– Chapter 3 –

Lemon Dropping: Do Physicians Respond to Incentives?

David Benson*

April 26, 2018

Abstract

Observing geographic variation in output price regulations and input costs, I exploit physician migration to identify the effects of financial incentives on patient acceptance, scale, and practice patterns. I find that physicians are more likely to accept more profitable patients. However, omnibus environmental factors subsuming prices explain less than half of the change in physician behavior upon migration. I use a structural supply model to estimate the idiosyncratic financial incentives affecting production choices and acceptance of Medicare and Medicaid Dual Eligible patients. I find that Dual Eligibles have lower marginal cost of primary care compared to Medicare patients. Estimated marginal cost is dispersed across suppliers within a given market. The regression results and structural cost estimates together imply that equalizing prices for Dual Eligible and Medicare patients would erase the primary care access gap for low income elders. Allowing price competition would lower prices compared to the regulated status quo, but reduce access for Medicare patients due to lemon dropping.

*Department of Economics, Northwestern University. Contact: dbenson@u.northwestern.edu. Thanks to my thesis committee for helpful comments and advice: Igal Hendel, Robert Porter, and Gaston Illanes. Thanks also to Alexander Torgovitsky, Leemore Dafny, Aviv Nevo, and the participants of seminars at Northwestern University, the Federal Reserve Bank of Chicago, and the University of Memphis for valuable feedback.

1 Introduction

In this paper, I examine "lemon dropping" and "cherry picking"[1] with respect to Medicare and Medicaid Dual Eligible patients in the U.S. primary care physician industry. Physicians have financial incentives to favor high payment and low cost patients, especially when capacity is scarce. Medicare and Medicaid are public health insurance programs who regulate payments less generously than private payors, and differentially across locations. How do physicians respond to these incentives?

I analyze a unique dataset on the primary care physician population over 2012-2015. I study lemon dropping and cherry picking in two ways with these data. First, I use migration as a source of variation in the physician's Medicare-Medicaid regime, and measure the migrant's response to the change in financial incentives using reduced form regressions. Second, in a structural supply model I use variation in input-output price ratios and variation in the marginal products of primary care production inputs to estimate the physician cost heterogeneity which drives lemon dropping and cherry picking.

In the reduced form, I consider three sources of identifying variation from migration. First, in an event study framework which includes physician and year fixed effects, I identify the effects of a migrant's exposure to geographic variation in financial incentives. Second, in a model which includes location fixed effects in addition to physician and year fixed effects, I identify the effects of relative changes in financial incentives on a migrant's behavior, holding between-location variation in the levels of incentives constant. Finally, I recognize that when physicians move, they inherit aggregate market factors beyond regulated payment rates and the local prices of production inputs. A response to the overall environment includes the effects of these financial incentives, but subsumes them. I extend the first event study framework, using the implied structure on market aggregates from the saturated fixed effects

[1] *"A highly colloquial term for the acceptance of patients based solely on their ability to pay–i.e., with insurance or cash–while turning away the indigent or poor."* Segen's Medical Dictionary, 2012.
In an online survey of the industry, 18 percent of Family Practice and Internal Medicine physicians reported dropping lemons and picking cherries (Page, 2017).

model, to examine an event study for a migrant's response to the overall environment.

From the event study for financial incentives and the saturated fixed effects regressions, I find that physicians respond to financial incentives by selectively accepting patients based on their insurance coverage and ability to pay. Physicians are more likely to accept more profitable patients. Private sector customers are preferred to elders on Medicare or Medicaid. Given Medicare acceptance, acceptance of Medicaid Dual Eligibles is increasing in the overall price level, increasing in relative changes to the Medicaid price, and decreasing in relative changes to the Medicare price. The evidence also suggests that scale is largely inherited from the physician's patient acceptance decision.

Conditional on accepting Medicare or Medicaid Dual Eligible patients, I find that total capacity for these payors is increasing in the overall price level and decreasing in the overall cost of operating a medical practice. However, except for sole proprietors, scale is not affected by relative changes in price or cost holding between-location variation in incentive levels constant. In contrast, I find that physician practice patterns do not vary with the overall levels of price and cost, but do respond to relative changes in these financial incentives holding their aggregate levels constant. Physicians who accept new Medicare patients as the Medicare price increases relative to other incentives are heavier users of equipment, without adjusting total output. Relative changes in financial incentives also affect the number of procedures offered by physicians, with larger responses from sole proprietors. When the Medicare price increases by one percent, physicians on the margin of accepting elder patients have 1.8 more procedures on offer.

Results from the event study with respect exposure to the overall environment provide an upper bound on the importance of prices and aggregate cost shifters, and quantify the relative importance of idiosyncratic cost incentives to a physician's acceptance policy, scale, and practice patterns. The results indicate that environmental factors explain only 45 percent of the change in extensive margin acceptance following migration, and only 22 percent of the change of the intensive margin. Idiosyncratic physician-patient factors are most important for lemon dropping and cherry picking. Unlike in the event study for financial incentives, results for scale and practice pattern outcomes from the event study for overall environmental exposure are not mixed. Exposure to the broader environment has a similar effect across all outcomes: scale, output per patient, facility usage, procedural specialization, the number of procedures offered, and the physician's labor to capital ratio. However, responses to environment can only explain 21-32 percent of the scale and practice pattern adjustment upon migration. As with patient acceptance, results imply that idiosyncratic factors including cost heterogeneity are more important than the incentives from regulated prices for physician treatment of the patients they accept.

In light of the theory of discrimination from Chapter 2, these empirical results for acceptance suggest that patients with public insurance come with greater fixed or sunk costs compared to patients with private insurance. Results also imply that Dual Eligibles have marginal cost efficiencies compared to Medicare patients, which compensates for the regulated price wedge. Results for the response of scale to financial incentives are consistent with this interpretation. Moreover, this implied cost relationship between Dual Eligibles and Medicare patients is exactly that which explained the empirical results from Chapter 2 regarding the relationships between discrimination, scale, and productivity.

Together, this motivates an analysis of heterogeneity in physician financial incentives. I employ a structural supply model to estimate physician-specific opportunity costs of labor, equipment capital, medical supplies, clinician labor, and effort exerted to abate medical malpractice risk. Importantly, I allow these costs to vary across Medicare and Medicaid Dual Eligible patients for each physician. The model exploits several features of the fee-for-service payment contracts used by Medicare and Medicaid.

I model the primitives of the physician's cost function, the primary care output technology, and assume physicians are price takers who maximize profit. I estimate the primary care production function with input-output data by adapting the methodology of Ackerberg, Caves, and Frazer (2015). The supply model provides physician specific marginal products and necessary conditions from profit maximization which identify heterogeneity in the opportunity costs of variable factors across physicians and across patient types. This identification approach is similar to that of De Loecker and Warzynski (2012), and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016). The results of the structural production function indicate that the marginal products of physician labor and effort are most important in primary care.

In the structural cost estimates, I find that the marginal cost of physician labor for Medicaid Dual Eligibles is greater than the cost of labor for Medicare. However, the Medicare patient's marginal costs of effort, equipment, medical supplies, and clinician labor are much larger than for Dual Eligibles. Thus, as expected from the regression evidence and the theory of capacity discrimination, physicians enjoy net marginal cost efficiencies from accepting low income Dual Eligibles in comparison to Medicare patients. Estimated cost heterogeneity is pronounced, both across patients and across physicians. The measure of inframarginal physicians earning economic rent is large.

I use the model to estimate counterfactual outcomes under alternative price regimes. I consider a competitive market counterfactual under which primary care is a monopolistically competitive industry. I find that this form of price competition is worse than regulated prices for Medicare and Dual Eligible access because of lemon dropping and the profitability of private payor patients: the probability of accepting elders falls by 2.38 percentage points. I then consider a competitive market counterfactual under which primary care is a perfectly competitive industry with inframarginal firms. I find the forces of perfect competition nearly erase price discrimination against Dual Eligibles. The competitive price for Dual Eligibles rises by 6 percent on average, increasing access for low income elders. However, competition drives down the output price for Medicare patients, which reduces these consumes' access because of lemon dropping. Last, I examine an alternative regulatory regime with equal Dual Eligible and Medicare payments. I find that Medicaid Dual Eligible acceptance rates would rise from 62 percent to 73.5 percent, matching the 71 percent acceptance rate for Medicare. This result implies that equal regulated payments would erase the primary care access gap for low income elders, without affecting access for Medicare patients.

There is a large literature on physician incentives. My contributions are unique and comprehensive data, a novel identification strategy, and results on physician heterogeneity. My results complement a broader literature on healthcare supplier responses to financial incentives, as in Alexander (2017) and Clemens and Gottlieb (2014). The data and findings also contribute to a growing literature on the supply side consequences of geographic variation in healthcare utilization, such as Finkelstein, Gentzkow, and Williams (2016) and Molitor (2016). My structural approach relies on unique microdata, but it complements the literature on the estimation and decomposition of firm heterogeneity using supply-side conditions. This paper also contributes to the empirical literature on discrimination, for instance Charles and Guryan (2008). To the knowledge of this author, this literature has yet to offer a structural approach for estimation of discrimination coefficients, the cost primitives which drive the economic analysis of discrimination.

There are several other reasons why these results are interesting. First, expansion of Medicaid is debated by economists and policy makers as one solution to the failure of health insurance markets, as a means to achieve full population coverage. My results imply that, while Medicaid patients are well insured, the poor design of payment regulations impede these consumers' access to basic healthcare services. As long as physicians have the right to accept or reject individual patients, my results suggest they will respond to financial incentives and avoid taking on new Medicaid patients, expansion or no.

My results also cast light on the objectives of primary care physicians. The main effect of financial incentives is on patient acceptance, and not on physician practice patterns. Indeed, the estimates for practice patterns can be explained entirely by physician heterogeneity on an acceptance margin. My evidence suggests that primary care physicians first do no harm, but also have opportunity costs associated with limited capacity which motivation them to discriminate access. Physicians respond to those incentives by dropping lemons and picking cherries at the beginning of the physician-patient relationship.

Finally, my results are interesting from the standpoint of market design. I find that regulated payments affect the types of patients a physician accepts, and I find evidence of spillovers from private insurance on who is accepted. Medicare can make lemons out of low income Medicaid Dual Eligible patients, and private payor cherries can make lemons out of both Medicare and Medicaid consumers. As different types of consumers search for access to primary care services, spillovers across these patients and from the private sector are associated with the rationing of scarce healthcare resources for elder and indigent consumers who place high value on these services.

1.1 Roadmap

The remaining sections of the paper are organized as follows. The next section provides an overview of the data on physician migration. In Section 3, I present the reduced form regression models, the role of physician migration and the assumptions necessary for identification, as well as their main results. Section 4 describes the structural supply model, and presents the estimates of the primary care production function and physician cost heterogeneity. I then present the price and profit counterfactuals. I conclude in Section 6 with a discussion of the positive and normative implications of these results for lemon dropping and cherry picking in the U.S. primary care industry.

2 Physician Migration

I now describe patterns in primary care physician geographic mobility from the microdata. There are three novel facts to document. First, Physician mobility is common, but is declining overtime. Second, the differences between movers and non-movers are small. Third, migration flows are bi-directional. Thus, the migrant's change in environmental factors is symmetrically distributed, mean zero for payment regulations and cost, and near mean zero for practice patterns and patient acceptance. Migrants do not disproportionately select more profitable Medicare or Medicaid destinations. Nor is there evidence that primary care movers select destinations with specific practice patterns. These data and the evidence in Chapter 1 suggest that incentives from the private payor market overwhelm those from Medicare and Medicaid in the location decision for physicians who move.

I illustrate the first of these facts in Table 1. Fifteen percent of physicians migrated across counties over 2011-2015, with mobility declining over time as in the broader labor market (Molloy, Smith, and Wozniak, 2014). There were 376,556 physicians designated as general Internal Medicine, Family Practice, General Practice, or an Internal Medicine sub-specialty in 2015. This supplier population grew 13.3 percent from 2011-2015, about 2.7 percent annually. There were 385,561 unique primary care physicians over the horizon, showing attrition and new entry in the population. Each year 3.4-4.3 percent of primary care physicians move counties, a small fraction moved twice. The mobility rate fell each year, declining twenty percent over the horizon. While 15 percent of the population moved counties, not every move was across a Medicare Locality or state line.

Table 1: Primary care physician population and migration by year

	2011	2012	2013	2014	2015	Total
Population	332,219	343,874	355,178	366,112	376,556	385,561
Movers	14,172	14,065	13,383	12,865	12,843	58,756
% movers	4.27	4.09	3.77	3.51	3.41	15

Notes: Author's tabulations from physician microdata.

With few exceptions, migrant physicians do not differ significantly from non-movers. In Table 2 I report summary statistics across the two groups. Because these are populations, there is no sampling error in the means, and the true standard deviations are known.[2] To examine whether the differences are meaningful, I normalize means by standard deviations and compute the difference in standardized scores. None of the scores are statistically significant at a usual level. The representative mover is within 60 percent of a standard deviation of the representative non-mover for every measure. Most mean differences are within ten

[2]A caveat applies to disease fractions, which are censored for privacy concerns from above at 0.75 and from below if fewer than 11 of the physician's patients were diagnosed.

percent of a standard deviation.

Table 2: Summary statistics for movers and nonmovers

	Nonmovers		Movers		
	Mean	St. Dev.	Mean	St. Dev.	Δz [a]
Physician characteristics:					
Female	0.32	0.47	0.41	0.49	0.14
Sole proprietor	0.23	0.42	0.22	0.41	-0.02
General Internal Medicine [b]	0.36	0.48	0.41	0.49	0.09
Family Practice	0.32	0.47	0.33	0.47	0.01
Accept Medicare	0.69	0.46	0.68	0.46	-0.04
Accept Medicaid	0.61	0.49	0.60	0.49	-0.02
Medicaid fraction	0.27	0.21	0.29	0.20	0.22
Medicare revenue ($)	169,226.6	296,106.8	114,737.6	219,886.3	-0.05
Patients	416.05	466.79	332.92	363.11	0.03
Facility fraction	0.40	0.42	0.52	0.44	0.23
Patient demographics:					
Average age	71.60	4.93	70.80	5.08	-0.59
Fraction female	0.55	0.15	0.54	0.16	-0.34
Fraction black	0.08	0.16	0.07	0.16	-0.02
Health risk (HCC score)	1.73	0.90	1.86	0.88	0.19
Alzheimers	0.14	0.13	0.15	0.14	-0.01
Arthritis	0.38	0.14	0.39	0.14	-0.08
Asthma	0.08	0.07	0.09	0.07	0.03
Atrial fibrillation	0.13	0.10	0.14	0.11	-0.06
Cancer	0.11	0.10	0.11	0.11	-0.10
Depression	0.25	0.13	0.29	0.14	0.19
Diabetes	0.37	0.15	0.38	0.16	-0.02
Heart failure	0.27	0.18	0.30	0.20	0.05
Hyperlipidemia	0.56	0.15	0.56	0.16	-0.07
Hypertension	0.68	0.12	0.69	0.12	-0.02
Ischemia	0.42	0.19	0.43	0.20	-0.01
Kidney disease	0.30	0.19	0.33	0.20	0.09
Obstructive pulmonary	0.20	0.14	0.23	0.15	0.08
Osteoporosis	0.08	0.07	0.08	0.07	-0.04
Schizophrenia	0.05	0.07	0.06	0.07	0.08
Stroke	0.06	0.07	0.07	0.07	-0.03

Notes: Author's calculations from physician microdata.
[a] Difference of normalized means: $\Delta z := \frac{\text{mean movers}}{\text{st. dev. movers}} - \frac{\text{mean nonmovers}}{\text{st. dev. nonmovers}}$.
[b] Omitted category is General Practice + Internal Medicine sub-specialties.

By these metrics, movers and non-movers are similar in most dimensions. However, there are four differences to highlight. Movers were more often women, by a nine percentage point difference in population gender means. They were more likely to be general Internists. Movers also had smaller businesses, by 83 patients and about $54,500 in revenue. Finally, migrants were more likely to use facilities.

When physicians migrate, they inherit new environmental factors in their destination. This includes new financial incentives in payment regulations and input factor costs. It also includes new county average practice patterns, and the Medicare-Medicaid acceptance policies of other local physicians. I now document facts on the distribution of changes in environmental factors. Table 3 reports summary statistics on destination-origin changes for the migrant subpopulation.

Table 3: Summary of mover's change in environmental factors

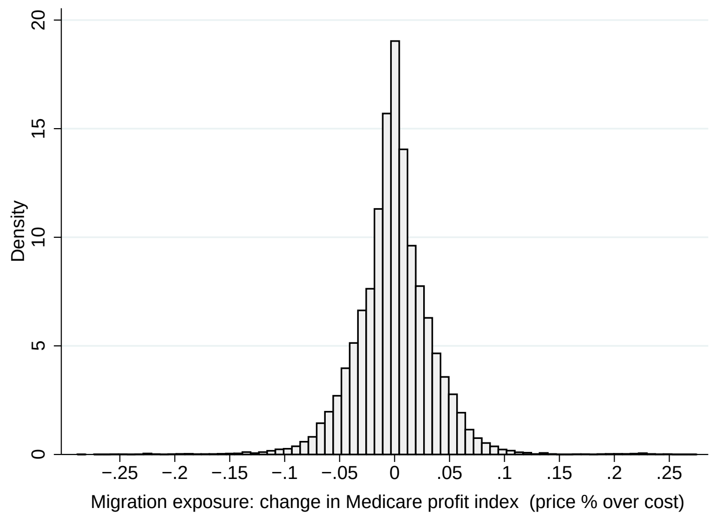
	Mean	St. Dev.	10th %	90th %
Practice patterns:				
Δ log(# procedures)	0.04	0.34	-0.36	0.47
Δ log(labor/capital)	-0.03	0.21	-0.24	0.19
Δ log(bundle HHI)	-0.02	0.23	-0.30	0.25
Δ facility fraction	-0.02	0.18	-0.25	0.20
Δ log(output (RVUs))	0.04	0.54	-0.58	0.66
Δ log(output per consumer)	0.01	0.26	-0.25	0.26
Business environment:				
Δ log(Medicare revenue)	0.01	0.26	-0.28	0.30
Δ Medicare acceptance rate	0.01	0.08	-0.09	0.10
Δ Medicaid fraction	-0.01	0.09	-0.12	0.10
Δ log(health risk)	-0.03	0.16	-0.23	0.17
Δ log(Medicare price)	0.00	0.07	-0.09	0.07
Δ log(Medicaid price)	0.00	0.16	-0.17	0.18
Δ log(cost index)	0.00	0.07	-0.09	0.08

Source: Author's calculation from physician microdata.
Variables are the destination - origin difference in county means.

The distributions are symmetric, due to bi-directional migration flows. The mean percentage change in financial incentives was zero. Mean changes in patient acceptance and other county business environment measures are near zero. The distribution of changes in the county profit index illustrates exposure to new financial incentives. In Figure 1, I plot the histogram of changes upon migration in the Medicare price level minus county cost. This figure shows that migrant destination choices are not driven by Medicare rent seeking. If

Medicare's financial incentives were an important factor in location choice, the density would be left skewed. There is no evidence of skew.

Figure 1: Empirical density of mover's change in Medicare financial incentives



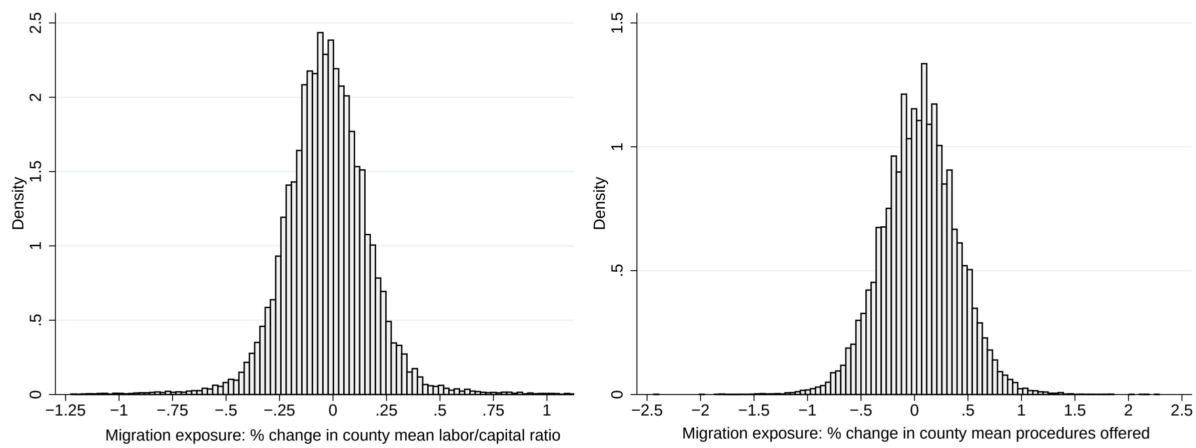
Notes: Change across county moves in the Medicare price as a percent of local cost, measured by (payment − cost)/$\mathbb{E}$(payment). Source: Author's calculations from physician microdata.

Mean changes in the practice pattern environment were larger, but still negligible. On average, movers entered counties where physicians offered four percent more procedures and supplied four percent more total output to Medicare and Medicaid. The destination county's labor to capital ratio was three percent lower, and average bundle specialization two percent lower, on average.

Growth in mean county practice patterns can suggest relocations are rent seeking. However, this interpretation is not supported by mean growth in regulated payment rates and local costs. Nor is this argument supported by mean growth in county average revenues, by mean differences in Medicare and Medicaid patient acceptance rates, or by mean growth in county output per patient.

The distribution of changes in environmental practice patterns illustrates this point for financial incentives. In Figure 2, I plot empirical densities of the percent change in county average labor to capital ratios, and of the percent change in county average procedures offered. For the interested reader, the appendix provides histograms for Medicare acceptance and for other practice patterns. Figure 5 shows that changes in environmental practice patterns are symmetrically distributed and, while not mean zero, are near mean zero.

Figure 2: Empirical density of mover's change in environmental practice patterns



Notes: Change across county moves in log labor minutes per equipment use (left), and log number of procedures offered (right). Source: Author's calculations from physician microdata.

There is geographic variation in financial incentives and market size, both in the private payor market and in the smaller regulated price markets of Medicare and Medicaid. Primary care physicians are dispersed geographically, and vary in their acceptance rates for Medicare and Medicaid Dual Eligible patients. Medicare and Medicaid are small fractions of the representative primary care physician's business. These facts from Chapter 1 provide an interpretation for the symmetric exposure distributions presented in Figures 1 and 2.

Migrants do not select more profitable Medicare or Medicaid destination counties, likely because these types of patients are second order in the location decision. For market selection, incentives from the private payor market overwhelm those from Medicare and Medicaid. Regulation matters after entering a location. Price controls might affect the decisions to accept Medicare and Medicaid Dual Eligible patients, to fill slack capacity left from serving the privately insured. Since primary care physician mobility is not associated with financial incentives from Medicare and Medicaid, I now ask whether changes in the mover's regulated prices and cost affect acceptance rates for these patients.

3 Regression Models

How do physicians respond to financial incentives? The facts and results from chapters 1 and 2 regarding on patient acceptance, dispersion in scale and productivity, and variation in financial incentives motivate a more formal analysis of the microdata. I examine lemon dropping and cherry picking through the lens of a linear regression model, with outcomes for both the extensive and intensive margins of patient acceptance, as well as physician scale and practice patterns.

I first estimate the migration event study, a regression including physician, year, and move year fixed effects which identifies the response to between-location variation in aggregate cost shifters and payment regulations from Medicare and Medicaid by a form of difference-in-differences. I secondly estimate regressions with location, physician, year, and move year fixed effects, which hold constant persistent between-market variation in financial incentives and market size. This saturated fixed effects specification holds constant any persistent between-market variation in the explanatory variables, meaning the physician's response to financial incentives is identified off year-over-year relative changes in Medicare and Medicaid regulations and in market factor prices. Finally, I return to the event study framework and use the model's implied structure on market aggregates to estimate the migrant's response to overarching, potentially unobserved environmental factors. In each specification I include a rich set of controls for patient demographics and diseases, as well as a control function for migrant selection. Throughout, standard errors are clustered at the physician level.

3.1 Migration Event Study for Financial Incentives

I index physicians by j, county locations by r, and time periods by t. Markets are thus indexed (r, t), and the set of suppliers therein by $j \in J_{rt}$. I assume a linear relationship between a physician's outcome measure y_{jrt} and measures of financial incentives $(p_{rt}^{MCR}, p_{rt}^{MCD}, \bar{C}_{rt})$

for Medicare prices, Medicaid prices, and cost shifters respectively.[3] Formally, I assume

$$(1)\ y_{jrt} = \sum_{\tau=-3}^{3} 1(t = \tau_j)\Big(\gamma_\tau + \rho_\tau^{MCR} p_{rt}^{MCR} + \rho_\tau^{MCD} p_{rt}^{MCD} + \rho_\tau^{C} \bar{C}_{rt}\Big) + x_{jt}\beta + \gamma_j + \gamma_t + e_{jrt}$$

I notate observable controls x_{jt} and unobservable physician and year fixed effects (γ_j, γ_t). I estimate the physician's response for each period of "event time" $\tau \in \{-3, -2, -1, 0, 1, 2, 3\}$, which tracks the calendar year relative to the date of migration, and let γ_τ denote the move year fixed effect.[4] Structure must be placed on the model's residual is e_{jrt} for formal identification of (1).

3.1.1 Identification and Interpretation of ρ_τ

Potential endogeneity between $(p_{rt}^{MCR}, p_{rt}^{MCD}, \bar{C}_{rt})$ and e_{jrt} arises due to unobserved variables affecting patient acceptance, output capacity, and practice patterns, including the physician's unobserved private payor business. Physicians may select locations with foresight of these unobserved variables. While the facts on physician migration indicate selection bias should be small because migrant exposures to environmental changes are symmetrically distributed and near mean zero, firm migration is still a choice to exit one market and enter another. Ignoring this confounds a causal interpretation of ρ. I assume the event study residual in (1) takes the form

$$e_{jrt} = B(\mathcal{P}_{jt}) + \tilde{e}_{jrt}$$

allowing a control function $B(\mathcal{P})$ to address selection bias in physician mobility.

The arguments of $B()$, $\mathcal{P}_{jt}$, are nonparametric estimates of the conditional probability a physician moved from one market to another. These are sufficient data to purge (1) of selection bias because the expectation of the residual e_{jrt} given r is a possibly unknown but monotone function of the conditional probability r was chosen, an insight due to Lee (1983)

[3] See Chapter 1 for details on the Geographic Adjustment Factor (GAF) measure of financial incentives.

[4] Though outcomes data are available only for four periods, 2012-2015, I can estimate three pre and post migration period effects because the physician locations panel is longer, 2011-2017.

and related to Hotz and Miller (1993). Because county migration flows are voluminous, as documented in Section 2, I can estimate $\mathcal{P}(r', r)$ nonparametrically for each county pair (r', r). I then assign $\mathcal{P}_{jt} = \mathcal{P}(d(j,t), o(j,t-1))$ given the physician's observed destination $r' = d(j,t)$ and origin $r = o(j,t-1)$ markets. The function $B(\mathcal{P})$ is nonparametrically identified. However, I take a semiparametric approach similar to Dahl (2002), and estimate a flexible sieve for B simultaneously with the linear parameters.[5]

The identifying assumption for (1) is then strict exogeneity of $\tilde{e}_{jrt}$. Formally, I assume

$$\mathbb{E}(\tilde{e}_{jrt} | p_{rt}^{MCR}, p_{rt}^{MCD}, \bar{C}_{rt}, x_{jrt}, \mathcal{P}_{jt}, \gamma_j, \gamma_t, \tau_j) = 0$$

In words, I assume exposure to new cost shifters and regulatory environments are exogenous conditional on observables, the move year, fixed effects, and the aggregate flow of physicians from one market to another. Two subtle assumptions are embedded in this condition: the migration year and any physician heterogeneity in responses are also conditionally exogenous.

The reduced form event study (1) generalizes a difference-in-differences identification strategy. Any systematic differences between movers and non-movers will be apparent in the pre and post move dynamics of ρ_τ. A parallel trends assumption is not required. However, if the data do exhibit parallel trends then ρ_τ will be constant over $\tau < 0$, will jump to a new level upon migration, and remain at this level over $\tau > 0$. Pre and post-move adjustments in ρ_τ are evidence of physician adaptation to environment, or evidence of systematic differences between movers and non-movers.

The coefficients ρ_τ thus provide a means to flexibly estimate the expected effect of financial incentives on the outcome variable y. The causal effect is given by their pre - post move difference, i.e. for the effect of regulated prices

$$\frac{\partial \mathbb{E}(y)}{\partial p^k} = \mathbb{E}(\rho_\tau | \tau > 0) - \mathbb{E}(\rho_\tau | \tau < 0)$$

[5] I estimate a 5th order polynomial sieve for each year relative to the migration year and non-migration, resulting in 40 parameters to be estimated for $B()$.

where $k \in \{MCR, MCD\}$. This reduces to the difference-in-differences estimate of the slope of y if $1(t = \tau)$ were replaced by a simple pre-post move indicator variable. Hence, the object of empirical interest is not the level of ρ_τ but its jump upon migration.

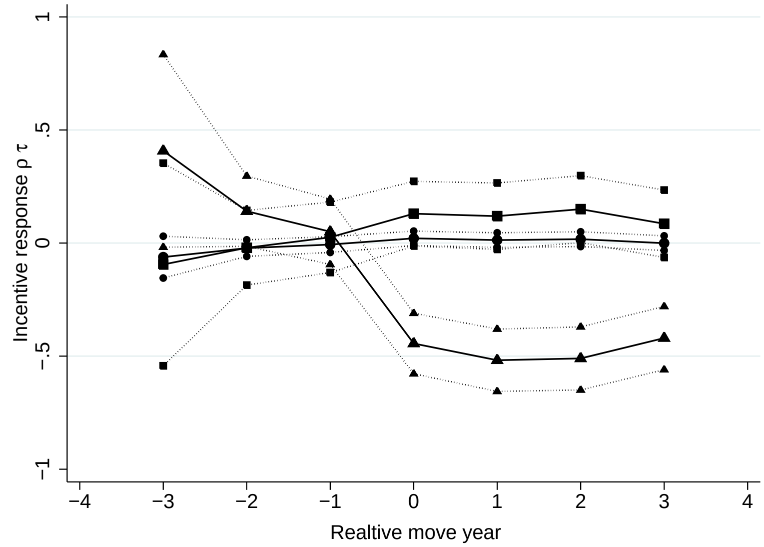
In the case of patient acceptance outcomes, I place levels of the GAF financial incentive measures into (1). For the extensive margin of acceptance, the outcome is an indicator variable for the acceptance of any Medicare patients, $y = 1(\text{N Medicare} > 0)$, or for acceptance of any Medicaid Dual Eligible patients given some elders are accepted $y = 1(\text{N Dual} > 0 \mid \text{N Medicare} > 0)$. Thus, (1) is a linear probability model for patient acceptance, and the jump upon migration $\frac{\partial \mathbb{E}(y)}{\partial \text{GAF}} = \mathbb{E}(\rho_\tau | \tau > 0) - \mathbb{E}(\rho_\tau | \tau < 0)$ is interpreted as the percentage point change in the probability of acceptance due to moving from zero to the mean financial incentive.

To examine the intensive margin of acceptance, I measure the log-odds ratio of the physician's patient mix, $y = \log(\text{N Duals}/\text{N Medicare})$. Thus, the jump in ρ_τ upon migration estimates the coefficient on a financial incentive in a logit model, i.e. the comparative static of the Dual Eligible patient mix is given by $\frac{\partial l^s}{\partial \text{GAF}} = l^s(1 - l^s) * \Big(\mathbb{E}(\rho_\tau | \tau > 0) - \mathbb{E}(\rho_\tau | \tau < 0)\Big)$. For outcome measures of physician scale and practice patterns, I place log outcomes as well as log financial incentives into (1). Hence, the jumps upon migration $\frac{\partial \mathbb{E}(y)}{\partial \text{GAF}} = \mathbb{E}(\rho_\tau | \tau > 0) - \mathbb{E}(\rho_\tau | \tau < 0)$ estimate the elasticity of the outcome variable with respect to financial incentives.

3.1.2 Event Study Results for Financial Incentives

The first result regards the extensive margins of discrimination. I plot the response of the extensive margin of Medicare acceptance to changes in financial incentives over relative move years in Figure 3. Clearly, the most salient financial incentive is cost: increases in the cost of running a medical practice decrease the probability of accepting any Medicare patients. The second salient financial incentive is the Medicare - Dual Eligible price level. An increase in the price level increases the probability of accepting these types of patients, though the

Figure 3: Response of Extensive Margin of Medicare Acceptance to Financial Incentives



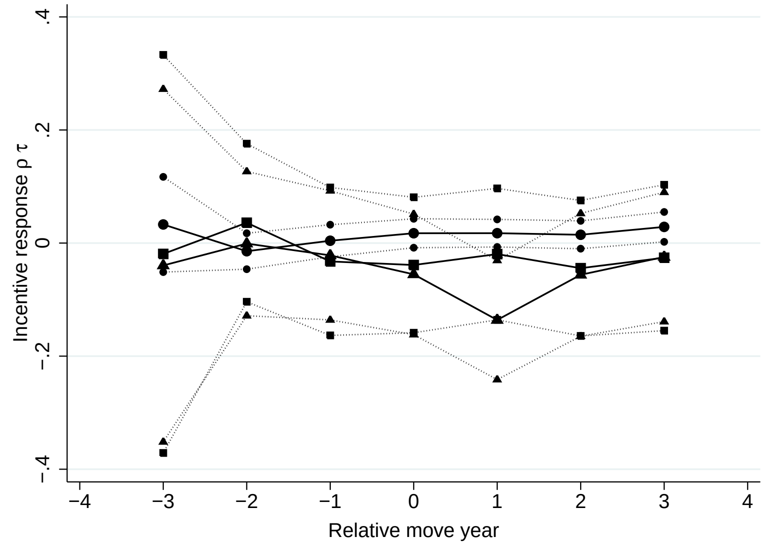
Notes: Estimated responses of Medicare patient acceptance by relative move year, for the Medicare and Dual Eligible price level $\hat{\rho}_\tau^{MCR}$ (squares), Medicaid prices $\hat{\rho}_\tau^{MCD}$ (circles), and the cost index $\hat{\rho}_\tau^{C}$ (triangles). Dotted lines are $\pm$ 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

effect is smaller than that of cost.

The estimated jump in the physician's cost response $\hat{\rho}_\tau^{\mathrm{C}}$ upon migration is -0.682, and the cost GAF has standard deviation 0.074 across the population. The event study results imply a one standard deviation increase in the cost index reduces the probability of accepting Medicare by 5 percentage points, an economically significant effect. The estimated jump upon migration in the physician's response to Medicare price levels $\hat{\rho}_\tau^{\mathrm{MCR}}$ is smaller, 0.148. The Medicare GAF price index has standard deviation 0.067 across the population. Thus, a one standard deviation increase in the Medicare price increases the probability of accepting Medicare patients by 1 percentage point at the extensive margin.

This result is interpretable through the lens of the theory of discrimination via Propositions 2 and 3 from Chapter 2, where Medicare and Medicaid Dual Eligibles are viewed as type 0 customers and private payor patients are viewed as type 1, and where the cost index is viewed as a negative productivity shock. If the Medicare price increases and private payor prices are held constant across locations, then discrimination against Medicare pa-

Figure 4: Response of Extensive Margin of Dual Eligible Acceptance to Financial Incentives



Notes: Estimated responses of Medicaid Dual Eligible patient acceptance by relative move year, for the Medicare and Dual Eligible price level $\hat{\rho}_{\tau}^{MCR}$ (squares), Medicaid prices $\hat{\rho}_{\tau}^{MCD}$ (circles), and the cost index $\hat{\rho}_{\tau}^{C}$ (triangles). Dotted lines are $\pm$ 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

tients should fall with the Medicare price. However, if private payor prices also increases as Medicare prices rise, then discrimination against Medicare patients will increase if younger private sector patients have fixed or sunk cost efficiencies compared to elders.

The estimated effect of a change in Medicaid prices on the extensive margin of elder acceptance is negligible. In fact, given acceptance of Medicare, the estimated effect of all financial incentives on the extensive margin of Dual Eligible acceptance is also negligible. I plot these coefficients in Figure 4. The effects identified off between-location variation have no impulse response to either financial incentive at the Dual Eligible extensive margin, given Medicare acceptance, and standard errors on the estimated coefficients are large.

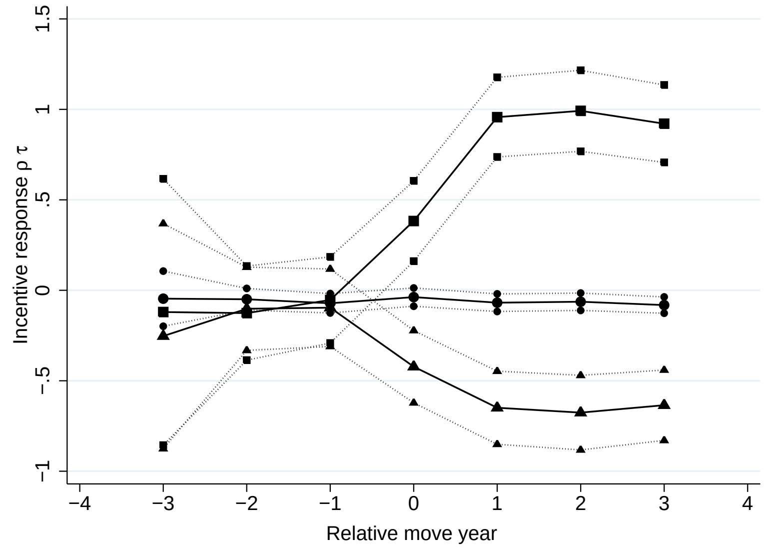
The second result from the event study model is that the intensive margin of Dual Eligible acceptance does respond to financial incentives, unlike its extensive margin. I plot the estimated coefficients ρ_τ for the log odds ratio $y = \log(\text{N Duals/N Medicare})$ in Figure 5 across relative move years. On the intensive margin, the most salient incentive is not cost, it is the overall Medicare - Dual Eligible price level $\hat{\rho}_{\tau}^{MCR}$, which jumps 1.056 upon migration.

The implied mean elasticity of the intensive margin of Dual Eligible acceptance with respect to the price level is 0.78; a one standard deviation increase in the overall price level increases the Dual Eligible patient mix by 1.37 percentage points. The second salient incentive is cost: the migration impulse in $\hat{\rho}_{\tau}^{C}$ is -0.503. The implied elasticity of the intensive margin of Dual Eligible acceptance with respect to cost is -0.372 at the mean patient mixture; a one standard deviation increase in the cost index reduces the intensive margin of Dual Eligible acceptance by 0.71 percentage points.

Recalling that as the Medicare price increases the price for Dual Eligibles also increases, the effect of an increasing price level on the intensive margin of Dual Eligible acceptance can be interpreted using Proposition 5 of the theory of discrimination in Chapter 2. An increase in the overall price level decreases discrimination if and only if Dual Eligibles' marginal cost efficiencies are greater than their average cost efficiencies, they must have sunk cost inefficiencies compared to regular Medicare patients. Thus, the empirical relationships between physician discrimination and prices are consistent with the empirical results regarding the discrimination and scale from Chapter 2.

The third result of the event study for financial incentives is that physician scale, as measured by log RVU output supplied, responds to financial incentives similarly as the extensive margin of Medicare acceptance. For this outcome, the jump upon migration in $\hat{\rho}_{\tau}$ measures the elasticity of scale with respect to prices and cost, I plot these estimates across relative move years in Figure 6. The implied elasticity of scale with respect to cost is -1.685. The implied elasticity with respect to the overall price level is smaller in comparison, 0.293. One feature of this empirical result is subtle: the impulse in $\hat{\rho}_{\tau}$ for scale occurs one year after the impulse in the extensive margin of patient acceptance plotted in Figure 3. Moreover, the jumps in $\hat{\rho}_{\tau}$ for the intensive margin of acceptance also proceed the jump for scale, but exhibits an elongated impulse compared to the extensive margin of acceptance. This suggests that output supplied is inherited from the physician's patient acceptance decision, and that the intensive margin of capacity discrimination is slower to adjust to a new environment

Figure 5: Response of Intensive Margin of Dual Eligible Acceptance to Financial Incentives



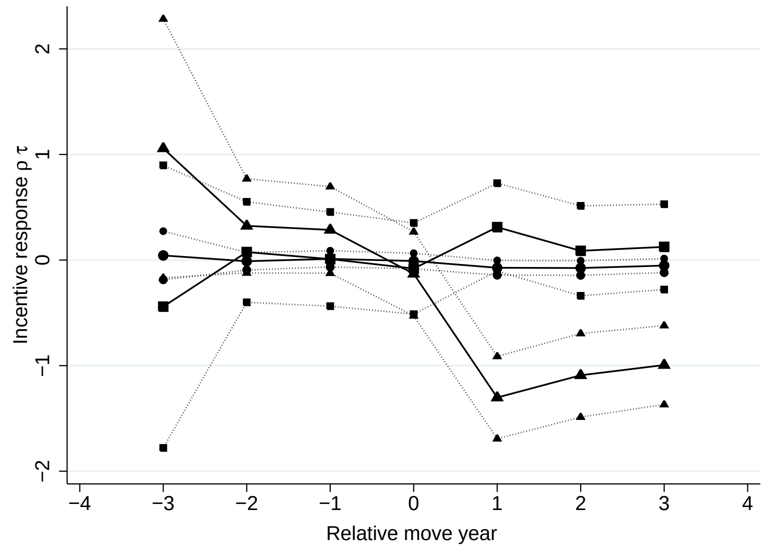
Notes: Estimated responses of the log-odds ratio of Medicaid Dual Eligibles by relative move year, for the Medicare and Dual Eligible price level $\hat{\rho}_{\tau}^{MCR}$ (squares), Medicaid prices $\hat{\rho}_{\tau}^{MCD}$ (circles), and the cost index $\hat{\rho}_{\tau}^{C}$ (triangles). Dotted lines are $\pm$ 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

than the extensive margin.

In light of Proposition 5 from Chapter 2, the price level result for scale is to be expected. Viewing the cost index as a negative productivity shock, that scale is empirically decreasing in cost while discrimination against Dual Eligibles is increasing is consistent with Proposition 4 of the theory. Moreover, since the price level effect on scale is muted in comparison to the cost index effect, and vice versa for the intensive margin of Dual Eligible acceptance, Propositions 4 and 5 of the theory strongly suggest that marginal cost efficiencies from Dual Eligibles are greater than average cost efficiencies, and that the cost wedge between these patients is shrinking with productivity (or increasing in the cost index). Once again, the empirical results for financial incentives and discrimination confirm the theoretical implications of the empirical results from Chapter 2.

Finally, estimates of (1) for physician practice pattern outcomes are surprisingly negative. All jumps in $\hat{\rho}_{\tau}$ upon migration for the number of procedures on offer, facility usage, labor to capital ratios, or procedural specialization are weak and imprecise. To the extent that

Figure 6: Response of Scale to Financial Incentives



Notes: Estimated responses of the log-odds ratio of Medicaid Dual Eligibles by relative move year, for the Medicare and Dual Eligible price level $\hat{\rho}_\tau^{MCR}$ (squares), Medicaid prices $\hat{\rho}_\tau^{MCD}$ (circles), and the cost index $\hat{\rho}_\tau^{C}$ (triangles). Dotted lines are $\pm$ 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

there is any effect of financial incentives for practice patterns identified by (1), the impulse occurs one year post migration as was found for physician scale.

The event study for financial incentives has positive results for patient acceptance and scale. However, results are negative for the extensive margin of Dual Eligible acceptance conditional on Medicare acceptance, and are negative for all practice patterns. The key source of identifying variation in the event study model (1) is between-location variation in financial incentives. However, an obvious weakness of (1) is potentially unobserved location specific factors, for instance unobserved variation in consumer preferences, which are correlated with the financial incentives. The model of the next section addresses this concern.

3.2 Saturated Fixed Effects Model

Purging sources of bias from unobserved market factors is feasible via market fixed effects, and these fixed effects are identified separately from the physician fixed effect due to physician migration. As with (1), I assume a linear model relates the physician's outcome y_{jrt} to financial incentives, observable controls, and separable fixed effects, letting γ_r notate location fixed effects. Formally, I assume

$$(2) \qquad y_{jrt} = \rho^{MCR} p_{rt}^{MCR} + \rho^{MCD} p_{rt}^{MCD} + \rho^{C} \bar{C}_{rt} + x_{jt}\beta + \gamma_j + \gamma_r + \gamma_t + e_{jrt}$$

where e_{jrt} is the model's residual. I use county fixed effects for r. This strategy leaves exposure to financial incentives from within physician changes in GAF financial incentive measures which result from migration *and* changes over time in regulation or factor market equilibria. The response to incentives is now a homogeneous coefficient ρ.

3.2.1 Identification and Interpretation of ρ

Model (2) includes a full set of fixed effects $(\gamma_j, \gamma_r, \gamma_t)$, which are ideal controls for unobserved market factors broader than financial incentives, unobserved physician idiosyncrasies, including productivity, and secular trends in outcomes. However, a saturated fixed effects model also purges all identifying variation for financial incentives arising from persistent between-location variation. Care must then be taken for the interpretation of the causal effect ρ, since all identifying variation is now from relative changes in observed financial incentives over time. For instance, the effects of increased cost is identified off year-over-year changes in input factor market equilibria, and the effects of Medicare prices are identified off changes in regulation over time.

As with (1), physician migration provides the identifying power for ρ in (2). Secular variation in policy is small, and factor market equilibria are persistent. For physicians who migrate, there is within-physician variation in $p_{rt} - p_{r't-1}$ and $\bar{C}_{rt} - \bar{C}_{r't-1}$ due to the differential policy and factor price shocks in markets r and r' between periods. Identification of demographic coefficients β is otherwise standard, noting that migration contributes important variation in x_{jt} when new physician-patient matches form.

I assume the saturated fixed effect model's residual in (2) takes the form

$$e_{jrt} = B(\mathcal{P}_{jt}) + \sum_{\tau=-3}^{\tau=3} \gamma_\tau 1(t = \tau_j) + \tilde{e}_{jrt}$$

once again allowing a control function $B(\mathcal{P})$ to address selection bias in physician mobility, but now including the fixed effect for the physician's move year within the control function. The identifying assumption for (2) is then strict exogeneity of the residual $\tilde{e}_{jrt}$. Formally, I assume

$$\mathbb{E}(\tilde{e}_{jrt} | p^{MCR}, p^{MCD}, \bar{C}, x, \mathcal{P}, \tau, \gamma) = 0$$

where omission of indices for $(p^{MCR}, p^{MCD}, \bar{C}, x, \mathcal{P}, \gamma)$ is intentional, because the conditioning set for (2) must include all terms of within-physician variation. This makes explicit that $\tilde{e}_{jrt}$ is mean orthogonal to contemporaneous observables as well as to leads, lags, spatial perturbations, and all fixed effects. This is a strong but common assumption in linear panel regression, see Wooldridge (2010). It is not directly testable. However, individual primary care physicians are insignificant in the regulator's calculus, and are likely unimportant to the equilibrium of local factor markets. In words, I assume that changes in price regulations and input factor prices are exogenous conditional on the fixed effects, migration year, observables, and physician flows from one market to another. Then, as with (1), any first order bias is thus selection bias, and purged by $B(\mathcal{P})$.

3.2.2 Effects of Relative Changes in Incentives on Patient Acceptance and Scale

I estimate (2) with the same outcome measures y examined with (1). In addition to GAF financial incentive measures, I now interact the indicator for sole proprietorship with the price and cost indices. I also report the effects of Medicare's ten percent revenue "bonus" for comparison. I report the main results in Table 4.

The first result regards the extensive margin of acceptance of Medicare and Dual Eligible patients, and the physician's tradeoff between accepting private payor patients versus elders.

In comparison to the price effect estimated in (1) and illustrated in Figure 3, the estimated effect from (2) reported in column 1 of Table 4 indicates that acceptance of elders is *declining* in the Medicare price. The difference between these two results is simple: the event study is identified off between-location variation in price levels, while the saturated fixed effects model absorbs the price level effect and leaves only relative price changes for identification. The price level effect is observable in the results of (2) only through the 10 percent revenue bonus effect, which carries the same sign as the price level effect identified in (1). A ten percent increase in the price level increases the probability and physician accepts elders by 0.7 percentage points at the extensive margin.

To understand the relative price changes driving these regression results, an institutional feature of the market is needed. Prices paid by private insurers to physicians are negotiated each year as a multiple of the local Medicare rate, and the literature has found a large pass-through effect of Medicare prices, 1.3 elasticity, on the physician's private payor price (Clemens and Gottlieb, 2017). Thus, holding between-location variation in the price levels constant, all leftover variation in the Medicare price for estimation of (2) reflects a relative increase private sector prices above and beyond any increase in the price for elders' care.

In light of this, the economic theory of discrimination is once again helpful for interpretation, now viewing elders as type 0 customers and private payor customers as type 1 in Proposition 3 from Chapter 2. The event study result for the overall price level illustrated in Figure 3 implies that elders have marginal cost efficiencies compared to private payor patients, but have fixed or sunk cost inefficiencies: $c_{ql} \leq c_l/q$. By the same Proposition, the result from Table 4 regarding an increase in private sector prices relative to the Medicare price, holding levels constant, suggests that $c_{ql} + c_{qq}q \geq c_l/q$. Together, this implies that physicians at the extensive margin of Medicare acceptance face significant opportunity costs of capacity, $c_{qq}q \geq 0$, exacerbated by decreasing returns to scale.

Table 4: Effects of Relative Changes in Incentives on Patient Acceptance and Scale

Outcome:	1(Medicare > 0)	1(Dual > 0)	$\log(\frac{\text{N Dual}}{\text{N Medicare}})$	log(RVUs)
Financial incentives:				
Medicare payment	-0.390***	-0.0452	-1.937***	0.228
	(0.0955)	(0.0850)	(0.126)	(1.265)
Medicaid payment	0.141	0.573***	3.540***	0.611
	(0.170)	(0.148)	(0.217)	(1.256)
Cost index	0.0897***	-0.00729	-0.384***	0.0164
	(0.0321)	(0.0288)	(0.0405)	(0.0769)
SP-financial interactions:				
Medicare payment	0.276***	0.0175	-0.173*	0.406**
	(0.0706)	(0.0635)	(0.100)	(0.184)
Medicaid payment	-0.0435*	0.0122	0.00242	-0.00776
	(0.0240)	(0.0186)	(0.0313)	(0.0431)
Cost index	-0.268***	-0.0120	0.141*	-0.258
Selected covariates:				
(SP) Sole proprietor	-0.0353	-0.0279	0.0466	-0.0740***
	(0.0492)	(0.0388)	(0.0689)	(0.0203)
10% payment bonus	0.00770**	-0.00101	0.0114***	-0.00196
	(0.00316)	(0.00269)	(0.00413)	(0.00916)
Elder health risk	NA	-0.0648***	0.0948***	-0.201***
		(0.00163)	(0.00297)	(0.00612)
Observations	1,340,159	866,343	774,470	855,800
R-squared	0.011	0.257	0.614	0.267
Mean outcome	.709	.871	-1.019	3.601

Notes: Author's calculations of (2) from physician microdata. All regressions include physician, county, year, and move year fixed effects. Columns 2 and 3 include consumer disease and demographic controls, these controls are unavailable for physicians who do not accept Medicare. The extensive margin of Dual Eligible acceptance in column 2 conditions on acceptance of Medicare. Financial incentives are levels of GAF indices for acceptance outcomes, and log GAF indices for the scale outcome. Std. errors clustered by physician in parenthesis; ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

An alternative economic interpretation does not require the theory of discrimination or physician capacity constraints. If there is imperfect information in the negotiation with private payors, then physicians can improve their bargaining position by keeping costs private information. Since the literature has found that bargained profits are elastic in the Medicare price, the information incentive presents another increasing opportunity cost of Medicare acceptance. Private insurers receive a public cost signal when a physician accepts Medicare and Dual Eligible patients, because average cost must be at least as low as the regulated prices. Which incentive dominates depends on Medicare as a fraction of the physician's business. Since the evidence from Chapter 1 implies this faction is small, the opportunity cost may outweigh potential gain and physicians will be less likely to accept elders.

Other results in column 1 are consistent with an opportunity cost of capacity interpretation. Consider the interaction of financial incentives with the physician's sole proprietor status.[6] For solo physicians, who typically have smaller scale practices, the net price effect is still negative but is muted. Hence, the theory suggests that for solo physicians, an elder's opportunity cost of capacity due to returns to scale, $c_{qq}q$, is smaller in comparison to physicians in group practice, so that $c_{ql} + c_{qq}q \geq c_l/q$ is closer to equality. Finally, the signs on the cost index in column 1 suggests that the cost wedge between elders and private payor patients is shrinking with productivity for physicians in group practice, while this wedge is growing with productivity for sole proprietors.

The Medicaid extensive margin is measured conditional on acceptance of Medicare Part B. Referring to column 2, relative increases in Medicaid prices, holding price level variation between locations constant, serve to increase acceptance of Dual Eligibles. Recall that prices for Dual Eligibles are regulated at 80 percent of Medicare, unless local Medicaid regulations are more generous than this threshold. Hence, the lone significant effect of financial incentives on the extensive margin of Dual Eligible acceptance is capturing exposure to increases in Dual Eligible prices relative to Medicare for physicians flowing in and out of states above the 80 percent threshold.

On the intensive margin of Dual Eligible acceptance in Table 4, physicians are elastic with respect to relative changes in payment rates. The results in column 3 are interpretable as a logit model for acceptance of Dual Eligibles, where due to location fixed effects only the remaining effect of relative changes in prices are identified, holding the overall level of prices fixed. The implied own and cross price elasticities for the intensive margin of

[6] Because of fixed effects, identification of the sole proprietor effect is off physician flows in and out of group practice.

the Dual Eligible patient mix are once again consistent with predictions from the theory of capacity discrimination. By Proposition 5, when the Dual Eligible price increases holding the Medicare price constant, physicians are more likely to accept Duals if they have marginal cost efficiencies which outweigh their fixed or sunk cost inefficiencies. By now, this is an expected cost relationship: it is the same interpretation explaining empirical results from the event study, results for the extensive margins of acceptance estimated by (2), and the results regarding scale and discrimination from Chapter 2. Moreover, the comparative statics results of the theory predict that relative price changes have symmetric effects on the intensive margin of the patient mix. From column 3 of Table 4, the estimated mean own price elasticity is 1.77, while the estimated mean cross price elasticity is -1.46. The 93rd percentile of the cross price elasticity distribution is -1, and the 96th percentile of the own-price elasticity distribution is +1.

The cost effect in the logit model is an order of magnitude smaller than the payment effect, and has the opposite sign compared to the extensive margin of acceptance. The estimated cost index effect for sole proprietors is once again muted, as with the extensive margin result. This is as predicted by the theory of discrimination if the cost wedge between these consumers is closing as productivity increases (or as the cost index falls) for physicians in group practice, while the wedge potentially grows with productivity for solo physicians. Once again, this is the exact theoretical relationship explaining robustness results for scale and productivity. The effect of an increase in the overall price level is identified by Medicare's physician bonus program, which has a statistically significant and positive effect. In reference to Proposition 5 from Chapter 2, this is further evidence that Dual Eligible marginal cost efficiencies are greater than average cost efficiencies.

The final result of estimating (2) for acceptance and total capacity outcomes is summarized in column 4 of Table 4. In contrast to the event study results illustrated in Figure 6 for the effects of increasing the overall levels of prices and cost, relative changes in financial incentives holding all between-location variation in levels constant have no additional effect on scale for physicians in group practices. The scale of sole proprietors is on average 7.4 percent lower than physicians in group practice, and a relative increase in the Medicare price weakly increases scale. Holding price levels fixed across locations, no other relative change in financial incentive is salient for the physician's choice of scale.

3.2.3 Effects of Relative Changes in Incentives on Practice Patterns

Next, I estimate (2) for outcomes of physician practice patterns. I report results for the (log) number of procedures offered, the (log) labor to capital ratio of treatments supplied, the physician's degree of procedural specialization as measured by the (log) HHI of the bundle supplied, and an indicator variable for whether the physician ever supplied primary care in an outpatient facility instead of an office setting. For procedures, labor to capital ratios, and specialization outcomes I measure financial incentives with log payment and cost GAFs, and so interpret the effects as an elasticity to a relative shock in the financial incentive. For the facility use outcome I measure financial incentives in levels of the GAF indices. As with the acceptance outcomes, for each practice pattern I include interactions of incentives with sole proprietorship, and also include a dummy variable for Medicare physician bonus zip codes. I summarize the results in Table 5.

I find that the number of procedures offered is increasing in relative changes in the Medicare price, decreasing in the Medicaid price, and decreasing in overall cost, holding between-location variation in the levels of these incentives fixed. Payment responses are elastic, and more so for sole proprietors. As the Medicare GAF rises by one percent holding other prices fixed, physicians place 1.8 additional procedures on offer at the mean. Practice patterns are less elastic with respect to cost. The cost effect is not statistically significant for physicians in group practices, but is significant for solo physicians. Sole proprietors are smaller operations, on average having 6.4 percent fewer procedures on offer compared to group practices. The ten percent revenue bonus has no effect on the number of procedures offered.

Table 5: Effects of Relative Changes in Incentives on Practice Patterns

Outcome:	# Procedures offered	Labor/Capital ratio	Bundle HHI	Facility Use
Financial incentives:				
Medicare payment	4.674***	-3.034**	-1.058	0.0489
	(1.064)	(1.278)	(1.374)	(0.100)
Medicaid payment	-4.061***	3.100**	0.910	-0.190
	(1.059)	(1.266)	(1.370)	(0.182)
Cost index	-0.0500	-0.354***	0.247***	-0.0661**
	(0.0498)	(0.0812)	(0.0522)	(0.0331)
SP-financial interactions:				
Medicare payment	0.223*	-0.126	0.132	-0.0806
	(0.122)	(0.159)	(0.122)	(0.0741)
Medicaid payment	-0.0560**	0.00839	0.0143	0.0549**
	(0.0268)	(0.0291)	(0.0266)	(0.0233)
Cost index	-0.200*	0.0970	-0.220**	0.129**
	(0.104)	(0.136)	(0.103)	(0.0653)
Selected covariates:				
Sole proprietor	-0.0635***	0.0299**	0.0318**	-0.0818
	(0.0127)	(0.0137)	(0.0127)	(0.0499)
10% payment bonus	-0.000274	-0.00272	0.0103**	0.00714**
	(0.00522)	(0.00706)	(0.00520)	(0.00348)
Health risk	-0.139***	0.168***	0.130***	0.00165
	(0.00314)	(0.00399)	(0.00318)	(0.00190)
Observations	869,403	831,514	861,465	866,343
R-squared	0.115	0.342	0.091	0.124
Mean Outcome	1.591	1.342	-0.538	0.653

Notes: Author's calculations of (2) from physician microdata. All regressions include consumer disease and demographic controls, as well as physician, county, year, and move year fixed effects. Columns 1-3 are log-log specifications; column 4 is a linear probability model with levels of financial incentives. Std. errors clustered by physician in parenthesis; ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The primary care labor ratio elastically relates to relative changes in financial incentives. However, the most significant determinants of the labor to capital ratio are the diseases and demographics of a physician's patients. These factors, consumer health risk in particular, contribute most to the exceptional fit of the linear regression. Referring to column 2, when the Medicare price goes up, primary care physicians accepting new patients are those who use more equipment, lowering the labor to capital ratio 3-to-1. When the Medicaid price increases, physicians accepting new Dual Eligibles use less capital relative to labor, raising the labor to capital ratio 3-to-1. However, the results imply that a one percent increase in Medicaid payment is only worth 40 additional seconds of physician labor per equipment billing at the mean, an economically small effect.

As local area costs increase, holding price levels fixed, primary care physicians are more likely to use equipment than their own labor. The cost elasticity is an order of magnitude smaller than both payment elasticities. Additionally, sole proprietors are more likely to use labor relative to capital, also worth 40 seconds of labor per equipment billing at the mean. However, solo physician labor ratios do not relate differently to financial incentives than group practice physicians. Again, the ten percent revenue bonus has no effect, indicating that it is relative changes in prices and cost and not the overall level of these incentives which alter physician practice patterns.

Referring to column 3 in Table 5, to the extent that primary care physicians specialize, sole proprietor are more likely to specialize. Otherwise, relative increases in cost matter, and only for group practice physicians. Relative shocks to payment regulation have no association with procedural specialization, consistent with the findings of Gottlieb et al (2010). Physician bonuses have a statistically significant but economically small, positive effect on specialization.

Finally, physician facility use is primarily responsive to relative changes in overall cost, but not to payment incentives. Facility use is decreasing in shocks to the cost index for group practice physicians. The estimated effect is opposite for solo physicians. A one standard deviation increase in cost, holding geographic variation in cost and prices levels constant, reduces a group practice physician's probability of facility use by half a percentage point, but increases the sole proprietor's probability of facility use by 1 percentage point. This result may reflect differences in equipment and other capital on hand between group practice and solo physicians. Medicare reimbursement is higher in the physician's office setting than in the facility, in order to compensate for these capital expenses. Since group practices are more likely to have all the capital inputs needed to service patients, as costs rise they practice more heavily in their office to capture the additional reimbursement. Since solo physicians are less likely to have the requisite equipment on hand, as costs rise they practice more heavily in the facility setting to avoid the fixed costs of capital investment.

3.3 Migration Event Study for Overall Environment

Using the structure on market aggregate outcomes implied by model (2), I extend the event study model (1) in order to estimate the physician's response to overall, potentially unobserved environmental factors subsuming price regulations and input factor costs. There are two purposes for this exercise. First, the physician's response to broader environmental factors will place an upper bound on the effects of prices and cost shifters on behavior, and will allow inference on the contribution of unobserved or idiosyncratic factors to physician decisions. Second, the exercise is way to validate the instrumental variables approach used in Chapter 2 for endogenous scale and productivity measurements.

To derive the effect of the overall environment on a migrant's outcome, first notice that additivity of fixed effects from (2) teases the first order contribution of broader environmental factors, all mean variation in the market cross-section r, from individual effects, all mean variation in the j cross-section. The expectation of (2) conditional on county r is

$$\overline{y}_r = \overline{x}_r\beta + \mathbb{E}(\gamma_j|j \in J_r) + \overline{\mathbf{p}}_r\rho + \gamma_r = \overline{\gamma}_r^p + \overline{\gamma}_r$$

where $\overline{\mathbf{p}}_r = \mathbb{E}((p_{rt}^{MCR}, p_{rt}^{MCD}, \bar{C}_{rt})|r)$ is the vector of location mean incentives, and $\mathbb{E}(\gamma_j|j \in J_r)$ is the mean physician fixed effect given the distribution of physicians who locate in market r. There are two aggregate terms. $\overline{\gamma}_r^p := \overline{x}_r\theta + \mathbb{E}(\gamma_j|j \in J_r)$ is derived from the market's distribution of physician idiosyncrasies and patient match effects. Environmental factors are together in $\overline{\gamma}_r := \overline{\mathbf{p}}_r\rho + \gamma_r$.

The migrant's exposure to all environmental factors is measured via the destination-origin difference in market aggregates:

$$\delta_j := \overline{y}_{d(j)} - \overline{y}_{o(j)} = (\overline{\gamma}^p_{d(j)} - \overline{\gamma}^p_{o(j)}) + (\overline{\gamma}_{d(j)} - \overline{\gamma}_{o(j)})$$

I use leave-out means when measuring δ to avoid a mechanical form of measurement error. As in (1), I then estimate the physician's response to δ_j for each relative migration year $\tau \in \{-3, -2, -1, 0, 1, 2, 3\}$. I define S_τ as the mover's mean response to exposure δ in relative move year τ:

$$S_\tau = \mathbb{E}(\frac{\partial y_{jt}}{\partial \delta_j} | t = \tau)$$

By substituting the definition of the exposure δ_j into (2) through the county fixed effect γ_r, a special version of the event study model (1) is obtained:

$$y_{jrt} = \sum_{\tau=-3}^{3} 1(t = \tau_j)\Big(\gamma_\tau + S_\tau \delta_j\Big) + x_{jt}\beta + \gamma_j + \gamma_t + e_{jrt} \tag{3}$$

The residual e_{jrt} now includes physician specific heterogeneity in the environmental response, $S_{j\tau}$, and physician fixed effects γ_j now absorb the fixed effect for the migrant's origin county.

3.3.1 Identification and Interpretation of S_τ

Because y and δ have identical units, the physician's response to changes in the overall environment S_τ is a unitless effect. Thus, the pre-post migration difference $\mathbb{E}(S_\tau | \tau > 0) - \mathbb{E}(S_\tau | \tau < 0)$ is interpretable as the expected percentage change in a mover's outcome due to the change in environmental factors upon migration. In the case of logged outcome variables, the impulse in S_τ is thus interpretable as the elasticity of an outcome with respect to its market level mean.

As with (1), I assume the special event study residual in (3) takes the form

$$e_{jrt} = B(\mathcal{P}_{jt}) + \tilde{e}_{jrt}$$

where $B(\mathcal{P})$ is the familiar control function to address selection bias in physician mobility. The identifying assumption for (3) is also strict exogeneity. Formally,

$$\mathbb{E}(\tilde{e}_{jrt}|x, \delta, \tau, \mathcal{P}, \gamma) = 0$$

As (3) is a special case of both (1) and (2), omission of subscripts on the conditioning set $(x, \delta, \tau, \mathcal{P}, \gamma)$ is intensional since I must assume strict exogeneity of $\tilde{e}_{jrt}$ conditional on all leads, lags, and spatial perturbations of the observables. In words, I assume exposure to environmental factors is exogenous conditional on the relative migration year, observables, and fixed effects. Two assumptions are embedded in this condition: the migration year and physician heterogeneity in the environment response are conditionally exogenous.

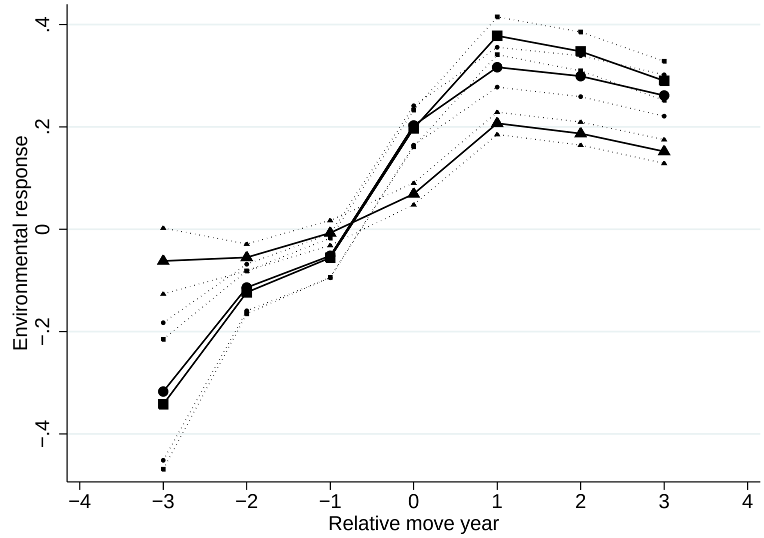
3.3.2 Event Study Results for Response to Overall Environment

I first examine the response of patient acceptance to the broader market environment. I estimate (3) with indicator outcomes for acceptance of Medicare and Medicaid Dual Eligible patients, and for the Medicaid fraction of accepted patients. I also analyze the response of the physician's average patient health risk and the physician's facility usage. I present estimates and confidence intervals of $\hat{S}_\tau$ in Figure 7, all standard errors are clustered at the physician level. I plot the environmental response for each relative move year.

The migrant's percent change in Medicare acceptance attributable to the change in environment is 45 percent, measured by the average height of the jump in $\hat{S}_\tau$ pre and post migration. This implies that location factors, including aggregate financial incentives, determine about half of a primary care physician's Medicare acceptance policy. Heterogeneity in physician productivity and cost are included in the remaining 55 percent. Results for Medicaid acceptance are similar at the extensive margin, standard errors do not reject equality.

For the extensive margins of Medicare and Medicaid, there is an upward trend in $\hat{S}_\tau$ before migration. There is a weak declining post migration trend. Standard errors reject

Figure 7: Response of patient acceptance to the market environment



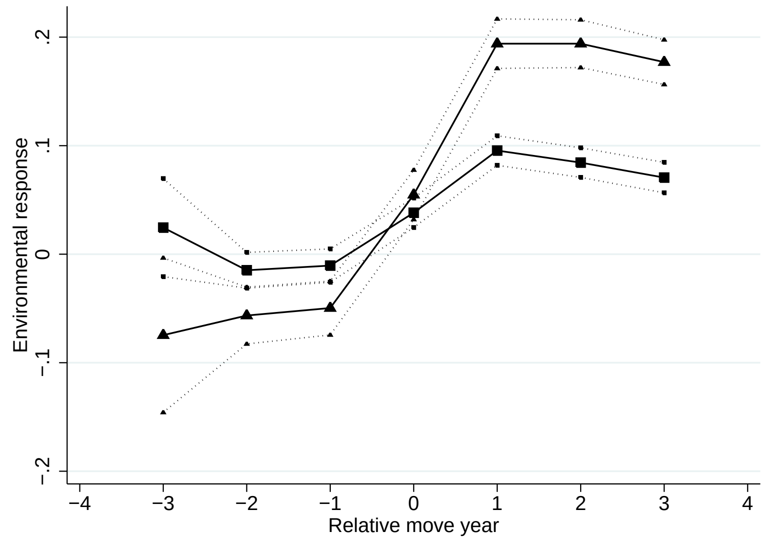
Notes: Estimated percent change ($\hat{S}_\tau$) in Medicare acceptance (circles), Medicaid acceptance (squares), and Medicaid patient fraction (triangles) due to the change in aggregate outcome (δ_j). Dotted lines are $\pm$ 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

pre move stationarity at the five, but not one percent level. Stationarity cannot be rejected post move. Pre migration estimates are also negative. This implies acceptance of patients is adaptive. Movers' acceptance of Medicare and Medicaid increasingly mirrors the market environment prior to migration, they move and adjust, and then do not trend differently from non-movers in their destination.

The intensive acceptance margin is less responsive to the broader environment, as measured by the Medicaid patient share conditional on acceptance of Medicare. The estimates $\hat{S}_\tau$ suggest the Medicaid intensive margin trends in parallel with non-movers both before and after a move. The height of the jump between pre and post migration periods is 22 percent. Location factors, including payment regulations, do not determine the majority of the physician's intensive margin acceptance policy. Physician productivity, capacity for patients from private payors, and other idiosyncratic factors are included in the remaining 78 percent.

To further examine the intensive margin, I notice that patient disease diagnoses and

Figure 8: Response of accepted patient's demographics to the market environment



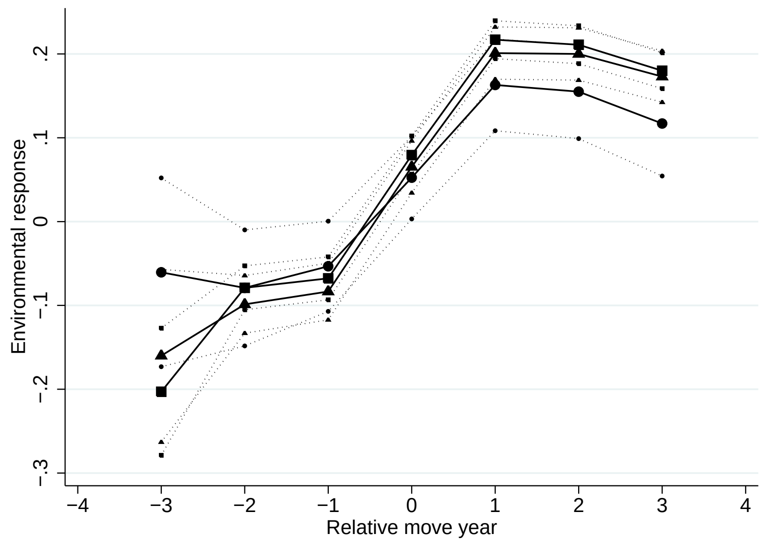
Notes: Estimated percent change ($\hat{S}_\tau$) in accepted patient health risk (squares) and facility use (triangles) due to the change in aggregate demographic rate (δ_j). Dotted lines are ± 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

demographic variables are the most important determinates of the physician's Medicaid acceptance policy. While there are geographic patterns to such factors, results suggest the most salient variation is specific to the physician-patient match. Thus, a mover's reaction to location is a smaller faction of the total response. The responses of accepted-patient's average health risk and of the physician's use of outpatient facilities provide indirect evidence for this. These state variables are thought to be given to the physician by the consumer's health needs and demographics, and thus only controlled through the physician's patient acceptance policy.

Referring to Figure 8, the estimated response of accepted patient health risk to market average health risk is modest. The jump upon migration is eight percent, which indicates that physician acceptance of high risk patients is determined by idiosyncratic factors, and not market average healthiness of elders. In contrast, the facility fraction response to the market environment is 29 percent. This difference is expected: use of an outpatient facility requires spare hospital or other institutional capacity, necessarily a factor of the physician's environment. Standard errors do not reject pre and post move stationarity in $\hat{S}_\tau$ for these outcomes.

Next I analyze the response of physician practice patterns to the market environment. I estimate (3) for the physician's labor to capital ratio, the number of procedures offered, the specialization of the supplied bundle, total capacity for Medicare and Medicaid, and average output per patient. Though the causal effects of financial incentives differed across these outcomes, surprisingly, their responses to broader environmental factors are similar.

Figure 9: Response of practice patterns to the market environment

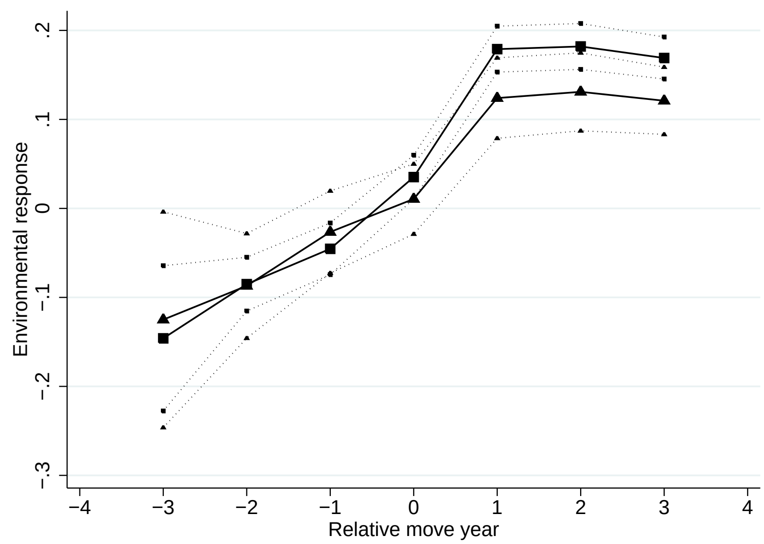


Notes: Estimated percent change ($\hat{S}_\tau$) in labor to capital (circles), number procedures (squares), and specialization HHI (triangles) due to the change in aggregate outcome (δ_j). Dotted lines are ± 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

In Figure 9, I plot estimates and confidence intervals of $\hat{S}_\tau$ for the labor to capital ratio, the number of procedures, and the bundle specialization measure (HHI). Standard errors do not reject pre and post migration stationarity. However, a weak pre trend is evident in point estimates for the number of procedures offered and for the HHI. The mover's percent change procedures offered due to the change in environment is 32 percent. The jump in the labor to capital ratio is 21 percent. The response of physician specialization is 31 percent. Physician productivity and other idiosyncratic incentives are collected in the remaining 68-79 percent.

Market trends affect how accepted patients are treated. However, practice patterns respond less to the market's environment than do the extensive margins of patient acceptance. Moreover, the environmental effect on practice patterns can be similar, while the effects of financial incentives differ. Results for the number of procedures offered versus the physician's degree of specialization illustrates this point. Regulated payments affect total offered procedures, with smaller cost effects. In contrast, payment does not affect the procedural HHI, while cost effects are significant. Yet, equality of the environment responses for these outcomes cannot be rejected statistically, nor rejected visually in Figure 9.

Figure 10: Response of output (RVUs) to the market environment



Notes: Estimated percent change ($\hat{S}_\tau$) in output (squares) and output per patient (triangles) due to the change in aggregate outcome (δ_j). Dotted lines are $\pm$ 2 std. error confidence intervals. Source: Author's calculations from physician microdata.

Finally, I examine effects for total Medicare and Medicaid output and for output per patient. These are plotted in Figure 10. The percent change in total capacity attributable to the change in market average capacity upon migration is 26 percent. The jump in output per patient is 21 percent. A pre trend in $\hat{S}_\tau$ is evident, though standard errors do not reject pre move stationarity for either outcome. There is no evidence of physician adaptation post migration.

The physician's output response to changes in the environment is more similar to that of the intensive patient acceptance margin than to other practice patterns. Excepting a small payment effect for sole proprietors, financial incentives had no effect on total output. This supports the opportunity cost of capacity interpretation.

4 Structural supply model

Unlike variation in regulated payments and local factor prices, idiosyncratic financial incentives are not data. However, there are microdata on inputs and output, and market level data on prices. I apply these data in a structural supply model to examine the idiosyncratic financial incentives of physicians as they determine the fraction of Dual Eligible versus Medicare patients who are accepted. Structural assumptions on payoffs are needed to define and identify these incentives, and behavioral assumptions on suppliers are needed to estimate them. I take a supply side approach. Throughout, I use notation from the previous sections for individual physicians j, in county markets r, over years t.

4.1 Setup

4.1.1 Stylized CMS Fee-For-Service Contracts

The model draws from the structure of CMS fee-for-service (FFS) contracts, details on which were offered in Chapter 1. There are two key features to exploit. The first feature of FFS contracts to exploit is the regulation of output prices, and the additive separability of revenue across the three categories of RVU output: physician work value added Q_L, practice expense value added Q_{PE}, and malpractice risk adjustment value added Q_M. I model the latter as a measure of physician effort abating the risk of medical malpractice liability. Let the physician's fraction of capacity chosen for Medicaid Dual Eligibles be notated l_{jt}. For each output $i \in \{L, PE, M\}$ let the regulated Dual Eligible price be p^0_{irt} and the Medicare price be p^1_{irt}. Define the physician's effective price per RVU given l_{jt} patient mix as

$$\bar{p}_{ijrt}(l_{jt}) = l_{jt}p^0_{irt} + (1 - l_{jt})p^1_{irt}$$

I notate the physician's effective, or average output price vector as a function of the patient mixture $\bar{\mathbf{p}}_{rt}(l_{jt})$.

The physician's FFS revenue for supplying to Medicare and Dual Eligible patients is

$$(4) \quad R(\bar{\mathbf{p}}_{rt}(l_{jt}), Q_{jt}) = \bar{p}_{Lrt}(l_{jrt})Q_{Ljt} + \bar{p}_{PErt}(l_{jrt})Q_{PEjt} + \bar{p}_{Mrt}(l_{jrt})Q_{Mjt}$$

CMS output prices are different across the three output categories, but are highly correlated.[7] Each value added subcomponent is given in the same "Relative Value Units" (RVUs), and the FFS input-output schedule for RVUs does not vary across locations r. Thus, the CMS payment contract is structured as if physicians were a three-product firm where outputs have common units.

The unit commonality of outputs and a set of exclusion restrictions from the input-output schedule is the second advantageous feature of FFS contracts I exploit. In words, in the FFS contract physician labor is excludible from the calculus of practice expense and malpractice value added. Similarly, equipment use is excluded from physician work and malpractice RVUs. I assume these exclusion restrictions over inputs and the additivity of outputs are expressed formally as

$$(5) \quad Q_{jt} = Q_{Ljt}(L_{jt}) + Q_{PEjt}(K_{jt}, MS_{jt}, CL_{jt}, F_{jt}) + Q_{Mjt}(M_{jt})$$

Total RVU output supplied is Q_{jt}, the log of which was a measure of scale used in Chapter 2 and in the regression analysis of the previous section. Physician labor L is measured in minutes. Practice expense factors are notated K for medical equipment, MS_{jt} for medical

[7] These prices are the CMS "conversion factor" times the GPCI index, the same GPCI indices used to construct the GAF financial incentives measure for the regression analysis.

supplies, CL_{jt} for clinician labor, and F_{jt} for other facility or office resources, the latter capturing indirect capital inputs and other overhead. The malpractice risk adjustment Q_{Mjt}, my measure of physician effort, is observed. However, true effort M_{jt} is unobserved, but is known to be excluded from practice expense value added and physician work value added.

4.1.2 Primary care production function

The key technical feature of the supply model is a structural production function for total output. This function represents the physician's physical production technology, subject to the additivity and excludibility conditions of the FFS contract expressed in (5). The pragmatism of (5) is now apparent: knowing the exclusion restrictions and the additivity property of RVUs, I avoid the complications of estimating a joint production function for three-products. Then, I only need to place structure only on a single mapping for the physician's total output $Q : L \times K \times MS \times CL \times F \times M \times \Theta \longrightarrow RVU \in \mathbb{R}_{+}$, where the productivity state is Θ.

I model the log of total RVUs $q_{jt} = \log(Q_{jt})$ as a function of log inputs. I allow the productivity state Θ_{jrt} to have two components: a scalar factor and scope neutral productivity ω_{jrt}, and a vector of shocks to variable factors $\mathbf{v}_{jrt}$. The latter are a formalism for patient health shocks. The former captures effects from unobserved capacity constraints and other variation in scale. Input factors are as given in the previous subsection.[8] I assume the mapping $Q()$ is continuously differentiable and strictly concave.

4.1.3 Prices

I assume physicians are price takers in the output market. I also assume physicians are price takers in input markets. Input prices are notated $\mathbf{w}_{rt} := (w_{Lrt}, w_{PErt}, w_{Mrt})$ for the

[8] For capital factors I do observe facility resources and the use of medical equipment. However, I cannot directly measure the true capital stock or total overhead, which includes the physician's human capital, the value of the office building, computers and other office equipment, and the total value of medical equipment on hand. These are small in the RVU calculation. To that extent, they have a negligible marginal revenue product in the FFS contract.

physician's wage, a price index for practice expense inputs, and finally the local price of medical malpractice insurance. These prices are measured by CMS using a market's average input factor prices, such as average rent on office space or the average employee wage.

The output price taking assumption is without loss since prices per RVU are regulated for Medicare and Dual Eligible patients. Price taking is a weak assumption for practice expense inputs and effort abating malpractice risk. Medical supplies are commodities sold on a national market. Equipment, too, is supplied on a national market where individual physicians have little monopsony power. Clinicians and administrative staff likely have employment alternatives, even in markets with few physicians. Malpractice insurance prices are set by large insurers at the state level.

That physicians are price takers with respect to their own wage is a stronger assumption. Physicians, especially sole proprietors, may be the only demander of their labor. This assumption is necessary without placing additional structure via a labor supply function, which requires knowledge of unobserved capacity for private payor patients. As noted by Becker (1957), monopsony power in wages is one of many microfoundations for the opportunity costs I model next.

One shortcoming to address is with respect to measurement of equipment and effort factor bills. In the necessary conditions of the model, as in De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), the factor bill to revenue ratio helps identify input-specific opportunity costs. The issue is twofold. First, as true effort is unobserved and there is no market price of effort, measurement of the physician's "effort bill" is difficult. Second, the prices of observed practice expense and malpractice factors are only available in ratios, denominated in RVUs relative to the input's national mean factor bill. Given the observed price ratios, computing the factor bill to revenue share requires the factor's value-added ratio.[9] Measuring the effort

[9] Thankfully, the physician's wage bill is well measured, either through use of the CMS GPCI wage index or the local hospital wage for primary care physicians. The latter is observed because of a regulated hospital wage survey. Clinician wages and the medical supplies bill are also more exactly measured in the data. Each market's output/input price ratio and the physician's Medicaid share are data. However, the price ratios are given in output units. As a result, in the necessary conditions the estimated marginal product elasticity must be scaled by the input's observed valued added share to obtain the level of the marginal product. This

bill additionally requires the use of local malpractice insurance prices.

4.1.4 Structural Discrimination Coefficients: Payoffs and Cost

To model variation in physician heterogeneity across patient types, I introduce a measure of the opportunity cost of an input as a percent deviation d from market factor prices $\mathbf{w}$. A physician's primitive factor cost for each patient is thus an opportunity cost parameter, or "discrimination coefficient", times the observed market price, as in the frameworks of Becker (1957) and Arrow (1971). The physician's opportunity cost for each variable input $i \in \{L, K, MS, CL, F, M\}$ given a patient mixture l_{jt} is the average discrimination coefficient

$$\bar{d}_{ijt}(l_{jt}) = l_{jt} d^0_{ijt} + (1 - l_{jt}) d^1_{ijt}$$

I define the vector of these opportunity costs $\bar{\mathbf{d}}_{jt}(l_{jt})$. For example, the physician's primitive cost per unit of labor is $w_{Lrt}\bar{d}_{Ljt}(l_{jt})L_{jt}$, given l_{jt} Medicaid patients are accepted. Likewise, the primitive cost of equipment is the practice expense $w_{PErt}\bar{d}_{Kjt}(l_{jt})K_{jt}$.

Each discrimination coefficient is a physician-year specific parameter. There are distinct discrimination coefficients for Medicare and Medicaid Dual Eligible patients and for each input, or 12 cost parameters per physician. However, given only the capacity share l_{jt} is observed, there is but one degree of freedom per physician to identify a discrimination coefficient for Medicare separately from that of Dual Eligibles. Thus, save for one input, I can only estimate the discrimination coefficient for Medicare patients *relative* to that of Dual Eligibles. I fully identify the discrimination coefficients for physician labor, since this is the most significant variable input to primary care.

allows use of the reported factor bill, which is only provided in the data in units of output added. For example, Q/Q_L scales the labor marginal product elasticity. For labor and effort these output shares are well measured in the FFS contract data. For equipment, the value added share is not directly measured, and must be imputed from the broader value of all practice expenses. I draw from two industry white papers to perform this imputation, Pope and Burge (1993) and Mackinney et al (2003). Each paper finds that "direct practice expenses" account for 33-35 percent of the total practice expense, putting an upper bound on the equipment bill. For a subset of physicians in the data I observe the share of equipment expenses to direct practice expenses, and I use their average share times 1/3 times the total practice expense bill to impute equipment value added.

Given total inputs, the Medicaid fraction, factor prices, and discrimination coefficient parameters, I model the physician's cost function as

$$(6)\quad C(L, K, MS, CL, F, M, \mathbf{w}_{rt}, \bar{\mathbf{d}}_{jt}(l)) = w_{Lrt}\bar{d}_{Ljt}(l)L + w_{PErt}\bar{d}_{Kjt}(l)K$$

$$+w_{PErt}\bar{d}_{MSjt}(l)MS + w_{PErt}\bar{d}_{CLjt}(l)CL + w_{PErt}\bar{d}_{Fjt}(l)F + w_{Mrt}\bar{d}_{Mjt}(l)M + d_{FC}$$

where d_{FC} is an unobserved fixed cost. The data provide measures of baseline average factor bills, i.e. $w_L L$ or $w_{PE} K$. The structural assumption (6) is an empirical counterpart to the theoretical cost function $c(q, l, \theta)$ analyzed in the theory of discrimination from Chapter 2.

I assume that physicians value profit. Given equations (4), (5), and (6) along with the usual definition of static profits, payoffs are modeled

$$(7)\quad \pi_{jrt} = R(\mathbf{p}_{rt}(l), Q_{jt}(L, K, MS, CL, F, M, \Theta_{jrt})) - C(L, K, MS, CL, F, M, \mathbf{w}_{rt}, \bar{\mathbf{d}}_{jt}(l))$$

The profit function is stylized in the sense that physicians choose inputs and output separately for Medicare and Medicaid Dual Eligible patients in reality, while in the data only input-output totals and the Dual Eligible patient fraction are observed. Representing the physician's payoff and choices in the observed variables is restrictive, though without much loss of generality. I assume physicians choose the Medicaid capacity share and production inputs taking prices and their production technology represented as given. Since prices are given and the production technology is continuously differentiable and concave, the payoff function is also concave and continuously differentiable in the arguments to be optimized.

4.2 Necessary Conditions from Optimization

The final element of the model is a behavioral assumption of profit maximization. The necessary conditions for choices of labor, practice expense inputs $i \in \{K, MS, CL, F\}$, effort abating malpractice risk, and the Medicaid capacity share are

$$(8) \quad \bar{p}_{Lrt}(l_{jt})\frac{\partial Q_{jt}}{\partial L} \leq w_{Lrt}\bar{d}_{Ljt}(l_{jt})$$

$$\bar{p}_{PErt}(l_{jt})\frac{\partial Q_{jt}}{\partial i} \leq w_{PErt}(1-l_{jt})d^1_{ijt}$$

$$\bar{p}_{Mrt}(l_{jt})\frac{\partial Q_{jt}}{\partial M} \leq w_{Mrt}(1-l_{jt})d^1_{Mjt}$$

$$(p^0_{rt}-p^1_{rt})Q_{jt} \leq w_{Lrt}L_{jt}(d^0_{Ljt}-d^1_{Ljt}) - w_{Mrt}M_{jt}d^1_{Mjt} - \sum_i w_{PErt}i_{jt}d^1_{ijt}$$

where the degrees of freedom restriction for identification of discrimination coefficients d^1_{ijt} relative to d^0_{ijt} is operationalized by setting $d^0_i = 0$ for practice expense inputs and for effort abating malpractice risk. These conditions hold with equality if solutions are interior, which can be observed in the data. Since 87 percent of physicians in the data have interior choices, I estimate the structural model on that subpopulation.

Observing (8), the average discrimination coefficients $\bar{\mathbf{d}}$ are identified by variation between physicians in factor marginal products, together with the price data. The marginal product elasticities are estimated from the structural production function, and converted to physician specific marginal products via the observed input-output ratio. Variation in the Dual Eligible patient mix then identifies the individual heterogeneity parameters $(d^0_L, \mathbf{d}^1)$, using the first order condition with respect to l.

4.3 Production Function Estimation

I assume the primary care production technology is given by

$$(9) \quad q_{jrt} = f(\tilde{L}_{jt}, \tilde{K}_{jt}, \tilde{M}S_{jt}, \tilde{C}L_{jt}, F_{jt}, \tilde{M}_{jt}) + \omega_{jrt} + \epsilon_{jrt}$$

where q is log total output and log inputs are $(\tilde{L}_{jt}, \tilde{K}_{jt}, \tilde{M}S_{jt}, \tilde{C}L_{jt}, F_{jt}, \tilde{M}_{jt})$. I consider various flexible sieves for $f()$ in (9), but following the literature my preferred specification is a simple Cobb-Douglas relationship.

I estimate (9) by adapting the methodology of Ackerberg, Caves, and Frazer (2015, henceforth ACF) to address simultaneity bias due to unobserved productivity. I also examine OLS estimates of (9), as well as the methodology of Levinsohn and Petrin (2003, henceforth LP). Factor neutral productivity in (9) has the interpretation of unobserved capacity constraints, and unseen shocks to patient health which result in variation in labor or other inputs to produce the same RVUs. The ACF approach is to assume a data generating process with judicious timing of productivity shocks and primary care input choices.

4.3.1 Identification without Independent Health Shocks

Define the physician's information set at time t as I_{jt}. Following ACF, I assume that facility resources F_{jt} are a perfect complement in the production of RVUs. This "value added" framework implies that facility resources are excluded from the marginal products of other factor inputs. The facility choice is deliberate, since in the FFS contract physicians are not compensated for their use facility resources when performing services in that setting, hence using more (or less) of these inputs does not affect their production of RVUs. A similar argument can be applied to medical supplies. A procedure requiring a pair of rubber gloves produces no more RVUs if three rubber gloves are used instead.

I assume input demand for facility resources is strictly increasing in the scalar productivity ω.

$$(A1) \qquad F_{jt} = D_F(L_{jt}, K_{jt}, MS_{jt}, CL_{jt}, M_{jt}, \omega_{jrt})$$

I further assume factor neutral productivity evolves by a first order markov process

$$(A2) \quad \omega_{jrt} = g(\omega_{jrt-1}) + \xi_{jrt}$$

where the function $g()$ is a conditional expectation $\mathbb{E}(\omega_{jrt}|\omega_{jrt-1})$ and the productivity innovation ξ_{jrt} is the structural residual. I assume the error is strictly exogenous given the present information set

$$(A3) \qquad \mathbb{E}(\epsilon_{jrt}|I_{jrt}) = 0$$

I assume innovations ξ_{jt} are unknown in the previous period

$$(A4) \qquad \mathbb{E}(\xi_{jrt} + \epsilon_{jrt}|I_{jrt-1}) = 0$$

Finally, following ACF I assume there are "dynamic" and "static" inputs. The distinction is motivated by a time to build assumption for capital and other sticky inputs. I assume that medical equipment, supplies, and clinician labor are dynamic inputs: K_{jt}, MS_{jt}, and CL_{jt} are chosen given information I_{jt-1}. I assume physician labor and and effort are static inputs: L_{jt} and M_{jt} are chosen given contemporaneous information I_{jt}.

The ACF argument for identification of (9) given these assumptions is as follows. Input demands for (L, K, MS, CL, M) depend on productivity, so the function $f()$ is not identified separately from ω. However, (A3) separates production $f() + \omega$ from the true error ϵ. Since intermediate inputs captured in facility resources are monotone in ω, the function $D_F()$ is surely invertible. Furthermore, by design of the FFS contract this input is a perfect complement to production of RVUs and thus excludible from the marginal products of other inputs. Together with (A2), excludibility and (A4) provides a set of moment conditions to identify the parameters of $f()$ and $g()$ given their implied structural residual $\hat{\xi}(f, g)$. For example, the static input assumption provides a moment condition for physician labor

$$\mathbb{E}(\hat{\xi}_{jrt}(f, g)L_{jrt-1}) = 0$$

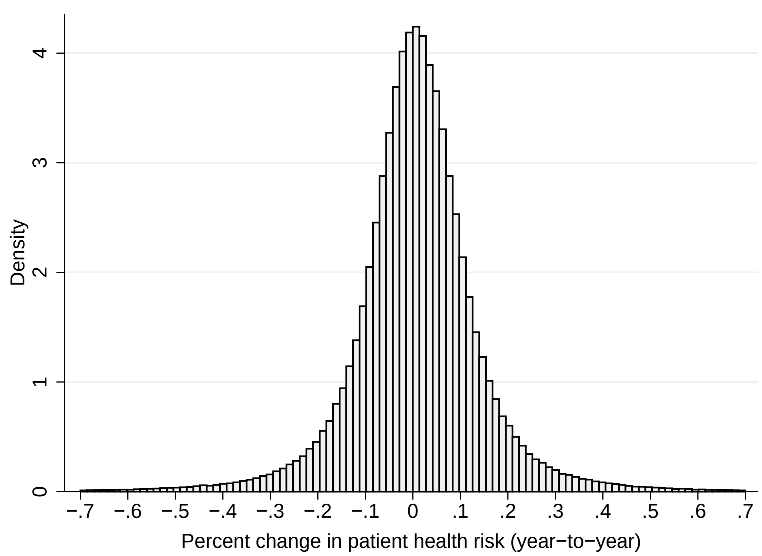
The time to build assumption provides a moment condition for medical equipment

$$\mathbb{E}(\hat{\xi}_{jrt}(f, g)K_{jrt}) = 0$$

I exploit further lags of inputs to build additional moments with the implied productivity innovation given a set of guessed parameters, and estimate the functions $\hat{f}()$ and $\hat{g}()$ by nonlinear GMM.

Identifying power in this method comes from within-physician variation in production choices. Unlike many industrial applications of production function estimation, I do observe substantial variation over time in the physician's production inputs. An intuitive reason for this variation is shocks to patient health. In Figure 11 below I plot within-physician year-over-year percent changes observable patient health, using the HCC score measure described in Chapter 1. This evidence suggests the health of a physician's accepted patients is highly volatile over time. Moreover, the empirical distribution is strikingly normal.

Figure 11: Empirical density of patient health shocks



Notes: Health shocks measured as year-over-year change in the log of a physician's average patient health risk (HCC) score. Source: Author's calculations from physician microdata.

4.3.2 Identification with Independent Health Shocks

ACF show that (A1)-(A3) is sufficient to estimate the variable factor parameters of $f()$ if there are shocks $\mathbf{v}_{jrt}$ to variable factor productivity that are independent of ξ, and that are learned after intermediate inputs are chosen by the physician. This assumption "saves" the first stage of LP. Shocks to patient health are an intuitive microfoundation for $\mathbf{v}_{jrt}$. For

(A1)-(A3) to be sufficient to estimate input marginal products, physicians must choose their facility policy, clinical staff, and/or store room of supplies prior to learning their patient's present health state. After observing the patient's health, the physician then chooses labor, which equipment on hand to use, and how much effort is required to address those needs.

While ACF argue this latter data generating process is a very special case, it is plausible in the present setting. There is evidence in the data supporting this assumption. Referring to Figure 11, the physician's observed average patient health varies widely from year to year. Unobserved patient health is likely also volatile.

4.4 Structural Model Results

In this section, I report the estimated parameters of the primary care production function and the opportunity cost estimates identified from variation in marginal products. I denote the marginal product elasticity of input i $\hat{\alpha}_{ijt}$. I present the estimated means of these structural parameters in Table 6. I treat facility resources as the complementary (Leontief) intermediate input. Adding over the marginal product elasticities provides a measure of returns to scale in primary care production for Medicare and Medicaid Dual Eligible patients. Estimated returns to scale are 0.964 in the preferred ACF specification, implying a decreasing but near-constant returns to scale technology.

The standard errors indicate precise estimates at the mean, significant to three or more digits. The marginal products of physician labor and effort are paramount in primary care, and their prominence is not sensitive to the estimation method. The physician labor elasticity is 0.469 in the baseline ACF estimator. The OLS and LP estimators provide a smaller physician labor elasticity, 0.403 and 0.329 respectively. The main distinction between ACF and LP methods is apparently that physician labor and effort exchange rolls: the ACF effort elasticity is 0.31 while the LP estimate is 0.44. The labor product is sensitive to the exclusion of effort from the production function. When effort is excluded, I find that the labor elasticity nearly doubles in size, combining the underlying products of effort and labor. This is evidence supporting the interpretation of the malpractice risk adjustment as physician effort.

The marginal products of equipment, clinician labor, and medical supplies are secondary to those of physician labor and effort in primary care. In the baseline ACF estimator, the equipment elasticity is 0.067. Moving to LP estimator only changes this estimate beyond its significant digits. The OLS estimator implies a slightly larger equipment product, 0.086. The ACF estimate of the product of clinician labor is 0.046, but rises to 0.138 with the LP estimator. Medical supplies exhibit an opposite trend across estimation methods. The marginal product elasticty of supplies is 0.072 using ACF, but falls to 0.035 in LP methodology.

Table 6: Estimated Structural Production Parameters

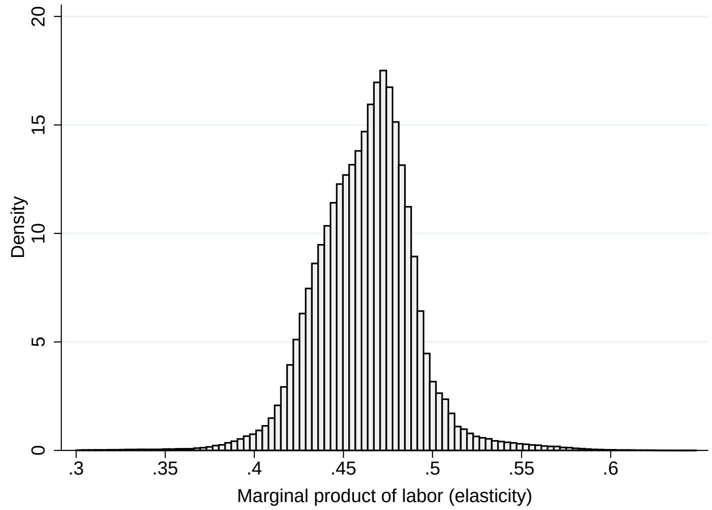
Input	ACF	OLS	LP
Physician labor $\hat{\alpha}_L$	0.469	0.403	0.329
	(0.138)	(0.003)	(0.003)
Equipment $\hat{\alpha}_K$	0.067	0.086	0.067
	(0.009)	(0.001)	(0.001)
Effort (malpractice) $\hat{\alpha}_M$	0.310	0.409	0.440
	(0.144)	(0.003)	(0.003)
Clinic labor $\hat{\alpha}_{CL}$	0.046	0.064	0.138
	(0.007)	(0.001)	(0.001)
Medical supplies $\hat{\alpha}_{MS}$	0.072	0.041	0.035
	(0.021)	(0.000)	(0.000)

Notes: Estimated marginal product elasticities $\alpha_k = \frac{\partial Q}{\partial k}\frac{k}{Q}$. Observations: 864,323. Author's calculations from physician microdata using subsample of physicians with interior inputs. ACF is Ackerberg, Caves, and Frazer (2005) estimation method, LP is Levinsohn and Petrin (2003) estimation method. Delta method standard errors clustered by physician in parenthesis for OLS and LP estimators. ACF standard errors are from a bootstrap with 100 replications.

4.4.1 Robustness: Marginal Product Heterogeneity

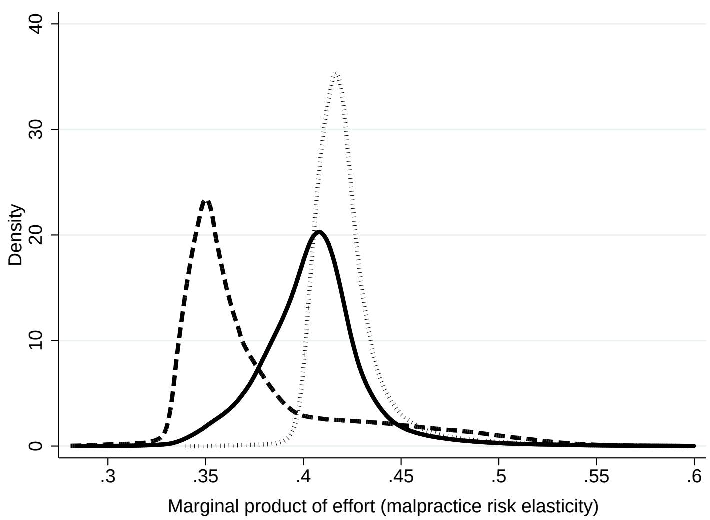
In the preferred ACF Cobb-Douglas estimator, there is no variation across physicians in the marginal product elasticity. Thus, all variation in marginal products across physicians arises from variation in their observed input-output ratio. Using the LP estimator, I consider robustness to this assumption and estimate a translog production function $f()$.

Figure 12: Empirical Distribution of Marginal Product Elasticity of Physician Labor $\hat{\alpha}_{Ljt}$



Notes: Histogram of estimated RVU marginal product elasticity of labor $\frac{\partial Q}{\partial L}\frac{L}{Q}$ from the structural production function. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

Figure 13: Empirical Distribution of Marginal Product Elasticity of Physician Effort $\hat{\alpha}_{Mjt}$

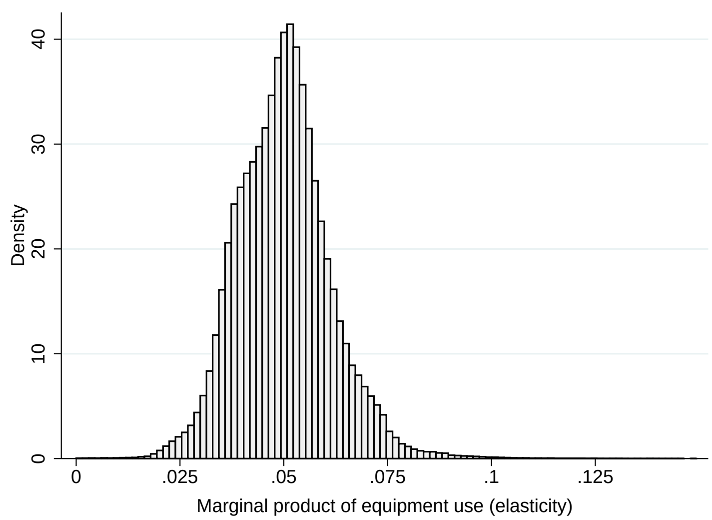


Notes: Office setting (dashed, left) versus facility setting (dotted, right) versus mixed facility/office (bold, center) kernel density estimated RVU marginal product elasticity of effort $\frac{\partial Q}{\partial M}\frac{M}{Q}$ from the structural production function. A rectangular kernel function was used. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs,observations 864,323.

Labor is the most important factor in the primary care production function. I find that the marginal product of labor varies little cross physicians. The density of the estimated marginal product elasticities $\hat{\alpha}_{Ljt}$ is plotted in Figure 12. The average and median elasticities of labor are near 0.46. The distribution has minor skew, with support over [0.19,0.68]. I find the distribution of labor marginal products is similar whether the primary care physician supplies in an office or a in facility setting.

The effort marginal product elasticity has a similar distribution to labor. There is little variation in this elasticity across physicians. I plot the estimated distribution of $\hat{\alpha}_{Mjt}$ across office/facility settings in Figure 13. The mean and median effort elasticities are near 0.41. Unlike labor minutes, I find that the marginal product of effort abating medical malpractice risk falls with facility use. Physicians are jointly rather than solely liable in the facility setting, there is less incentive to exert effort.

Figure 14: Empirical Distribution of Marginal Product Elasticity of Equipment $\hat{\alpha}_{Kjt}$



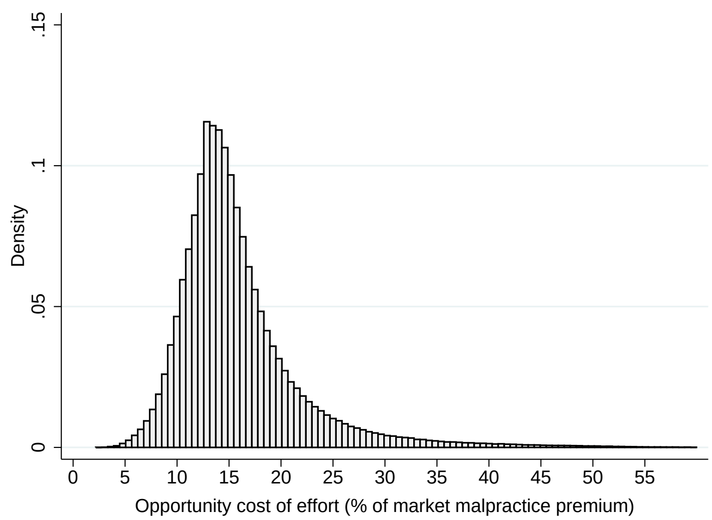
Notes: Histogram of estimated RVU marginal product elasticity of equipment capital $\frac{\partial Q}{\partial K}\frac{K}{Q}$ from the structural production function. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

The estimated marginal products of equipment capital $\hat{\alpha}_{Kjt}$, clinical labor $\hat{\alpha}_{CLjt}$, and medical supplies $\hat{\alpha}_{MSjt}$ remain small compared to the products of physician labor and effort in primary care. The dispersion in their marginal contributions across suppliers is also slight. I plot the empirical density of the marginal product elasticity of equipment in Figure 14. The mean and median of the distribution is near 0.05, an order of magnitude below labor and effort. The mean and median of the estimated clinician labor elasticity is also 0.05; for medical supplies, like rubber gloves, the mean and median are near 0.02.

4.4.2 Estimated Discrimination Coefficients

The necessary conditions for an interior optimum profit relate variation in the physician marginal product of labor to variation in the physician's average opportunity costs of labor $\hat{\bar{d}}_{Ljt}(l_{jt})$, given the Medicaid capacity share l. The average is identified by the estimated marginal product and price ratio data. Together with the necessary condition for the capacity share, this in turn identifies the decomposition of cost across Medicare and Medicaid Dual Eligibles $(\hat{d}^0_{Ljt}, \hat{d}^1_{Ljt})$.

Figure 15: Heterogeneity in the Opportunity Cost of Physician Effort

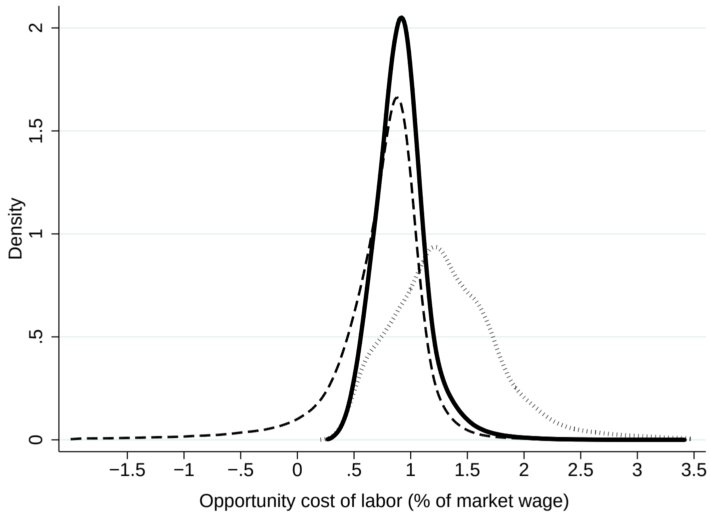


Notes: Histogram of estimated heterogeneity in the opportunity cost of physician effort for Medicare $\hat{d}^1_M$ patients relative to Dual Eligible patient $\hat{d}^0_M = 0$. Opportunity cost is measured as a percent deviation from the county average medical practice insurance premium. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

I find that the opportunity cost of effort abating malpractice risk for Medicare patients is greater than for Dual Eligibles. In Figure 15, I plot the density of estimated heterogeneity across physicians in this cost, measured relative to the market malpractice insurance price. The results indicate that the average opportunity cost of effort is 16 times the insurance premium. Measuring in reference to insurance prices is the best that can be done, but is nonetheless opaque.

A different metric is available by conversion of the implied cost into output units using the measured factor bill and price index. In output units, the mean effective cost of effort is worth 2,051 RVUs, and is worth 1,163 RVUs at the median. By comparison, in the FFS payment contract only 138 RVUs are the average product of malpractice risk adjustments. This suggest physicians view the effective cost of effort avoiding malpractice liability as at least as great as their effective cost of labor.

Figure 16: Heterogeneity in the Opportunity Cost of Physician Labor



Notes: Kernel density of estimated heterogeneity in the opportunity cost of physician labor for Medicare $\hat{d}_L^1$ (dashed, left), Medicaid $\hat{d}_L^0$ (dotted, right), and the physician's average $\hat{\bar{d}}_L$ (bold, center), measured as a percent deviation from the county wage. A rectangular kernel with bandwidth 0.1 was used. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

I plot the empirical distributions of estimated labor cost heterogeneity in Figure 16. This cost is measured relative to the market average physician wage. The median and mean of physician expected opportunity cost of labor, given the Medicaid capacity share, are 91 and 93 percent of the market wage, respectively. The distribution of Medicare discrimination coefficients is dominated by the Medicaid distribution. The estimates imply the physician's marginal cost of labor for Dual Eligible patients is greater than for Medicare patients, the average being 32 percent above the market price. Denominating this cost in output units, I find the mean effective cost of labor is worth 1,649 RVUs, and that the median is worth 1,097 RVUs. In contrast to effort, in the FFS payment contract physician labor produces 1,847 RVUs at the mean. This suggests the effective cost of labor is well compensated by Medicare, while the cost of effort is mostly uncompensated.

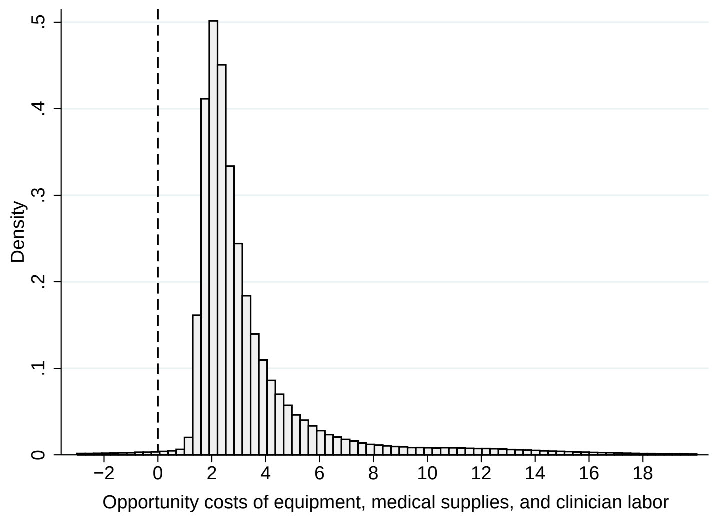
The theory of discrimination together with the regression results from this chapter and from Chapter 2 suggest the physician's marginal cost for Dual Eligibles must be lower than for Medicare patients. On one hand, the estimated marginal costs of effort for Medicare patient relative to Dual Eligibles supports this interpretation. On the other hand, the marginal costs of physician labor do not. The remaining costs are due to practice expenses: medical equipment, supplies, and clinician labor. Since these costs are estimated relative to the same input price index, I can compute a composite discrimination coefficient for practice expenses by adding over these inputs' first order conditions

$$\hat{d}^1_{PEjt} = \frac{\bar{p}_{PErt}(l_{jt})}{w_{PErt}(1-l_{jt})}\left(\hat{\alpha}_K \frac{Q_{jt}}{K_{jt}} + \hat{\alpha}_{MS}\frac{Q_{jt}}{MS_{jt}} + \hat{\alpha}_{CL}\frac{Q_{jt}}{CL_{jt}}\right)$$

I plot the empirical distribution of $\hat{d}_{PE}$ in Figure 17. As with effort, the marginal cost of practice expense inputs is higher for Medicare patients than for Dual Eligibles.

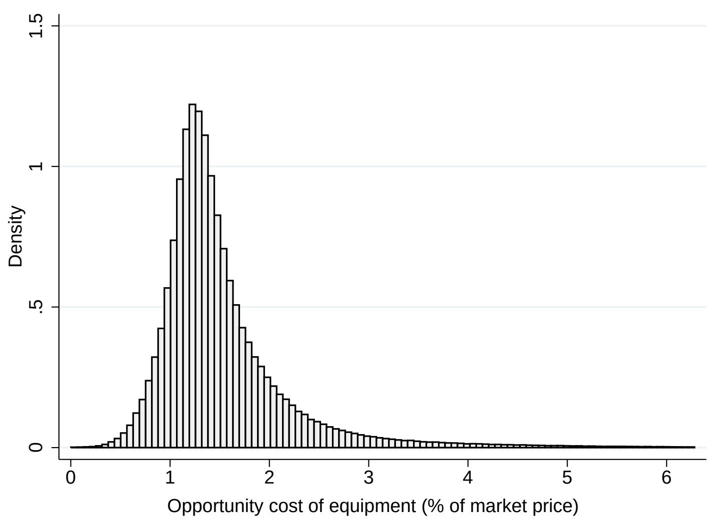
The practice expense marginal cost efficiency for Dual Eligibles is driven largely by saving in the use of medical equipment. To illustrate this, I plot the distribution of discrimination coefficients for equipment in Figure 18. The average opportunity cost of equipment is 60 percent above the market rate, the median 36 percent above the market. I find wide dispersion across physicians in the cost of equipment, and that the effective equipment bill is small. Denominating this bill in output units, I find that the average effective equipment bill is worth only 235 RVUs, and that the median is worth 152. In the FFS contract, equipment produces 181 billable RVUs on average and 106 billable RVUs at the median. This suggests that, even accounting for the large opportunity cost relative to the market, equipment is not

Figure 17: Heterogeneity in Practice Expense Discrimination Coefficients



Notes: Histogram of estimated heterogeneity in the opportunity cost of equipment use for Medicare patients $\hat{d}^1_{PE}$ relative to Dual Eligible patients $d^0_{PE} = 0$. Opportunity cost is measured as a percent deviation from the mean county practice expense. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

Figure 18: Heterogeneity in the Opportunity Cost of Equipment



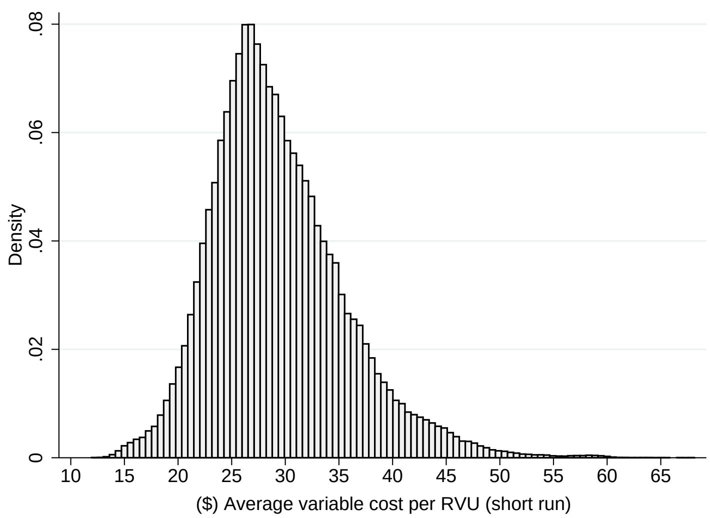
Notes: Histogram of estimated heterogeneity in the opportunity cost of equipment use for Medicare patients $\hat{d}^1_K$. The Medicaid cost is normalized, $d^0_K = 0$. Opportunity cost is measured as a percent deviation from the mean county practice expense. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

comparable to the effective costs of physician labor and effort in primary care. The combined bill for practice expenses is larger, but still below the bills for physician labor and effort.

4.4.3 Estimated Cost, Profit, and Dual Eligible Efficiencies

Using the estimated discrimination coefficients, I estimate each physician's average variable cost and average variable profit per RVU supplied. Since the estimated production technology for Medicare and Dual Eligibles is approximately constant returns to scale, the average variable cost estimates are very close to the physicians underlying marginal cost of output. I then use the discrimination coefficients to estimate the lower bound of marginal cost efficiencies from supplying to Dual Eligibles versus Medicare patients, the empirical counterpart to c_{ql} and c_l/q from the theory of discrimination.

Figure 19: Empirical Density of Estimated Average Variable Cost



Notes: Histogram of estimated physician-year heterogeneity in average variable cost per Relative Value Unit (RVU). Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

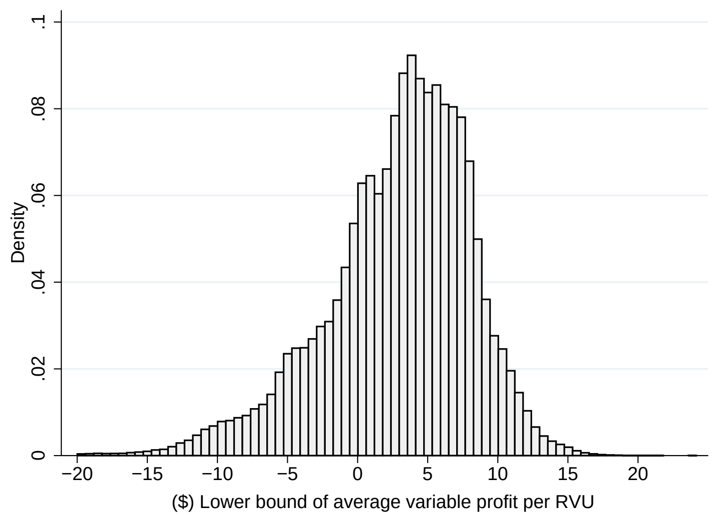
Average variable cost, approximately marginal cost due to near-constant returns to scale in production, is dispersed across primary care physicians. The average across the population is \$29.33 per RVU supplied, well below the regulated national average Medicare price per RVU ranging from \$34.02-\$35.82 over 2012-2015. Figure 19 provides empirical evidence for productivity dispersion in primary care, resulting in a sizable mass of inframarginal physicians earning rent.

To quantify these inframarginal rents, I use the cost estimates together with output and price data to obtain provide a lower bound on average variable profit. Recalling the structure of revenue in the FFS contract has three RVU categories, the implied lower bound of variable profit from the necessary conditions of optimization is

$$\bar{\pi}^{lb}_{jrt} = \bar{p}_{Lrt}(l_{jt})(\frac{Q_{Ljt}}{Q_{jt}} - \hat{\alpha}_L) + \bar{p}_{PErt}(l_{jt})(\frac{Q_{PEjt}}{Q_{jt}} - \hat{\alpha}_K - \hat{\alpha}_{MS} - \hat{\alpha}_{CL}) + \bar{p}_{Mrt}(l_{jt})(\frac{Q_{Mjt}}{Q_{jt}} - \hat{\alpha}_M)$$

I estimate profit for each physician each year and plot the distribution of estimated profit in Figure 20.

Figure 20: Empirical Density of Average Variable Profit

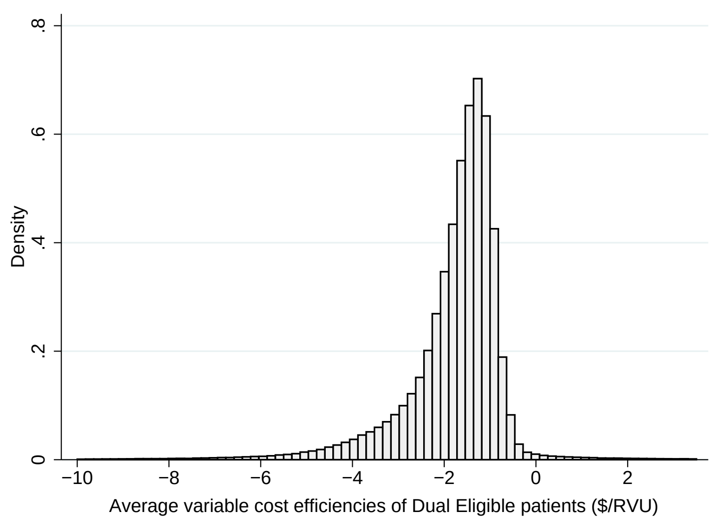


Notes: Histogram of estimated physician-year heterogeneity in average variable profit per Relative Value Unit (RVU). Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

The estimates plotted in Figure 20 illustrate that 76 percent of primary care physicians have a positive lower bound on variable profit. To provide context for these estimates. This implies a lower bound markup of ten percent at the median, and eight percent at the mean. In the right tail of the estimated profit distribution the markup is 43 percent. However, many physician-year observations have a lower bound markup estimated below zero. Interestingly, the majority of physicians with a negative markup in one year, have positive markup in other years. This suggests the left tail may not be the results of empirical error, but rather the result of natural fluctuations in variable profit and the long-run incentives inherent to a physician's market entry/exit decision.

Finally, I compute the implied average cost efficiencies from Dual Eligible patients. The theory and empirical results thus far suggest the marginal cost of Dual Eligibles should be lower than the marginal cost for Medicare patients, and that marginal cost efficiencies should outweigh average cost efficiencies. Thus, these estimates are an lower bound on Dual Eligible marginal cost efficiencies. However, since production is near constant returns to scale, average cost efficiencies should closely approximate marginal cost efficiencies. As expected, I find that the average physician's average variable cost of low income Dual Eligibles is 5.76 percent below the average variable cost of Medicare patients. The expected savings from a Dual Eligible is $1.78 per RVU supplied. I plot the empirical distribution of these efficiencies in Figure 21.

Figure 21: Estimated Wedge in Average Variable Cost for Dual Eligible Patients



Notes: Histogram of estimated physician-year heterogeneity in the derivative of average variable cost per Relative Value Unit (RVU) with respect to the Dual Eligible patient mixture, $c_l/q \approx c_{ql}$. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs, observations 864,323.

The products of physician labor and effort are most important in primary care. From the estimated production function, there is not much dispersion across physicians in the marginal product of each of these factors. The products of equipment, and intermediate factors like clinician labor or medical supplies, are an order of magnitude smaller than those of labor and effort.

The discrimination coefficients for physician labor vary considerably across Medicare and Medicaid Dual Eligible patients. Dual Eligibles imposes a higher opportunity cost on physician time. However, the evidence suggests the opportunity costs of effort and equipment are much lower Dual Eligibles compared to Medicare patients. One explanation for this is if Dual Eligibles are receiving care from physicians during sub-periods of slack capacity. If equipment is on hand and there are few other patients in the office demanding effort, the marginal cost of these inputs is effectively zero.

A cost efficiency from Dual Eligibles was expected from the regression results and the theory of capacity discrimination. Together, the discrimination coefficients imply that the marginal cost of a Dual Eligible is indeed lower than the marginal cost of a Medicare patients, despite a higher estimated marginal cost of physician labor. Marginal cost efficiencies arising from medical equipment, clinician labor, medical supplies, and effort are overwhelming. Dispersion in the estimated discrimination coefficients results in dispersion in physician average costs as well as in the cost wedge between Medicaid Dual Eligibles and Medicare Patients. I now exploit these results to analyzes several important counterfactuals for the primary care industry.

5 Price Counterfactuals

5.1 Competitive Prices

I consider two free market counterfactuals to the present regulated price environment. I analyze the short run partial equilibrium consequences of price competition, since full long run analysis requires knowledge of the consumers' demand curves. The partial equilibrium closely approximates the true counterfactual if elders' demand for primary care is inelastic, a plausible assumption. Throughout, in order to predict counterfactual physician behavior resulting from the new equilibrium, I use the estimated responses of acceptance and scale to a changes in price levels from the event study regressions, and the estimated responses to changes in relative prices from the saturated fixed effects models. I first predict the short run distribution of prices should the primary care industry be monopolistically competitive, and the resultant effects on patient acceptance and scale. The second counterfactual predicts perfectly competitive prices when there are inframarginal firms and potential price discrimination against Dual Eligibles, as well as the resultant patient acceptance rates and scale.

5.1.1 Monopolistic Competition

Monopolistic price competition results in average cost pricing, dispersed not only across locations but also across physicians within a location. Using the estimates of average cost from the structural model for physicians who earn a positive profit, I find that monopolistically competitive prices are largely within the support of observed regulated prices. As expected, prices fall under price competition compared to regulation. The average price under monopolistic competition falls to \$29.32 per RVU, a reduction of 16.2 percent from the presently regulated price level.

Monopolistically competitive prices would result in a 4.7 percent reduction in output capacity for Medicare and Dual Eligible patients. At the extensive margin of acceptance of both types of elders, the new lower price level would reduce the probability of acceptance by 2.38 percentage points. On the intensive margin of acceptance, the declining price level would reduce the Dual Eligible patient mixture by 3.26 percentage points. If consumer access is the criterion, competition in this scenario is not welfare improving for patients.

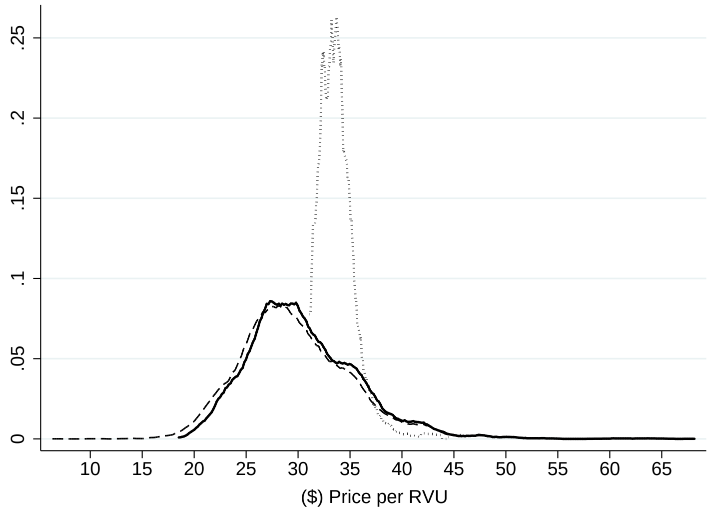
5.1.2 Perfect Competition with Inframarginal Firms

Next I consider the model of a homogeneous product industry with inframarginal firms, and apply the results of Proposition 7 from Chapter 2 to compute counterfactual prices. If the cost wedge between patients is shrinking with productivity, then whether there is price discrimination between customers in equilibrium depends on the mass of highly productive physicians in each market. If this mass is sufficiently large, price competition will erase price discrimination even if there is heterogeneity in cost across patient types. In this setting, the equilibrium Medicare price is the average variable cost of the marginal firm in each county each year, since Dual Eligibles are marginal cost efficient compared to Medicare elders. To compute counterfactual short run competitive equilibrium prices p^{1CE}, I take the empirical maximum average variable cost from the structural model, market by market. To compute the price for Dual Eligibles, I use Proposition 7, $p^{0CE} = p^{1CE} + \max\{c_l/q\}$, and find the smallest average cost efficiency market-by-market to compute the price wedge. If the mass of productive physicians is significant, then $\max\{c_l/q\} \approx 0$ and there will be little or no price discrimination. I plot the distribution of perfectly competitive prices across counties, alongside the prevailing regulated prices, in Figure 22.

The perfectly competitive price counterfactual differs from the monopolistically competitive counterfactual in three important ways. First, the average price across markets falls to \$30.5 per RVU for Medicare patients, a reduction of only 12.8 percent from the average of regulated prices. Second, the mass of productive physicians is nearly sufficient to erase price discrimination against Dual Eligibles, but serves to *increase* low income patient prices relative to current regulation. The average Dual Eligible price is \$29.81 per RVU under price competition, compared to \$27.87 per RVU under status quo regulation, a relative increase of 6.95 percent. Finally, there is a long thin upper tail of the distribution of both competitive prices which *increase* relative to present regulation.

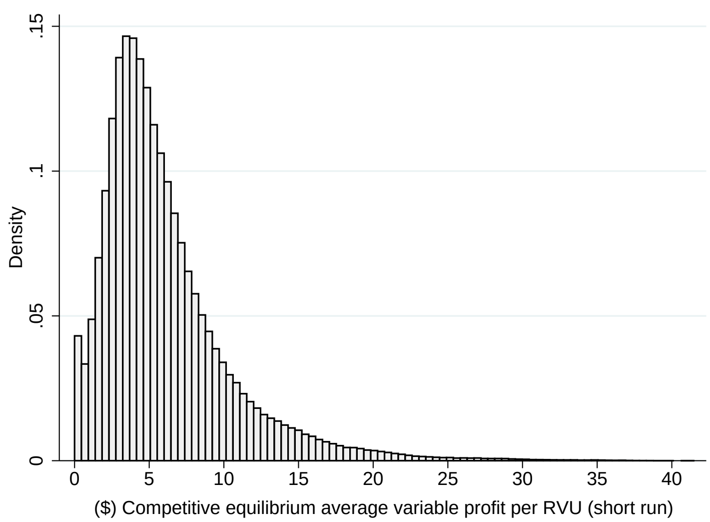
The new lower price level would, on average, reduce physician scale for elders by 3.72 percent. On the extensive margin of discrimination, the probability an elder is accepted as

Figure 22: Densities of Current Prices versus Estimated Short Run Competitive Prices



Notes: Kernell density estimates of counterfactual price per Relative Value Unit (RVU) under short run competition with inframarginal firms for Dual Eligible patients (dashed), Medicare Patients (solid), and current regulated Medicare prices (dotted). A rectangular kernel was used with a bandwidth of $1 per RVU. Source: Author's calculations from county aggregates of physician microdata panel.

Figure 23: Empirical Density of Average Variable Profit in Competitive Equilibrium



Notes: Histogram of estimated physician-year heterogeneity in average variable profit per Relative Value Unit (RVU), assuming short run competitive equilibrium with inframarginal firms. Source: Author's calculations from physician microdata panel, from the subpopulation with interior inputs.

a patient subsequently falls by 1.88 percentage points as a result of the lower perfectly competitive price level. However, conditional on physician acceptance of elders, the probability of Dual Eligible acceptance goes up as a result of their relative price increase, an effect worth 3.99 percentage points at the extensive margin for a net increase of 2.11 percentage points. On the intensive margin of discrimination, the lower price level would on average reduce the Dual Eligible capacity share by 2.57 percentage points. However, the relative increase in their price would serve to increase their capacity share 3.2 percentage points, for a net increase of 0.64 percentage points on the intensive margin of acceptance.

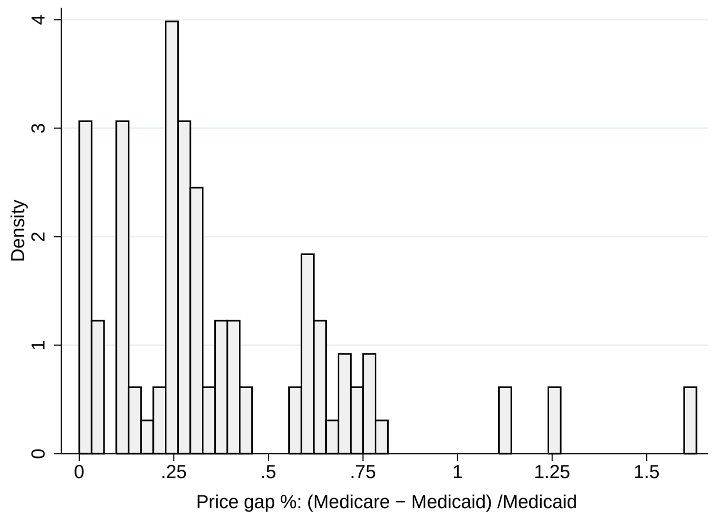
If patient access and reduced discrimination is the welfare criterion, price competition when there are inframarginal firms is welfare diminishing for Medicare patients because of lemon dropping, but welfare improving for low income Dual Eligibles because of cherry picking of low cost patients. However, if prices are part of the welfare calculus, then price competition may be welfare improving for those Medicare patients who do obtain access to primary care, while welfare diminishing for low income Dual Eligibles who face now higher prices.

Also notice the effect on total welfare is ambiguous, because physicians away from the break even margin enjoy profit from their productivity. While the results imply some consumers are hurt by competition, others might gain, and a subset of physicians in the industry may be better off in the perfectly competitive scenario. The distribution of counterfactual average variable profits is strictly positive. In contrast, the estimated average variable profit under regulation was negative for 24 percent of firms. The difference is driven by markets in the far right tail of cost, where prices increase relative to regulation. I plot the counterfactual competitive profit distribution in Figure 23. The point mass at zero is the estimated density of marginal physicians. The density of positive profits are driven by inframarginal cost heterogeneity.

5.2 Equal Medicare and Medicaid prices

Leaving the Medicare price unchanged, I last consider the effect of increasing the Medicaid Dual Eligible price equal to Medicare's across locations. I plot the estimated percent increase in Medicaid prices in Figure 24. I compute the counterfactual increase in acceptance along the Dual Eligible extensive and intensive margin, accounting for the regulated floor in the payment gap at 80 percent of Medicare.

Figure 24: Percent Price Increase Closing the Medicare-Medicaid payment gap



Notes: Histogram of the countferfactual percent increase in Medicaid prices required to erase the Medicare-Medicaid payment wedge. Source: Author's calculations from county financial incentives panel.

The mean percent increase in Medicaid Dual Eligible payments is 20.17 percent, accounting for the Dual Eligible price floor, leaving the overall price level fixed. The estimated acceptance response on the extensive margin due to the relative price increase raises the probability of Dual Eligible acceptance by 11.56 percentage points. At the population mean, this implies counterfactual Medicaid Dual Eligible acceptance would rise from 62 percent to 73.5 percent under equal payments. The new acceptance rate more than matches the observed acceptance rate of Medicare, at 71 percent, with additional acceptance due potentially to these patients' cost efficiencies. On the intensive margin, the 20.17 relative percent increase in price, holding overall level constant, increases the Dual Eligible patient mix by

9.28 percentage points.

These results suggest two way to address discrimination in primary care access. First, the access gap might be closed by allow price competition, at the cost of decreased access for regular Medicare patients. This is the result of relatively high private sector prices, and physician lemon dropping of elder patients. Second, discrimination could be addressed with regulation. The results imply that merely equating prices for Medicare and Medicaid Dual Eligible patients would more than close the access gap for low income patients. Moreover, such a policy may increase efficiency, due to Proposition 1 from Chapter 2.

6 Conclusion

Physicians have financial incentives to accept high payment patients, especially when capacity is scarce. The Code of Medical Ethics allows discrimination against consumers from unprotected classes, and for reasons related to physician profits. The terms "lemon dropping" and "cherry picking" have entered industry colloquial, suggesting financial incentives are most salient at the formation of the physician-patient relationship. In this chapter, the data and analysis suggest that physicians are more likely to acceptance more profitable patients, and that price wedges and cost heterogeneity can explain discrimination against one type of patient over another.

I find little evidence that financial incentives affect practice patterns in a nefarious way. Overall levels of financial incentives apparently have no effect on practice patterns. Only relative changes in payment and cost are associated with practice pattern variation, but these effects can be interpreted entirely as physician heterogeneity acting through an extensive margin of patient acceptance. Output capacity or scale for these patients may even be inherited from this patient acceptance decision. Seemingly, once the physician-patient relationship is formed, primary care physicians first do no harm. However, facing capacity constraints, they allocate capacity favorably to the most profitable patients. Private payors take priority, Medicare receives capacity next, and I find that Medicaid Dual Eligible patients are most often the lemons who are dropped.

This ranking of profitability arises despite the relatively low marginal costs of primary care for Dual Eligibles. The payment generosity of private payors ensures those patients are prioritized, and the payment regulations of CMS ensure there is less room for Medicare and Medicaid Dual Eligible patients in a physician's day. Between Medicare and low income Dual Eligible patients, the empirical evidence implies a similar relationship due to the low income price wedge. Together with the theory, these empirical results for the effects of increasing price levels, increasing cost levels, and relative changes in price and cost on discrimination and scale imply that low income elders have lower marginal costs than other Medicare patients, and that the cost wedge between these consumers is shrinking in physician productivity. The regression evidence and theory support the empirical findings of Chapter 2.

The structural results confirm the expected cost relationships from the regression results and theory: Dual Eligible patients have a lower marginal cost of primary care. Low income patients impose a high opportunity cost of physician labor, but come with efficiencies in the costs of all other primary care production inputs. Because the structural model can only estimate variable costs, this work cannot speak to the fixed and sunk cost inefficiencies for Dual Eligibles that are also expected from the regression evidence and theory. However, industry heuristics suggest Medicaid Dual Eligible patients come with a substantial paperwork burden, and thus require significant time from administrative staff in the physician's hire. The added burden of complying with Medicaid regulation is an intuitive annual sunk cost per Dual Eligible able to generate the expected average cost - marginal cost relationship. I leave quantification of these sunk costs for future work.

This study finally is an exercise in the unintended consequences of price regulation and price competition in the healthcare industry. The counterfactual analysis suggests that allowing the free market and competition to set prices would be welfare reducing for Medicare consumers, but potentially welfare improving for Medicaid Dual Eligibles. If the industry is perfectly competitive, the results imply that competition would erase prevailing price discrimination against low income patients, but at the expense of raising these patients' price for healthcare.

In the regulatory counterfactual, the results imply that the low income access gap would be erased if payment for treating Dual Eligible patients were set equal to that of Medicare, without the adverse consequence of reduced access for Medicare patients which was observed under competition. According to the efficiency result from the theory of capacity discrimination in Chapter 2, this policy would also result in a cost-efficient patient mixture. Since competitive prices also erase most of the price wedge between consumers, the results imply a perfectly competitive market would also be more cost efficient that the status quo, but at the expense of reducing access for middle and high income Medicare patients. Only an omniscient principle could weigh the cost efficiencies resulting from equitable price regulation or price competition with their public finance and consumer welfare costs. Yet, a remedy for lemon dropping may be as simple as eliminating regulated price discrimination for low income consumers.

7 Bibliography

Ackerberg, Daniel A., Kevin Caves, Garth Frazer. "Identification Properties of Recent Production Function Estimators." Econometrica. 83(6). November 2015. pp. 2411-2451.

Alexander, Diane. "How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs." Working paper. April 2017.

Becker, Gary. "The Economics of Discrimination." 2nd Edition, University of Chicago Press. 1957.

Bresnahan, Timothy F., Peter C. Reiss. "Entry and Competition in Concentrated Markets." Journal of Political Economy. 99(5). Oct. 1991. pp. 977-1009.

Charles, Kerwin Kofi, and Jonathan Guryan, "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination." Journal of Political Economy 116, no. 5, (October 2008): 773-809.

Clemens, Jeffry, Joshua D. Gottlieb. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" American Economic Review. 104(4). Apr 2014. pp. 1320-49.

Clemens, Jeffry, Joshua D. Gottlieb. "In the Shadow of a Giant: Medicare's Influence on Private Physician Payments." Journal of Political Economy. 125(1). Feb 2017: pp. 1-39.

Dahl, Gordon B. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." Econometrica. 70(6). November 2002. pp. 2367-2420.

De Loecker, Jan, Frederic Warzynski. "Markups and Firm-Level Export Status." American Economic Review. 102(6). Oct 2012. pp. 2437-71.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K. and Pavcnik, N. "Prices, Markups, and Trade Reform." Econometrica 84. 2016. pp. 445-510.

Finkelstein, Amy, Matthew Gentzkow Heidi Williams "Sources of Geographic Variation in

Health Care: Evidence From Patient Migration." The Quarterly Journal of Economics, 131(4) November 2016: Pages 1681-1726.

Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn G. Andrews, Jonathan S. Skinner, and Jason M. Sutherland. 2010. "Prices Don't Drive Regional Medicare Spending Variations." Health Affairs, 29(3): pp. 537-543.

Hotz, Joseph V., Robert A. Miller. "Conditional Choice Probabilities and the Estimation of Dynamic Models." The Review of Economic Studies. 60(3). Jul., 1993. pp. 497-529.

Lee, Lung-Fei. "Generalized Econometric Models with Selectivity." Econometrica. 51(2). Mar., 1983., pp. 507-512.

MacKinney AC, McBride TD, Shambaugh-Miller MD, and Mueller KJ. "Medicare physician payment: Practice expense." RUPRI Rural Policy Brief. 8(9). 2003.

Molitor, David. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration" American Economic Journal: Economic Policy (forthcoming)

Molloy, Raven, Christopher L. Smith, and Abigail K. Wozniak. "Declining Migration within the U.S.: The Role of the Labor Market." NBER Working Paper No. 20065. April 2014.

Page, Leigh. "Are More Doctors Cherry-Picking and Lemon-Dropping Patients?" Medscape. Feb 15, 2017.

Pope, Gregory C., and Russel T. Burge. "Allocating Practice Expense Under the Medicare Fee Schedule." Health Care Financing Review. 14(3). 1993. pp. 139-162.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Health Care." In Handbook of Health Economics. Vol. 2, Elsevier. Eds. Mark V. Pauly, Thomas G. McGuire and Pedro P. Barros, pp. 45-93.

Wooldridge, Jeffrey M. "Econometric Analysis of Cross Section and Panel Data." Second Edition, 2010. The MIT Press.