Are firms' cost predictions accurate? Evidence from Medicare Advantage

Keaton Miller
Univ. of Oregon*

Sichao Jiang
Univ. of Oregon†

February 2022

Abstract

Many firms employ forecasters to predict future market conditions. We investigate the accuracy of these predictions in the context of Medicare Advantage, where insurers receive subsidies from the government and compete to provide health insurance to seniors. As part of the "bidding" process, firms must submit forecasts of their costs. Insurers have incentives to report accurately, as these predictions are used to determine both the level of the subsidy and (implicitly) the degree to which that subsidy can be spent on various plan features. We collect data on predictions and realized expenses per member per month at the plan-service-category level from 2008-2015, and document three stylized facts. First, on average firms overestimate future costs. Second, this overestimation decreases with the experience of the firm. Firms in more competitive markets (as measured by the number of other firms present) form more accurate estimates. We show that firms with higher costs than expected generally offer plans that feature greater patient cost sharing (i.e. higher deductibles and copays).

*keatonm@uoregon.edu

†sichaoj@uoregon.edu

1 Introduction

How well can firms predict their costs? As costs play a key role in decisions about prices, benefits, and the adoption of innovative technologies (Dranove and Satterthwaite, 2000, Jena and Philipson, 2008, Baicker and Goldman, 2011), firms which consistently forecast their own costs inaccurately may make decisions that are consistently suboptimal *ex post* for both themselves and for their customers. While it is common knowledge that firms use the data they collect on their customers and specific market environment to make forecasts, it is difficult to evaluate performance on this dimension as forecasts and resulting performance evaluations are generally considered highly valuable trade secrets. As a consequence, previous empirical work on firm learning has largely focused on macroeconomic conditions. For instance, Coibion et al. (2018) use survey data from New Zealand and find that firms share a pervasive inattention to macroeconomics and have a wide set of beliefs especially regarding inflation; competition provides incentives to collect more accurate macroeconomic information.

We examine forecasts made by insurers participating in the Medicare Advantage (MA) market in the United States. MA is a "managed competition" (Miller et al., 2021) environment in which the government provides subsidies to firms who then compete with each other to provide health insurance plans to seniors that replace traditional fee-for-service Medicare benefits (TM) with (generally speaking) privately-administered managed care benefits. MA is substantial in its own right; the federal government spends more than $125 billion annually on payments to MA insurers in addition to the premiums paid by enrollees. As part of the regulatory "bidding" process detailed in Section 2, firms report detailed data on realized past medical costs and forecasted future costs across a number of categories. Firms then choose cost-sharing product characteristics relevant to those categories. Crucially, firms are incentivized to report accurate forecasts (i.e. the best forecasts they have) by the bidding system, as the forecasts constrain the set of possible products the firm may choose to offer. If a firm reports that it expects costs to be higher than it truly believes, it may be forced to offer an inferior product (i.e. higher copays and fewer benefits) than is optimal given its true beliefs.

We begin by combining the reported costs and forecasts of future costs to construct forecast errors. We construct the error both in total medical expenses and in the three largest individual categories: inpatient facility, surgery, and professional services. We show that forecast errors in these categories are smaller than the overall error but are positively correlated with each other. Thus, for a given plan-year, firms tend to "miss" in the same direction across service categories. We then document changes in the average forecast error as firms gain experience in the market and explore heterogeneity across firm experience (as measured by the number of enrollees in the firm's plans) and competitive environment. Using simple models of the prediction error, we find that the magnitude of the forecast error decreases as the firm gains experience and competitors, consistent with theories of learning.

Finally, we examine the relationship between these forecast errors and product characteristics. We show that, after accounting for time- and location-specific factors, firms which under-predict their costs offer plans with greater patient cost-sharing—i.e. products that are 'worse' from the consumer's perspective.

We contribute to an extensive literature studying firm learning and the evolution of equilibria over time (Milgrom and Roberts, 1991, Fudenberg and Kreps, 1993, Börgers and Sarin, 1997, Hart and Mas-Colell, 2000). This literature has branched in several directions, including using lab experiments to test and verify theoretical models (see e.g. Erev and Roth, 1998) and increasing the sophistication of the modelled learning process (see e.g. Camerer et al., 2002, Crawford, 2003). Such refinements include adding a belief elicitation process into the model; Rutström and Wilcox (2009) find that eliciting beliefs can change games and that there is a significant difference between the strong-elicitation treatment, but not between the no-elicitation treatment and the less intrusive weak-elicitation procedure. In addition to the empirical work citied above, others have also studied firm learning and adaptation. Graham et al. (2005) find that upper managers main focus lies on meeting or beating earning benchmarks; specifically the quarterly earnings for the same quarter last year and the analyst consensus estimate. Most managers prefer to have smooth earnings and to maintain the stability of future predictions. Their information disclosure coincides with the effort to stability where they strive for clarity but also select news in a strategical manner. Doraszelski and Markovich (2007) use frequency response data from the new UK electricity system to see how firms compete and respond with prices and how the price converge to its equilibrium using an adaptive learning model. They find that especially during the middle phase, the best-fitting models are those in which firms more heavily weight recent rival behavior in forming beliefs about their rivals' bids and adaptively learn about the price elasticity of demand.

This work also builds upon a literature examining the behavior of firms and seniors within the MA system. Much of the work on the bidding system has focused on how changes in the subsidy rates offered by the government passes through to benefits (see e.g. Song et al., 2013, Duggan et al., 2016, Cabral et al., 2018). We complement that work by investigating other potential drivers of firm behavior. Other work has focused on patient outcomes with a particular focus on comparisons between MA and the traditional Medicare system (see e.g. Figueroa et al., 2020, Park et al., 2020, Meyers et al., 2021). We shed additional light on the way in which MA firms play a role in seniors' health.

The remainder of this paper is organized as follows: Section 2 contains some general background of the MA market and its bidding system; Section 3 describes the data and methodology of our practice; Section 4 presents the results in two different settings and then discusses our findings; Section 5 concludes.

2 Bidding in Medicare Advantage

Medicare Advantage (MA) was developed as a response to rising costs in the Traditional Medicare (TM) system.[1]Under TM, the government pays service providers according to a fee-for-service (FFS) schedule. Under MA, the government pays insurers a per-enrollee subsidy that is adjusted for observable risk factors (i.e. demographic characteristics and diagnoses) but that does not vary by realized medical expenditures. The enrollee experience is substantially different: TM enrollees may generally choose any provider for any service, whereas MA enrollees generally face restricted provider networks and referral requirements

[1]One of the biggest goals to implement MA is to increase efficiencies of managed care and to save money: specifically to minimize the inefficiencies induced by the inevitable errors in TM's administered price system, by allowing the health plans and providers to negotiate prices or, in some cases, to integrate the finance and delivery functions (McGuire et al., 2011).

for specialists.

Firms offer plans on an annual basis. The bidding process for plans which will cover medical expenses[2] in year t begins in the spring of the previous year. First, the Centers for Medicare and Medicaid Services (CMS) releases 'benchmark' subsidy rates—the per-enrollee subsidy that CMS will pay firms for an 'average' risk enrollee—that vary at the county level. Firms then submit detailed proposals to provide MA plans. These 'bids' begin with detailed information about the firm's realized medical expenses per-member-per-month in year $t-1$ across more than 15 categories. Firms then report their projected expenses for each category for year t using the 'Bid Pricing Tool' released by CMS.[3] These estimated costs are to provide the minimal number of required benefits for an average mix of risks against the county-level benchmark. If a plan's bid was greater than the benchmark, it was required to collect the difference from its enrollees through a monthly premium. If it was lower, 75 percent of the difference was to be returned to enrollees in the form of supplemental coverage or lower premiums, and in an effort to reduce Medicare's obligations, the remaining 25 percent of the savings was to be returned to the Medicare program (McGuire et al., 2011). Thus, it is important for plan providers to have a reasonable estimate to plan ahead and hedge against any uncertainties for MA enrollees.

Figure 1 illustrates an example of these submissions for a plan offered for the 2015 benefit year as entered into the Bid Pricing Tool in 2014. The left-hand columns represent the plan's realized costs for the 2013 plan year, and the right-hand columns represent the firm's projections of their costs for the 2015 plan year. This plan added dental coverage between 2013 and 2015, and so the firm did not have previous data on dental costs for this particular plan with which to form projections. Ultimately, this firm projected that the 'per-member per-month' cost of this plan would increase by \$39.65 from 2013 to 2015, or 7.4% of the realized costs in 2013. In dollars, this increase is driven mostly by projected increases in inpatient and skilled nursing facility costs, as well as professional service costs.

[2]These bids are based on the plans' estimates of the cost of providing required Medicare Part A, which covers most medically necessary hospital; skilled nursing facility; home health and hospice care; and Part B services, which cover medically necessary services by providers and other services deemed medically necessary, to cover an average beneficiary (Berenson et al., 2015).

[3]While firms are required to provide projections based on their past experience, these projections may be manually overridden.

A key question with respect to any reported firm projections, particularly those which the firm knows will eventually be released to the public, concerns the extent to which these reports reflect the actual beliefs of the firm i.e. the beliefs which underlie the firm's actions in the market (see e.g. Graham et al., 2005, Rutström and Wilcox, 2009, Coibion et al., 2018). In this environment, incentives for accurate reporting come from the payment system. The projected costs are use to form a final 'bid amount' which is then compared to the benchmark subsidy rates. Plans with a bid amount that is higher than the benchmark must charge premiums to enrollees; supplemental premiums may also be charged if plans include benefits beyond those offered by TM. Firms that bid below the benchmark receive a portion of the difference as a 'rebate' that must be passed on the consumers through plan benefits. The rebate payment varies across firms and over time based on the CMS 'star rating' measure of insurer quality, which is a summary of multiple measures of service quality such as the fraction of members receiving influenza vaccinations, the 30-day hospital readmittance rate, and enrollee qualitative assessments of care quality. After taking into account risk adjustment, we can write the rebate payment as a function of the bid b_{jt} and the plan-level benchmark $B_{jt} = B_{mt} \times \phi_{ft}$ with

$$reb(b_{jt}; B_{jt}, \lambda_{ft}) = \begin{cases} \lambda_{ft}(B_{jt} - b_{jt}) & \text{if } b_{jt} < B_{jt} \\ 0 & \text{if } b_{jt} \geq B_{jt} \end{cases}, \tag{1}$$

where λ_{ft} is the rebate percentage.

Crucially, any reductions in patient cost-sharing, relative to TM, *must* be 'paid for' with rebate funds. Cost-sharing plan features such as copays and deductibles have previously been shown to be determinants of demand for MA plans (see e.g. Curto et al., 2021, Miller et al., 2021). As a consequence, the projections made by firms directly constrain the set of possible product characteristics offered to consumers along dimensions that are relevant for their (the firms') success in the market. If a firm reports that it expects to have much higher costs than it truly believes it will face, then it may be constrained to offer an less-than-optimally generous plan (i.e. higher copays/deductibles and fewer additional benefit categories) and therefore lose market share. If a firm reports that it expects to have lower costs than it truly believes it will have, it may not receive enough revenue (in the form of the government subsidy plus premium payments from enrollees) to pay for the care that it covers under the plan.

One of the major differences between MA and TM is that MA plans must provide all of the mandated insurance benefits of TM in exchange for a capitated monthly payment (Abaluck et al., 2021). In general, MA plans typically offer more generous benefits and lower cost-sharing than TM, whereas MA plans tend to have limited physician networks and require higher cost-sharing for costly services, perhaps because of these risk adjustment practices (Meyers and Trivedi, 2021). According to the Medicare Payment Advisory Commission (MedPAC), MA payers plans to continue to increase enrollment by offering extra benefits that beneficiaries find attractive. According to their 2021 report, bids slightly decreased to 87 percent of TM, a record low (Commission et al., 2021).

3 Data and Methods

We collect data on all bids and plans offered from 2007 to 2015 from public CMS files, including data generated by submissions made through the Bid Pricing Tool and (separately) detailed data on the provision of benefits. Following Miller et al. (2021), we focus on the market for individual insurance. We drop plans sponsored by employers and plans designed for individuals who are "dual-eligible" for Medicare and Medicaid.

For each plan, we collect enrollment, the average per-capita payment, the star rating, the deductible, the out of pocket limit, and copays for primary care visits, specialist visits, and 7-day hospital stays.[4] We construct the forecast error for each plan j offered in year t by comparing the projected costs reported at $t-1$ to the actual costs reported at $t+1$: $Err_{jt} = actual_{j,t+1} - projected_{j,t-1}$. In other words, positive prediction errors indicate that firms underestimated their costs, while negative prediction errors indicate that firms overestimated their costs.

We construct the overall forecast error (i.e. the forecast error in the Total Medical Expenses line in Figure 1), as well as forecast errors in the Inpatient Facility, Professional,

[4] A small fraction of plans use coinsurance cost-sharing mechanisms. We convert these to copayments using the Medicare Physician Fee Schedule and the American Hospital Association Annual Survey. Details and code are available upon request.

and OP Facility - Surgery service categories. We choose these three categories as the largest components of plans' medical expenses. We also construct the percent forecast error as $Err_{jt}/projected_{j,t-1}$.

We construct four additional covariates from these data. First, we identify the state in which the plan has the most enrollees. Second, we construct the total previous enrollment at the contract-level to capture variation in the experience of the firm submitting the bid.[5] Finally, we construct two variables designed to capture the extent of competition: the share-weighted average number of contracts and plans offered by competitors in each plan's service area. That is, we weight the number of competitor contracts/plans offered in each county in which the plan offered by the share of the plan in that county.

Table 1 provides summary statistics on our $11,670$ plan-year observations offered from 2008-2013 (i.e. after calculating forecast errors). On average, firms over-estimate their costs by \$270 per-member-per-month (PMPM), though the interquartile range include firms who underestimate their costs. The average prediction error in inpatient facility costs is \$103 PMPM, 38% of the average overall prediction error. Professional and outpatient surgery facility costs are overestimated on average by \$77 and \$16 PMPM, respectively. Over 75% of the plan-years in our data feature zero deductible and a limit on out-of-pocket expenses. A few plans feature zero copays; most feature positive copays for primary care visits, specialist visits, and hospital stays. Roughly one quarter of plan-years did not receive a star rating from CMS, either because they were too new or because not enough data was available for CMS to evaluate the plans. The median plan is in a competitive environment featuring more than 10 firms offering approximately two plans each.

Table 2 reports correlations between the overall forecast error and the three components discussed above. All correlations are positive and greater than 0.7, indicating that errors in the prediction process tend to compound across cost categories, rather than offsetting.

Table 3 details the distribution of the overall forecast error across quartiles of our measure of experience in the top panel and our contract measure of competition in the bottom

[5] Firms sign contracts with CMS to offer potentially several plans; plans under the same contract generally offer similar provider networks and are available in similar geographies. Large firms (e.g. Aetna, Blue Cross Blue Shield, etc.) may have multiple contracts with CMS to offer plans in different areas; we calculate enrollment at the contract-level instead of the firm level to capture the possibility that the firm's experience may vary locally in ways that affect their ability to accurately forecast costs.

panel. With respect to experience, the biggest difference is between the first (lowest experience) quartile and other quartiles; the difference between (for example) the third and fourth (highest experience) quartile is minimal. This suggests that while there may be some benefit to experience, the benefit is quickly realized. This is perhaps reasonable if firms are using regression techniques to estimate future costs, as the power of such techniques does not scale linearly. With respect to competition, variation across quartiles is minimal – if anything, this cut of the data suggests firms' prediction errors increase in magnitude as the number of competitors increases. While one might expect firms to invest more in accurate predictions when the competitive environment is more saturated, the presence of other firms may increase the variance in costs as insurers compete to add providers to their network (Gaynor et al., 2015).

Figure 2 illustrates the evolution of the distribution of forecast errors over time. While the interquartile range shrinks dramatically from 2008 to 2011, along with a decrease in the absolute value of the mean and median, these trends reverse in 2012 and 2013. We note that several provisions of the Affordable Care Act affecting MA went into effect in around this time, which may have increased the aggregate uncertainty experienced by MA firms.

Our analyses broadly fall into two categories. First, we study the relationship between the prediction error and covariates of interest including our measures of experience and competition, taking into account the possibility that prediction errors may be serially correlated. We model prediction error outcomes y_{jt} as

$$y_{jt} = \alpha_y y_{j,t-1} + X_{jt}'\alpha_x + FX_t + FX_f + FX_s + \epsilon_{y,jt}, \tag{2}$$

where α_x is the vector of coefficients of interest (where X_{jt} includes a constant), FX are fixed effects (t indicates time, f indicates the insurer sponsoring plan J, and s indicates the primary state in which j is offered), and $\epsilon_{y,jt}$ are unobservable factors influencing the forecast error.

Second, we study the relationship between the last-period prediction error and the product characteristics chosen by the firm. We model product characteristic x_{jt} as

$$x_{jt} = \beta_y y_{j,t-1} + Z'_{jt}\beta_z + FX_t + FX_f + FX_s + \epsilon_{x,jt}, \tag{3}$$

where β_y is the coefficient of interest, Z'_{jt} includes the star rating and the average per-enrollee payment from the government, FX are the fixed effects described above, and $\epsilon_{x,jt}$ are unobservable factors influencing product characteristics.

We estimate the parameters of Equations (2) and (3) using OLS and report heteroskedasticity-robust standard errors.

4 Results

In this table we report the results of our analyses. We begin by exploring the relationship between prediction errors and plan-level observables and then continue by examining the relationship between product characteristics and prior prediction errors.

4.1 The relationship between prediction errors and plan-level observables

Table 4 reports estimates of the parameters of Equation (2) when the prediction error is measured in levels. Column (1) reports estimates for the overall prediction error. The error is increasing in the previous forecast error, the number of competitors, and the age of the plan (measured as the number of years the plan has been in the market since 2007, the plan year that Medicare Advantage implemented risk adjustment). The error is decreasing in the number of competing plans, though this is estimated with slightly more noise. These patterns generally continue across the major cost components reported in Columns (2)-(4), though the sign on the estimated relationship between the prediction errors and the log of the total previous contract enrollment varies.

In isolation, these results suggest that prediction errors may worsen over time, due to the positive coefficients on both the past prediction error and the age of the plan. However, we note that the mean prediction error is significantly negative; thus these results may instead indicate that firms in our data are becoming more accurate. To investigate this possibility, we re-estimate Equation (2) where y_{jt} is the absolute value of the prediction error. Table 5 reports the results. We estimate in Column (1) that last period's error, previous enrollment, the number of competitors, and the age of the plan are all associated with a decrease in the magnitude of the overall prediction error. This pattern is largely consistent across the major cost components in Columns (2)-(4). While the number of competing plans enters positively for several regressions, the effect size is smaller than that of the number of competitors, indicating that, overall, increased competition is associated with smaller forecast errors (in magnitude).

4.2 The relationship between prediction errors and product characteristics

Table 6 reports estimates of the parameters of Equation (3) when the prediction error is measured in levels. Each column represents a distinct product characteristic: the deductible, the out-of-pocket spending limit, and copays for primary care, specialist, and hospital visits. We include the star rating (and relevant indicators) and risk-adjusted payments as covariates to control for plan quality and observable differences in mean costs across geographies (Miller et al., 2021). At the point estimates, increases in the last-period prediction error (i.e. as firms actual costs are more and more above their prediction costs), the deductible, out-of-pocket expenditure limit, specialist visit copay, and hospital visit copay all increase (though the hospital visit copay is estimated imprecisely). In other words, as firms realize costs are higher than they believed, they seek to pass-through a portion of those costs onto consumers (see e.g. Butters et al., 2020, Kim, 2021). Indeed, the negative coefficient on primary care copays (indicating that first realizing higher costs than expected decrease the patient-facing cost of primary care visits) is consistent with this hypothesis, as increasing primary care service utilization is thought to reduce health care expenses overall (see e.g. Starfield et al., 2005).

As above, while these results are informative about the *direction* of the relationship, they are less information about the impact of the magnitude of the prediction error. We therefore re-estimate the parameters of Equation (3) with the absolute overall prediction error as an independent variable. The results are reported in Table 7. Across product characteristics, the estimated coefficients on the absolute prediction error are the opposite sign of those reported in Table 6. This is perhaps best interpreted in light of the results reported in Table 4. Firms with large prediction errors are likely to more accurately predict their costs in the future. If uncertainty had pushed firms to offer plans with high levels of cost sharing (i.e. large deductibles, etc.), a reduction in that uncertainty should push firms to offer more desirable plans (see e.g. De Vany and Saving, 1977).

5 Discussion and conclusion

The idea that firms can learn about and adjust to changes in market environments is at the heart of many common models of equilibrium (see e.g. Berry et al., 1995, Ifrach and Weintraub, 2016). While there is an extensive literature developing the theory of firm learning, the empirical literature has largely been limited by a lack of revealed preference data on firm beliefs. We fill that gap by studying the MA market, in which health insurance firms are incentivized to accurately disclose their best estimates of their own future medical costs as part of a mandatory regulatory process. These disclosures are then released publicly some years after the plan year.

We document the relationship between these estimates along two directions. First, we consider the "input" side: what factors influence the accuracy of the forecasts? We find that as the experience of the firm increases, as the age of the plan increases, and as the competitive environment intensifies, the magnitude of forecast errors decreases. In other words, those factors which we might reasonably expect to positively affect firms' ability to forecast their own costs do so in the direction consistent with reasonable priors.

We then turn to the "output" side: How do accurate and inaccurate forecasts influence product characteristic decisions. We document two stylized facts. First, the more firms underestimate their own costs, the more they seek to pass their costs on to consumers in the future through cost-sharing benefit structures. Second, the larger the magnitude of past forecast errors (in either direction), the more firms seek to improve their plans by offering reduced cost sharing.

Taken together, our results paint a picture of firms behaving as is often implicitly assumed in dynamic models of imperfect competition. Even though the firms in our data consistently overestimate costs on average, at the plan level the predictions increase in accuracy, allowing firms to offer better products (i.e. insurance plans with lower out-of-pocket costs) to consumers.

We conclude by pointing out that significant gaps in our understanding of firm expectations remain for future research to fill. For example, firms generally must predict not just supply (cost) conditions but also demand conditions. These predictions may behave differently than cost predictions, particularly as they are made in the context of competitors who are also making similar predictions.

References

Jason Abaluck, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. Mortality effects and choice across private health insurance plans. *The quarterly journal of economics*, 136 (3):1557–1610, 2021.

Katherine Baicker and Dana Goldman. Patient cost-sharing and healthcare spending growth. *Journal of Economic Perspectives*, 25(2):47–68, 2011.

Robert A Berenson, Jonathan H Sunshine, David Helms, and Emily Lawton. Why medicare advantage plans pay hospitals traditional medicare prices. *Health Affairs*, 34(8):1289–1295, 2015.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):pp. 841–890, 1995.

Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of economic theory*, 77(1):1–14, 1997.

R Andrew Butters, Daniel W Sacks, and Boyoung Seo. How do national firms respond to local shocks? *Available at SSRN 3535206*, 2020.

Marika Cabral, Michael Geruso, and Neale Mahoney. Does privatized medicare benefit patients or producers? evidence from the benefits improvement and protection act. *American Economic Review*, 2018.

Colin F Camerer, Teck-Hua Ho, and Juin-Kuan Chong. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic theory*, 104(1):137–188, 2002.

Olivier Coibion, Yuriy Gorodnichenko, and Saten Kumar. How do firms form their expectations? new survey evidence. *American Economic Review*, 108(9):2671–2713, 2018.

Medicare Payment Advisory Commission et al. The medicare advantage program: status report. In *Report to the Congress: Medicare Payment Policy*, 2021.

Vincent P Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149, 2003.

Vilsa Curto, Liran Einav, Jonathan Levin, and Jay Bhattacharya. Can health insurance competition work? evidence from medicare advantage. *Journal of Political Economy*, 129 (2):570–606, 2021.

Arthur S De Vany and Thomas R Saving. Product quality, uncertainty, and regulation: The trucking industry. *The American Economic Review*, 67(4):583–594, 1977.

Ulrich Doraszelski and Sarit Markovich. Advertising dynamics and competitive advantage. *The RAND Journal of Economics*, 38(3):557–592, 2007.

David Dranove and Mark A Satterthwaite. The industrial organization of health care markets. *Handbook of health economics*, 1:1093–1139, 2000.

Mark Duggan, Amanda Starc, and Boris Vabson. Who benefits when the government pays more? pass-through in the medicare advantage program. *Journal of Public Economics*, 141:50 – 67, 2016. ISSN 0047-2727. doi: https://doi.org/10.1016/j.jpubeco.2016.07.003. URL `http://www.sciencedirect.com/science/article/pii/S0047272716300767`.

Ido Erev and Alvin E Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, pages 848–881, 1998.

Jose F Figueroa, Rishi K Wadhera, Austin B Frakt, Gregg C Fonarow, Paul A Heidenreich, Haolin Xu, Barbara Lytle, Adam D DeVore, Roland Matsouaka, Clyde W Yancy, et al. Quality of care and outcomes among medicare advantage vs fee-for-service medicare patients hospitalized with heart failure. *JAMA cardiology*, 5(12):1349–1357, 2020.

Drew Fudenberg and David M Kreps. Learning mixed equilibria. *Games and economic behavior*, 5(3):320–367, 1993.

Martin Gaynor, Kate Ho, and Robert J Town. The industrial organization of health-care markets. *Journal of Economic Literature*, 53(2):235–84, 2015.

John R Graham, Campbell R Harvey, and Shiva Rajgopal. The economic implications of corporate financial reporting. *Journal of accounting and economics*, 40(1-3):3–73, 2005.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

Bar Ifrach and Gabriel Y. Weintraub. A Framework for Dynamic Oligopoly in Concentrated Industries. *The Review of Economic Studies*, 84(3):1106–1150, 09 2016. ISSN 0034-6527. doi: 10.1093/restud/rdw047. URL `https://doi.org/10.1093/restud/rdw047`.

Anupam B Jena and Tomas J Philipson. Cost-effectiveness analysis and innovation. *Journal of health economics*, 27(5):1224–1236, 2008.

Harim Kim. Heterogeneous impacts of cost shocks, strategic bidding and pass-through: Evidence from the new england electricity market. *American Economic Journal: Microeconomics*, 2021.

Thomas G McGuire, Joseph P Newhouse, and Anna D Sinaiko. An economic history of medicare part c. *The Milbank Quarterly*, 89(2):289–332, 2011.

David J Meyers and Amal N Trivedi. Medicare advantage chart reviews are associated with billions in additional payments for some plans. *Medical Care*, 59(2):96–100, 2021.

David J Meyers, Amal N Trivedi, Ira B Wilson, Vincent Mor, and Momotazur Rahman. Higher medicare advantage star ratings are associated with improvements in patient outcomes: Study examines how medicare quality ratings impact patient outcomes for those enrolled in medicare advantage. *Health Affairs*, 40(2):243–250, 2021.

Paul Milgrom and John Roberts. Adaptive and sophisticated learning in normal form games. *Games and economic Behavior*, 3(1):82–100, 1991.

Keaton S Miller, Amil Petrin, Robert Town, and Michael Chernew. Optimal managed competition subsidies. Technical report, National Bureau of Economic Research, 2021.

Sungchul Park, Lindsay White, Paul Fishman, Eric B Larson, and Norma B Coe. Health care utilization, care satisfaction, and health status for medicare advantage and traditional medicare beneficiaries with and without alzheimer disease and related dementias. *JAMA network open*, 3(3):e201809–e201809, 2020.

E Elisabet Rutström and Nathaniel T Wilcox. Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior*, 67(2): 616–632, 2009.

Zirui Song, Mary Beth Landrum, and Michael E Chernew. Competitive bidding in medicare advantage: Effect of benchmark changes on plan bids. *Journal of health economics*, 32(6): 1301–1312, 2013.

Barbara Starfield, Leiyu Shi, and James Macinko. Contribution of primary care to health systems and health. *The milbank quarterly*, 83(3):457–502, 2005.

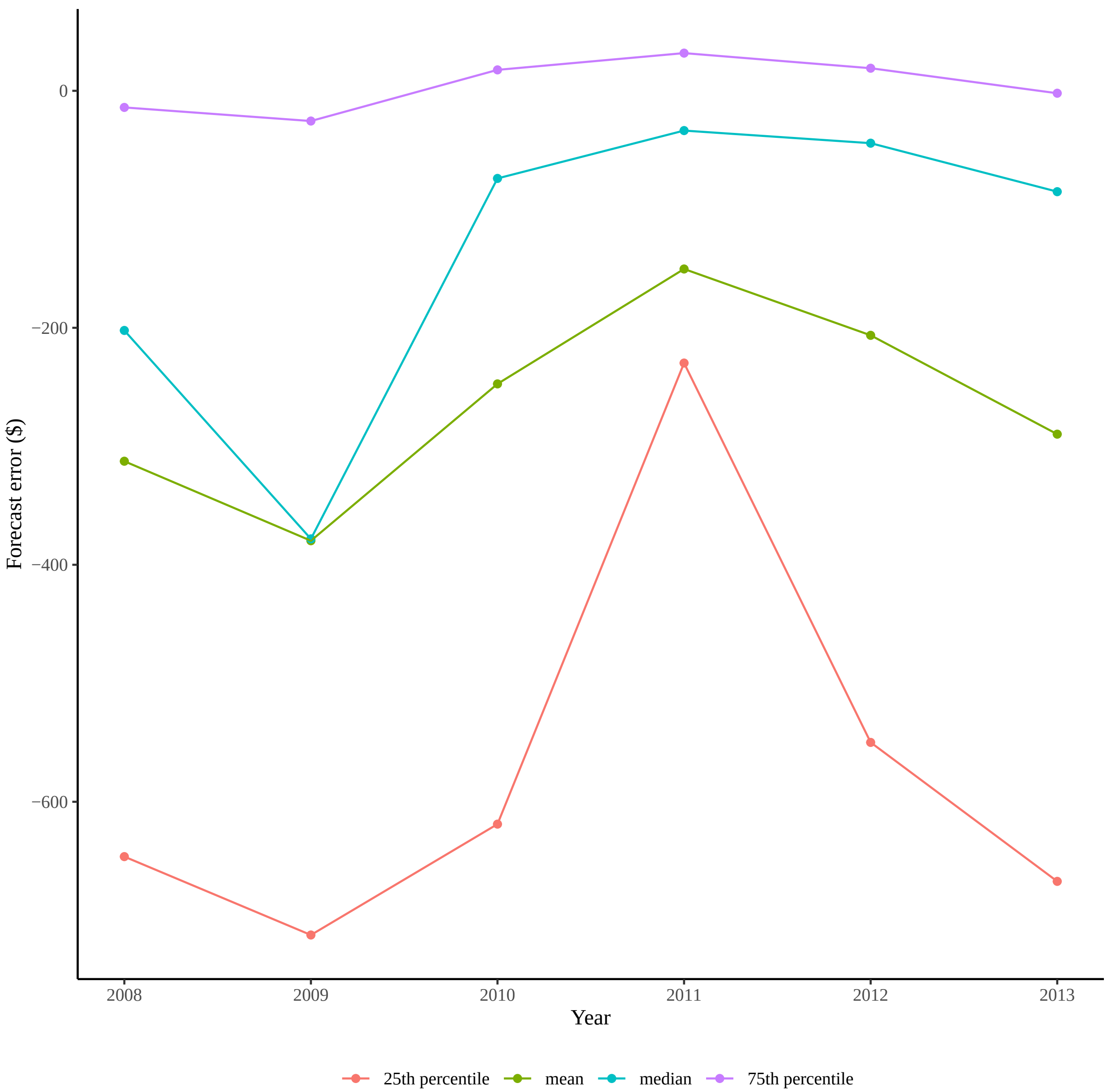
6 Tables and Figures

Figure 1: Example bid cost reporting

		Actual, 1/1/2013-12/31/2013			Projected, 1/1/2015-12/31/2015		
		Annualized		Allowed	Annual		Total Allowed
	Service Category	**Util/1000**	**Avg Cost**	**PMPM**	**Util/1000**	**Avg Cost**	**PMPM**
a.	Inpatient Facility	802	$2,358.54	$157.71	813	$2,486.41	$168.47
b.	Skilled Nursing Facility	672	427.90	23.97	800	473.02	31.54
c.	Home Health	580	192.78	9.32	674	211.30	11.86
d.	Ambulance	107	438.99	3.92	125	480.31	4.99
e.	DME/Prosthetics/Supplies	9,312	14.40	11.18	10,850	15.78	14.27
f.	OP Facility - Emergency	305	747.00	19.00	323	778.66	20.94
g.	OP Facility - Surgery	444	2,133.11	78.91	462	2,196.53	84.57
h.	OP Facility - Other	5,106	108.59	46.20	5,384	112.78	50.60
i.	Professional	14,225	122.25	144.92	15,009	120.85	151.14
j.	Part B Rx	1,765	216.40	31.83	1,861	201.07	31.18
k.	Other Medicare Part B	0	0.00	0.00	0	0.00	0.00
l.	Transportation (Non-Covered)	0	0.00	0.00	0	0.00	0.00
m.	Dental (Non-Covered)	0	0.00	0.00	195	88.10	1.43
n.	Vision (Non-Covered)	0	0.00	0.00	0	0.00	0.00
o.	Hearing (Non-Covered)	0	0.00	0.00	0	0.00	0.00
p.	Health & Education (Non-Covered)	19,435	4.92	7.96	6,791	5.53	3.13
q.	Other Non-Covered	5,347	1.17	0.52	7,256	1.64	0.99
r.	COB/Subrg. (outside claim system)						0.00
s.	**Total Medical Expenses**			**$535.47**			**$575.12**

Notes: This figure is reconstructed from the 2015 bid data for plan H0028-01-0, "Humana Gold Plus", offered in Cedar Rapids, Iowa, and surrounding counties, using the CMS public bid data and the 2015 Bid Pricing Tool (BPT). The 'Actual' columns are excerpted from the "MA Base Period Experience" portion of the BPT. The 'Projected' columns are excerpted from the "MA Projected Allowed Costs" portion.

Figure 2: The distribution of the overall prediction error over time



Notes: An observation is a plan-year. Statistics are unweighted. The overall prediction error is calculated per-member-per-month (PMPM). Negative prediction errors indicate firms overpredicted their costs.

Table 1: Summary statistics

	Min	25th	Mean	Median	75th	Max
Prediction errors:						
Overall	-1515.93	-638.00	-270.20	-89.29	5.89	2228.70
Inpatient Facility	-638.93	-243.72	-103.09	-43.10	1.80	778.10
Professional	-596.44	-167.62	-77.03	-37.73	1.03	2168.77
Outpatient Facility - Surgery	-244.07	-38.12	-15.61	-8.42	3.27	158.45
Firm-chosen product characteristics:						
Deductible	0	0	96.15	0	0	5000
Out-of-pocket limit	0	2000	3437.44	3400	5000	12000
Primary care copay	0	5	11.83	10	15	47.5
Specialist copay	0	20	26.09	30	35	60
Hospital copay	0	400	893.79	875	1400	2800
CMS-determined quality rating variables:						
Star rating	0.0	0.0	2.48	3.0	3.5	5.0
‘Too new’ indicator	0	0	0.08	0	0	1
‘No data’ indicator	0	0	0.18	0	0	1
Other covariates:						
Risk-adjusted payment	312.75	683.11	731.49	726.17	775.25	1134.19
Number of competitors	1.06	7.41	11.72	10.21	15.02	38
Number of competing plans	0	12.38	23.42	20.10	29.90	89
Log previous enrollment	2.48	8.34	9.47	9.69	10.73	13.30
Observations			11,670			

Notes: An observation is a plan-year. Statistics are unweighted. Prediction errors and risk-adjusted payments are per-member-per-month (PMPM). Negative prediction errors indicate firms overpredicted their costs. Hospital copays are for seven day stays. The number of competitors is calculated at the contract level, weighted by the number of contract enrollees in each county covered by the contract. The number of competing plans is calculated at the plan level, weighted by the number of plan enrollees in each county covered by the contract.

Table 2: Correlations between the overall forecast error and its major components

	Overall	Inpatient Facility	Professional	Outpatient Facility Surgery
Overall	1			
Inpatient Facility	0.97	1		
Professional	0.84	0.82	1	
Outpatient Facility - Surgery	0.94	0.86	0.73	1

Notes: This table presents correlations between the overall forecast error and its three largest (on average) components. An observation is a plan-year; there are 11,670 observations total. Statistics are unweighted.

Table 3: The distribution of the overall prediction error across quartiles of measures of experience and competition

	Experience			
	First quartile	Second quartile	Third quartile	Last quartile
Outcome: Overall prediction error				
Min	-1515.93	-1358.61	-1247.43	-1308.29
25th percentile	-667.15	-632.43	-570.92	-649.47
Mean	-308.36	-257.80	-227.50	-287.14
Median	-191.04	-79.68	-58.62	-86.17
75th percentile	0	12.96	9.73	0.55
Max	1276.81	924.99	792.77	2228.70
	Competition			
	First quartile	Second quartile	Third quartile	Last quartile
Outcome: Overall prediction error				
Min	-1494.49	-1309.37	-1358.61	-1515.93
25th percentile	-590.81	-639.82	-685.87	-614.90
Mean	-230.48	-287.67	-307.58	-255.08
Median	-56.11	-124.33	-131.33	-81.24
75th percentile	13.27	7.53	-2.87	7.77
Max	1276.81	2278.70	792.77	1023.31

Notes: An observation is a plan-year. Statistics are unweighted. The overall prediction error is calculated per-member-per-month (PMPM). Negative prediction errors indicate firms overpredicted their costs. Our measure of experience is the log of the total number of previous enrollees at the contract level. Our measure of competition is the number of competitors calculated at the contract level, weighted by the number of contract enrollees in each county covered by the contract.

Table 4: The association between prediction errors and plan observables

	Dependent variable: Prediction errors			
	Overall	Inpatient facility	OP Facility - Surgery	Professional
	(1)	(2)	(3)	(4)
Overall prediction error$_{t-1}$	0.002***	0.001***	0.0001***	0.001***
	(0.0001)	(0.00003)	(0.00001)	(0.00002)
Log previous enrollment	2.945	$-$1.907*	$-$0.572***	3.122***
	(2.461)	(1.018)	(0.205)	(0.756)
Number of competitors	6.413***	3.199***	0.267**	1.661***
	(1.403)	(0.580)	(0.117)	(0.431)
Number of competing plans	$-$0.902**	$-$0.430***	$-$0.059*	$-$0.229**
	(0.367)	(0.152)	(0.031)	(0.113)
Plan age	46.284***	17.676***	4.218***	11.924***
	(9.357)	(3.869)	(0.779)	(2.872)
Constant	$-$811.990***	$-$295.974***	$-$27.565	$-$292.505***
	(206.648)	(85.458)	(17.196)	(63.439)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	11,530	11,530	11,530	11,530
R^2	0.257	0.238	0.239	0.265
Adjusted R^2	0.232	0.213	0.214	0.241

Notes: An observation is a plan-year. Estimates are obtained using unweighted OLS. Stars indicate p-values: *p<0.1; **p<0.05; ***p<0.01.

Table 5: **The association between absolute prediction errors and plan observables**

	Dependent variable: Prediction errors			
	|Overall|	|Inpatient facility |	|OP Facility - Surgery |	|Professional |
	(1)	(2)	(3)	(4)
Overall prediction error$_{t-1}$	-0.001^{***}	-0.001^{***}	-0.0001^{***}	-0.0005^{***}
	(0.0001)	(0.00003)	(0.00001)	(0.00002)
Log previous enrollment	-6.270^{***}	-1.580^{*}	-0.083	-1.883^{***}
	(2.170)	(0.855)	(0.158)	(0.640)
Number of competitors	-5.345^{***}	-2.363^{***}	-0.281^{***}	-0.956^{***}
	(1.236)	(0.487)	(0.090)	(0.365)
Number of competing plans	0.634^{*}	0.336^{***}	-0.023	0.316^{***}
	(0.324)	(0.128)	(0.024)	(0.096)
Plan age	-43.944^{***}	-15.107^{***}	-4.044^{***}	-7.898^{***}
	(8.248)	(3.252)	(0.602)	(2.434)
Constant	888.563^{***}	314.236^{***}	68.322^{***}	246.858^{***}
	(182.162)	(71.815)	(13.303)	(53.755)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	11,530	11,530	11,530	11,530
R^2	0.261	0.259	0.218	0.282
Adjusted R^2	0.236	0.234	0.191	0.258

Notes: An observation is a plan-year. Estimates are obtained using unweighted OLS. Stars indicate p-values: *p<0.1; **p<0.05; ***p<0.01.

Table 6: The association between plan benefits and the overall prediction error

	Dependent variable:				
	Deductible	OOP limit	Primary care copay	Specialist copay	Hospital copay
	(1)	(2)	(3)	(4)	(5)
Overall prediction error$_{t-1}$	0.012^{*}	0.373^{***}	-0.001^{***}	0.001^{***}	0.008
	(0.007)	(0.052)	(0.0002)	(0.0002)	(0.013)
Star rating	-11.060	-1.900	0.319^{*}	-0.882^{***}	-37.499^{***}
	(6.746)	(52.015)	(0.182)	(0.246)	(12.696)
Star 'too new' indicator	-43.882^{*}	784.931^{***}	0.545	-3.279^{***}	-67.160
	(22.888)	(176.471)	(0.616)	(0.835)	(43.072)
Star 'no data' indicator	-56.314^{**}	388.196^{**}	1.477^{**}	-3.496^{***}	-150.724^{***}
	(21.984)	(169.503)	(0.592)	(0.802)	(41.372)
Risk-adjusted payment	0.369^{***}	0.588^{*}	0.009^{***}	-0.006^{***}	0.010
	(0.041)	(0.314)	(0.001)	(0.001)	(0.077)
Constant	-194.165	1,169.898	-1.320	29.584^{***}	983.983^{***}
	(141.780)	(1,093.145)	(3.816)	(5.174)	(266.812)
Year FE	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Mean Dependent Variable	96.14722	3437.44	11.83048	26.0924	893.788
Observations	11,670	11,670	11,670	11,670	11,670
R^2	0.566	0.514	0.476	0.475	0.447
Adjusted R^2	0.551	0.498	0.459	0.458	0.429

Notes: An observation is a plan-year. The overall prediction error is defined as actual costs for the year minus the projected costs at the time of plan submission. Estimates are obtained using unweighted OLS. Stars indicate p-values: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 7: The association between plan benefits and the absolute overall prediction error

	Dependent variable:				
	Deductible	OOP Limit	Primary care copay	Specialist copay	Hospital copay
	(1)	(2)	(3)	(4)	(5)
$\lvert$Prediction Error$\rvert_{t-1}$	0.002	-0.483^{***}	0.0004^{*}	-0.002^{***}	-0.053^{***}
	(0.008)	(0.059)	(0.0002)	(0.0003)	(0.014)
Star rating	-10.551	-5.256	0.306^{*}	-0.917^{***}	-39.096^{***}
	(6.748)	(51.987)	(0.182)	(0.246)	(12.690)
Star 'too new' indicator	-43.005^{*}	780.005^{***}	0.521	-3.335^{***}	-69.828
	(22.892)	(176.355)	(0.616)	(0.834)	(43.048)
Star 'no data' indicator	-55.382^{**}	390.855^{**}	1.445^{**}	-3.528^{***}	-152.776^{***}
	(21.985)	(169.373)	(0.592)	(0.801)	(41.344)
Risk-adjusted payment	0.355^{***}	0.636^{**}	0.010^{***}	-0.005^{***}	0.050
	(0.041)	(0.314)	(0.001)	(0.001)	(0.077)
Constant	-191.285	$1,215.398$	-1.455	29.616^{***}	981.348^{***}
	(141.791)	(1,092.351)	(3.818)	(5.167)	(266.643)
Year FE	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Mean Dependent Variable	96.14722	3437.44	11.83048	26.0924	893.788
Observations	11,670	11,670	11,670	11,670	11,670
R^2	0.566	0.515	0.475	0.477	0.448
Adjusted R^2	0.551	0.499	0.458	0.459	0.429

Notes: An observation is a plan-year. The overall prediction error is defined as actual costs for the year minus the projected costs at the time of plan submission. Estimates are obtained using unweighted OLS. Stars indicate p-values: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.