - \alpha)c_X + (1 - 2\alpha)p_X = 0$$

Define  $D$  as the second derivative for this maximization problem. This allows us to write the comparative static of the problem as:

$$\frac{\partial X}{\partial p_X} = \frac{(2\alpha - 1)}{D}$$

If the problem is strictly concave at the optimum, then  $D < 0$ . As a result, if  $\alpha > \frac{1}{2}$ , the own-price elasticity is negative, because the physician is sufficiently altruistic that her decision problem resembles that of the fully informed patient. We call these "patient-driven pricing dynamics." If, on the other hand,  $\alpha < \frac{1}{2}$ , the opposite dynamics prevail: the own-price elasticity is positive. We call these "physician-driven pricing dynamics."<sup>3</sup>

## 2.2 General model

The derivation above assumed physicians and patients are risk-neutral over consumption. It also abstracted from the existence of health insurance. To generalize the simple model, suppose the representative patient derives utility  $u(I + vF(X) - \pi - \sigma(X; p_X))$ , where  $I$  is income,  $\pi$  is an ex ante insurance premium, and  $\sigma$  represents patient out-of-pocket expenditures. For the bulk of the analysis, we assume that patients face some cost-sharing in the sense that  $\sigma_X > 0$ , and in the sense that they bear costs when prices rise (i.e.,  $\sigma_{Xp_X} > 0$ ). Later, we discuss the polar case of zero cost-sharing. Here and elsewhere, we abstract from effects of physician decision-making on the insurance

---

<sup>3</sup> This model can be easily extended to identify the effects of a substitute good or procedure. When the own-price elasticity is positive, the cross-price elasticity is negative, and vice versa.

premium. This assumption sacrifices little generality in a public insurance scheme or when studying a relatively small set of procedures.

Now suppose physicians derive utility from a weighted average of patient utility and their own utility over consumption,  $z(\cdot)$ , where  $u$  and  $z$  are weakly concave utility functions. Physicians may also earn some non-labor income  $N \geq 0$ . Assume the physician utility function satisfies the assumptions of monotonicity, risk-aversion, and weak prudence, as in  $z' > 0$ ,  $z'' < 0$ , and  $z''' \geq 0$  (Felder & Mayrhofer, 2011). The generalized physician objective function can then be written as:

$$\max_X (1 - \alpha)z(N + p_X X - c(X)) + \alpha u(I + vF(X) - \pi - \sigma(X; p_X))$$

The first-order conditions now become:

$$\alpha u' * (vF_X - \sigma_X) + (1 - \alpha)z' * (p_X - c_X) = 0$$

The optimality conditions are weighted averages of physician profit-maximization and patient utility-maximization.

To simplify the analysis, we follow the convention adopted in much of the insurance literature and abstract from the direct income effects associated with patient out-of-pocket payments (Lakdawalla & Sood, 2013). This amounts to holding  $u'$  fixed when prices change. The comparative static now become:

$$\frac{\partial X}{\partial P_X} = \frac{(\alpha u' \sigma_{Xp_X} - (1 - \alpha)z' - (1 - \alpha)z''(p_X - c_X)X)}{D}$$

This comparative static suggests a simple empirical test for the presence of physician altruism. Absent altruism, own-price elasticities will always be positive. To see this, observe from the physician's optimality condition that  $p_X = c_X$  in the absence of physician altruism. Therefore, without altruism, it follows that  $\frac{\partial X}{\partial p_X} = -\frac{(1-\alpha)z'}{D} > 0$ . If price increases lead to quantity reductions in real-world data, this necessarily signals the presence of altruism.

To develop further the implications of the comparative statics, we investigate how changes in exogenous parameters change the likelihood of patient-driving pricing dynamics. Before beginning the formal derivation, we impose one final assumption, namely that the patient's own private marginal benefit exceeds her marginal out-of-pocket cost, or  $vF_X \geq \sigma_X$  at the optimum. The asymmetry of information means this is not a trivial

assumption, but – at least for insured consumers -- it would only be violated in fairly extreme cases of overuse. This assumption, coupled with the first-order condition for  $X$ , implies that  $p_X \leq c_X$ .

Pricing dynamics are patient-driven if and only if  $\alpha u' \sigma_{Xp_X} > (1 - \alpha)z' + (1 - \alpha)z''(p_X - c_X)X$ . Thus, it is straightforward to show that they are more likely to be patient-driven if:

1. Physician altruism is higher – i.e.,  $\alpha$  is higher;
2. Physician non-labor income is higher – i.e.,  $N$  is higher, which implies that  $z'$  and  $z''(p_X - c_X)$  are both lower;
3. Patient income is lower – i.e.,  $I$  is lower and thus  $u'$  higher;
4. Patient out-of-pocket spending is higher – i.e.,  $\pi + \sigma$  is higher and thus  $u'$  higher;
5. The physician's price-cost margin,  $p_X - c_X$ , is lower.

Intuitively, pricing is more likely to be patient-driven if: physicians care more about their patients (#1); physicians are richer and willing to pay more to purchase patient welfare (#2); patients are more sensitive to spending growth (#3 and #4); and the opportunity cost to physicians of boosting utilization is lower (#5). Finally, note that in a zero cost-sharing environment,  $\sigma_{Xp_X} = 0$ , which implies that pricing is always physician-driven. Intuitively, even altruistic physicians have no incentive to worry about patients' financial impacts when there is no cost-sharing.

Finally, it is worth discussing the forces that move the slope of the equilibrium demand,  $\frac{\partial X}{\partial P_X}$ . This can be written as:<sup>4</sup>

$$\frac{\partial X}{\partial P_X} = \frac{(\alpha u' \sigma_{Xp_X} - (1 - \alpha)z' - (1 - \alpha)z''(p_X - c_X)X)}{D}$$

In words, the responsiveness of input usage to price is equal to the marginal return to input usage divided by the second-order condition,  $D$ . Above, we demonstrated that this marginal return: 1) falls with physician altruism; 2) falls with physician non-labor income; 3) rises with patient income; 4) falls with patient out-of-pocket spending; and 5) rises with the physician's price-cost margin. Thus, holding the second-order condition constant, the

---

<sup>4</sup> The determinant  $D$  is equal to  $\alpha u''(vF_X - \sigma_X)^2 + \alpha u'(vF_{XX}) + (1 - \alpha)z''(p_X - c_X)^2 + (1 - \alpha)z'(-c_{XX})$ .



same factors that make patient-driven pricing more likely also push the slope of the demand curve downwards.

The effects on the slope of the demand curve become ambiguous if movements in the second-order condition overwhelm changes in the marginal return function. Thus, it becomes an empirical question as to whether the same forces that make patient-driven pricing more likely also reduce the slope of the demand function in each individual case.

### 2.3. Policy implications

From a welfare perspective, the degree of inefficient input overuse depends on moral hazard and on the over- (or under-) reimbursement of physicians. In patient-driven markets, moral hazard is relatively more important to address, while physician reimbursement is more important in physician-driven markets.

To understand these results, observe that Pareto-efficiency requires the standard input efficiency conditions,  $vF_X = c_X$ . Thus, we can characterize the degree of inefficient overuse by quantifying  $c_X - vF_X$ . By inspecting the first-order conditions for physician decisionmaking, we can derive:

$$c_X - vF_X = \frac{\alpha u'}{\alpha u' + (1 - \alpha)z'} \overbrace{\frac{1}{(c_X - \sigma_X)}}^{\text{Moral hazard}} + \frac{(1 - \alpha)z'}{\alpha u' + (1 - \alpha)z'} \overbrace{\frac{1}{(p_X - vF_X)}}^{\text{Over-reimbursement}}$$

This condition demonstrates that both moral hazard and physician over-reimbursement play a role in input efficiency. The overall degree of input inefficiency is the weighted average of these two sources, with the weights given by the relative importance of patient versus physician consumption. If physicians are perfectly altruistic, the over-reimbursement effect vanishes. On the other hand, if they are perfectly self-interested, the moral hazard effect vanishes. In addition, note that increases in physician consumption levels place more weight on the moral hazard effect, because richer physicians place more value on their patients' consumption than their own.

The relative importance of physician versus patient consumption has implications for which policy levers are most efficient at reducing distortions. If the degree of altruism is high, reimbursement reforms aimed at patients will be relatively more effective. If low, on the other hand, reforms aimed at physician reimbursement will be correspondingly more effective. Put differently, policymakers should focus more on moral hazard in

patient-driven markets, but on physician reimbursement in physician-driven markets. More formally, holding all patient and physician incentives constant, reimbursement reforms that compress  $p_X - vF_X$  will contribute less to efficiency when  $\alpha u' > (1 - \alpha)z'$ , and vice-versa.

Factors that promote physician altruism tend to promote the importance of moral hazard, while factors promoting physician self-interest do the opposite. Thus, increases in patient wealth will tend to increase the importance of aligning physician incentives through payment reform. In contrast, aligning patient incentives becomes more important among poorer populations, where physicians pay less attention to their own financial incentives.

Our analysis also has implications for global reimbursement reforms that affect many markets or procedures at once. The effect of price on quantity may be positive or negative. Thus, uniform reimbursement changes – either global increases or global decreases in price – may have unintended consequences that depend on the mix of patient-driven versus physician-driven markets or procedures. Targeted reforms that change reimbursement for some markets, but not for others, might be dramatically more effective. We return to this point in the empirical analysis.

### **3. Empirical Analysis**

The theoretical analysis generates at least five testable implications:

1. Both the size and even the sign of the price elasticity may vary when there is joint decision making between patients and physicians.
2. When patient income is lower, price elasticities are more likely to reflect patient-driven pricing behavior.
3. When patient cost-sharing is higher, price elasticities are more likely to reflect patient-driven pricing behavior.
4. When the physician's price-cost margin,  $p_X - c_X$ , is lower, price elasticities are more likely to reflect patient-driven pricing behavior.

5. Physician payment reforms have a larger effect in market segments where pricing is physician-driven than elsewhere.

### 3.1. Data

To test these implications, we rely on data from 1993 to 2002 from the Center for Medicaid and Medicare (CMS) Medicare Carrier Claims File (CCF) and the Medicare Current Beneficiary Survey (MCBS). The CCF data contains the fee-for-service Physician/Supplier Part B claims for a random 5% sampling of Medicare enrollees. For each service provided, we have information on the price, including the co-pay, deductible, physician submitted charge, and Medicare allowed amount.<sup>5</sup> All prices are converted to 2010 dollars using the Consumer Price Index (CPI) for medical expenditures. The CCF also provides information on patient diagnoses and basic demographics, such as age, race, and gender. We define a market-area using the Dartmouth Atlas' Hospital Referral Region (HRR).

The MCBS data consists of a smaller, but still nationally representative, sample of 12,100 Medicare beneficiaries. By combining patient surveys with administrative payment files, the MCBS data provides a richer set of covariates that allow us to identify patient income and education. While neither dataset provides information on whether copayments are paid by the patient or a third party payer, the MCBS allows us to exclude those enrolled in a Medigap policy that covers patient co-pays.<sup>6</sup> This exclusion allows us to avoid the case of zero patient cost sharing. Due to MCBS' small sample size, we rely on the CCF dataset when possible and use the MCBS when considering patient income and cost sharing.

### 3.2. Medicare Payments and Policy Shocks

For each HCPCS, CMS calculates a payment based on three factors: (1) a relative value unit (RVU), (2) a geographic adjustment factor (GAF), and (3) a conversion factor (CF).<sup>7</sup> RVUs are procedure specific, and they reflect differences in the time, skill, training,

---

<sup>5</sup> The submitted charge is the amount physicians bill Medicare. The allowed amount is what Medicare actually pays for the procedure, which is described in more detail in Section 3.2.

<sup>6</sup> Eight of the ten Medigap policies cover 100% Part B co-pays. The other two policies cover 50% to 75% of Part B copays.

<sup>7</sup> The exact formula for calculating Medicare payments is given by

$$Pay = [RVU_W GPCI_W + RVU_{PE} GPCI_{PE} + RVU_{MP} GPCI_{MP}] \times CF,$$

and costs required to perform different procedures. GAFs are region-specific, so they account for geographic variation in the cost of providing services.<sup>8</sup> Finally, the CF is a nationally uniform adjustment factor that converts RVUs into a dollar amount. This factor is updated annually by CMS according to a formula specified by statute, but Congress can and has overridden the statutorily defined formula.<sup>9</sup>

To measure price elasticities, we need to identify payment changes within a market that are independent of patient demand, technological change, and supply. First, we consider changes to the overall Medicare payment rate, which will include variation from RVUs, GAFs, and the CF. Since GAFs are set across several different markets and CF is one number set nationally, these two components of Medicare pricing are likely exogenous to the dynamics within any one given market. However, variation in RVUs may not be exogenous within a market over time. At least once every five years, about 138 physicians from the Specialty Society Relative Value Committee (RUC) and its advisory committee convene to re-evaluate and assign RVUs. Their main objective is to adjust the work component of RVUs to reflect procedural differences in physician time, skill, and training. If adjustments in RVUs are systematically correlated with demand for a procedure, then price elasticity estimates based on RVU variation may be biased.

While changes in work RVUs may be non-random in theory, the practical case for bias is less clear. The assignment of relative weight is complex and political with battle lines and alliances drawn between specialties (Eaton, 2010). Deliberations are complicated by the fact that the size of the Medicare payment pie is fixed. As such, the final weights have been viewed as being somewhat arbitrary. For example, after the first major review of RVUs, the Health Care Financing Administration (HCFA) received “voluminous identical comments from family practitioners stating that [the HCFA had...] used an arbitrary method for revising the work RVUs” (Department of Health and Human Services, 1996).<sup>10</sup>

---

where  $W$  indexes the work component,  $PE$  indexes the practice expense component, and  $MP$  indexes the malpractice expense component.  $GPCI$  represents the geographic practice cost indices, and  $CF$  is the conversion factor.

<sup>8</sup> GAF is a weighted sum of the work, practice expense, and malpractice GPCIs. Details can be found in MaCurdy et al. (2012).

<sup>9</sup> The CF in 2013 was \$36.61 per RVU. Congress has overridden this formula in 1998, 2009, and 2011.

<sup>10</sup> Between 199 to 2002, work RVUs experienced two major reviews which became effective in 1997 and 2002. The change in average work RVU is depicted in Appendix Figure 1.

To address potential endogeneity, we rely on two policy shocks in Medicare pricing. The first major policy shock occurred in 1997 when the Healthcare Financing Administration (HCFA) consolidated the number of geographic payment regions from 210 distinct payment regions to only 89 distinct regions in 1997. Discussed in Clemens and Gottlieb (2014), this consolidation generated differential price shocks across county groupings within a state. While some states were unaffected by this policy, in about 26 states, the variation in reimbursement rates across counties was either significantly reduced or eliminated because multiple regions were collapsed into one single payment area.

While the 1997 policy shock affected payments across geographic areas, a second major policy shock in 1999 created differential changes across services. Prior to 1999, practice expense RVUs (PE-RVUs) were measured using prevailing charges. However, Section 121 of the Social Security Amendments of 1994 and the Balanced Budget Act of 1997 mandated that PE-RVUs be determined by relative costs, instead of prevailing charges. Phased in over a four-year period from 1999 to 2002, the modified PE-RVU calculations better differentiated between the costs of performing a procedure in a facility setting—such as a hospital, skilled nursing facility, or ambulatory surgical center—and a non-facility setting, such as an office or clinic.<sup>11</sup>

Figure 2 depicts the variation in the GAF and PE-RVU components of Medicare reimbursements over time. Using data from Federal Register reports, plot (a) depicts the change in GAF among counties that were affected by the 1997 consolidation versus those that were unaffected. It is clear that much of the pre-1997 differentiation across counties was eliminated post-1997. Plot (b) shows the change in average facility and non-facility PE-RVUs across HCPCS over time. While the transition from charge- to resource-based estimations was phased in over a four-year period, the differentiation between facility and non-facility RVUs created a large drop in average PE-RVUs over time. As Appendix Figure A1 depicts, much of the observed drop in PE-RVUs in 1999 comes from changes in the non-facility estimates. Changes in the other components of Medicare reimbursements are discussed in Appendix A.

---

<sup>11</sup> Prior to 1999, the non-facility PE-RVU was simply 50% of the facility PE-RVU (Maxwell and Zuckerman, 2007).

### 3.3. Empirical Approach

In our baseline specification, we use data at the HCPCS-HRR-year level to estimate the following equation:

$$\log(Q_{iht}) = \beta^i \log(P_{iht}) + \Gamma^i X_{iht} + \gamma_h^i + \eta_t^i + \xi_{ht}^i + \epsilon_{iht}. \quad (1)$$

$Q_{iht}$  is the count of claims recorded for HCPCS  $i$  in HRR  $h$  in year  $t$ .  $P_{iht}$  measures the allowed Medicare payment for the service.  $X_{iht}$  are HRR-specific determinants of quantity that change over time, including the Charlson Comorbidity Index (CCI) calculated according to Quan et al. (2005), beneficiary age, Black and Hispanic dummies, and gender.  $\gamma_h$  are market fixed effects,  $\eta_t$  are year fixed effects,  $\xi_{ht}^i$  is a market by year time trend, and  $\epsilon_{iht}$  is an idiosyncratic error term. The remaining variation used to estimate  $\beta$  comes from changes in the reimbursement rate for a procedure within a market over time. Robust standard errors are clustered by HRR.

Assuming that variation in Medicare payments for a specific HCPCS within a given market over time is plausibly exogenous to other unobserved changes in local health demand and supply, the  $\beta^i$  estimate denotes the own-price elasticity of HCPCS  $i$ , identified by variation in pricing within a market over time.<sup>12</sup> However, there are a number of reasons why exogeneity might fail. For example, given the political nature of RVU changes, more popular procedures may draw a higher Medicare payment increase. Alternatively, changes in payments may reflect recent or contemporaneous changes in the cost of performing a given procedure. Although CMS uses the decennial census to determine certain indices, such as employee wage indices, it also uses the most recent retrospective data to determine other indices, such as office rental expenses. If costs are serially correlated, then changes in overall payment may be correlated with changes in costs. Finally, CMS updates RVUs based on comments submitted by physicians, health care workers, and professional associations and societies, increasing the likelihood of payment changes being correlated with other local supply factors (Federal Register).

In light of the potential threats to exogeneity, we rely on two policy shocks for identification. The 1997 geographic shock and the 1999 PE\_RVU procedure-specific shock

---

<sup>12</sup> An alternative to the HRR level is to use either the pre-1997 CMS geographic regions or counties to identify market areas. Results are similar when using both alternative measures.

generated exogenous variation in Medicare reimbursements that is arguably unrelated to the local demand for and supply of services. We use them as instruments for observed Medicare payments. Specifically, our first stage identifies the predictability of PE-RVU and GAF changes on overall Medicare payment changes within a market while controlling for the covariates specified in Equation (1). The PE-RVU policy shock differentially changes reimbursements for services performed in facility versus non-facility settings. The PE-RVU instrument is equal to the PE-RVU that would have resulted from the policy change alone, holding the share of facility and non-facility procedures fixed. We deliberately exclude changes in the share of facility procedures that may have resulted, since these could reflect physician preferences, altruism, or other confounding factors. Practically speaking, the PE-RVU instrument for HCPCS  $i$  in HRR  $h$  in year  $t$  is given by:

$$PE - RVU_{iht} = \begin{cases} s_{iht} * PERVU_{it}^f + (1 - s_{iht}) * PERVU_{it}^{nf} & \text{if } t < 1999 \\ \bar{s}_{ih} * PERVU_{it}^f + (1 - \bar{s}_{ih}) * PERVU_{it}^{nf} & \text{if } t \geq 1999 \end{cases}$$

where the  $f$  and  $nf$  superscripts denote facility and non-facility components, respectively.  $s_{iht}$  is the share of services performed in a facility setting for a given HCPCS-HRR-year. For post-1999 policy years, we use the average share  $\bar{s}_{ih}$  of services performed in a facility setting over the pre-policy years 1996 to 1998. The GAF instrument is simply the GAF for a given HRR-year. The second stage uses the instrumented variation to estimate price elasticities.

When using the two-stage least squares (2SLS) regression model to estimate Equation (1), we rely on price variation for each HCPCS that comes from within-market changes in the procedure-specific PE-RVU component over time and within-market changes in the geographic component over time as the 210 to 89 GAF units do not align directly to the 306 HRR markets. Because Medicare payments are based on both PE-RVUs and GAFs, these instruments will be highly correlated with Medicare payments. These two policy shocks are also conditionally independent of other sources of change in quantity, strengthening the case for instrument validity.

## 4. Results

### 4.1. Prediction 1: Heterogeneity in Elasticities

First, we show that the size and sign of price elasticities may vary. Ordering HCPCS by their price elasticities, we plot the price elasticities estimated via OLS in Figure 3a and 2SLS with both instruments in Figure 3b. To counteract the problem of multiple comparisons, we apply a Bonferroni correction and show only estimates that are statistically significant at the level  $(0.05/3,691)$  level. Both subplots clearly indicate that there are two types of HCPCS: (1) patient-driven HCPCS with negative price elasticities, and (2) physician-driven HCPCS with positive price elasticities.<sup>13</sup>

To check if the observed heterogeneity in price elasticities is driven by weak instruments, we examine the distribution of first stage F-statistics for the HCPCS shown in Figure 3. As Panel A in Table 1 indicates, the 25<sup>th</sup> percentile of the F-statistic distribution is 9.63, suggesting that the PE-RVU and GAF instruments are sufficiently strong for at least 75% of the HCPCS. The 75<sup>th</sup> percentile F-statistic is 64.81. To test whether we should use one instrument or two, we perform a Sargan-Hansen test where the joint null hypothesis is that the overidentifying restrictions are valid. The first row of Panel B in Table 1 indicates that only 4% of HCPC-estimates have a J-Statistic with  $p\text{-value} < 0.10$ . In other words, for the majority of estimates, the p-value is large, and we cannot reject the null. Therefore, we rely on both instruments in subsequent IV estimates presented in this paper.

By comparing Figure 3a with Figure 3b, it becomes evident that the OLS estimates tend to be more negative than the 2SLS estimates. This is consistent with a story where higher costs, which are unaccounted for in the regressions, decrease utilization. It is also consistent with RUC showing preferential payment increases for less common procedures, perhaps because those services were considered to be undervalued. However, despite the differences between OLS and 2SLS, we cannot reject OLS as a valid approach. We test the endogeneity of the Medicare payment variable by examining the difference of two Sargan-Hansen statistics: one where payments are treated as endogenous (i.e., 2SLS) and another

---

<sup>13</sup> The presence of physician- and patient-driven HCPCS appear across all types of services. For example, when examining elasticities by Berenson-Eggers Type of Service (BETOS) codes, we find the presence of both positive and negative elasticities among each category of service (i.e., major procedures, minor procedures, imaging services, evaluation and management services, etc.).



where payments are treated as exogenous (i.e, OLS).<sup>14</sup> Unlike the Durbin-Wu-Hausman test, this statistic is robust to violations of homoskedasticity (Sargan, 1958; Hansen, 1982). Panel B of Table 1 lists the summary statistics for the p-value of the endogeneity test. Because p-values are large, we cannot reject the use of OLS in favor of IV.

While we have provided evidence that both the size and even the sign of the price elasticity may vary at the HCPCS level, this exercise can easily be replicated at the market level. We demonstrate the heterogeneity in price elasticities by HRR in Appendix Figure B1. In scattered areas across the US, increasing price reduces quantity, while in other areas, increasing price increases quantity.

It is important to note that estimating one average price elasticity across all procedures or markets will mask some notable differences in pricing dynamics between physicians and patients that we have highlighted. To further demonstrate this point, we collapse the data to the HRR-year level and estimate the price elasticity using a model akin to Equation (1). We show in Appendix Table B2 that the overall impact of price changes on quantity is positive. Column (2), which uses the GAF instrument, yields an aggregate elasticity of 1.251, which is comparable to conceptually similar estimates found in Clemens and Gottlieb (2014).

## **4.2. Prediction 2: Patient Income**

Next, we test the conjecture that lower patient income increases the likelihood of patient-driven pricing. To evaluate the effects of patient income, we rely on data from MCBS and consider socioeconomic status more broadly. Panel A of Table 2 shows the sample means when dividing the MCBS samples between patient- and physician-driven HCPCS. The means are weighted by the number of times each HCPCS is performed, which accounts for the relative importance of each HCPCS. Panel A demonstrates that on average, patient-driven HCPCS tend to contain claims for patients with lower incomes, fewer years of schooling, and a smaller likelihood of having employer-sponsored insurance. These conclusions hold true regardless of whether we split the sample using the OLS or 2SLS price elasticity estimates, and the difference in means between patient- and physician-driven HCPCS are statistically different at the 5% level.

---

<sup>14</sup> Under homoskedasticity, this test is numerically equivalent to a Hausman test (Hayashi, 2000).

To better identify whether patient-driven behavior is more likely when patients have lower incomes, we perform a two-step estimation at the patient-physician-year level. The purpose is to measure how the exogenous factors in our theoretical model change the fraction of physicians that behave in a patient-driven fashion.

First, we calculate price elasticities for each physician across different terciles of patient income. Second, for each tercile of patient income, we calculate the share of patient-driven physicians. If patient-driven behavior is more likely among patients with lower income, then the share of patient-driven physicians will be highest in the lowest tercile of income. More specifically, for each physician  $j$ , we estimate three elasticities using:

$$\log(\text{RVU}_{ijt}) = \beta_j^n \log(P_{ijt}) + \Gamma_j X_{ijt} + \epsilon_{ijt}. \quad (2)$$

Here,  $i$  indexes the patient and  $t$  indexes the year, and for each  $\beta_j^n$  estimate where  $n \in \{1,2,3\}$ , only patients in the  $n^{\text{th}}$  tercile of income are included. Note that income terciles are estimated across all patients, not within each physician. We account for differences in the intensity of services across patients by using the total RVUs consumed per patient as a measure for quantity. Prices are estimated as the average allowed charge per RVU.  $X_{ijt}$  is a vector of demographic characteristics for patient  $i$  treated by physician  $j$  in year  $t$ , including age, CCI, and three dummies for being male, white, or black. Elasticities are identified by variation across patients over time. In the second step, we calculate the share of elasticities that are negative. All standard errors are bootstrapped using 500 iterations.

The results are presented in Panel A of Table 3. When treating patients in the lowest tercile of income, 21.9% to 38.9% of physicians appear to be patient-driven. However, in the highest terciles of patient income, this rate ranges from 13.3% to 21%. For both the OLS and 2SLS estimates, the share of patient-driven physicians in the lowest tercile of patient income is statistically different from shares in the other two terciles at the 5% level.

Next, we consider how income affects the magnitude of price elasticities, rather than the probability of positive or negative elasticities. Again using the data at the patient-

physician-year level, we estimate the following model for physician  $j$  treating patient  $i$  in year  $t$ :

$$\log(\text{RVU}_{ijt}) = \beta \log(P_{ijt}) + \alpha \log(P_{ijt}) \times Z_{ijt} + \phi Z_{ijt} + \Gamma X_{ijt} + \xi_j + \eta_t + \gamma_i + \epsilon_{ijt}. \quad (3)$$

Here,  $Z_{ijt}$  represents the log of patient income. The coefficient on the interacted term ( $\alpha$ ) identifies the responsiveness of price elasticities to patient income. In our baseline specification, we include physician ( $\xi_j$ ) and year ( $\eta_t$ ) fixed effects. In a more stringent, subsequent specification, we add patient ( $\gamma_i$ ) fixed effects so that estimates are identified off of variation in a patient's income over time and variation in the physician-patient interaction over time.

Results are presented in Panel A of Table 4. All columns show that the interacted coefficient is positive, and most of the columns are statistically significant. The positive interaction term indicates that when patient or market-area income increases, elasticities are more likely to be higher, or more physician-driven. This finding presents evidence that the slope of the equilibrium demand ( $\frac{\partial X}{\partial P_X}$ ) does indeed rise with patient income.

### 4.3. Prediction 3: Patient Cost Sharing

Our third prediction is that patient-driven behavior is more likely when patient cost sharing is higher. Because this empirical implication applies only to patients with non-zero out-of-pocket costs (OOP), we rely on the MCBS data so that we can exclude patients with Medigap coverage. In Panel B of Table 2, we show summary means of patient cost sharing variables by patient- versus physician-driven HCPCS. Patient-driven HCPCS are correlated with higher OOP payments and higher coinsurance payments. Although the difference in deductible payments is not statistically different across procedure types, this finding is not surprising. Part B deductibles—set at \$100 per year in 2003—are constant across HCPCS. On the other hand, coinsurance payments are set at 20% of the Medicare specified-fee and therefore vary by HCPCS. OOP costs are defined as the sum of the deductible and coinsurance.

In Panel B of Table 3, we estimate Equation (2) at the physician-HCPCS-level data. While we divided income into terciles at the patient level, we divide OOP into terciles at the HCPCS level because the impact of OOP costs on price elasticities is more likely occur at the

service level, instead of the patient level. The MCBS indicates that patient-driven elasticities are more prevalent in services with higher OOP costs. While differences in the share of patient-driven elasticities in the second and third terciles of OOP costs are not statistically different, the lowest tercile of OOP cost has a smaller share of patient-driven elasticities (55%) that is statistically different from the two higher terciles (64% to 66%).

In Panel B of Table 4, we consider the effect of OOP costs on the magnitude of the price elasticity. We find that the interaction term between log price and the fraction of payments which are OOP is negative, which indicates that when patients are responsible for a larger share of the physician payment, price elasticities tend to look more patient-driven. This finding is true even when controlling for patient fixed effects.

#### **4.4. Prediction 4: Physician Price-Cost Margin**

Fourth, we test the conjecture that HCPCS are patient-driven when the physician's price-cost margin is lower. This is equivalent to testing that HCPCS are physician-driven when the physician's price-cost or profit margin is higher. Because we do not have data on costs, we construct two proxies to measure profitability using the allowed amount—which is what Medicare pays physicians—and the submitted amount—which is what physicians say they should be paid. The ratio between the allowed and submitted charge should indicate the percent of a physician's charges that are covered by CMS. Alternatively, the difference between the submitted and allowed charges should indicate the shortfall or the remaining cost to physicians that they must “cover” themselves because Medicare reimburses less than their proposed charges.

The first row of Panel C of Table 2 shows that physician-driven HCPCS are associated with procedures where Medicare covers a larger share of their requested payment. The second row shows that physician-driven HCPCS are associated with procedures where physicians incur a smaller cost from performing the procedure. Panel C of Table 3 indicates that patient-driven behavior is more prominent in the lowest tercile of profitability, and these differences are statistically different at the 5% level. Finally, Panel C of Table 4, which looks at the interaction between log prices and log physician profitability, also shows that when Medicare covers a larger share of a physician's requested payment, HCPCS tend to be more positive, i.e., more physician driven.

One may argue that submitted charges are biased by measurement error if the charges physicians submit have no bearing on the actual payment received. To address this concern, we note that changes in the submitted charge for a given procedure over time are less likely to be driven by random noise. Thus, we approach this conjecture by administering another test. For each HCPCS, we calculate two elasticities: one that uses profitability changes above the median and another that uses changes below the median. Results are shown in Appendix Table B1. We find that when changes in physician profitability are larger, HCPCS have a 0.06 to 0.09 higher probability of being physician-driven.

#### 4.5. Prediction 5: Policy Implication

One of our normative implications is that physician payment reforms – such as reductions in reimbursement— will have larger effects among physician-driven HCPCS. To empirically assess this hypothesis, we use the 1999 PE-RVU payment change as our physician payment reform policy and perform a two-step estimation procedure. First, we establish whether HCPCS are physician- or patient-driven using pre-policy data from 1993 to 1998. For each HCPCS  $i$ , we follow Clemens and Gottlieb (2014) and estimate:

$$\log(Q_{ht}^i) = \beta^i \Delta GAF_h * 1(t \geq 1997) + \Gamma^i X_{ht}^i + \eta_t + \delta_h + \epsilon_{iht}, \quad (4)$$

where  $\Delta GAF$  is calculated using the change in GAF from 1996 to 1997. Then, we run a second regression at the HCPCS-year level that examines whether the post-1999 PE-RVU shock led to larger quantity changes for physician-driven HCPCS. Specifically, the second regression utilizes data from 1998-2002, and we estimate separately for physician-driven HCPCS ( $\hat{\beta} > 0$ ) and patient-driven HCPCS ( $\hat{\beta} < 0$ ):

$$\log(RVU_{it}) = \beta \log(P_{it}) + \Gamma X_{it} + \eta_t + \delta_i + \epsilon_{it}, \quad (5)$$

We construct bootstrap standard errors.

One complication for interpreting the results is that coinsurance rates are 20% of Medicare reimbursement rates, making it difficult to separate the effects of physician payment changes from the effects of patient cost sharing changes. We address this issue by focusing on the Qualified Medicare Beneficiaries (QMBs). QMBs are Medicare beneficiaries who are also eligible for Medicaid, and they are not responsible for paying either the

Medicare deductible or Part B. Thus, changes in quantity among this population will only reflect responses to a physician payment change.<sup>15</sup>

Results are shown in Table 6. In Panel A, we consider all Medicare beneficiaries. In Panels B and C we consider dual- and non-dual eligibles. In all three panels, the response to a price change is larger for physician-driven HCPCS. However, the difference in quantity response between physician- and patient-driven HCPCS is largest and most statistically significant among the dual-eligible population.<sup>16</sup> One drawback of the data is that it does not allow us to differentiate between QMBs and other dual-eligibles. For example, Service Limited Medicare Beneficiaries (SLMBs) are not responsible for their deductible, but they are still required to pay the copay. Because we cannot isolate QMBs from other dual-eligibles, the difference between Panels B and C will likely be understated.

#### **4.6. Implications for Medicare Reimbursement Changes**

Our analysis suggests that changes in physician reimbursement rates may not always have the intended effects. To further illustrate this point, we consider four types of Medicare payment changes and show their corresponding responses in total quantity.

Suppose Medicare seeks to restrain utilization. Recall our finding and those of prior studies that Medicare's aggregate price elasticity is positive. Thus, the natural approach is to lower reimbursements. Suppose specifically that Medicare lowers reimbursements by 10% across the board. We use the IV-estimated elasticities from Equation (1) to calculate the change in total quantity when prices change by 10%. Depicted in Scenario 1 of Figure 7, the result is a 34% decline in total quantity.

Policies that reflect the systematic variation in price elasticities may achieve more precise outcomes. For example, consider a policy that lowers payments by 10% only for physician-driven HCPCS. The result is a 52% reduction in total quantity, larger than the broader across-the-board reimbursement cut. Moreover, policymakers can also exploit knowledge of specific price elasticities to restrain use in patient-driven HCPCS, where

---

<sup>15</sup> Also in 1997, the Balanced Budget Act reduced QMB cost-sharing rates. Post-1997, states were only required to cover cost-sharing rates up to the Medicaid reimbursement rate, instead of the Medicare reimbursement rate (Mitchell and Haber, 2003). This reduced the payments that physicians received, but it did not affect the zero cost-sharing policy among QMBs.

<sup>16</sup> Estimates with an OLS second stage model are shown in Appendix Table B3. The results are similar in this table.

reimbursement cuts led only to more waste. Raising reimbursements by 10% for those categories reduces overall utilization by 17%. Targeting reimbursement changes can achieve larger utilization declines for a given level of reimbursement reform.

An alternative is to implement policies that are market specific, as opposed to service specific. Scenario 3 indicates that a targeted 10% decrease in payments in physician-driven markets and a 10% increase in payments in patient-driven markets can lead to overall reductions in care. Finally, Scenario 4 contemplates the impact of future growth in patient income by increasing income across all markets. Changing income can magnify the total cost reduction stemming from reimbursement cuts in physician-driven markets.

## **5. Conclusion**

In this paper, we present a model of joint physician and patient decision-making. By examining how altruism interacts with profit-maximizing incentives, our model demonstrates that the effect of price changes on quantity can vary not only in magnitude, but also in direction. Economic theory provides insights into the likelihood of positive or negative price elasticities, which we identify as physician-driven or patient-driven pricing behavior. Specifically, patient-driven behavior is more common when patient income is low, patient health care spending is high, and when the physician's price-cost margin is low. We provide empirical evidence in support of these conjectures. The theory suggests two remaining implications that could be tested in future work: patient-driven behavior is more common when physician altruism is high and when physician income is high.

Our model also offers an important policy implication: physician reimbursement reforms that move reimbursements closer to the social value of inputs used will be more effective in reducing social inefficiency when pricing is physician-driven. While we do not structurally estimate the degree of social inefficiency in our data, we provide empirical evidence that suggests physician reimbursement reforms have a larger effect on physician-driven HCPCS.

The health economics literature has long recognized the tension between physician altruism and physician profit-maximization. In other healthcare contexts, economists have

developed elegant and tractable models accounting for this tension. We exploit these tools to generate novel testable predictions about pricing and utilization behavior in healthcare markets. Our analysis demonstrates that the unique preferences and objectives of physicians create pricing dynamics in healthcare that depart from those in other product markets.

These implications seem consistent with the data and provide useful guidance for policymakers and researchers. First, physicians are systematically more “altruistic” – in the sense of pursuing patient interests – when treating more vulnerable and disadvantaged patients. Second, heterogeneity in the effect of reimbursement changes is to be expected, and can be exploited to increase the effectiveness of reimbursement reforms.

Reimbursement reductions might be useful tools for containing costs when physicians are largely profit-maximizing, but they may be counterproductive when they are more altruistic. Being able to differentiate when a service or market is physician- vs. patient-driven will allow policy makers to more effectively target supply- and demand-side incentives. More generally, economic theory provides policymakers with guidance on the source and nature of variation in price elasticities. Suitably directed empirical analysis can help inform more targeted approaches to reforming reimbursement policy, particularly when the goal is to restrain or boost the quantity of healthcare utilized.

## References

- Andrews, D., M. Moreira, and J. Stock. “Optimal Two-Sided Instrumental Variables Regression.” *Econometrica* 74(3): 116-132.
- Arrow, K.J. (1963). “Uncertainty and the Welfare Economics of Medical Care.” *American Economic Review* 53: 941-973.
- Becker, G. (1957). *The Economics of Discrimination*. Chicago: The University of Chicago Press, 2<sup>nd</sup> edition.
- Blomqvist, A. (1991). “The Doctor as Double Agent: Information Asymmetry, Health Insurance and Medical Care.” *Journal of Health Economics* 10:411-432.

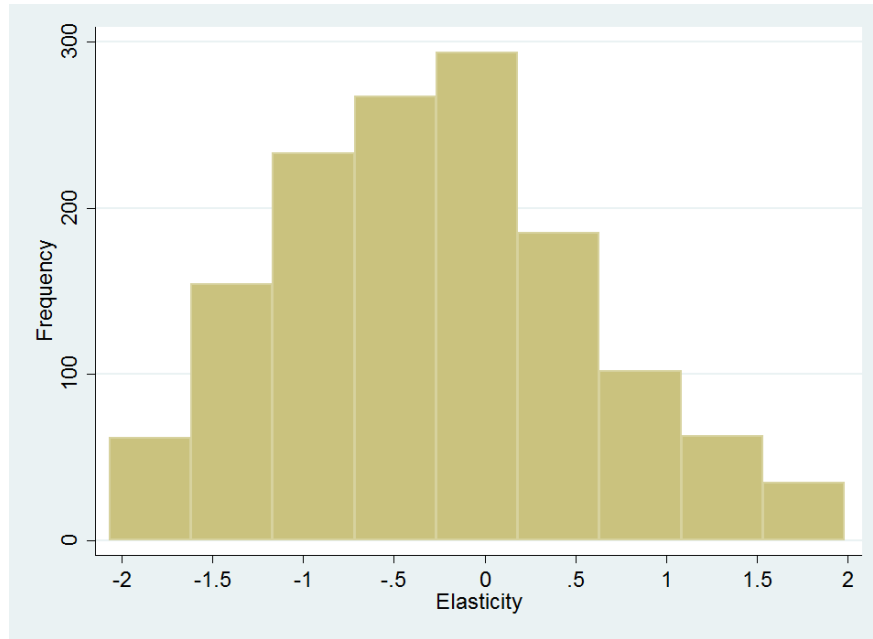


- Clemens, J. and Gottlieb, J. (2014). Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *American Economic Review*, 104(4): 1320-49.
- Clemens, J. and Gottlieb, J. (2013). Bargaining in the Shadow of a Giant: Medicare's Influence on Private Payment Systems. NBER Working Paper No. 19503.
- Chone, P. and Ma, C. (2011). Optimal Health Care Contract Under Physician Agency. *Annals of Economics and Statistics* 101/102: 229-256.
- Dickstein, M. (2014). Physician vs. Patient Incentives in Prescription Drug Choice. *Draft*.
- Dranove, D. (1988). Demand Inducement and the Physician-Patient Relationship. *Economic Inquiry*, 26(2): 281-296.
- Dranove, D. and W.D. White (1987). "Agency and the Organization of Health Care Delivery." *Inquiry* 24:405-415.
- Eaton, J. Little-known AMA Groups has Outsized Influence on Medicare Payments. *The Center for Public Integrity*. <<http://www.aafp.org/fpm/2008/0200/p36.html>>
- Ellis, R. and McGuire, T. (1985). Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply. *Journal of Health Economics* 5:129-151.
- Ellis, R. and McGuire, T. Optimal Payment Systems for Health Services. *Journal of Health Economics* 9:375-396.
- Emanuel, E. and Emanuel, L. (1992). Four Models of the Physician-Patient Relationship. *Journal of the American Medical Association*, 267(16), 2221-2226.
- Department of Health and Human Services. "Medicare Program; Revisions to Payment Policies and Five-Year Review of and Adjustments to the Relative Value Units Under the Physician Fee Schedule for Calendar Year 1997. *Federal Register* 61(227): Report 59490.
- Farley, P.J. (1986). "Theories of the Price and Quantity of Physician Services." *Journal of Health Economics* 315-333.
- Felder, S., & Mayrhofer, T. (2011). *Medical decision making : a health economic primer*. Heidelberg: Springer.
- Gaynor, M. (1994). "Issues in the Industrial Organization of the Market for Physician Services." *The Journal of Economics & Management Strategy* 211-255.
- Glied, S. and Zivin, J.G. (2002). How do Doctors Behave When Some (But Not All) of Their

- Patients are in Managed Care? *Journal of Health Economics* 21:337-53.
- Gruber, J. (2013). Proposal 3: Restructuring Cost Sharing and Supplemental Insurance for Medicare. *Brookings: The Hamilton Project*.
- Hadley, J. and R. Lee (1978). "Toward a Physician Payment Policy: Evidence From the Economic Stabilization Program." *Policy Sciences* 10:105-120.
- Hansen, L.P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50: 1029-1054.
- Hayashi, F. (2000). *Econometrics*. Princeton, NJ: Princeton University Press.
- Lakdawalla, D. and Sood, N. (2013). Health Insurance as a Two-Part Pricing Contract. *Journal of Public Economics*, 102, 1–12.
- Liu, T. and Ma, C. (2013). Health Insurance, Treatment Plan, and Delegation to Altruistic Physician. *Journal of Economic Behavior & Organization* 85:79-96.
- Lu, M. (1999). "Separating the 'True Effect' from 'Gaming' in Incentive-based Contracts in Health Care." *Journal of Economics and Management Strategy* 8:383-432.
- MacCurdy T., Sood, J., DeLeire, T., DeVaro, J., Bounds, M., Pham, D., and Chia, A. (2012). Geographic Adjustment of Medicare Payments to Physicians: Evaluation of IOM Recommendations. Acumen, LLC: Burlingame, CA.
- Maxwell, S. and Zuckerman, S. Impact of Resource-Based Practice Expenses on Medicare Physician Volume. *Health Care Financing Review* 29(2): 65-79.
- McGuire, T. (2000). Physician Agency. In: *Handbook of Health Economics, Volume I*. Culyer, A.J. and Newhouse, J.P. editors. Elsevier Science: 462-536.
- Mitchell, J.B. and Haber S.G. (2003). State Payment Limitations on Medicare Cost-Sharing: Impacts on Dually Eligible Beneficiaries and Their Providers. *Health Economics Research, Inc.* CMS Contract No. 500-95-0058.
- Mooney, G. and M. Ryan (1993). "Agency in Health Care: Getting Beyond First Principles." *Journal of Health Economics* 12:125-135.
- Pauly, M. (1980). Doctors and their Workshops: Economic Models of Physician Behavior." *National Bureau of Economic Research Monograph*.
- Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J., Saunders, L., Beck, C., Feasby, T. (2005). Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care*, 43(11): 1130-9.

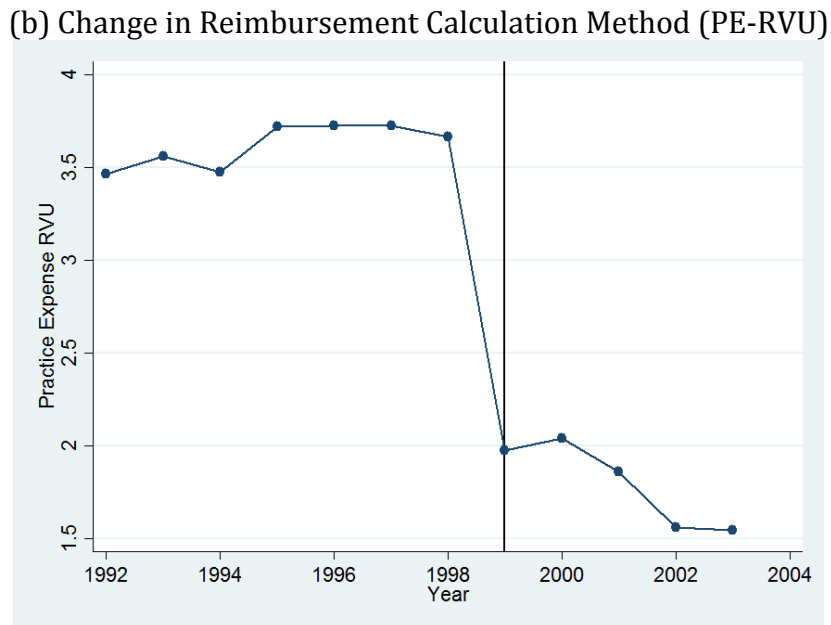
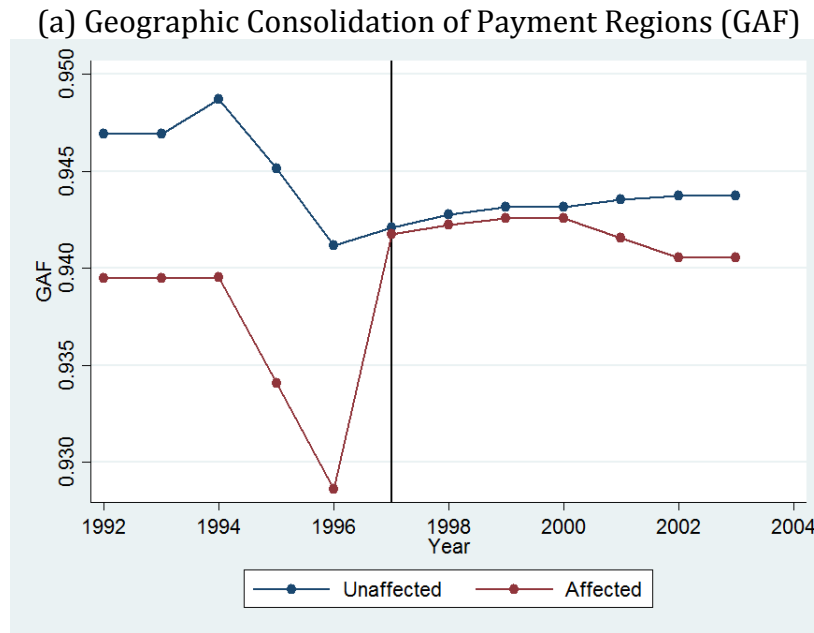
- Sargan, J.D. (1958). The Estimation of Economic Relationships Using Instrumental Variables. *Econometrica* 26: 393-415.
- Schwartz, A.L., Lnadon, E. Elshaug, A., Chernew, M., McWilliams, J. (2014). Measuring Low-value Care in Medicare. *JAMA Internal Medicine* 174(7):1067-76.
- The National Commission on Fiscal Responsibility and Reform. (2010). The Moment of Truth. Accessed February 26, 2015. <[http://www.fiscalcommission.gov/sites/fiscalcommission.gov/files/documents/TheMomentofTruth12\\_1\\_2010.pdf](http://www.fiscalcommission.gov/sites/fiscalcommission.gov/files/documents/TheMomentofTruth12_1_2010.pdf)>.
- U.S. Government Accountability Office (2005). *Medicare Physician Fees Geographic Adjustment Indices are Valid in Design, But Data and Methods Need Refinement*. Pub No. GAO-05-119. Washington DC.
- Zwefel, P. and F. Breyer (1997). *Health Economics*. Oxford University Press: New York.
- Zuckerman, S., Shang, B., and Waidman, T. (2010). Reforming Beneficiary Cost Sharing to Improve Medicare Performance. *Inquiry*, 47(3): 215-225.

Figure 1: Histogram of Elasticities When Prices Increase Significantly



*Notes:* Data from CMS Medicare 5% claims, 1992-2003. This figure shows the elasticities (calculated simply as the annual percent change in quantity divided by the annual percent change in price) for HCPCS with annual physician payment increases ranging from 45% to 55%. It is evident that quantity increases for about half of the HCPCS, while quantity falls for the other half. The long right tail has been truncated.

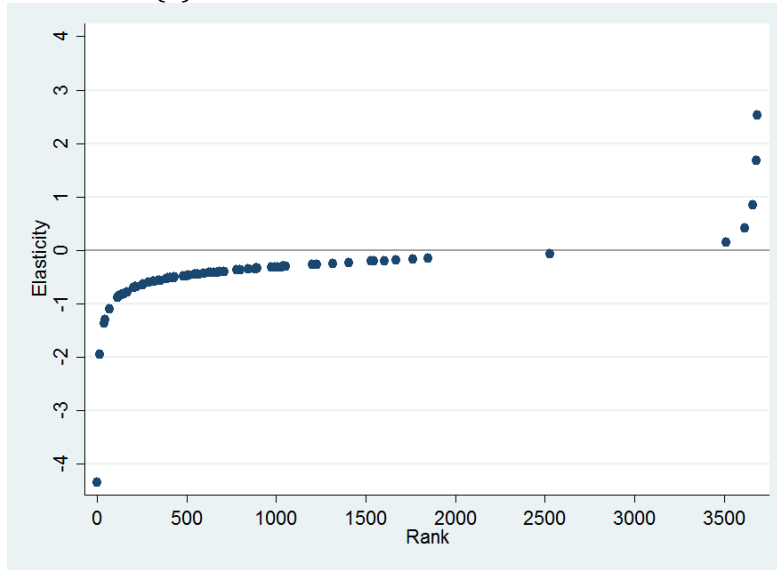
Figure 2: Shocks in Components of Medicare Payments Over Time



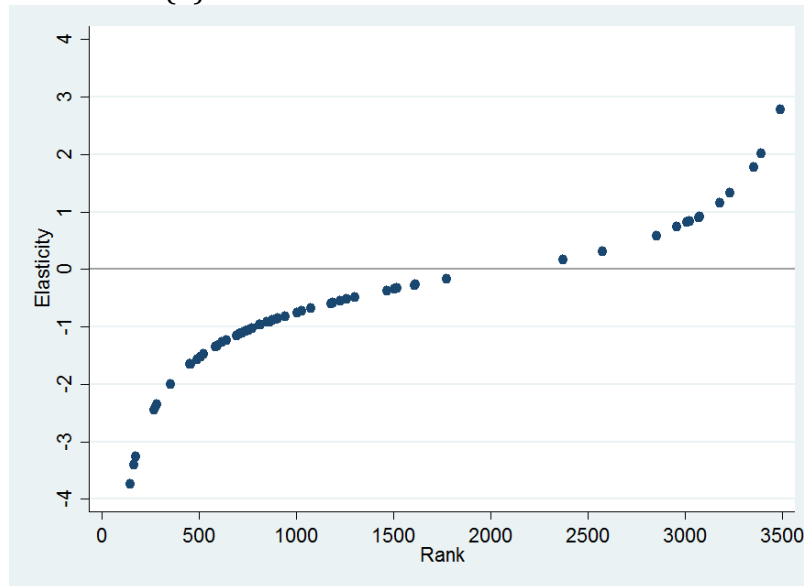
Notes: Data from the Federal Register 1992-2003. The sample is limited to HCPS observed in all years. Plot (a) shows the average GAF across counties that were or were not affected by the 1997 consolidation of payment regions from 210 to 89 payment regions. Plot (b) depicts the change in average of facility and non-facility PE-RVUs across HCPCS. In 1999, HCFA more accurately priced non-facility services and phased in a new methodology of calculating PE-RVUs.

Figure 3: Estimated Elasticities by HCPCS

(a): OLS-Estimated Price Elasticities



(b): 2SLS-Estimated Price Elasticities



*Notes:* Data comes from the CCF. Each dot comes from a separate regression of Equation (1); elasticities are ordered and plotted. To account for the multiple comparisons, a Bonferroni correction has been applied; for both plots, only HCPCS with statistically significant price elasticities with  $p\text{-value} < (0.05/3,691)$  are shown. In plot (b), the instruments are PE-RVU and GAF.

Table 1: Summary of IV Related Statistics

IV: PE-RVU + GAF	(1)	(2)	(3)
Panel A: First Stage F-Statistics			
	25 <sup>th</sup> percentile	Median	75 <sup>th</sup> percentile
First Stage F-Stat	9.63	20.89	64.81
Panel B: Endogeneity and Over-identification			
	Fraction p-value<0.10	Fraction p-value<0.05	Fraction p-value<0.01
Hansen J-Statistic	0.041	0.016	0.004
Endogeneity Test	0.14	0.078	0.019
No. of Regressions	73		

*Notes:* This table shows the IV related summary statistics used to estimate the statistically significant elasticities shown in Figure 3(b). Panel A shows the first-stage F-statistics when using both PE-RVU and GAF as instruments. Panel B shows (1) the distribution of p-values for the test that the Medicare price variable used in OLS is endogenous, and (2) Panel C shows the Hansen J-statistic for the test for the validity of using both instruments, instead of one or the other.

Table 2: Summary Statistics, by Sign of Own-Price HCPCS Elasticity

	OLS		2SLS	
	Patient-Driven (1)	Physician-Driven (2)	Patient-Driven (3)	Physician-Driven (4)
Panel A: Patient Socioeconomic Status (MCBS)				
Gross income (\$1000s)	21.28	23.06***	21.15	23.35***
Employer coverage	0.33	0.35***	0.33	0.36***
Less than high school	0.43	0.39***	0.37	0.38***
HS graduate	0.16	0.17***	0.43	0.17***
Some college	0.13	0.14***	0.16	0.14***
College grad or more	0.13	0.15***	0.13	0.15***
Black	0.11	0.094***	0.11	0.092***
Hispanic	0.043	0.042	0.043	0.041*
Panel B: Patient Cost-Sharing (MCBS)				
OOP (\$)	51.83	20.45***	48.38	31.01***
Coinsurance (\$)	49.87	18.41***	46.45	28.98***
Deductible (\$)	1.96	2.04	1.94	2.03
Panel C: Profitability (CCF)				
Percent Reimbursed (%)	55	62***	54	63***
Shortfall (\$)	419.19	143.40***	387.61	237.26***
No. of Obs. (MCBS)	675	929	884	794
No. of Obs. (CCF)	538	742	562	718

*Notes:* Data from 1993-2002 at HCPCS level. Data for Panels A is from MCBS. Data from Panels B and C are from CCF. Summary statistics are weighted by number of observations per HCPCS. Columns (1) and (2), or Columns (3) and (4) are statistically different at the \* 10% level, \*\* 5% level, \*\*\* 1% level. In Panel A, the insurance coverage and education variables are measures of the fraction of patients with each characteristic. In Panel B, *Any OOP* is the fraction of patients who had OOP>0. *Fraction OOP* is the average fraction of total payments attributed to out of pocket costs. In Panel C, *Fraction Reimbursed* is calculated by the share of payments CMS allows relative to the physician submitted charge (i.e., Allowed/Submitted). The *Shortfall* is the amount providers bill CMS minus the actual CMS payment (i.e., Submitted-Allowed).



Table 3: Physician Elasticities Among Various Patient Subgroups

	Fraction of Physicians with $\epsilon < 0$	
	OLS (3)	IV (4)
Panel A: Patient Income (MCBS)		
Tercile 1 Patients	0.389** (0.0172)	0.219** (0.0211)
Tercile 2 Patients	0.211 (0.0160)	0.112 (0.0191)
Tercile 3 Patients	0.210 (0.0155)	0.133 (0.0186)
No. of Physicians	1,538	1,538
Total Observations	25,996	25,996
Panel B: Patient Cost Sharing (MCBS)		
Tercile 1 HCPCS	0.577** (0.0164)	0.550** (0.0186)
Tercile 2 HCPCS	0.658 (0.0177)	0.637 (0.0244)
Tercile 3 HCPCS	0.666 (0.0161)	0.666 (0.0259)
No. of Physicians	9,827	9,827
Total Observations	126,070	126,070
Panel C: Physician Profitability (CCF)		
Tercile 1 Patients	0.768*** (0.058)	0.623*** (0.0417)
Tercile 2 Patients	0.431 (0.0717)	0.441 (0.0447)
Tercile 3 Patients	0.418 (0.0919)	0.499 (0.0457)
No. of Physicians	2,534	2,534
Total Observations	2,939,691	2,939,691

*Notes:* Each estimate comes from a separate two-step estimator. First, we estimate Equation (2) for each physician treating patients (Panels A and C) or performing HCPCS (Panel B) within a tercile group. Second, we calculate the share of physician elasticities that are negative (i.e., patient-driven). Bootstrapped standard errors are displayed in parentheses. Patients and services with zero out-of-pocket costs are excluded from the analysis. The MCBS columns additionally exclude patients with Medigap coverage. Estimates that are statistically different from all other numbers in the panel-column group is denoted by \*\* 5%-level and \*\*\* at the 1% level. For tractability, the CCF sample is restricted to physicians with at least 50 patients in each of the 10 years of data.

Table 4: Effect of Patient Characteristics on Measured Elasticities

† Dep. Var.: Log(RVU) $\mu = 4.09$ or $2.79$	OLS (1)	OLS (2)	2SLS (3)	2SLS (4)
Panel A: Patient Socioeconomic Status (MCBS)				
Log(Price)	-1.037*** (0.0195)	-1.045*** (0.0227)	-2.099*** (0.075)	-2.305*** (0.0942)
Log(Price) x Log(Income)	0.0173*** (0.00656)	0.0163*** (0.0247)	0.0210* (0.0136)	0.0517* (0.0312)
First stage F-stat	---	---	1,247	916
R-squared	0.892	0.923	0.663	0.639
Panel B: Patient Cost Sharing (MCBS)				
Log(Price)	-0.979*** (0.0149)	-0.947*** (0.0167)	-0.452*** (0.0711)	-0.587*** (0.0779)
Log(Price) x Fraction OOP	-0.00252 (0.00328)	-0.0131*** (0.0190)	-0.345*** (0.0190)	-0.316*** (0.0206)
First stage F-stat	---	---	516	437
R-squared	0.892	0.923	0.647	0.661
Panel C: Physician Profitability (CCF)				
Log(Price)	-0.242*** (0.000671)	-0.360*** (0.000852)	-2.960*** (0.00511)	-3.379*** (0.00658)
Log(Price) x Allow/Submit	0.0776*** (8.4E-5)	0.0796*** (9.45E-5)	0.0140*** (7.03E-5)	0.0122*** (7.51E-5)
First stage F-stat	---	---	137,458	96,952
R-squared	0.423	0.705	0.625	0.645
Year FE	Y	Y	Y	Y
Physician FE?	Y	Y	Y	Y
Patient FE?	---	Y	---	Y
No. of Obs (MCBS)	64,816	64,816	64,816	64,816
No. of Obs (CCF)	16,055,188	16,055,188	16,055,188	16,055,188

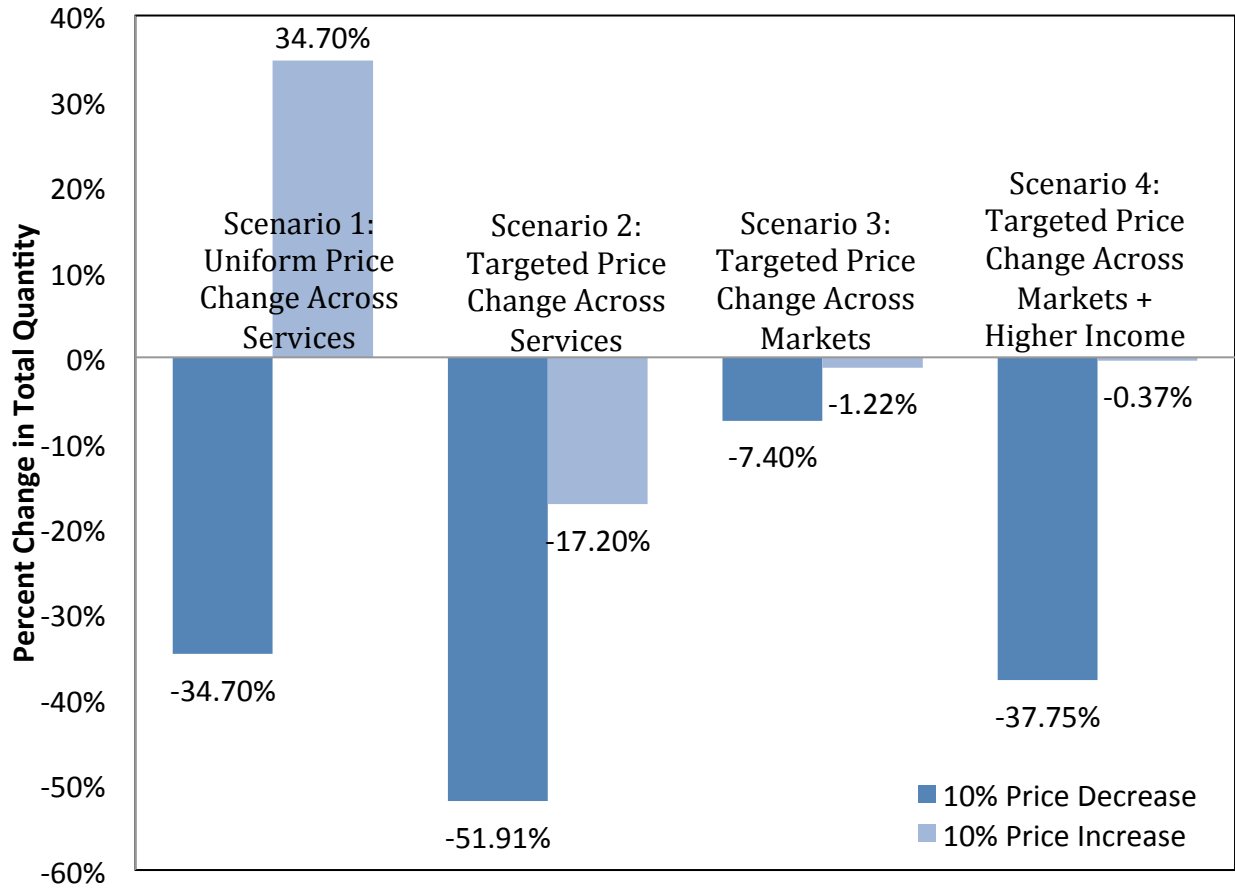
Notes: Each panel and column represents a separate regression at the patient-year level. Data for Panel A is from MCBS. Data for Panels B and C are from CCF. † First reported mean is from MCBS; second reported mean is for CCF. The dependent variable is log(total RVU). All regressions include the relevant characteristic (income, cost-sharing, or profitability). Columns (1) and (3) control for patient's CCI, age, male, white, black, and year by HRR fixed effects. Columns (2) and (4) control for patient's CCI, age, and person fixed effects. Robust standard errors are shown in parentheses. \* 10% level, \*\* 5% level, \*\*\* 1% level.

Table 5: Differential Effect of a Physician Reimbursement Reform

	Dependent Var: Log(RVU)		
	Physician- Driven HCPCS (1)	Patient- Driven HCPCS (2)	X <sup>2</sup> and P- value for H <sub>0</sub> : (1)=(2) (3)
	Panel A. All Beneficiaries		
Log(Price)	1.755*** (0.142)	1.000*** (0.320)	2.41 0.125
	Panel B. Dual Eligibles		
Log(Price)	1.630*** (0.150)	0.796*** (0.245)	5.07 0.024
	Panel C. Non-Dual Eligibles		
Log(Price)	1.784*** (0.138)	1.217*** (0.400)	3.89 0.067
First Stage F-stat	[60.3, 89.5]	[6.3, 17.1]	---
No. of Obs.	3,396	735	---

*Notes:* The physician-and patient-driven HCPCS are determined using CCF data from 1993 to 1998 and the GAF policy change. Each cell contains data from a separate regression using CCF data from 1998 to 2002. The dependent variable is Log(Total RVU) and independent variables include CCI, age, race, and gender dummies, year, and HCPCS fixed effects. Bootstrapped errors shown in parentheses. Column (3) shows the two-sided chi-squared and p-values for the hypothesis test that the elasticity estimates in Columns (1) and (2) are the same. \* 10% level, \*\* 5% level, \*\*\* 1% level.

Figure 4: Effects of Counterfactual Price Changes on Quantity



Notes: Scenario 1 shows the percent change in total RVUs performed when 2000 prices uniformly decrease by 10% (dark blue) or uniformly increase by 10% (light blue). Scenario 2 shows the percent change in total RVU when 2000 prices decrease by 10% for only the physician-driven HCPCS (dark blue) or increase by 10% for only the patient-driven HCPCS (light blue). Scenarios 3 and 4 show the targeted and uniform changes by HRRs, instead of HCPCS. Scenario 4 adds a 10% income increase across all HRRs. The IV-elasticity estimates, as shown in Figure 3b, are used to calculate the percent change in total RVU.

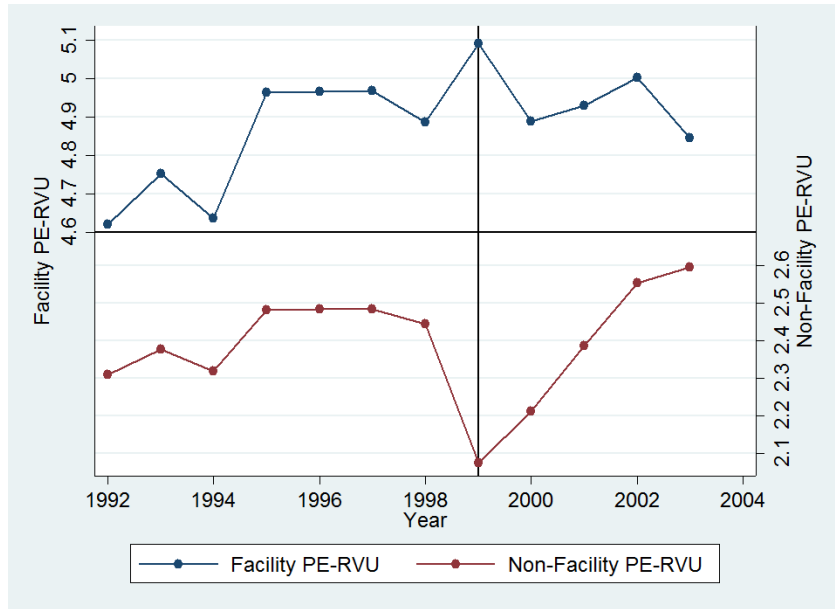
## Appendix A

In this section, we discuss the policy changes that affected the remaining Medicare components. As discussed in Section 3.2, we use the 1997 GAF and 1999 PE-RVU policy shocks as instruments for Medicare prices. The remaining variation in Medicare payments come from variation in the work RVU, malpractice RVU, and CF. On average, work accounts for 52% of total physician payments, practice expenses represent 44%, and liability insurance represents 4% (US Government Accountability Office, 2005). Because the malpractice component accounts for a small share of payments, we do not focus on that component.

During this time period, work RVUs experienced two major reviews which became effective in 1997 and 2002. Plot (a) of Figure A1 shows the average work RVU over time for HCPCS that are observed in each year of the study period. After the RUC committee met to re-assess work RVUs, we see clear jumps in the RVU. However, with competing political pressures and physician incentives, it is unlikely that RUC committee changes are exogenous to local demand and supply factors.

The CF also experienced a major change during the study period. Prior to 1998, there were three different CFs: one for surgery, primary care, and non-surgical services. The CF for surgical procedures led to surgeons earning a 17% bonus payment relative to all other procedures. This generated political discontent and led to a budget-neutral merger of CFs in 1998 (Clemens and Gottlieb, 2013). Plot (b) shows the CFs over time. After 1998, the CF for surgical procedures fell by about 11%, whereas the CF for non-surgical procedures increased by about 6%. We do not use this policy shock as another instrument for two reasons. First, CFs are constant across all geographic regions and all procedures, so their explanatory power for payment changes within in market area for a given HCPCS is weak. Second, the shock in CF payments occurs mainly for surgical procedures, while changes in CF for non-surgical and primary care procedures are much less pronounced.

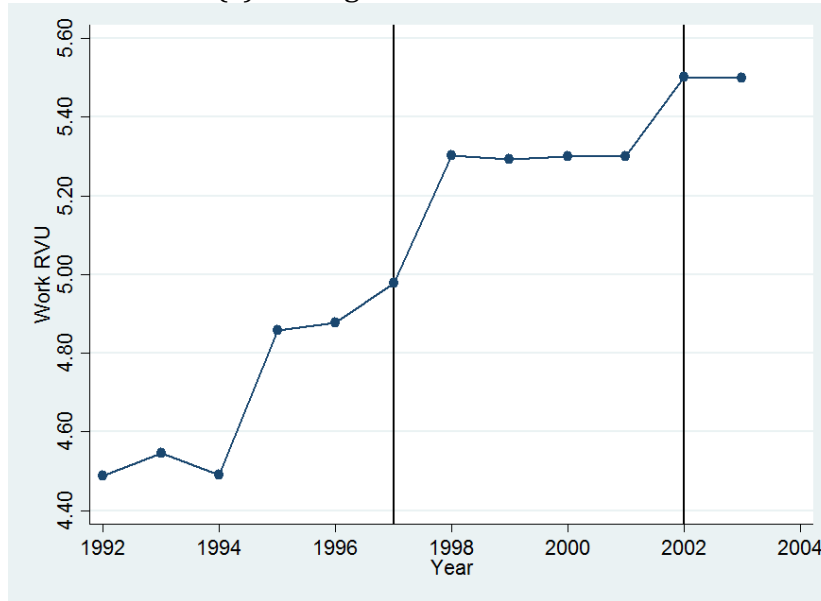
Figure A1: Practice Expense RVU, by Facility Over Time



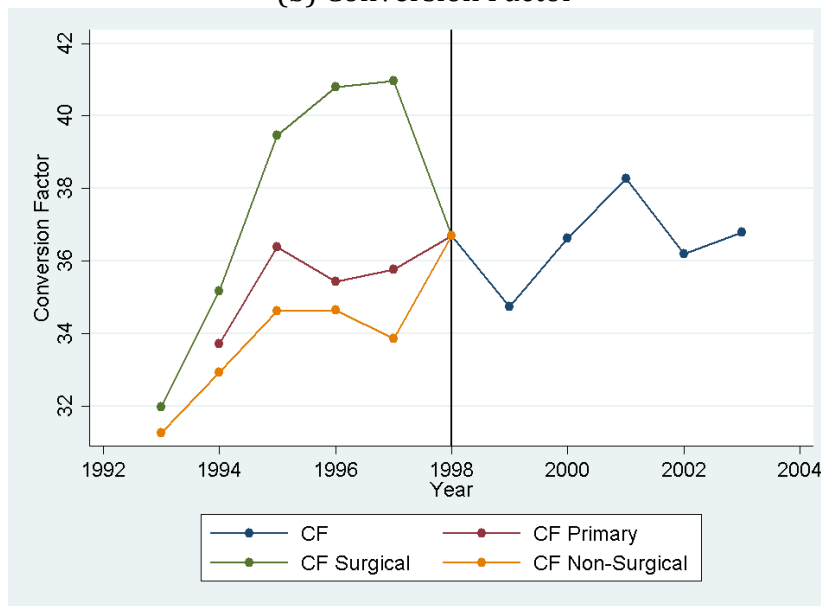
*Notes:* Data from the Federal Register 1992-2003. The top line shows changes in the facility PE-RVU. The bottom line shows changes in the non-facility PE-RVU. Sample restricted to HCPCS observed in all years.

Figure A2: Remaining Variation in Medicare Payments

(a) Average Work RVU Over Time



(b) Conversion Factor

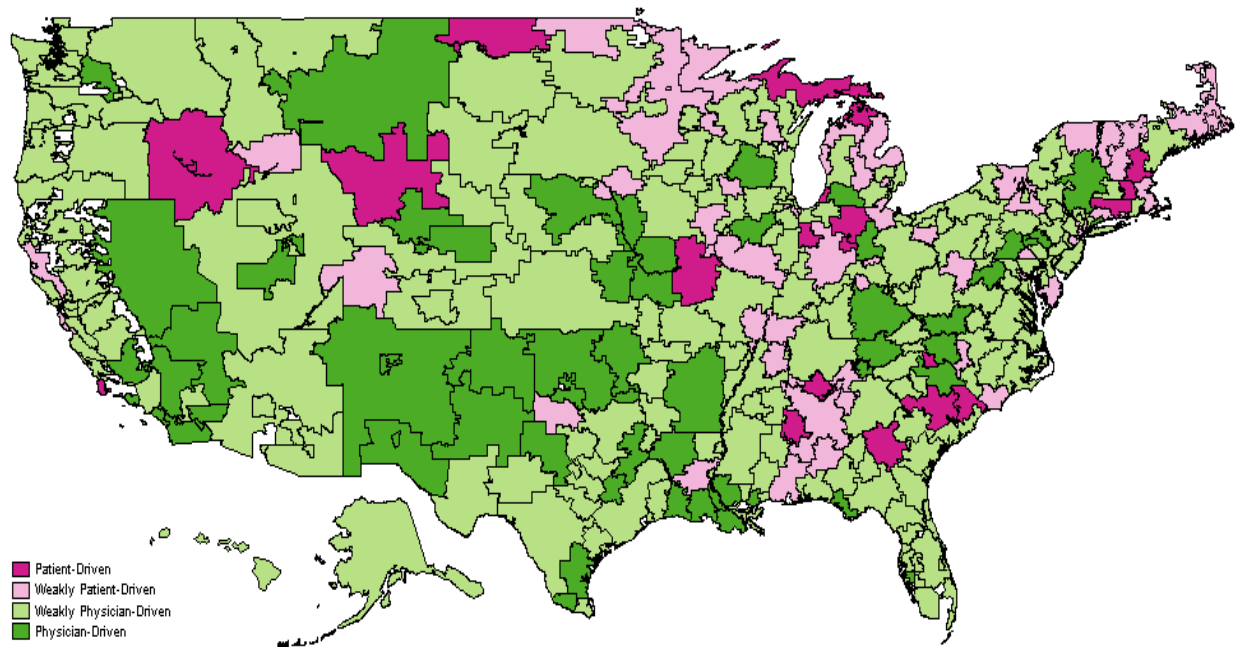


Notes: Data from Federal Register 1992-2003. Plot (a) show the change in work-RVUs. Evident from the graph are the two major reviews by the RUC committee in 1997 and 2002. The sample is restricted to HCPCS observed in all years. Plot (b) shows the change from three CFs (primary care, surgical, and non-surgical) to a single budget-neutral CF in 1998.

## Appendix B

In this section, we show additional results for the paper. Appendix Figure B1 shows the heterogeneity in price elasticities estimated by HRR. Table B1 shows the overall supply response to care, akin to Clemens and Gottlieb (2014). In Table B2, we administer another test that physician-driven HCPCS are associated with higher price-cost margins. For each HCPCS, we calculate two elasticities: one that uses profitability changes above the median, and another that uses changes below the median. Table B2 indeed shows that larger increases in profitability are associated with more physician-driven behavior. Specifically, when the physician incurs a lower cost of administering a procedure, HCPCS have a 0.04 to 0.14 higher probability of being physician-driven. In Table B3, we show the OLS estimates for the analysis presented in Table 5 of the paper.

Appendix Figure B2: Patient- and Physician-Driven HRRs



*Notes:* Data from CCF. For each HRR, we calculate the average price elasticity using data at the HCPCS-year level and a 2SLS model. We include HCPCS fixed effects, year fixed effects, BETOS by year trend, CCI, age, and gender and race dummies. HCPCS are weighted by the national usage. Green areas represent physician-driven HRRs. Pink areas represent patient-driven HCPCS. The lighter shades indicate HRRs where the price elasticity estimate is not statistically significant at the 10% level.



Appendix Table B1: Overall Impact of Price Changes on Quantity

	OLS (1)	IV=GAF (2)	IV=PE- RVU (3)	2SLS (4)
Log(Price)	0.035 (0.0839)	1.251*** (0.853)	0.714*** (0.0944)	0.744*** (0.0927)
First Stage F-Stat	---	96.97	2080	1120
No. of Observations	3,012	3,012	3,012	3,012
R-squared	0.944	0.648	0.678	0.677

*Notes:* Data from CCF at the HRR-year level. The dependent variable is log(total RVU). Covariates included are patient age, CCI < gender, and race dummies, and we control for year fixed effects, HRR fixed effects, and HRR by year trends. Robust standard errors are shown in parentheses \* 10% level, \*\* 5% level, \*\*\* 1% level.

Appendix Table B2: Probability of Being Physician-Driven, by Changes in “Profitability”

	OLS		2SLS	
	$\Delta\pi$ Below Median (1)	$\Delta\pi$ Above Median (2)	$\Delta\pi$ Below Median (3)	$\Delta\pi$ Above Median (5)
1(Physician-Driven)	0.525	0.589*	0.777	0.639***

*Notes:* Column (1) and (3) shows the probability that the own-price elasticity, calculated using changes in annual profitability that are below the median, is positive. Columns (2) and (4) show the probability that the elasticity, calculated using changes in profitability above the median, is positive. Above- and below- median are identified according to the data for each HCCPS-HRR. The means in columns (1) and (2) or (3) and (4) are statistically different at the \*\* 5% level or \* 10% level. Profitability is measured using the “allowed-submitted” measure.

Appendix Table B3: Differential Effect of a Physician Reimbursement Reform, OLS

Dependent Var: Log(RVU)			
	Physician- Driven HCPCS (1)	Patient- Driven HCPCS (2)	X <sup>2</sup> and P- value for H <sub>0</sub> : (1)=(2) (3)
Panel A. All Beneficiaries			
Log(Price)	0.496*** (0.105)	-0.0019 (0.0273)	21.12 0
Panel B. Dual Eligibles			
Log(Price)	0.389*** (0.105)	-0.0200 (0.0766)	9.90 0.0017
Panel C. Non-Dual Eligibles			
Log(Price)	0.542*** (0.103)	-0.0148 (0.0293)	26.81 0
No. of Obs.	3,351	723	----

*Notes:* Data from CCF. Standard errors are bootstrapped. OLS estimates are shown. 2SLS counterpart shown in Table 6.