Designing Physician Incentives and the Cost-Quality Tradeoff: Evidence from Accountable Care Organizations

Kole Reddig*

October 27, 2022

Abstract

This paper estimates a structural model of multitasking agents to investigate the cost-quality tradeoff in health care and design contracts for a large physician incentive program. The setting involves Medicare's Accountable Care Organizations (ACOs), which are groups of health care providers that receive incentive pay for spending below a cost target on shared patients. Estimation of the structural model and counterfactual simulations reveal a modest tradeoff between reducing cost and increasing quality: a contract that incentivizes a one standard deviation increase in ACO cost savings (equivalent to a decrease in spending of $357.45 per beneficiary) also decreases ACO quality of care by 0.14 standard deviations.

1 Introduction

In the United States health care sector, public and private insurers often implement physician incentive programs and pay-for-performance initiatives to control the cost of care. Designing payment contracts for these programs requires facing a fundamental challenge: physicians may decrease the quality of care they provide in order to reduce cost. This issue is an example of agent multitasking, which plays a critical role in decision-making and contract design in all sectors of the economy. In this paper, I estimate a structural model of multitasking agents to identify the extent to which health care providers decrease quality of care in order to reduce cost. I use the structural model to conduct counterfactual analysis that highlights the role of multitasking in incentive design.

The setting of this study is the Medicare Shared Savings Program (MSSP), a large incentive program that involves 11 million Medicare beneficiaries and over $2 billion in provider incentives each year.[1] The MSSP gives incentive pay to Accountable Care Organizations (ACOs), which are joint ventures of physicians, group practices, and hospitals that form to coordinate care of their shared patients. An ACO earns incentive pay through the MSSP if its members collectively reduce expenditure on health services. Because providers might decrease the quality of care they provide in order to reduce expenditure, both tasks (monetary savings

*John Chambers School of Business and Economics, West Virginia University. Email: kole.reddig@mail.wvu.edu. An earlier version of this paper titled "Designing Contracts for Multitasking Groups: A Structural Model of Accountable Care Organizations" was a chapter of my doctoral thesis at Carnegie Mellon University. I would like to thank Martin Gaynor and Maryam Saeedi for support throughout this project. I would also like to thank Chris Cronin, Dennis Epple, Akshaya Jha, Rebecca Lessem, Albert Ma, Hakkı Özdenören, and Amanda Starc, as well as participants at several seminars and conferences for their feedback.

[1]Source: The Centers for Medicare and Medicaid Services. "Shared Savings Program Fast Facts." `https://www.cms.gov/files/document/2022-shared-savings-program-fast-facts.pdf`.

and quality of care) determine ACO payment. Moreover, because the earnings of a provider in an ACO depend heavily on the decisions of others, free-riding within ACOs may severely limit performance.

What role does multitasking play in the decisions of Medicare providers? How much is lost to free-riding in ACOs? Answers to these questions are central the design of the MSSP, and furthermore will inform incentive and contract design throughout the health care sector. I answer these questions by building and estimating a structural model of Medicare providers in ACOs. In the model, providers participating in an ACO act strategically and choose efforts to put towards cost saving and quality of care in order to maximize their own payoff. The choices of efforts of each ACO member form a Nash equilibrium that describes the ACO's overall performance.

In counterfactual analyses, I solve for the contract between ACOs and Medicare that maximizes the quality-weighted cost savings of providers in the MSSP.[2] In this principal agent problem, Medicare is restricted by federal regulation in such a way that it can only choose contracts that pay groups of providers (the ACO), and not individual providers (the ACO participants). Accordingly, for a given contract, ACO performance depends not only on participant characteristics, but also the economic consequences of moral hazard on teams and multitasking agency.

Federal regulation also restricts the specific ways that Medicare can define incentive pay for ACOs. First, Medicare can set the generosity of the contract for a given level of quality of care and expenditure, and second, Medicare can require ACOs to make penalty payments to Medicare if their expenditure exceeds its target. Knowing the magnitude of the tradeoff between cost reduction and quality of care is critical for choosing these contract parameters. Depending on the structural relationship between cost reduction and quality of care in health care, enforcing penalties with low-powered contracts has the potential to greatly diminish quality of care.

My research design exploits observed variation in the well-defined contracts between Medicare and ACOs in the MSSP to identify structural parameters. In the MSSP, an ACO is assigned a benchmark expenditure for the health care services provided to its participants' patients. If a year's actual Medicare expenditure on those beneficiaries is less than the benchmark amount, an ACO earns a portion of the difference, adjusted by a quality score, as incentive pay (hence "sharing savings" with Medicare).

The form of these contracts is chosen by Medicare and is public information, so I observe cross-ACO variation in the marginal dollar of group incentive pay for a given level of cost savings and quality of care. Under an equilibrium assumption, this identifies a function describing the marginal cost of reducing expenditure and improving quality at the ACO level. Intuitively, I estimate marginal cost functions for the representative participant in each ACO. This ultimately yields an empirical estimate of the magnitude of the tradeoff between cost savings and quality—a key factor driving multitasking choices and a crucial component to computing contracts that make a combination of monetary savings and quality of care the objective.

The conclusions of this work apply to both economics and policy audiences. I estimate a modestly-sized

[2]For comparison, I also solve for contracts that are optimal by other definitions, including per-beneficiary and non-quality-weighted objectives.

structural tradeoff between Medicare savings and quality of care. In the context of the MSSP, I provide evidence through counterfactual simulations that the savings-quality tradeoff is not large enough to create significant quality decreases when ACOs are incentivzed to reduce cost. I find imposing penalties on ACOs for having expenditure larger than benchmark expenditure, compared to contracts without penalties, increases average ACO savings by $305.25 per beneficiary (0.85 standard deviations). These "two-sided" contracts, however, effectively incentivize quality of care reductions (as measured by lower ACO quality scores) in favor of cost saving (higher ACO savings rates), so simulations also show ACO quality of care drops by 0.12 standard deviations.

Under "one-sided" contracts, where ACOs are not penalized for overspending, simulations show per-beneficiary savings is capped at $118.40. The optimal sharing rate needed to achieve this savings is rather high at 85%, while existing MSSP one-sided contracts offer a sharing rate of just 50%. ACO quality scores also increase slightly, by 0.12 standard deviations, when sharing rates increase from 50% to 85%.

Two-sided contracts, where ACOs are penalized for having an expenditure larger than benchmark expenditure, are able to incentivize ACOs to have a higher savings rate for a lower sharing rate. Because of the penalties, simulations show that two-sided contracts maximize savings at a much weaker 46% sharing rate. ACO quality scores are 0.16 standard deviations lower at the optimal two-sided contract compared to the baseline contract, however. Quality-weighted ACO savings is maximized at a more generous 75% sharing rate, where ACO savings is $212 per beneficiary larger and quality scores are nearly unchanged compared to the baseline of a 50% sharing rate and one-sided contracts.

Because the earnings of a provider in an ACO depend on group performance, providers act strategically when choosing savings effort and quality effort. As more providers join an ACO, any one provider's influence on ACO outcomes diminishes. The result of this is incentive dilution and free-riding, and the optimal effort choices of providers are less than the effort choices that would maximize the total surplus to all providers. I find that average per-beneficiary ACO savings would increase by $456 without free-riding within ACOs. In other words, the efficiency gains from provider coordination within ACOs must be quite large (about $500 per beneficiary) for the MSSP group incentive scheme to be self-financing. Because of free-riding, Medicare must pay more to ACOs: if free-riding was eliminated in all ACOs, the program would maximize its monetary savings by sharing about 30% savings with ACOs.

Ultimately, the results show that there is little Medicare can do, within the scope of the contract parameters analyzed, to meaningfully reduce costs. This paper empirically verifies the premonitions of Frandsen & Rebitzer (2015) that organizational incentives within ACOs impose inefficiency larger than the efficiency introduced by coordination. This paper goes a step further: even if free-riding is eliminated in all ACOs **and** the optimal contract is implemented, average ACO savings is only $602.91 per beneficiary. Medicare Parts A and B spending per enrolled beneficiary was $10,210 in 2020, and increased by an average of $165.62 every year since 2013.[3] This means applying the *ceiling* level of ACO performance to the entirety of FFS Medicare

[3]Source: Kaiser Family Foundation. "Medicare Spending per Beneficiary." https://tinyurl.com/a56w8pbp.

would yield only a 6% cost savings. If the rate of increase in spending is not also decreased, spending would return to pre-ideal ACO levels in roughly three and a half years. And, if the population of health care providers in ACOs is representative of providers the in the general health care market, this paper shows both public and private insurers should expect small efficiency gains from shared savings programs.

Related Literature. This paper contributes to economics literature concerning evidence of multitasking and agent response to incentive pay (Slade, 1996; Bai & Xu, 2005; Dumont, Fortin, Jacquemet, & Shearer, 2008; Mullen, Frank, & Rosenthal, 2010; Feng Lu, 2012; Hong, Hossain, List, & Tanaka, 2018). Along with the study by Kim, Sudhir, & Uetake (2022), this paper is among the first to estimate a structural model of multitasking agents. I also contribute to economics literature concerning health care provider payment systems and provider behavior in organizations (Gaynor, Rebitzer, & Taylor, 2004; Encinosa, Gaynor, & Rebitzer, 2007; Choné & Ma, 2011; Rebitzer & Votruba, 2011; Ho & Pakes, 2014; Frandsen & Rebitzer, 2015; Grassi & Ma, 2016; Frandsen, Powell, & Rebitzer, 2019). More generally, this paper aligns with the literature that studies the supply-side of health care, and examines the incentives faced and decisions made by physicians, hospitals, and insurers (Gaynor, 2006; Chandra, Cutler, & Song, 2011; Gaynor, Ho, & Town, 2015; Foo, Lee, & Fong, 2017; Einav, Finkelstein, & Mahoney, 2018; Eliason, Grieco, McDevitt, & Roberts, 2018; Hackmann, 2019; Gaynor, Mehta, & Richards-Shubik, 2020).

Few studies in economics have discussed ACOs directly. Frandsen & Rebitzer (2015) calibrate a model of ACO performance to examine the size-variance tradeoff in group payment mechanisms like the MSSP, and they argue that ACOs will be unable to self-finance. That is, there is no contract with strong enough incentives to overcome the incentive to free-ride among a group of physicians. The authors conclude with a skeptical look at the MSSP, and mention the untenability of integrated organizations in the now very fractured US health care market. Frech et al. (2015) study county-level entry of private and public ACOs. The authors find small markets generally discourage ACO entry, and that public ACO entry is largely predicted by higher Medicare spending, higher population, and lower physician site concentration. Frandsen et al. (2019) discuss the MSSP's impact on health care in the United States in the context of common agency, where several payers motivate the same agent to improve care delivery and integration. The authors find that unique equilibrium contracts from payers are lower powered in the presence of shared savings payments, and ACO entry can possibly inspire other shared savings contracts in the private sector if they do not already exist.

Two studies evaluate actual or simulated ACO performance with results comparable to those of this paper. Aswani, Shen, & Siddiq (2019) study how to design MSSP ACO contracts. The authors focus on asymmetric information between Medicare and ACOs, and write contracts such that ACO payment is a function of underlying ACO type (benchmark expenditure per beneficiary). Unlike this paper, Aswani et al. (2019) do not consider multitasking or free-riding within ACOs, and their simulations are for expenditure at the ACO level. Aswani et al. (2019) do however include endogenous participation of ACOs, which is not considered in this study. Their results, in simulations that are comparable, are very similar. For example, optimal sharing rates for two-sided contracts in both papers are found to be about 50%.

Wilson et al. (2020) is a rapid review of studies addressing ACO performance. Projections of MSSP ACO cost savings from the reviewed studies are on the same magnitude of this paper, ranging between -$107 to $857 per beneficiary.

This paper continues as follows: Section 2 gives a brief overview of the MSSP and ACOs, including descriptive ACO statistics. I outline my model of performance in ACOs in Section 3. I describe identification and estimation of model primitives in Section 4, and estimation results and model fit are in Section 5. I present counterfactual analyses, including computation of savings-maximizing contracts between ACOs and Medicare and performance losses to free-riding, in Section 6, and Section 7 concludes.

2 Background and Data

The MSSP, a part of the Patient Protection and Affordable Care Act of 2010 (ACA), is a policy response to increasing health care costs in the United States. The premise of the program is that the United States is inefficient at providing health care because care delivery is *fragmented*. That is, unique to the United States, patients tend to see several distinct providers that belong to separate businesses with little incentive to coordinate care. Patients therefore receive haphazard and/or redundant care, implying increased utilization, cost, and risk of adverse health outcomes. A recent empirical analysis by Agha, Frandsen, & Rebitzer (2019) examined the fragmentation of health care in the United States. Using Medicare claims data, the authors show utilization increases in areas with more fragmented care, and patients substitute low cost services for high cost services in areas with high fragmentation.

The MSSP gives providers financial motivation to integrate care delivery. To overcome institutional boundaries to care integration, the program explicitly evaluates and pays Medicare providers based on group performance. First, providers join Accountable Care Organizations, or ACOs, which are joint ventures of Medicare providers created to earn payment through the MSSP. Nearly any Medicare provider, including individual physicians, group practices, and large hospital systems, can start or participate in an ACO. Medicare fee-for-service (FFS) beneficiaries are then assigned to ACOs by Medicare according to their primary care provider (PCP).[4]

The first ten rows of Table 1 display statistics describing ACO participants and beneficiaries assigned to ACOs.[5] There is substantial heterogeneity in the number of providers that join an ACO—some large hospitals are able to form an ACO independently by employing enough PCPs to be assigned the legally required minimum of 5000 beneficiaries (these ACOs have one participant), and others are joint ventures of hundreds of providers. Every state has beneficiaries assigned to an ACO, and most ACOs concentrate

[4]Appendix A gives a more detailed description of MSSP ACO formation, beneficiary assignment, and payment.

[5]The data for this table, and all analysis in this section, is from MSSP ACO Public Use Files, MSSP Participant Lists, MSSP ACO Performance Year Results, and Number of ACO Assigned Beneficiaries by County Public Use Files. In short, the data consists of ACO expenditures, benchmark expenditures, quality scores (along with every quality sub-measure), various assigned beneficiary demographics, and various participant and provider statistics. Little public information is available on the characteristics of specific ACO participants or providers.

Table 1: **Summary ACO Statistics: Providers and Beneficiaries**

Variable	Mean	Std. Dev.	Min.	Med.	Max.
Number of participants[a]	37.79	58.03	1.00	20.00	840.00
Total number of individual providers (1000s)	0.59	0.84	0.00	0.28	7.28
Proportion of providers PCP	0.41	0.18	0.03	0.36	1.00
Proportion of providers specialist	0.41	0.21	0.00	0.44	0.88
Number of states where beneficiaries reside	1.52	0.97	1.00	1.00	10.00
Number of assigned beneficiaries (1000s)	17.78	17.49	0.15	11.87	149.63
Average risk score	1.06	0.11	0.81	1.04	2.09
Percent of beneficiaries over age 75	39.10	6.04	13.20	39.18	66.25
Percent of beneficiaries male	42.66	2.06	34.57	42.69	57.50
Percent of beneficiaries nonwhite	16.88	15.34	1.49	12.45	94.98
Sharing rate 50%, one-sided	0.96	0.19	0.00	1.00	1.00
Sharing rate 60%, two-sided	0.01	0.11	0.00	0.00	1.00
Sharing rate 75%, two-sided	0.03	0.16	0.00	0.00	1.00
Benchmark Expenditure ($ billions)	0.19	0.18	0.00	0.13	1.97
Expenditure ($ billions)	0.19	0.18	0.00	0.12	1.97
Benchmark Expenditure - Expenditure ($ millions)	1.46	10.18	-72.49	0.67	89.13
Savings Rate	0.01	0.05	-0.44	0.01	0.30
Quality Score	0.87	0.12	0.07	0.90	1.00
$\mathbf{1}\{$Savings Rate $\geq$ Min. Savings Rate$\}$	0.31	0.46	0.00	0.00	1.00
Earned shared savings or losses	1.50	3.64	-4.66	0.00	41.91
Earned shared savings, given qualified	4.95	5.11	0.00	3.48	41.91
Proportion of expenditure on inpatient services[a]	0.31	0.03	0.22	0.31	0.43
Proportion of expenditure on outpatient services[a]	0.20	0.06	0.08	0.19	0.49
Number of primary care services (1000s)	10.29	1.76	5.39	9.98	26.16
Number of inpatient admissions (1000s)	0.33	0.09	0.17	0.32	1.86

Note: $N = 1849$. This table shows summary statistics for ACOs for years 2013-2017. The superscript ^{a} indicates statistics are for 2014-2017 (due to data availability). “Quality Score” is computed by the author from ACO quality sub-measures (public data codes Quality Score as 1 or “P4R” in an ACO’s first performance year.)

on beneficiaries within only one state. "Average risk score" is the average Hierarchical Condition Category (HCC) risk score of non-dual eligible beneficiaries assigned to an ACO. A beneficiary's risk score increases as predicted health care costs of that beneficiary increase.

Payment of an ACO depends on a calendar year's Medicare expenditure on beneficiaries assigned to the ACO, a quality of care score, and the contract the ACO has with Medicare. Upon formation of an ACO, Medicare assigns a "benchmark expenditure" by forecasting Medicare expenditure for beneficiaries assigned to the ACO. After operating for a year, the ACO's payment is determined by the difference between the benchmark expenditure and realized expenditure on assigned beneficiaries and a composite quality score between 0 and 1.[6] If the ACO's savings rate, defined as $\frac{\text{Benchmark Expenditure}-\text{Expenditure}}{\text{Benchmark Expenditure}}$, exceeds a predetermined minimum called the Minimum Savings Rate (MSR), and if the ACO meets minimum quality of care standards, it earns and distributes to its members the amount

$$\text{Shared Savings} = \text{Sharing Rate} \cdot \text{Quality Score} \cdot (\text{Benchmark Expenditure} - \text{Expenditure}) \tag{1}$$

where "Sharing Rate," a number between 0 and 1, is determined by the type of contract the ACO has with Medicare.

The overwhelming contract choice of ACOs from 2013 to 2017, "Track 1," has a sharing rate of 50%. Under this contract, if a hypothetical ACO with a benchmark expenditure of \$186 million and an MSR of 0.02 had an expenditure of \$180 million with a quality score of 0.90, it would earn

$$0.5 \cdot 0.9 \cdot (\$186 \text{ million} - \$180 \text{ million}) = \$2.7 \text{ million} \tag{2}$$

in shared savings. Its savings rate is $(186 - 180)/186 = 0.03$, so the MSR is exceeded. Though paying a subsidy, Medicare saves money as well: on net, this ACO contributed a \$3.3 million decrease in Medicare expenditure, as it was paid \$2.7 million for saving \$6 million.

While uncommon in the first few years of the MSSP, some contracts also penalize ACOs for having expenditure *larger* than benchmark expenditure. These are called "two-sided" contracts. "Track 2" and "Track 3" ACOs have two-sided contracts, and pay Medicare "Shared Losses" if their savings rate is less than the Maximum Loss Rate (MLR):

$$\text{Shared Losses} = (1 - \text{Sharing Rate} \cdot \text{Quality Score}) \cdot (\text{Benchmark Expenditure} - \text{Expenditure}) \tag{3}$$

Note that the value of this equation is negative when expenditure exceeds benchmark expenditure, and that the value can be diminished by having a higher sharing rate or quality score. During the early years of the

[6] An ACO's overall quality score is determined by the combination of 30-40 sub-measures of care quality. These sub-measures fall into the domains of "Patient/Caregiver Experience," "Care Coordination/Patient Safety," "Preventative Health," and "At-Risk Population." Some sub-measures are survey responses (e.g., "ACO2: How Well Your Doctors Communicate"), while others are computed from Medicare Claims and aggregated to the ACO-level (e.g., "ACO21: Proportion of Adults who had blood pressure screened in past 2 years"). See `https://go.cms.gov/2xHy7Uo` for a full list of ACO quality scores for every performance year.

MSSP, two-sided contracts came with higher sharing rates: 60% for Track 2 and 75% for Track 3 ACOs.

For example, if a Track 3 ACO with a benchmark expenditure of \$30 million and an MLR of -0.03 had an expenditure of \$33 million and quality score of 0.80, then they would have shared losses

$$(1 - 0.75 \cdot 0.8) \cdot (\$30 \text{ million} - \$33 \text{ million}) = -\$1.2 \text{ million}\,. \tag{4}$$

This hypothetical ACO would offset the \$3 million in expenditure above the benchmark with a \$1.2 million payment to Medicare.

The last fifteen rows of Table 1 contain statistics on ACO performance. ACO benchmark expenditures and realized expenditures are large: the mean is approximately \$190 million, with several ACOs having expenditure over \$1 billion. From 2013 to 2017, ACOs saved money on average. However, less than one third of ACOs had a savings rate at least as large as their minimum savings rate, meaning most ACOs do not actually earn incentive pay. Average earned incentive pay is \$1.5 million per ACO, and given an ACO earns incentive pay, incentive pay is nearly \$5 million. Per participant, average earned incentive pay is \$189,108 unconditionally and \$609,654 among ACOs that qualify.

3 Model of Multitasking ACO Participants

This section lays out a model of health care provider decision-making within ACOs. I model health care provider actions in ACOs as a simultaneous move game. ACO participation and incentive contracts are taken as given, and each ACO participant chooses unobservable efforts to put towards cost reduction and quality of care.[7] An ACO's overall savings rate and quality score is the outcome of the Nash equilibrium strategies chosen by its participants. Though this model is written in a way such that decisions are made by individual participants, underlying structural parameters are identified and estimated with aggregate, ACO level data. The environment is static with an exogenous set of ACOs $\mathcal{J}$ where each ACO $j \in \mathcal{J}$ has a heterogeneous set of participants I_j, each indexed by i.

3.1 Contract Definition and Multitasking Game

In the MSSP, Medicare's incentive contracts pay ACOs a fixed proportion of the cost reduction of an ACO, weighted by a quality of care score. A one-sided contract pays an ACO as long as their savings rate $S \in [-1, 1]$ is larger than the minimum savings rate $\underline{S}_j$ and as long as the ACO's quality score $Q \in [0, 1]$ exceeds the

[7] Participation of Medicare providers in ACOs is deliberately not made endogenous in the model. This simplification is necessary, as a formal model of endogenous ACO formation that addresses optimal group size, provider networks, and multiplicity of equilibria is prohibitively complicated and not within the scope of this paper given the available data. I discuss the consequences of this simplification and the others made throughout the paper after presenting simulation results at the end of Section 6.

quality threshold $\underline{Q}$. Formally, the one-sided contract for ACO j is the function

$$P_j^{OS}(S,Q) = F_j \cdot B_j \cdot S \cdot Q \cdot \mathbb{1}\{S \geq \underline{S}_j\} \cdot \mathbb{1}\{Q \geq \underline{Q}\}, \tag{5}$$

where the sharing rate is $F_j \in [0,1]$, the benchmark expenditure is B_j, and $\mathbb{1}\{\cdot\}$ is the indicator function. Two-sided contracts offer the same incentive pay as one-sided contracts, but also penalize ACOs for savings rates less than the maximum loss rate:

$$P_j^{TS}(S,Q) = P_j^{OS}(S,Q) + (1 - F_j \cdot Q) \cdot B_j \cdot S \cdot \mathbb{1}\{S \leq -\underline{S}_j\}. \tag{6}$$

For ease of notation, let the payment contract of ACO j be the parameterized function

$$P_j\left(S,Q;F_j,T_j\right) = F_j \cdot B_j \cdot S \cdot Q \cdot \mathbb{1}\{S \geq \underline{S}_j\} \cdot \mathbb{1}\{Q \geq \underline{Q}\} + T_j \cdot (1 - F_j \cdot Q) \cdot B_j \cdot S \cdot \mathbb{1}\{S \leq -\underline{S}_j\}, \tag{7}$$

where $T_j \in \{0,1\}$ indicates a two-sided contract. The parameters of this function, F_j and T_j, are the only variables that Medicare can control to influence ACO participant choices.

ACO performance is a stochastic function of the multitasking choices made by ACO participants. All ACO participants $i \in I_j$ simultaneously choose savings effort and quality effort $s_{ij} \in [-1,1]$ and $q_{ij} \in [0,1]$ in order to maximize their share of incentive pay.[8] The ACO's realized savings rate, $\mathring{S}_j$ and realized overall quality score $\mathring{Q}_j$ are the weighted averages of these efforts and idiosyncratic shocks to savings and quality:

$$\mathring{S}_j = \sum_{i \in I_j} w_{ij}\left(s_{ij} + \eta_{ij}^S\right) \qquad \mathring{Q}_j = \sum_{i \in I_j} w_{ij}\left(q_{ij} + \eta_{ij}^Q\right). \tag{8}$$

The terms η_{ij}^S and η_{ij}^Q are independent and normally distributed, each with mean 0 and standard deviations σ_S and σ_Q, respectively. The variables $\{w_{ij}\}_{i \in I_j}$ are exogenous influence weights such that $w_{ij} \geq 0$ for all $i \in I_j$ and $\sum_{i \in I_j} w_{ij} = 1$. These weights account for heterogeneous influence of participants efforts on ACO performance.[9]

Each participant $i \in I_j$ simultaneously chooses efforts s_{ij} and q_{ij}, and each recieves the stochastic payoff

$$\mathring{\pi}_{ij} = w_{ij} P_j\left(\mathring{S}_j, \mathring{Q}_j; F_j, T_j\right) - c_{ij}\left(s_{ij}, q_{ij}\right) \tag{9}$$

where $w_{ij} P_j\left(\mathring{S}_j, \mathring{Q}_j; F_j, T_j\right)$ is provider i's portion of shared savings earned by ACO j with a realized savings rate of $\mathring{S}_j$ and quality score $\mathring{Q}_j$. The function $c_{ij} : [-1,1] \times [0,1] \to \mathbb{R}$ is the strictly convex and

[8] Savings effort s_{ij} is restricted to the domain $[-1,1]$—this implicitly restricts an ACO's total expenditure to be between zero and twice its benchmark expenditure. The upper bound on expenditure is arbitrary, and exists only so that strategy spaces of agents are compact. Quality effort q_{ij} is restricted to $[0,1]$ so that overall quality score also falls between $[0,1]$ (which is always the case in the MSSP).

[9] For example, consider an ACO with two participants: a hospital with savings effort analogous to saving 2%, and an individual provider with savings effort analogous to saving 4%. These effort choices would have a simple mean of 3%, but because the hospital has a relatively large share of expenditure, the ACO's overall savings rate would be smaller than 3%.

twice-continuously differentiable participant effort cost function. Specifically, $c_{ij}(s_{ij}, q_{ij})$ is the explicit and implicit costs incurred by $i \in I_j$ when choosing s_{ij} and q_{ij}. For example, a physician that chooses very large values of s_{ij} and q_{ij} would incur significant cost—both in operational expenses as well as opportunity cost from forgone services to reduce expenditure on assigned beneficiaries. There is no direct utility gained for quality of care through provider altruism, which is common in canonical and modern models of physician decision-making (Glied & Hong, 2018; Hackmann, 2019; Gaynor et al., 2020). In this model, altruistic preference for increasing quality of care is absorbed by the cost function, effectively decreasing the marginal cost of quality effort. Ultimately, c_{ij} places a natural restriction on how well participants, and hence ACOs, can perform.

In specification above, I assume that ACOs split their earned incentive pay with their participants according to influence weights w_{ij}, and not evenly between participants. Actual contracts between ACOs and ACO participants (known as "ACO Participant Agreements") are generally not publicly available. However, splitting shared savings according to influence on ACO outcomes is likely a good approximation of how ACOs actually split earnings.[10] For example, Gaynor et al. (2004) make the similar assumption that HMO group incentive pay is allocated among the group according to physician patient shares.

3.2 Equilibrium in the Multitasking Game

I assume that equilibrium savings and quality effort choices of ACO participants maximize the expected value of their respective payoffs, given the choices of other participants and conditional on the known distributions of realized ACO savings and quality scores. Like Frandsen & Rebitzer (2015), I assume ACO participants are risk-neutral, due to the relatively small amount of additional income earned by participating in ACOs. Therefore, ACO participants simultaneously solve

$$\max_{s_{ij} \in [-1,1], q_{ij} \in [0,1]} \mathbb{E}\left[\mathring{\pi}_{ij} \,\middle|\, \{s_{i'j}, q_{i'j}\}_{i' \in I_j} \right]. \tag{10}$$

Under the assumption that η_{ij}^S and η_{ij}^Q are normally distributed, $\mathring{S}_j$ and $\mathring{Q}_j$ are also normally distributed with means equal to the weighted average effort choices of participants and standard deviations equal to scaled standard deviations of the idiosyncratic effort shocks. That is,

$$\mathring{S}_j \sim N\left(S_j, W_j \sigma_S^2\right) \qquad \mathring{Q}_j \sim N\left(Q_j, W_j \sigma_Q^2\right) \tag{11}$$

where

$$S_j = \sum_{i \in I_j} w_{ij} s_{ij} \qquad Q_j = \sum_{i \in I_j} w_{ij} q_{ij} \qquad W_j = \sum_{i \in I_j} w_{ij}^2. \tag{12}$$

[10] See https://go.cms.gov/2HiHgus for more detail.

This means that expected ACO pay conditional on the aggregated choices S_j and Q_j is

$$\mathbb{E}\left[P_j\left(\mathring{S}_j, \mathring{Q}_j; F_j, T_j\right)\middle| S_j, Q_j\right] = F_j \cdot B_j \cdot E_j^S(S_j) \cdot E_j^Q(Q_j) - T_j \cdot (1 - F_j \cdot Q_j) \cdot B_j \cdot E_j^S(-S_j) \tag{13}$$

where

$$E_j^S(S_j) = S_j \Phi\left(\frac{S_j - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) + \sqrt{W_j}\sigma_S \phi\left(\frac{S_j - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) \tag{14}$$

$$E_j^Q(Q_j) = Q_j \Phi\left(\frac{Q_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) + \sqrt{W_j}\sigma_Q \phi\left(\frac{Q_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right). \tag{15}$$

Define the functions

$$E_j^P\left(S_j, Q_j; F_j, T_j\right) = \mathbb{E}\left[P_j\left(\mathring{S}_j, \mathring{Q}_j; F_j, T_j\right)\middle| S_j, Q_j\right] \tag{16}$$

$$\pi_{ij}\left(s_{ij}, q_{ij}; \mathbf{s}_{-ij}, \mathbf{q}_{-ij}\right) = w_{ij} E_j^P(S_j, Q_j; F_j, T_j \cdot) - c_{ij}(s_{ij}, q_{ij}). \tag{17}$$

Proposition 3.1 formally establishes existence and uniqueness in this game.

Proposition 3.1. *Let the 2×2 Hessian matrix $\boldsymbol{H}_{\pi_{ij}}$ be negative definite. Then, there exists a unique equilibrium $\left\{s_{ij}^*, q_{ij}^*\right\}_{i \in I_j}$ such that $\left(s_{ij}^*, q_{ij}^*\right)$ maximizes $\pi_{ij}(\cdot; \boldsymbol{s}_{-ij}^*, \boldsymbol{q}_{-ij}^*)$ for all $i \in I_j$.*

Proof. See Appendix B. □

The proof of this proposition merely shows that if expected payoff functions are strictly concave in participant's own choices s_{ij} and q_{ij}, then there is a unique equilibrium to the multitasking game. In Appendix B, I also show the more crucial condition for model identification and estimation that the assumption of Proposition 3.1 is met for virtually all ACOs and all simulated counterfactual contracts, given model estimates and data.

Whether this game is concave is conditional on the relationship between several model primitives—specifically, the magnitude of uncertainty (measured by W_j, σ_S, and σ_Q), and the values of $\frac{\partial^2 c_{ij}}{\partial s_{ij}^2}(s_{ij}, q_{ij})$, $\frac{\partial^2 c_{ij}}{\partial q_{ij}^2}(s_{ij}, q_{ij})$, and $\frac{\partial^2 c_{ij}}{\partial s_{ij} q_{ij}}(s_{ij}, q_{ij})$. As uncertainty increases, the expected payoff function of participants smooths. If uncertainty is large enough, and therefore expected payoff is sufficiently smooth, then there is only one intersection of marginal incentive pay and marginal cost of effort. However, when there is little uncertainty, expected incentive pay changes rapidly near the MSR $\underline{S}_j$ and quality standard $\underline{Q}$, meaning participants may have intersections of marginal cost and marginal incentive pay at two (or more!) locations. Sufficiently large second order derivatives guarantee global concavity of the objective functions.

4 Identification and Estimation

In this section, I walk through the identification and estimation of ACO participant effort cost functions c_{ij} and uncertainty parameters σ_S and σ_Q. To estimate model primitives, I first filter the data collected and discussed in Section 2. First, I use ACO-level data on only Track 1 ACOs from 2014 to 2017. The first performance year of the program, 2013, is omitted from analysis because some crucial variables (particularly the number of participants in each ACO) are not available that year. Certain lagged variables are used in estimation, meaning 2013 data and ACO observations in their first performance year cannot be used. Track 2 and 3 ACOs are omitted from the estimation sample because these ACOs choose to face downside risk when in the MSSP, so selection may bias estimates if these ACOs were included. Finally, cost function estimation is sensitive to several outlier observations of ACO savings rate, so I drop ACO observations in the bottom and top 2 percent of savings rate. This leaves a final estimation sample of $|\mathcal{J}| = 1486$ observations.[11]

Identification and estimation of this model is complicated by the limited data available on ACO participants. Only aggregate data are observed: decisions in the multitasking game are made by Medicare providers, and available data describes the outcome of these agents' decisions aggregated to the ACO level. To overcome the challenge imposed by data availability, I use functional form assumptions and an aggregation procedure to map observed ACO characteristics and performance to model primitives. The procedure can also be interpreted as estimating a model of representative ACO participant decision-making, where the choices of the representative participant of ACO j are the same as the aggregated multitasking equilibrium choices of participants in ACO j.[12]

4.1 Identification Overview

The key estimates in this paper are identified from co-variation in savings rates and overall quality scores across ACOs, *conditional on the observed marginal incentive pay for each measure of performance.* For example, the existence of a large positive correlation of savings rates and quality scores across ACOs would not be conclusive evidence of complementarity of savings and quality. In this example, if ACOs with large quality scores also tend to have large marginal incentive pay with respect to savings, then this positive correlation could still exist in the presence of a structural tradeoff. This is likely true for MSSP ACOs, since payment is multiplicative is savings effort and quality effort. Therefore, incorporating the structure of ACO participants' incentives into the empirical model is a requirement for identifying the underlying multitasking tradeoff.

To begin, I make an assumption about observed ACO savings rates and quality scores.

Assumption 4.1. Let $\{s^*_{ij}, q^*_{ij}\}_{i \in I_j, j \in \mathcal{J}}$ be equilibrium effort choices as defined in Proposition 3.1. Then, observed ACO savings rates and quality scores are equal to S^*_j and Q^*_j, where $S^*_j = \sum_{i \in I_j} w_{ij} s^*_{ij}$ and

[11] Filtering data to only Track 1 ACOs means that $F_j \equiv 0.5$ and $T_j \equiv 0$ in the estimation sample. I show identification and estimation for any value of these parameters in this section, however.

[12] The drawbacks to using this procedure is discussed in the context of the simulation results at the end of Section 6.

$Q_j^* = \sum_{i \in I_j} w_{ij} q_{ij}^*$.

This assumption has two implications: first, it implies there are no observations from ACOs out of equilibrium. Second, it implies observations are the actual aggregated choices, and not the realized aggregated choice (denoted $\mathring{S}_j$ and $\mathring{Q}_j$ in Section 3). The former implication is ultimately untestable, however, filtering observations as described at the beginning of this section likely aids in removing ACOs off equilibrium. The latter implication is of course not true for any specific observation, but it is on average true since $\mathbb{E}\left[\mathring{S}_j \,|S_j\right] = S_j$ and $\mathbb{E}\left[\mathring{Q}_j \,|Q_j\right] = Q_j$.

To characterize the equilibrium values of S_j^* and Q_j^* in such a way that facilitates estimation, first recall the ACO participant maximization problems in Equation 10. As a function of only deterministic variables, the problem is

$$\max_{s_{ij} \in [-1,1], q_{ij} \in [0,1]} w_{ij} E_j^P \left(S_j, Q_j; F_j, T_j\right) - c_{ij}\left(s_{ij}, q_{ij}\right). \tag{18}$$

where

$$E_j^P \left(S_j, Q_j; F_j, T_j\right) = F_j B_j E_j^S(S_j) E_j^Q(Q_j) - T_j \left(1 - F_j Q_j\right) B_j E_j^S(-S_j) \tag{19}$$

and

$$E_j^S(S_j) = S_j \Phi\left(\frac{S_j - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) + \sqrt{W_j}\sigma_S \phi\left(\frac{S_j - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) \tag{20}$$

$$E_j^Q(Q_j) = Q_j \Phi\left(\frac{Q_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) + \sqrt{W_j}\sigma_Q \phi\left(\frac{Q_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right). \tag{21}$$

In the proof of Proposition 3.1, I show the first order conditions to this problem are both necessary and sufficient to describe equilibrium choices $\left\{\left(s_{ij}^*, q_{ij}^*\right)\right\}_{i \in I_j}$. These conditions are

$$\frac{\partial c_{ij}}{\partial s_{ij}}\left(s_{ij}^*, q_{ij}^*\right) = w_{ij}^2 B_j \left[F_j E_j^Q(Q_j^*) \frac{\partial E_j^S}{\partial s_{ij}}\left(S_j^*\right) + T_j(1 - F_j Q_j^*) \frac{\partial E_j^S}{\partial s_{ij}}\left(-S_j^*\right)\right] \tag{22}$$

$$\frac{\partial c_{ij}}{\partial q_{ij}}\left(s_{ij}^*, q_{ij}^*\right) = w_{ij}^2 B_j \left[F_j E_j^S(S_j^*) \frac{\partial E_j^Q}{\partial q_{ij}}\left(Q_j^*\right) + T_j F_j E_j^S(-S_j^*)\right] \tag{23}$$

where

$$\frac{\partial E_j^S}{\partial s_{ij}}\left(S_j^*\right) = \Phi\left(\frac{S_j^* - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) + \frac{\underline{S}_j}{\sqrt{W_j}\sigma_S} \phi\left(\frac{S_j^* - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) \tag{24}$$

$$\frac{\partial E_j^Q}{\partial q_{ij}}\left(Q_j^*\right) = \Phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) + \frac{\underline{Q}}{\sqrt{W_j}\sigma_Q} \phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right). \tag{25}$$

That is, for any Nash equilibrium, the marginal cost of savings effort (the left hand side of Equation 22) is equal to the expected marginal incentive pay of savings effort (the right hand side of Equation 22) for all participants $i \in I_j$. Similarly, the marginal cost of quality effort is equal to the expected marginal benefit of quality effort (Equation 23).

The following step is critical, and it facilitates estimating the model and performing counterfactual simulations. First, pre-multiply each side of Equations 22 and 23 by w_{ij}. Next, sum each side of the equations over $i \in I_j$. The result is the aggregated first order conditions

$$\sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial s_{ij}} \left(s^*_{ij}, q^*_{ij}\right) = W^{(3)}_j B_j \left[F_j E^Q_j(Q^*_j) \frac{\partial E^S_j}{\partial s_{ij}} \left(S^*_j\right) + T_j(1 - F_j Q^*_j) \frac{\partial E^S_j}{\partial s_{ij}} \left(-S^*_j\right) \right] \tag{26}$$

$$\sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial q_{ij}} \left(s^*_{ij}, q^*_{ij}\right) = W^{(3)}_j B_j \left[F_j E^S_j(S^*_j) \frac{\partial E^Q_j}{\partial q_{ij}} \left(Q^*_j\right) + T_j F_j E^S_j(-S^*_j) \right], \tag{27}$$

where the term $W^{(3)}_j = \sum_{i \in I_j} w^3_{ij}$ discounts the marginal benefit of savings and quality at the ACO level—that is, $W^{(3)}_j$ represents the incentive dilution introduced by paying providers as groups.

The newly defined term $W^{(3)}_j$, and the related term $W_j = \sum_{i \in I_j} w^2_{ij}$ that appears in $E^S_j(\cdot)$ and $E^Q_j(\cdot)$, measure the dispersion of influence in an ACO, similar to the Herfindahl-Hirschman Index that measures market concentration and the Gini Coefficient that measures income inequality. These terms capture the main structural challenge of the MSSP according to Frandsen & Rebitzer (2015): averaging over many providers decreases uncertainty by a factor of W_j, but having many providers split earnings introduces incentive dilution on the order of $W^{(3)}_j$. Because the latter is always much smaller than the former, inefficiency from free-riding is likely to dominate any performance increase from risk mitigation. I compute W_j ($W^{(3)}_j$) from data as the sum of squared (cubed) shares of expenditure for each type of provider within an ACO. This process and the logic behind it is discussed in detail in Appendix C.

Equations 26 and 27 are aggregate analogs of Equations 22 and 23.[13] Equation 26 states that the weighted average of marginal cost of savings across ACO participants is equal to the weighted average marginal benefit of savings across ACO participants. Similarly, Equation 27 states that the weighted average of marginal cost of quality across ACO participants is equal to the weighted average marginal benefit of quality across ACO participants.

[13]In the case an ACO is in their first performance year, contracts are one-sided ($T_j = 0$) and the ACO's quality score does not impact ACO pay as long as the minimum quality standards are met. For these ACOs, expected pay is

$$\mathbb{E}\left[P_j\left(\mathring{S}_j, \mathring{Q}_j; F_j, T_j = 0\right) | S_j, Q_j\right] = \mathbb{E}\left[F_j B_j \mathring{S}_j \mathbb{1}\left\{\mathring{S}_j \geq \underline{S}_j\right\} \cdot \mathbb{1}\left\{\mathring{Q}_j \geq \underline{Q}\right\} | S_j, Q_j\right] = F_j B_j E^S_j(S_j) \Phi\left(\frac{Q_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right). \tag{28}$$

Differentiating with respect to s_{ij} and q_{ij} yields the first performance year analogs of Equations 26 and 27:

$$\sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial s_{ij}} \left(s^*_{ij}, q^*_{ij}\right) = F_j W^{(3)}_j B_j \frac{\partial E^S_j}{\partial s_{ij}} \left(S^*_j\right) \Phi\left(\frac{Q^*_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) \tag{29}$$

$$\sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial q_{ij}} \left(s^*_{ij}, q^*_{ij}\right) = F_j \frac{W^{(3)}_j}{\sqrt{W_j}\sigma_Q} B_j E^S_j(S^*_j) \phi\left(\frac{Q^*_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right). \tag{30}$$

Next, let the left hand sides of Equations 26 and 27 be denoted MC_j^S and MC_j^Q:

$$MC_j^S = \sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial s_{ij}} \left(s_{ij}^*, q_{ij}^*\right) \tag{31}$$

$$MC_j^Q = \sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial q_{ij}} \left(s_{ij}^*, q_{ij}^*\right) \,. \tag{32}$$

Also, let the right hand sides of Equations 26 and 27 be denoted $MB_j^S\left(\sigma_S, \sigma_Q\right)$ and $MB_j^Q\left(\sigma_S, \sigma_Q\right)$, respectively:

$$MB_j^S\left(\sigma_S, \sigma_Q\right) = W_j^{(3)} B_j \left[F_j E_j^Q(Q_j^*) \frac{\partial E_j^S}{\partial s_{ij}} \left(S_j^*\right) + T_j(1 - F_j Q_j^*) \frac{\partial E_j^S}{\partial s_{ij}} \left(-S_j^*\right)\right] \tag{33}$$

$$MB_j^Q\left(\sigma_S, \sigma_Q\right) = W_j^{(3)} B_j \left[F_j E_j^S(S_j^*) \frac{\partial E_j^Q}{\partial q_{ij}} \left(Q_j^*\right) + T_j F_j E_j^S(-S_j^*)\right] \,. \tag{34}$$

Note that the weighted average marginal benefits of savings and of quality, $MB_j^S\left(\sigma_S, \sigma_Q\right)$ and $MB_j^Q\left(\sigma_S, \sigma_Q\right)$, depend only on known contract values, such as benchmark expenditure B_j, and two unknown structural parameters, σ_S and σ_Q.

Next, I will assume a specific functional form for participant effort cost functions c_{ij} that will yield weighted average marginal costs characterizations that are a function of observed data and structural parameters.

Assumption 4.2. The effort cost function c_{ij} takes the quadratic form

$$c_{ij}\left(s, q\right) = \frac{\delta_S}{2} s^2 + \frac{\delta_Q}{2} q^2 + \left(\gamma_S' x_{ij} + \xi_{ij}^S\right) s + \left(\gamma_Q' x_{ij} + \xi_{ij}^Q\right) q + \kappa s q \,. \tag{35}$$

The parameters δ_S, δ_Q, and κ are quadratic coefficients that determine the overall curvature of the cost function. The linear coefficients of the cost function depend on a vector of participant- and ACO-specific characteristics, $x_{ij} \in \mathbb{R}^k$, as well as unobserved (to the econometrician) components ξ_{ij}^S and ξ_{ij}^Q. The k-dimensional coefficient vectors $\boldsymbol{\gamma}_S$ and $\boldsymbol{\gamma}_Q$ along with the quadratic coefficients δ_S, δ_Q, and κ will be estimated.

In this parameterized form, the cost function has the implicit property than ACO participants vary in the magnitude of their marginal costs of savings effort and quality effort. Variations in marginal costs across ACO participants and across ACOs—variation driven by x_{ij}, ξ_{ij}^S, and ξ_{ij}^Q—explains variation in observed ACO savings rates and quality scores. The curvature of effort cost is assumed constant across ACOs.

Data limitations become evident at this stage. Since within-ACO variation across participants is unavailable, its impossible to identify a specific marginal cost for each participant in each ACO. Weighted average marginal costs MC_j^S and MC_j^Q are the fullest description of marginal cost that can be identified from aggregate values. These values are still informative, however: because each is weighted by participant influence on ACO outcomes, they can be interpreted as the marginal costs of savings effort and quality effort for a representative participant in a given ACO.

Under Assumption 4.2, marginal cost is linear in savings and quality effort. This permits aggregation of participants' first order conditions within ACOs. Specifically,

$$\frac{\partial c_{ij}}{\partial s_{ij}}\left(s_{ij}^*, q_{ij}^*\right) = \delta_S s_{ij}^* + \boldsymbol{\gamma}_S' x_{ij} + \xi_{ij}^S + \kappa q_{ij}^* \tag{36}$$

$$\frac{\partial c_{ij}}{\partial q_{ij}}\left(s_{ij}^*, q_{ij}^*\right) = \delta_Q q_{ij}^* + \boldsymbol{\gamma}_Q' x_{ij} + \xi_{ij}^Q + \kappa s_{ij}^* \,. \tag{37}$$

Recall the definitions $S_j^* = \sum_{i \in I_j} w_{ij} s_{ij}^*$, $Q_j^* = \sum_{i \in I_j} w_{ij} q_{ij}^*$, and $\sum_{i \in I_j} w_{ij} = 1$. Using these, we can derive a functional form for MC_j^S:

$$MC_j^S = \sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial s_{ij}}\left(s_{ij}^*, q_{ij}^*\right) \tag{38}$$

$$= \sum_{i \in I_j} \left(w_{ij}\delta_S s_{ij}^* + w_{ij}\boldsymbol{\gamma}_S' x_{ij} + w_{ij}\xi_{ij}^S + w_{ij}\kappa q_{ij}^*\right) \tag{39}$$

$$= \delta_S \left(\sum_{i \in I_j} w_{ij} s_{ij}^*\right) + \boldsymbol{\gamma}_S' \left(\sum_{i \in I_j} w_{ij} x_{ij}\right) + \sum_{i \in I_j} w_{ij}\xi_{ij}^S + \kappa \left(\sum_{i \in I_j} w_{ij} q_{ij}^*\right) \tag{40}$$

$$= \delta_S S_j^* + \boldsymbol{\gamma}_S' X_j + \kappa Q_j^* + \Xi_j^S \,. \tag{41}$$

where $X_j \equiv \sum_{i \in I_j} w_{ij} x_{ij}$ is a k-dimensional vector of the weighted averages of provider characteristics, x_{ij}, and $\Xi_j^S \equiv \sum_{i \in I_j} w_{ij}\xi_{ij}^S$. Similarly for MC_j^Q:

$$MC_j^Q = \sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial q_{ij}}\left(s_{ij}^*, q_{ij}^*\right) \tag{42}$$

$$= \sum_{i \in I_j} \left(w_{ij}\delta_Q q_{ij}^* + w_{ij}\boldsymbol{\gamma}_Q' x_{ij} + w_{ij}\xi_{ij}^Q + w_{ij}\kappa s_{ij}^*\right) \tag{43}$$

$$= \delta_Q \left(\sum_{i \in I_j} w_{ij} q_{ij}^*\right) + \boldsymbol{\gamma}_Q' \left(\sum_{i \in I_j} w_{ij} x_{ij}\right) + \sum_{i \in I_j} w_{ij}\xi_{ij}^Q + \kappa \left(\sum_{i \in I_j} w_{ij} s_{ij}^*\right) \tag{44}$$

$$= \delta_Q Q_j^* + \boldsymbol{\gamma}_Q' X_j + \kappa S_j^* + \Xi_j^Q \,. \tag{45}$$

Define the parameterizations $MC_j^S(\boldsymbol{\theta}) = MC_j^S - \Xi_j^S$ and $MC_j^Q(\boldsymbol{\theta}) = MC_j^Q - \Xi_j^Q$, where $\boldsymbol{\theta} = \left\{\delta_S, \delta_Q, \boldsymbol{\gamma}_S, \boldsymbol{\gamma}_Q, \kappa\right\}$. Combining this with Equations 26 and 27, we get

$$MB_j^S(\sigma_S, \sigma_Q) = MC_j^S(\boldsymbol{\theta}) + \Xi_j^S \tag{46}$$

$$MB_j^Q(\sigma_S, \sigma_Q) = MC_j^Q(\boldsymbol{\theta}) + \Xi_j^Q \,. \tag{47}$$

I make the following assumption on the aggregated marginal cost shocks Ξ_j^S and Ξ_j^Q to identify the model parameters $\boldsymbol{\theta}$, σ_S, and σ_Q.

Assumption 4.3. $\mathbb{E}\left[\Xi_j^S, \Xi_j^Q | X_j, S^*_{j,PY-1}, Q^*_{j,PY-1}\right] = 0$, where $S^*_{j,PY-1}$ and $Q^*_{j,PY-1}$ are the previous performance year's savings rate and quality score of the ACO of observation j.

Assumption 4.3 permits identification of $\boldsymbol{\theta}$, σ_S, and σ_Q. I use a nested multiple equation Generalized Method of Moments (GMM) estimator (Hansen, 1982) to estimate the structural parameters. Appendix C gives a detailed description of the estimation procedure used.

This assumption on error terms is supported by the interpretation that Ξ_j^S and Ξ_j^Q are marginal cost shocks. Critically, these cost shocks are not assumed to be independent of S_j^* and Q_j^*. This is because an ACO experiencing a large marginal cost shock to their ability to reduce costs (high Ξ_j^S) chooses, *ceteris paribus*, a lower savings effort, all else constant. This means Ξ_j^S is negatively correlated with S_j^*, and likewise Ξ_j^Q is negatively correlated with Q_j^*.

The cross-relationships of Ξ_j^S with Q_j^* and Ξ_j^Q with S_j^* depend on the tradeoff between savings effort and quality effort. For example, if there is a tradeoff between savings effort and quality effort, then a positive marginal cost shock to savings effort causes a decrease in savings effort. However, because of the tradeoff between savings and quality, lower savings effort implies that the marginal cost of quality effort is decreased, so quality effort may optimally increase. This would imply a positive relationship between Ξ_j^S and Q_j^*. To cover all cases, I allow for $\mathbb{E}[\Xi_j^S | S_j^*, Q_j^*] \neq 0$ and $\mathbb{E}[\Xi_j^Q | S_j^*, Q_j^*] \neq 0$ in estimation.

The following question remains: what causes the exogenous variation in Ξ_j^S and Ξ_j^Q? The answer is well-established in health policy literature. First, some ACOs, even after conditioning on observed characteristics, are favored by a larger assignment of benchmark expenditure B_j, making it easier to reduce costs (McWilliams, 2014; McWilliams et al., 2018). Furthermore, beneficiary assignment and quality scoring are fine tuned year-to-year by CMS, causing exogenous variations in the difficulty of improving quality of care.

The lagged-variables of ACO savings rate and quality score are valid instruments because they correlate with contemporaneous savings rates and quality scores through idiosyncratic and time-invariant unobserved advantages and disadvantages of each ACO. The implicit assumption of using lagged-variables in this manner is that there is no autocorrelation of Ξ_j^S and Ξ_j^Q and the values of these variables in the previous performance year of observation j. The exclusion restriction is thus Ξ_j^S and Ξ_j^Q are not correlated with $S^*_{j,PY-1}$ or $Q^*_{j,PY-1}$, meaning an ACO's previous year's efforts do not impact the current year's unobserved marginal cost.

The elements of X_j, which can be interpreted as ACO-specific marginal cost shifters, appear in Table 5 in Appendix C. The vector X_j includes information concerning an ACOs assigned beneficiaries, participating Medicare providers, as well as expenditure and service statistics.

4.2 Identification Intuition and Illustrative Example

The functional form assumption on cost is necessary to identify a cost function from aggregate outcomes. Nonetheless, several results of this paper remain for any cost function. In particular, an empirical estimate of the savings-quality tradeoff is by definition the change in marginal cost of savings with respect to quality, or $\frac{\partial \hat{MC}_j^S}{\partial Q_j^*}$. While this is given a single parameter, κ, above, the tradeoff would nonetheless be identified for any cost function satisfying model assumptions. To gain intuition for identifying the shape of ACO participants' cost functions (including a savings-quality tradeoff), let's consider a simple example. This example is written without assuming an explicit functional form for c_{ij} so that it's clear that results regarding the shape of participant's cost functions stem from variation in ACO outcomes and not a specific functional form assumption.

First suppose there are two (conditionally identical) ACOs with weighted average marginal benefit of savings such that $MB_1^S < MB_2^S$, where $S_1^* < S_2^*$ and $Q_1^* = Q_2^*$. Since marginal benefit is (on average) equal to marginal cost in equilibrium, we have $S_1^* < S_2^*$ and $MC^S(S_1^*, Q_1^*) < MC^S(S_2^*, Q_2^*)$. Marginal cost is of savings increasing in savings, so cost is convex in savings. The average increase in this case would then be equal to

$$\frac{\partial^2 c_{ij}}{\partial s_{ij}^2} \approx \frac{MC^S(S_2^*, Q_2^*) - MC^S(S_1^*, Q_1^*)}{S_2^* - S_1^*} \equiv \hat{\delta}_S \,. \tag{48}$$

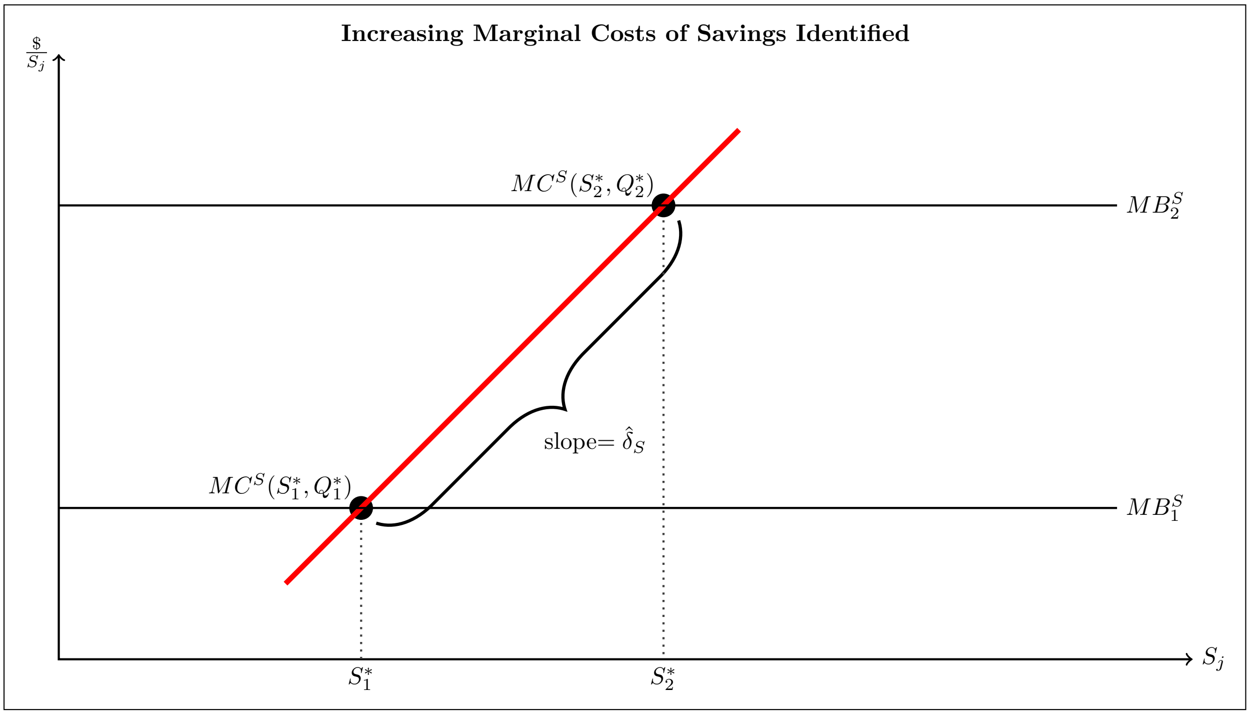
The functional form imposed in Assumption 4.2 therefore restricts the estimate of $\frac{\partial^2 c_{ij}}{\partial s_{ij}^2}$ to a constant value, rather than one that's a function of savings or quality.

The argument is pictured in Figure 1. Dollars are on the y-axis, and ACO savings rate is on the x-axis. The slope of the line connecting points at $\big(S_1^*, MC^S(S_1^*, Q_1^*)\big)$ and $\big(S_2^*, MC^S(S_2^*, Q_2^*)\big)$ is $\hat{\delta}_S$. The variation in weighted average marginal benefit identifies the slope (with respect to ACO savings rate) of the marginal cost of savings.

In this example, I've assumed identical quality scores, so variation across just these two hypothetical ACOs does not identify a savings-quality tradeoff. To show variation that identifies a savings-quality tradeoff, suppose there is another ACO with the same marginal benefit as ACO 2, but a different savings rate and quality score. Specifically, suppose ACO 3 is observed such that $MB_1^S < MB_2^S = MB_3^S$, $S_1^* < S_2^* < S_3^*$, and $Q_1^* = Q_2^* > Q_3^*$. Define $\hat{\delta}_S$ as before. The change in marginal cost of savings with respect to quality (the savings-quality tradeoff) is equal to

$$\frac{\partial^2 c_{ij}}{\partial s_{ij} \partial q_{ij}} \approx \frac{MC^S(S_2^*, Q_2^*) - MC^S(S_2^*, Q_3^*)}{Q_2^* - Q_3^*} \,. \tag{49}$$

Figure 1: **Identification of Marginal Cost**



Note: This figure shows how a convex cost function is identified from observed values of marginal benefit of savings. Given two ACOs with different marginal benefits of savings MB_1^S and MB_2^S (that are otherwise identical), the observed difference between their chosen savings rates S_1^* and S_2^* identifies the change in marginal cost of savings with respect to savings.

Then, applying the defintion of $\hat{\delta}_S$, we have

$$\frac{\partial^2 c_{ij}}{\partial s_{ij} \partial q_{ij}} \approx \frac{MC^S(S_2^*, Q_2^*) - MC^S(S_2^*, Q_3^*)}{Q_2^* - Q_3^*} \tag{50}$$

$$= \frac{MC^S(S_2^*, Q_2^*) - \left[MC^S(S_3^*, Q_3^*) - \hat{\delta}_S(S_3^* - S_2^*)\right]}{Q_2^* - Q_3^*} \tag{51}$$

$$= \frac{MC^S(S_2^*, Q_2^*) - \left[MC^S(S_2^*, Q_2^*) - \hat{\delta}_S(S_3^* - S_2^*)\right]}{Q_2^* - Q_3^*} \tag{52}$$

$$= \hat{\delta}_S \frac{S_3^* - S_2^*}{Q_2^* - Q_3^*} \equiv \hat{\kappa}\,. \tag{53}$$

Since $Q_2^* > Q_3^*$ and $S_3^* > S_2^*$, this means the marginal cost of savings is increasing in quality. Like above, Assumption 4.2 makes $\frac{\partial^2 c_{ij}}{\partial s_{ij} \partial q_{ij}}$ be a constant value in estimation.

Figure 2 shows this process for identifying a tradeoff in the top panel. The solid red line is the same that was found in Figure 1. Since ACOs 2 and 3 have the same marginal revenue of savings (and thus marginal cost), I can compute the marginal cost of savings for an ACO with the savings rate of ACO 2 and quality score of ACO 3, denoted by the point at $MC^S(S_2^*, Q_3^*)$. Then, the difference between $MC^S(S_2^*, Q_2^*)$ and $MC^S(S_2^*, Q_3^*)$ (the vertical difference between the red solid line and blue dashed line) is the increase in marginal cost of savings for an increase in quality from Q_3^* to Q_2^*.

The bottom panel of Figure 2 shows how complementarity of savings and quality can be identified. The setup remains the same, except the savings rate of ACO 2 is now greater than the quality score of ACO 3: $MB_1^S < MB_2^S = MB_3^S$, $S_1^* < S_2^* > S_3^*$, and $Q_1^* = Q_2^* > Q_3^*$. Note that $\hat{\kappa}$ is negative in this case, since increasing the quality score from Q_3^* to Q_2^* decreases marginal cost by $\hat{\kappa} \cdot (Q_2^* - Q_3^*)$.

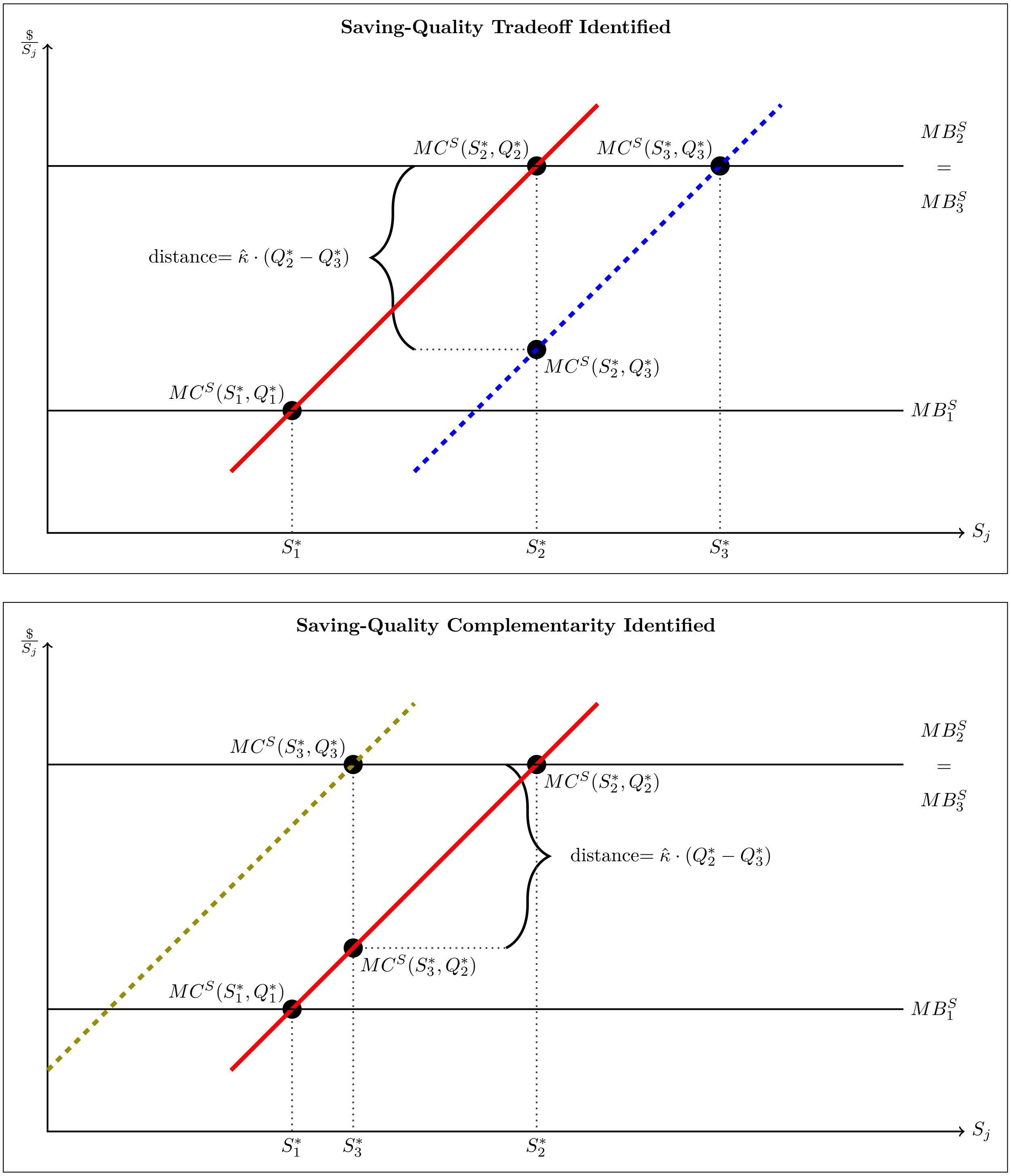
5 Estimation Results and Model Fit

Table 2 shows the estimates for the nonlinear cost function parameters δ_S, δ_Q, and κ as well as estimates for the uncertainty parameters σ_S and σ_Q. In order to generate the statistics presented, a single GMM estimation of parameters is performed first, and the results of this estimation are contained in the "Estimate" column. Next, the same estimation is performed, however with a re-sampled dataset with the same number of observations as the original. The estimation results are recorded, and the dataset is again re-sampled, and again estimation is preformed and recorded. After 42,327 estimation replications were performed, the "Standard Error" and "95% CI Lower (Upper) Bound" columns were generated from the empirical standard deviations and quantiles of the recorded estimates.[14] The "P-value" column tests the null hypothesis that population structural parameter in question is zero.

The estimated cost function and uncertainty parameters satisfy the requirements for existence and uniqueness of equilibrium derived in Appendix B. Ultimately, for equilibrium to exist and be unique, the estimates

[14]There were 50,000 total replications of estimation performed, 7,673 of which did not find a minimum in the GMM minimization problem, leaving 42,327 remaining valid estimates.

Figure 2: **Identification of Savings-Quality Tradeoff or Complementarity**



Note: In the top panel, this figure shows how a savings-quality tradeoff is identified from observed values of marginal benefit of savings. Given three ACOs such that marginal benefits of savings are $MB_1^S < MB_2^S = MB_3^S$, savings rates are $S_1^* < S_2^* < S_3^*$, and quality scores are $Q_1^* = Q_2^* > Q_3^*$, the observed difference between savings rates S_2^* and S_3^* identifies the change in marginal cost of savings with respect to quality. The bottom panel shows the analogous case for identifying complementarity between savings and quality, where $S_1^* < S_3^* < S_2^*$.

Table 2: **Cost Function and Uncertainty Parameter Estimates**

Parameter	Estimate	Standard Error	P-value	95% CI Lower Bound	95% CI Upper Bound
δ_S	185.400	71.225	0.009	70.289	345.420
δ_Q	8.820	1.637	0.000	6.062	12.513
κ	18.037	8.318	0.030	3.157	35.403
σ_S	0.160	0.010	0.000	0.150	0.190
σ_Q	0.420	0.030	0.000	0.340	0.430

Note: $|\mathcal{J}| = 1486$ observations. This table displays parameter estimates for the effort cost function $c(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \gamma_S s + \gamma_Q q + \kappa sq$ and for the standard deviations of the effort-shocks η_{ij}^S and η_{ij}^S. Standard errors, p-values, and confidence interval (CI) bounds are obtained by bootstrapping with 42,327 rep. Parameter estimates not shown here, including year and Census Division fixed-effects, are available in Tables 7 and 8 in Appendix C.

of parameters δ_S, δ_Q, σ_S and σ_Q must be sufficiently large relative to the magnitude of the estimate of κ. The estimates in Table 2 satisfy this requirement for all possible contracts for 99.8% of observations.

The estimates of the vectors of parameters $\boldsymbol{\gamma}_S$ and $\boldsymbol{\gamma}_Q$ are shown in Appendix D. Large and precisely estimated marginal cost of effort shifters are risk score ($\uparrow MC^S$, $\uparrow MC^Q$), the proportion of services in hospital settings ($\downarrow MC^S$, $\downarrow MC^Q$), ACO age ($\downarrow MC^S$, $\downarrow MC^Q$), and whether an ACO is comprised of only hospitals/group practices ($\uparrow MC^S$, $\uparrow MC^Q$). There is not any variable that increases the marginal cost of savings effort and decreases the marginal cost of quality effort (or vice versa).

5.1 Model Fit

I address model fit by simulating equilibrium ACO performance measures S_j^* and Q_j^* where all ACOs have the same contract as in estimation: $T_j \equiv 0$ and $F_j \equiv 0.5$.[15] Table 3 shows simulation results against data for several key measures. TSP is total program savings, defined as the sum of total savings and losses of all

Table 3: **Model Fit: ACO Performance**

	TPS	SPB	QWS	QWB	S_j^*	Q_j^*
Data "Track 1": $F = 0.5$, $T = 0$	0.120	19.754	-14.551	-11.186	0.008	0.884
Simulated "Track 1": $F = 0.5$, $T = 0$	0.202	11.092	-13.996	-11.871	0.015	0.916

Note: This table compares various measures of ACO performance calculated in data and in simulations. TPS is total program savings (in \$ billions) and SPB is average ACO savings per-beneficiary (in \$). QWS and QWB are quality-weighted total program savings average quality-weighted per-beneficiary savings. Columns S_j^* and Q_j^* contain the average equilibrium ACO savings rate and quality score. Larger values of each measure indicate better average performance.

ACOs minus the sum of all ACO incentive payments and penalties. SPB is average savings per beneficiary, which is the simple mean of ACO net savings (savings − incentive pay) divided by the ACO's number of assigned beneficiaries. QWS and QWB are quality-weighted total program savings and quality-weighted

[15]The process of simulating counterfactual equilibrium ACO performance values is described in detail in Section 6.

average savings per beneficiary, where one additional dollar of savings for an ACO increases both measures by $Q_j \in [0, 1]$, the ACO's quality score. These measures are defined formally and discussed in more detail in Section 6, but for now they can be interpreted as varying representations of Medicare's or a social planner's preferences.

The first row of Table 3 shows that total program savings over the estimation sample was \$120 million, with \$19.754 saved per beneficiary. Simulating choices with the same contract parameters as the data generating process (sharing rate $F = 0.5$, no penalties $T = 0$), we get predicted savings of \$202 million overall and \$11.092 average per beneficiary. These estimates are rather close: the error of \$82 million dollars of total program savings across all ACOs is considerably small relative to the average per-ACO expenditure of \$190 million. Also, while predictions of savings per beneficiary are off by \$8.22, the standard deviation of savings per beneficiary across ACOs is \$357.45, implying the simulated value is actually rather close to the value in the data.

For the quality-weighted measures, simulated values are even closer to data values. However, while TPS was overestimated and SPB was underestimated in the simulations, the quality-weighted versions of these variables are underestimated and overestimated, respectively. The reason is contained in the final two columns showing averages of ACO savings rates and quality scores: simulations slightly overestimate (by less than 0.1 of a standard deviation) average ACO savings rates S_j^* and average quality scores Q_j^*.

Because the relationship between ACO savings rates and quality scores is at the heart of the theoretical and empirical questions of this paper, Table 4 shows the covariance and correlation matrices of observed and simulated ACO savings rates and quality scores. In both matrices, cells that contain comparable pairs of data and simulated values are shaded the same color. The covariance of S_j^* in the data is 0.00186 and in simulations it is 0.00259—taking square roots, these values correspond to a standard deviation of equilibrium savings rates across ACOs of 0.043 to 0.051, respectively. Similarly, the values in the covariance matrix indicate the standard deviation of Q_j^* is 0.097 in the estimation sample and 0.119 in the simulation.

Comparing the green and yellow boxes in the covariance and correlation matrices shows how well the model simulation can to match the small positive relationship of savings and quality measures in the data. The bold numbers in the correlation matrix are simply the unconditional correlations of values in the data and values in simulations. Savings rates of ACOs are matched very closely in the model, and most variation in quality scores is explained by the model. The weaker model fit of quality scores likely stems from greater inherent uncertainty in quality of care (as measured by σ_Q) and from this model's shortcomings in explaining variation in ACO quality score at the extremes. Specifically, many ACOs have quality scores near 1, and variation in quality scores at that level is more likely to be driven by external factors than provider decisions.

Table 4: **Model Fit: Savings Rates and Quality Scores**

Covariance Matrix

	S_j^* Data	S_j^* Simulated	Q_j^* Data	Q_j^* Simulated
S_j^* Data	0.00186			
S_j^* Simulated	0.00186	0.00259		
Q_j^* Data	0.00018	0.0004	0.00953	
Q_j^* Simulated	0.00004	0.0004	0.00775	0.01408

Correlation Matrix

	S_j^* Data	S_j^* Simulated	Q_j^* Data	Q_j^* Simulated
S_j^* Data	1			
S_j^* Simulated	**0.84585**	1		
Q_j^* Data	0.04393	0.08079	1	
Q_j^* Simulated	0.00833	0.0661	**0.66923**	1

Note: This table compares various relationships of observed and predicted ACO savings rates S_j^* and quality scores Q_j^*. Model fit is indicated by numbers in same-colored cells sharing similar values.

6 Counterfactual Simulations

In this section, I use the estimated model of ACO performance to evaluate counterfactual contracts between ACOs and Medicare in the MSSP. The main goal of this section is to evaluate the cost-quality tradeoff within the context of this incentive program. This section begins with a simulation of ACO performance under all possible sharing rates ($F \in [0, 1]$) and with and without two-sided risk ($T \in \{0, 1\}$). Then, performance is compared—in particular, we are interested in whether imposing penalties on ACOs will lead to a substantial decrease in quality of care, as measured by quality score Q_j. I will find that penalties for overspending decrease quality of care by no more than 0.2 standard deviations, while these penalties incentivize savings increases of up to 1.5 standard deviations.

Following the comparison of ACO contracts, I will simulate counterfactual ACO performance when there is no free-riding in ACOs. This counterfactual is possible because the model endogenizes choice under group incentives. Not only will this allow us to see the performance losses from free-riding, but it also permits comparison of optimal contracts parameters depending on the presence of free-riding. As predicted by Frandsen & Rebitzer (2015), I find substantial losses from free-riding: per-beneficiary savings are nearly $500 higher without free-riding, and quality scores are one standard deviation higher. Medicare maximizes ACO savings with less generous contracts when there is no free-riding.

Note that unlike the simulations by Aswani et al. (2019), this model includes endogenous choice of quality and savings levels under group incentives. However, also unlike Aswani et al. (2019), it omits a model of provider participation. Reassuringly, the optimal contract parameters of this section are not very different from those found by Aswani et al. (2019). In particular, the optimal sharing rate F for one-sided contracts is found to be between 0.49 and 0.55 in both papers.

The counterfactual simulations follow the these steps.

1. Let the structural parameters estimated in Section 4 be denoted $\hat{\boldsymbol{\theta}} = \{\hat{\delta}_S, \hat{\delta}_Q, \hat{\kappa}, \hat{\gamma}_S, \hat{\gamma}_Q\}$, $\hat{\sigma}_S$, and $\hat{\sigma}_Q$. Let the estimated marginal cost shocks (residuals) be denoted $\hat{\Xi}_j^S$ and $\hat{\Xi}_j^Q$ for each $j \in \mathcal{J}$.

2. Specify $F \in [0,1]$ and $T \in \{0,1\}$ for the simulation.

3. Define the estimated (parameterized) weighted-average marginal cost and marginal benefit equations

$$\hat{MC}_j^S\left(S, Q; \hat{\boldsymbol{\theta}}\right) = \hat{\delta}_S S + \hat{\gamma}_S' X_j + \hat{\Xi}_j^S + \kappa Q \tag{54}$$

$$\hat{MC}_j^Q\left(S, Q; \hat{\boldsymbol{\theta}}\right) = \hat{\delta}_Q Q + \hat{\gamma}_Q' X_j + \hat{\Xi}_j^Q + \kappa S \tag{55}$$

$$\hat{MB}_j^S\left(S, Q; \hat{\sigma}_S, \hat{\sigma}_Q\right) = W_j^{(3)} B_j \left[F \hat{E}_j^Q\left(Q; \hat{\sigma}_Q\right) \frac{\partial \hat{E}_j^S}{\partial s_{ij}}\left(S; \hat{\sigma}_S\right) + T(1 - FQ) \frac{\partial \hat{E}_j^S}{\partial s_{ij}}\left(S; \hat{\sigma}_S\right)\right] \tag{56}$$

$$\hat{MB}_j^Q\left(S, Q; \hat{\sigma}_S, \hat{\sigma}_Q\right) = W_j^{(3)} B_j \left[F \hat{E}_j^S\left(S; \hat{\sigma}_S\right) \frac{\partial \hat{E}_j^Q}{\partial q_{ij}}\left(Q; \hat{\sigma}_Q\right) + TF \hat{E}_j^S\left(S; \hat{\sigma}_S\right)\right] \tag{57}$$

for all ACOs $j \in \mathcal{J}$, and where $\hat{E}_j^S(S; \hat{\sigma}_S) = S\Phi\left(\frac{S - \underline{S}_j}{\sqrt{W_j}\hat{\sigma}_S}\right) + \sqrt{W_j}\hat{\sigma}_S \phi\left(\frac{S - \underline{S}_j}{\sqrt{W_j}\hat{\sigma}_S}\right)$ (and similarly define $\hat{E}_j^Q(Q, \hat{\sigma}_Q \cdot))$.

4. For all ACOs $j \in \mathcal{J}$, solve for the values S' and Q', that satisfy

$$\hat{MC}_j^S\left(S', Q'; \hat{\boldsymbol{\theta}}\right) = \hat{MB}_j^S\left(S', Q'; \hat{\sigma}_S, \hat{\sigma}_Q\right) \tag{58}$$

$$\hat{MC}_j^Q\left(S', Q'; \hat{\boldsymbol{\theta}}\right) = \hat{MB}_j^Q\left(S', Q'; \hat{\sigma}_S, \hat{\sigma}_Q\right). \tag{59}$$

The proof in Appendix B shows that the solution of this system S' and Q' are the same values of Q_j^* and S_j^* such that $S_j^* = \sum_{i \in I_j} w_{ij} s_{ij}^*$, $Q_j^* = \sum_{i \in I_j} w_{ij} q_{ij}^*$, and $\left(s_{ij}^*, q_{ij}^*\right)_{i \in I_j}$ are the multitasking equilibrium choices of participants in ACO j, given contract parameters F and T. To emphasize the dependence of ACO choices on contract parameters, denote $S' = \hat{S}_j(F, T)$ and $Q' = \hat{Q}_j(F, T)$.

5. Compute ACO performance measures: total program savings $\mathscr{U}^{TPS}$, quality-weighted total program savings $\mathscr{U}^{QWS}$, per-beneficiary average ACO savings $\mathscr{U}^{SPB}$, and per-beneficiary average ACO quality-weighted savings $\mathscr{U}^{QWB}$. Theses values are taken to represent different specifications of Medicare's preferences for cost reduction and quality of care as the principal in the optimal contracting problem, and each is defined formally as a function of counterfactual performance values $\hat{S}_j(F, T)$ and $\hat{Q}_j(F, T)$ in the following subsection.

6.1 MSSP Contract Comparisons

Here, I model incentive design in the MSSP as an optimal contracting problem and solve for contracts that maximize a wide range of objectives. This model departs from typical models of contracting in several ways. First, Medicare faces several regulatory constraints that hinder its ability to impose contracts that achieve first-best outcomes. Specifically, Medicare is required to pay entire ACOs and not individual ACO participants. This results in free-riding due to incentive dilution (the complications of which are the focus of Frandsen & Rebitzer (2015)). Another difference is that Medicare is restricted to a very specific functional form that is multiplicative and discontinuous in agent effort choices.

The ideal objective function for evaluating counterfactual simulations would be economic welfare—however, this is out of the scope of this paper, as that would require an estimation of a demand model for health care and health care quality as measured by ACO quality score (in order to compute willingness-to-pay for different values of Q_j) and a common agency model to account for efficiency spillover to the private insurance market (as modeled in Frandsen et al. (2019)). Further complicating this prospective analysis, both of these additions would require data that does not exist at the ACO participant level.

Instead, let the objective functions I define be interpreted as representations of the preferences of Medcare, who is the principal in the optimal contracting problem that is designing ACO payment models in the MSSP. The first definition is total program savings, defined as the sum of all ACO savings (according to the defined benchmark expenditure B_j) minus the sum of all ACO incentive pay. Formally, let $\left\{\hat{S}_j\left(F,T\right),\hat{Q}_j\left(F,T\right)\right\}_{j\in\mathcal{J}}$ be the simulated equilibrium savings rates and quality scores of the sample of ACOs under counterfactual sharing rate $F\in[0,1]$ and risk model $T\in\{0,1\}$. Total program savings is therefore

$$\mathscr{U}^{TPS}\left(\left\{\hat{S}_j\left(F,T\right),\hat{Q}_j\left(F,T\right)\right\}_{j\in\mathcal{J}}\right)=\sum_{j\in\mathcal{J}}\left[B_j\cdot\hat{S}_j(F,T)-P_j\left(\hat{S}_j(F,T),\hat{Q}_j(F,T);F,T\right)\right]. \tag{60}$$

The first term in the summation, $B_j\cdot\hat{S}_j(F,T)$, is the predicted dollar amount of savings (or losses if negative) of ACO j. The second term, $P_j(\hat{S}_j(F,T),\hat{Q}_j(F,T);F,T\cdot)$, is the payment to the ACO. According to this definition, Medicare's objective is to maximize the total dollar savings to Medicare for the ACOs in the simulation sample $\mathcal{J}$.

There are two problems with Medicare's preferences defined as in $\mathscr{U}^{TPS}$. First, $\mathscr{U}^{TPS}$ depends on quality score $\hat{Q}_j(F,T)$ only through ACO payment $P_j(\cdot)$, and so it is *decreasing* in quality score. Second, the contract parameters (F,T) that maximize total program savings are *not* the same as the parameters that maximize *per-beneficiary* savings. This is because ACOs that have the largest cost reductions have disproportionately large number of assigned beneficiaries. Therefore, contract parameters chosen to maximize total savings favor ACOs that have large numbers of beneficiaries, while contract parameters chosen to maximize per-beneficiary savings (at the ACO level) may be more representative of optimal contracts for the average Medicare provider.

In order to solve these two problems, define the three following alternative objectives: quality-weighted

ACO savings $\mathscr{U}^{QWS}$, average per-beneficiary savings $\mathscr{U}^{SPB}$, and average quality-weighted per-beneficiary savings $\mathscr{U}^{QWB}$:

$$\mathscr{U}^{QWS}(\cdot) = \sum_{j \in \mathcal{J}} \left\{ \left[B_j \cdot \hat{S}_j(F,T) - P_j\left(\hat{S}_j(F,T), \hat{Q}_j(F,T); F, T\right) - \underline{\Psi} \right] \cdot \hat{Q}_j(F,T) + \underline{\Psi} \right\} \tag{61}$$

$$\mathscr{U}^{SPB}(\cdot) = |\mathcal{J}|^{-1} \cdot \sum_{j \in \mathcal{J}} \left[\frac{B_j \cdot \hat{S}_j(F,T) - P_j\left(\hat{S}_j(F,T), \hat{Q}_j(F,T); F, T\right)}{\#\text{ beneficiaries}_j} \right] \tag{62}$$

$$\mathscr{U}^{QWB}(\cdot) = |\mathcal{J}|^{-1} \cdot \sum_{j \in \mathcal{J}} \left\{ \left[\frac{B_j \cdot \hat{S}_j(F,T) - P_j\left(\hat{S}_j(F,T), \hat{Q}_j(F,T); F, T\right)}{\#\text{ beneficiaries}_j} - \underline{\psi} \right] \cdot \hat{Q}_j(F,T) + \underline{\psi} \right\}. \tag{63}$$

The quality-weighted measures multiply the total (per beneficiary) savings of each ACO by quality score Q_j before summing over the observations in the sample. The terms $\underline{\Psi} = \min_{j \in \mathcal{J}, F \in [0,1], T \in \{0,1\}} B_j \cdot \hat{S}(\cdot) - P_j(\cdot)$ and $\underline{\psi} = \min_{j \in \mathcal{J}, F \in [0,1], T \in \{0,1\}} \frac{B_j \cdot \hat{S}(\cdot) - P_j(\cdot)}{\#\text{ beneficiaries}_j}$ are normalizing values chosen so that only positive values are multiplied by $\hat{Q}_j(F,T)$. Normalization is necessary to make the quality-weighted objectives increasing in ACO quality score $\hat{Q}_j(F,T)$ for all values of $\hat{S}_j(F,T)$—without normalization, $\mathscr{U}^{QWS}$ and $\mathscr{U}^{QWB}$ would be decreasing in $\hat{Q}_j(F,T)$ if $\hat{S}_j(F,T) < 0$. The quality-weighted objectives $\mathscr{U}^{QWS}$ and $\mathscr{U}^{QWB}$ can be interpreted as the total (or mean per beneficiary) savings of ACOs such that every one dollar of cost reduction of ACO j is worth $Q_j \in [0,1]$ dollars to Medicare.

Finally, we can set up incentive design in the MSSP as a contracting problem. For preference specifications $M \in \{TPS, QWS, SPB, QWB\}$, I solve Medicare's optimal contracting problem:

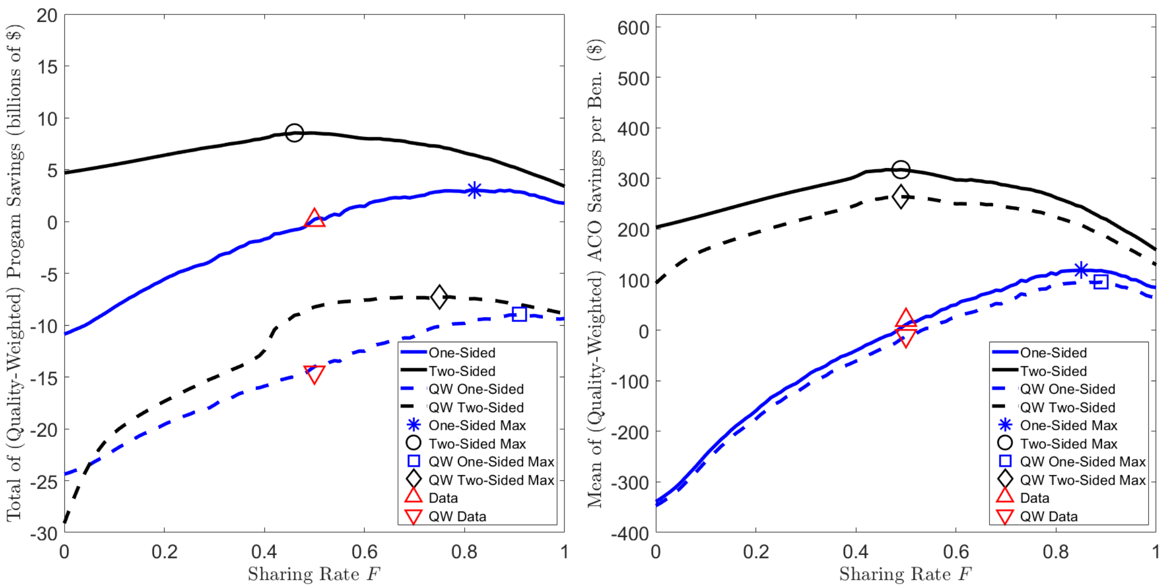
$$\max_{F,T} \; \mathbb{E}\left[\mathscr{U}^M \left(\left\{ \mathring{S}_j, \mathring{Q}_j \right\}_{j \in \mathcal{J}} \right) \middle| \left\{ \hat{S}_j, \hat{Q}_j \right\}_{j \in \mathcal{J}} \right] \tag{64}$$

$$\text{subject to } \; (\hat{s}_{ij}, \hat{q}_{ij}) = \arg\max_{s_{ij}, q_{ij}} \; w_{ij} \mathbb{E}\left[P_j\left(\mathring{S}_j, \mathring{Q}_j; F, T\right) \middle| S_j, Q_j \right] - c_{ij}(s_{ij}, q_{ij}) \text{ for all } i \in I_j, j \in \mathcal{J} \tag{65}$$

where $\hat{S}_j = \sum_{i \in I_j} \hat{s}_{ij}$ and $\hat{Q}_j = \sum_{i \in I_j} \hat{q}_{ij}$. The constraint of this problem is an incentive compatibility constraint requiring that for any chosen contract parameters, the ACO-level outcomes are determined by the Nash equilibrium strategies played in the multitasking game's equilibrium.

Figure 3 plots the objective functions for all four of of the different preference specifications for Medicare. Each line in each plot is a graph of F vs. $\mathbb{E}[\mathcal{U}^M(\cdot)]$ for $F \in [0,1]$, $T = 0$ and $T = 1$. The blue lines are simulations of behavior under one-sided contracts ($T = 0$), and the black lines are behavior under two-sided contracts ($T = 1$). Dashed lines are quality-weighted measures. On the left panel, we can assess model fit with the red triangle on the solid blue line (total savings) and the dashed blue line (quality-weighted savings). Likewise, the red triangles on the right panel of Figure 3 show model fit for the per-beneficiary measures of performance. As mentioned in Section 5, the model's simulations of ACO performance has very small relative error.

Figure 3: **Simulated Total Program Savings and Average Savings per Beneficiary**



Note: This figure plots simulated total program savings and quality-weighted total program savings on the left panel and simulated average per-beneficiary savings and average quality-weighted per-beneficiary savings on the right panel (in both panels, quality-weighted measures are the dashed lines). Simulations vary MSSP contracts by Sharing Rate (horizontal axis) and presence of penalties (blue line is without penalties, black line is with). In the legend, "QW" means "quality-weighted."

6.1.1 Optimal Sharing Rates

Optimal sharing rates for one-sided ACOs are marked in Figure 3 by the blue star and square. These values range from a sharing rate of 83% to a sharing rate of 92%, which is significantly higher-powered than existing MSSP one-sided contracts that offer a sharing rate of only 50%. Two-sided contracts, on the other hand, are optimally weaker-powered. All measures of performance, except for quality-weighted total program savings, have optimal sharing rates between 47% and 50%.[16] Also, there is a region of increasing returns to F around $F = 0.4$ on the black dashed line representing $\mathscr{U}^{QWS}$.

Weighting savings and per-beneficiary savings by quality score has the overall effect of shifting the levels of these objective functions, but the shape of the objective functions are mostly unchanged. Optimal sharing rates move from lower values (circles and stars in Figure 3) to higher values (diamonds and squares). But, the changes are small in magnitude, implying Medicare can do little to improve quality of care by increasing the sharing rate F.

[16] Maximization of quality-weighted total program savings (the black diamond on the left panel of Figure 3) occurs at $F = 0.76$. However, the curvature of $\mathscr{U}^{QWS}$ at this point is extremely flat relative to the other measures, so the presence of even very small statistical uncertainty in the simulated value of $\mathscr{U}^{QWS}$ will lead to overlapping values at $F \in [0.45, 0.9]$.

6.1.2 One-Sided vs. Two-Sided Contracts

Implementing two-sided contracts, where health care providers are penalized for overspending, is a primary goal of policy-makers involved with the MSSP. These simulations indicate why: ACOs are predicted to save substantially more when facing downside risk. Specifically, total program savings would increase by over eight billion dollars in the estimation sample if penalties were imposed on all providers. Per-beneficiary spending decreases by \$305.25 on average, which is a 0.85 standard deviation decrease. The savings-quality tradeoff manifests as a decrease in quality score of just 0.0115, which is a decrease in 0.12 standard deviations.

Our concerns about quality of care dropping in response to two-sided incentives are further ameliorated by comparing the levels of quality-weighted savings. Overall, quality-weighted total program savings for one-sided incentives is higher than that for two-sided incentives only when the sharing rate is unreasonably low ($F < 0.05$). At the per-beneficiary level, two-sided incentives strictly dominate one-sided incentives.

6.1.3 Policy Conclusions from Counterfactual Simulations

The policy conclusions regarding sharing rates and ACO penalties from these simulations are clear: one-sided ACO contracts are optimally more generous than their two-sided counterparts. This is because the presence of penalties in two-sided contracts shifts the behavior of lower performing ACOs to higher levels, and thus requiring less overall positive compensation. Furthermore, one-sided contracts are inherently limited in the cost savings it can generate compared to two-sided contracts.

While I find evidence of a cost-quality tradeoff made by providers, the magnitude of the tradeoff is not large enough to forgo the efficiency gains from penalizing overspending providers. In terms of policy, the results of these simulations indicate that quality of care should be measured and incentivized in physician incentive programs, but the incentives need not be strongly powered. Ultimately, competitive market forces and/or government oversight likely assure quality of care of these providers is satisfactory without careful, formal consideration in contract definition.

6.2 Computing Losses to Free-Riding in ACOs

Equilibrium ACO savings rates S_j^* and quality scores Q_j^* are the aggregated equilibrium strategies of ACO participants playing a game with incentive dilution, and therefore free-riding. We can solve for perfectly-coordinating levels of S_j and Q_j, where free-riding is absent, by maximizing the total surplus of the ACO participants:

$$\max_{\mathbf{s}_j, \mathbf{q}_j} E_j^P(S, Q; F, T) - \sum_{i \in I_j} c_{ij}(s_{ij}, q_{ij}) \, . \tag{66}$$

After taking first order conditions of this objective and performing a series of algebraic operations (similar to the operations in Equations 22 through 27), I arrive at the marginal benefit formulas

$$\hat{MB}_j^{S,PC}(S',Q';\hat{\sigma}_S,\hat{\sigma}_Q) = W_j B_j \left[F\hat{E}_j^Q(Q';\hat{\sigma}_Q)\frac{\partial \hat{E}_j^S}{\partial s_{ij}}(S';\hat{\sigma}_S) + T(1-FQ')\frac{\partial \hat{E}_j^S}{\partial s_{ij}}(S';\hat{\sigma}_S) \right] \tag{67}$$

$$\hat{MB}_j^{Q,PC}(S',Q';\hat{\sigma}_S,\hat{\sigma}_Q) = W_j B_j \left[F\hat{E}_j^S(S';\hat{\sigma}_S)\frac{\partial \hat{E}_j^Q}{\partial q_{ij}}(Q';\hat{\sigma}_Q) + TF\hat{E}_j^S(S';\hat{\sigma}_S) \right], \tag{68}$$

which is analogous to Step 3 above. The counterfactual savings rates and quality scores without free-riding are the values $\hat{S}_j^{PC}(F,T) = S'$ and $\hat{Q}_j^{PC}(F,T) = Q'$ that solve the system

$$\hat{MC}_j^S\left(S',Q';\hat{\boldsymbol{\theta}}\right) = \hat{MB}_j^{S,PC}(S',Q';\hat{\sigma}_S,\hat{\sigma}_Q) \tag{69}$$

$$\hat{MC}_j^Q\left(S',Q';\hat{\boldsymbol{\theta}}\right) = \hat{MB}_j^{Q,PC}(S',Q';\hat{\sigma}_S,\hat{\sigma}_Q), \tag{70}$$

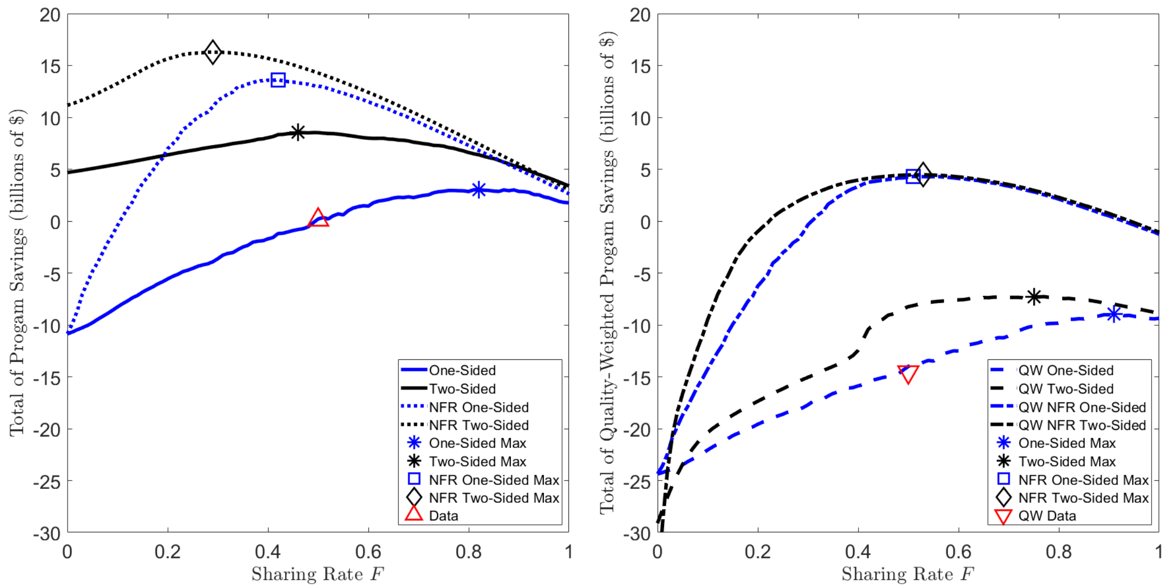
which is analogous to Step 4 above. Note that the lone difference between $\hat{S}_j(F,T)$ and $\hat{S}_j^{PC}(F,T)$ and between $\hat{Q}_j(F,T)$ and $\hat{Q}_j^{PC}(F,T)$ is that the latter values are chosen with a weighted marginal benefit equations that have first terms W_j, and the former values are chosen where the first terms are $W_j^{(3)}$. Therefore, the amount of incentive dilution and free-riding depends on the ratio $W_j/W_j^{(3)} \approx \#\ \text{participants}_j$. For that reason, this measurement is based heavily on our assumption that participants split earnings according to the influence weights w_{ij}. In theory, an ACO could impose sharing contracts that mitigate free-riding and incentive dilution. If this is the case, then the counterfactual estimates of performance loss from free-riding in this section are overstated. While there is very little evidence in ACO Participant Agreements that ACO earnings are distributed among participants in ways that are robust to strategic decision-making, a conservative interpretation of these estimates of the consequences of free-riding in ACOs is as upper bounds.

Figure 4 shows Medicare utility measures $\mathcal{U}^{TPS}$ and $\mathcal{U}^{QWS}$, and Figure 5 shows $\mathcal{U}^{SPB}$ and $\mathcal{U}^{QWB}$. Qualitatively, examining per-beneficiary savings and total savings offers the same conclusions. First, while there are substantial increases in all measures when imposing two-sided incentives, the increase is far smaller when there is no free-riding. This is because the absence of free-riding effectively increases the efficiency of a specific provider by a factor of $\frac{1}{w_{ij}}$, making any given ACO less likely to be impacted by penalties in equilibrium. Similarly, the increase in efficiency also weakens the optimal contract. Specifically, this counterfactual provides evidence that contracts incentivizing a certain level of *group* performance must necessarily be higher-powered than ones incentivizing the same level of performance at the individual level.

6.3 Addressing Simplifications and Assumptions

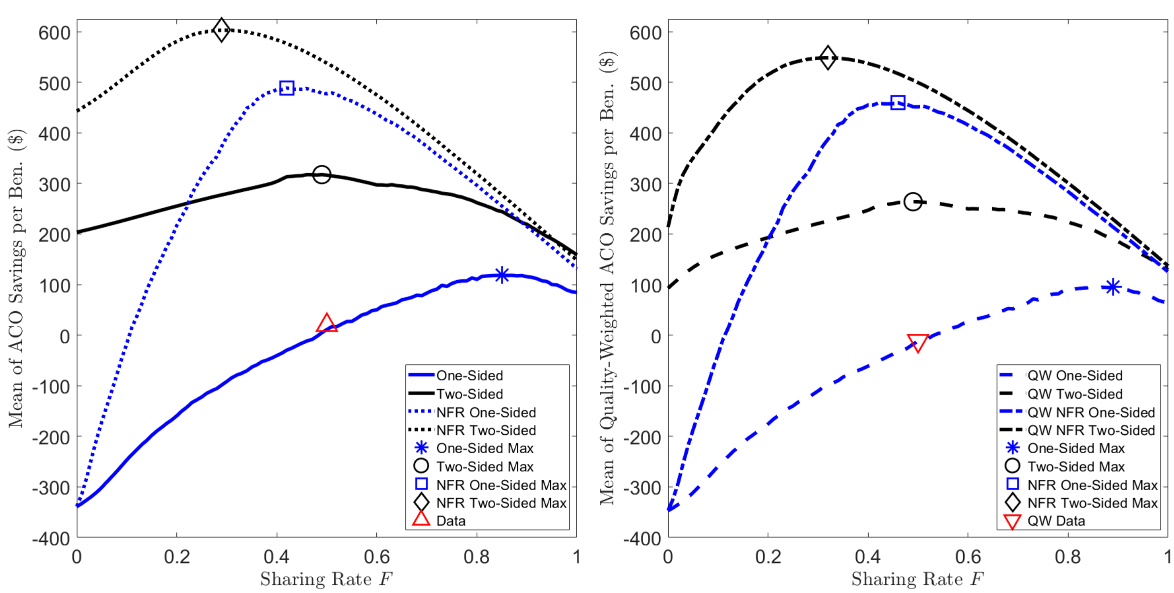
In Sections 3 and 4, I make a variety of assumptions and simplifications that may play a role in determining results. Here, I will briefly discuss these decisions, why they were made, and how changing them may change the results.

Figure 4: **Simulated Total Program Savings without Free-Riding**



Note: This figure plots simulated total program savings (left panel) and quality-weighted total program savings (right panel) with and without free-riding in ACOs. Simulations vary MSSP contracts by Sharing Rate (horizontal axis) and presence of penalties (blue line is without penalties, black line is with). In the legend, "QW" means "quality-weighted" and "NFR" means "no free-riding."

Figure 5: **Simulated Average Savings per Beneficiary without Free-Riding**



Note: This figure plots simulated average per-beneficiary savings (left panel) and simulated average quality-weighted per-beneficiary savings (right panel) with and without free-riding in ACOs. Simulations vary MSSP contracts by Sharing Rate (horizontal axis) and presence of penalties (blue line is without penalties, black line is with). In the legend, "QW" means "quality-weighted" and "NFR" means "no free-riding."

ACO Participation. An obvious omission from the contracting problem is a *participation constraint* which would require the contract specified by Medicare to be sufficiently generous such that the ACOs agree to join the MSSP in the first place. I omit incorporating a participation constraint into this model for several reasons. First, it's not clear if a binding participation is a desirable requirement for ACOs: in some cases, an ACO with participants that have difficulty reducing cost could exit the MSSP to increase savings. Second, the participation constraint places a large restriction on possible values of F and T that maximize CMS's objective function. Specifically, values of the sharing rate F that solve the optimal contracting problems would necessarily be close to one in order to get the least efficient ACO to not exit. Finally, as I mentioned in Section 3, extending the model of multitasking providers to endogenize ACO formation and participation, and estimating that expanded model, is not possible with the available data on ACOs.

ACO Participant Agreements and Distributing ACO Earnings. In Section 3, I assume that each participant earns a share of total ACO earnings equal to their influence weight, w_{ij}. While this likely approximates the existing sharing structure for many ACOs, it rules out situations where ACO management imposes formal sharing rules that incentivize different choices. In other words, splitting ACO earnings by influence share is only slightly more efficient than splitting earnings evenly, and ACOs might be more sophisticated than that. Moreover, if an ACO keeps a share of subsidy pay for reinvestment into operations, then ACO choices may be skewed. The impact of this assumption on estimation depends on the difference between incentive dilution factor $W_j^{(3)}$, which determines the extent of free-riding in ACOs, and the "true" amount of incentive dilution at j. If this difference is correlated with lagged ACO savings rates and quality scores, then estimates of δ_S, δ_Q, and κ would be different than the estimates in this paper.

Risk Aversion and Loss Aversion. Risk aversion and loss aversion likely play little roles in ACO performance. As mentioned, while the MSSP is a very large program, its average per-participant payment is less than \$200,000 per year. Given estimated values of σ_S and σ_Q, this small amount (relative to the regular earnings of these businesses) does not vary significantly enough to meaningfully drive decisions. Loss aversion only impacts performance for two-sided contracts, where it would further incentivize savings at the cost of quality of care. Again, since penalties are still low per-participant, the differences in performance with loss aversion would likely be small in magnitude. By omitting risk and loss aversion from the model, I preserve the linearity of the first order conditions of the participant's problems that permits aggregation, and therefore identification, estimation, and simulation.

Effort Cost Functional Form. Because the derivatives are linear, a quadratic form for effort cost is the highest-order parametric function that can be estimated with aggregate data. This omits the possibility of higher order cost terms. In simulations, this would mean the model is restricted in its ability to predict data variation at the extremes of savings rates and quality scores. Thus, using this functional form is a question of model fit, discussed in Section 5.

To see again that cost function specification is a matter of model fit, consider the following. Assume marginal cost of savings effort $\frac{\partial c_{ij}}{\partial s_{ij}}$ is defined such that the weighted average $\sum_{i \in I_j} w_{ij} \frac{\partial c_{ij}}{\partial s_{ij}}$ is a function of weighted average savings S_j^* and weighted average quality Q_j^*. Then, a first-order Taylor-series approximation of weighted average marginal cost would be identical to the linear function $MC_j^S(\boldsymbol{\theta})$. Including higher-order terms of S_j^* and Q_j^* would help fit, if needed, by being a closer approximation to the true effort cost function.

7 Conclusion

Incentive design is used by firms, governments, households, educators, and many others to achieve a variety of ends. Accordingly, designing effective incentives is a popular topic of study in all fields of economics. It is particularly important in the United States health care sector, where physician incentive programs and pay-for-performance initiatives impact the quality of life and spending of individuals in 3.5 trillion dollar industry. In this paper, I investigate the empirical role of multitasking in incentive design, specifically in the context of the Medicare Shared Savings Program and Accountable Care Organizations. I identify the cost-savings tradeoff experienced by ACO providers and accordingly design MSSP contracts that maximize the quality-weighted incentive program savings, while accounting for free-riding of health care providers induced by ACO payment.

I estimate a model of multitasking agents to simulate performance in ACOs. I find that provider face a modest tradeoff between increasing savings and increasing quality of care. Counterfactual policy analysis shows that if ACOs are required to pay penalties to Medicare for spending too much, savings increases by up to 1.5 standard deviations, though quality falls by no more than 0.2. The optimal proportion of savings to share with an ACO falls between 0.4 and 0.5 for two-sided contracts, but is higher at 0.85 for one-sided contracts.

Another counterfactual shows performance of ACOs is significantly lower due to free-riding. I argue that up to $500 per beneficiary in savings and one standard deviation of quality of care is lost to free-riding within ACOs. Optimal contracts that account for free-riding need to pay at least double the rate of contracts that don't. This means losses to the principal from free-riding agents comes from both under performance and over-payment relative to performance.

Appendix

Contents

A Institutional Details

This appendix gives detailed background information on the Medicare Shared Savings Program. ACOs began operating in the MSSP in 2012. Nearly any Medicare provider, including individual physicians, group practices, and large hospital systems, can start an ACO and recruit other Medicare providers to participate in their joint venture.[17] Once an ACO shows they have established a governing board that oversees clinical and administrative aspects of operation and shows the presence of formal contracts between itself and its member participants (including the distribution of any earned incentive pay), it then enters into a five year agreement with the Centers for Medicare and Medicaid Services (CMS).[18] Medicare fee-for-service (FFS) beneficiaries are assigned to ACOs by CMS: if a given Medicare beneficiary receives the plurality of primary care services from a primary care provider who is (or is employed by) an ACO participant, that beneficiary is assigned to that participant's ACO.[19]

There are two separate components of assessing ACO performance, and both determine the amount ACOs are paid. The first is an overall quality score, which is a composite score between 0 and 1 of several sub-measures of care quality. These sub-measures fall into the domains of "Patient/Caregiver Experience," "Care Coordination/Patient Safety," "Preventative Health," and "At-Risk Population." Some sub-measures

[17]A participant can be nearly any health care provider that accepts and bills Medicare. Participants are legally defined by their Tax ID Number (TIN) or CMS Certification number (CCN).

[18]Before July 2019, agreements lasted three years.

[19]When a Medicare beneficiary receives the plurality of primary care services from a primary care provider not associated with an ACO, they are not assigned to an ACO. This assignment methodology results in roughly one fifth to one third of all FFS beneficiaries assigned to ACOs each year. An ACO must be assigned at least 5000 beneficiaries to operate and earn shared savings payments.

are survey responses (e.g., "ACO2: How Well Your Doctors Communicate"), while others are computed from Medicare Claims and aggregated to the ACO-level (e.g., "ACO21: Proportion of Adults who had blood pressure screened in past 2 years").[20]

The second component is ACO savings. CMS first establishes an ACO's benchmark expenditure by forecasting per-beneficiary Medicare expenditure for beneficiaries that would have been assigned to the ACO in the three years prior to the agreement period. For performance years after the first, the benchmark is updated based on projected growth of per-beneficiary Medicare expenditure.[21] The savings rate of an ACO in a performance year is then the difference between its benchmark expenditure and the actual expenditure on assigned beneficiaries divided by its benchmark expenditure.

A.1 ACO payment from 2012 until June 2019

For the first six performance years of the MSSP, ACOs had a choice between four payment contracts called "Tracks." The contracts vary in power and exposure to downside risk. Track 1, available to ACOs only in their first six years of operation, is lowered powered and requires no loss sharing with CMS (i.e., it's one-sided). Accordingly, each performance year the shared savings paid by CMS to an ACO on Track 1 is

$$\frac{1}{2} \cdot (\text{Benchmark Expenditure} - \text{Expenditure}) \cdot \text{Quality Score} \tag{71}$$

when an ACO's savings rate meets or exceeds its minimum savings rate and its quality score meets or exceeds quality reporting standards. Otherwise, an ACO earns \$0 in shared savings. For example, consider an ACO with a benchmark expenditure of \$186 million (the average over 2012-2017) and a minimum savings rate of 0.02. If that ACO has an expenditure of \$160 million with a quality score of 0.90, it would earn

$$\frac{1}{2} \cdot (\$186 \text{ million} - \$160 \text{ million}) \cdot 0.90 = \$11.7 \text{ million} \tag{72}$$

in shared savings. Its savings rate is $(18.6 - 16)/16 = 0.1625$, and hence the minimum savings rate is exceeded. Though paying a subsidy, Medicare saves money as well: on net, it saves \$14.3 million, as it paid \$11.7 million to save \$26 million.

Like Track 1, Track 1+ offers ACOs up to 50% of savings as incentive pay. It differs by introducing downside risk, requiring ACOs to pay 30% of losses to Medicare if expenditure is much larger than benchmark expenditure and savings is below the minimum loss rate. Track 2 and Track 3 ACOs face both higher powered incentives and downside risk. Track 2 and Track 3 give 60% and 75% of savings back to ACOs, respectively. If savings is below the minimum loss rate, these ACOs must pay money back to Medicare at a rate of $\left(1 - \frac{3}{5} \cdot \text{Quality Score}\right) \cdot 100\%$ and $\left(1 - \frac{3}{4} \cdot \text{Quality Score}\right) \cdot 100\%$ of losses for Tracks 2 and 3, respectively.

Track 1 has been the overwhelming contract choice of ACOs. In 2013, 2014, and 2015, between 97% and

[20] See `https://go.cms.gov/2xHy7Uo` for a full list of ACO quality scores for every performance year.

[21] Regional adjustments to benchmarks were introduced in 2017 for ACOs in their fourth year of operation.

99% of the 200-400 operating ACOs chose Track 1. In 2016 and 2017, 95% and 92% of the 432 and 472 ACOs operating that year chose Track 1.

A.2 ACO payment from July 2019 until the present

The first six years of the Medicare Shared Savings Program produced modest decreases in Medicare expenditure (McWilliams et al., 2018). In an attempt to improve ACO performance, CMS made several changes to the MSSP with its final rule named "Pathways to Success" (or "Pathways").

Changes in Pathways pertinent to this paper regard the contracts between ACOs and Medicare. Tracks 1, 1+, 2, and 3 are replaced with two Tracks: "Basic" and "Enhanced." Under the Basic Track, there are five levels, "A" through "E." Under levels A and B, ACOs earn up to 40% of savings and do not pay shared losses if expenditure exceeds benchmark expenditure. Under levels C, D, and E, ACOs earn up to 50% of savings and pay an increasing amount of shared losses if expenditure exceeds benchmark expenditure. An ACO is automatically advanced one level (e.g., from level A to B) after each performance year. The Enhanced Track is equivalent to Track 3.

Various other changes were made to the MSSP in Pathways, including beneficiary assignment methodology, benchmark calculation, and assigning new ACO classifications ("low-renvenue" and "experienced") that impact the payment contracts available to an ACO.

In this paper, counterfactual predictions consider two dimensions of contracts: the fraction of savings shared with an ACO and the presence of downside risk. These dimensions broadly account for all previous (Tracks 1, 1+, 2, and 3) and current (Basic and Enhanced Tracks) contract options.

B On Existence and Uniqueness of Equilibrium

This appendix derives conditions under which the multitasking game defined in Section 3 has a unique equilibrium. As conditions are derived, I show that estimated parameters and observed data satisfy the conditions of a unique equilibrium. The intuition of the proof is simple: I show the multitasking game played by ACO participants is a concave game, as in Rosen (1965), if two inequalities of structural parameters and data are satisfied.

Formally, consider the game played by ACO j—the game is defined by a specification of players $i \in I_j$, strategy sets $s_{ij} \in [-1, 1]$ and $q_{ij} \in [0, 1]$, and payoff functions

$$\pi_{ij}\left(s_{ij}, q_{ij}, \mathbf{s}_{-ij}, \mathbf{q}_{-ij}\right) = w_{ij} E_j^P\left(S_j, Q_j; F_j, T_j\right) - c_{ij}\left(s_{ij}, q_{ij}\right) \tag{73}$$

for all $i \in I_j$, where

$$E_j^P\left(S_j, Q_j; F_j, T_j\right) = F_j \cdot B_j \cdot E_j^S(S_j) \cdot E_j^Q(Q_j) - T_j \cdot (1 - F_j \cdot Q_j) \cdot B_j \cdot E_j^S(-S_j) \tag{74}$$

and

$$E_j^S(S_j) = S_j \Phi\left(\frac{S_j - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) + \sqrt{W_j}\sigma_S \phi\left(\frac{S_j - \underline{S}_j}{\sqrt{W_j}\sigma_S}\right) \tag{75}$$

$$E_j^Q(Q_j) = Q_j \Phi\left(\frac{Q_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) + \sqrt{W_j}\sigma_Q \phi\left(\frac{Q_j - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right). \tag{76}$$

Our objective is to show that given $\mathbf{s}_{-ij}$ and $\mathbf{q}_{-ij}$, participant $i \in I_j$ has a unique solution to maximizing $\pi_{ij}(\cdot)$. The gradient vector $\nabla\pi_{ij}$ and Hessian matrix $\mathbf{H}_{\pi_{ij}}$ have elements

$$(\nabla\pi_{ij})_1 = w_{ij}^2 B_j \left[F_j \frac{\partial E_j^S}{\partial s_{ij}}(S_j) E_j^Q(Q_j) + T_j(1 - F_j Q_j)\frac{\partial E_j^S}{\partial s_{ij}}(-S_j)\right] - \frac{\partial c_{ij}}{\partial s_{ij}}(s_{ij}, q_{ij}) \tag{77}$$

$$(\nabla\pi_{ij})_2 = w_{ij}^2 B_j \left[F_j E_j^S(S_j)\frac{\partial E_j^Q}{\partial q_{ij}}(Q_j) + T_j F_j E_j^S(-S_j)\right] - \frac{\partial c_{ij}}{\partial q_{ij}}(s_{ij}, q_{ij}) \tag{78}$$

$$\left(\mathbf{H}_{\pi_{ij}}\right)_{11} = w_{ij}^3 B_j \left[F_j \frac{\partial^2 E_j^S}{\partial s_{ij}^2}(S_j) E_j^Q(Q_j) - T_j(1 - F_j Q_j)\frac{\partial^2 E_j^S}{\partial s_{ij}^2}(-S_j)\right] - \frac{\partial^2 c_{ij}}{\partial s_{ij}^2}(s_{ij}, q_{ij}) \tag{79}$$

$$\left(\mathbf{H}_{\pi_{ij}}\right)_{12} = \left(\mathbf{H}_{\pi_{ij}}\right)_{21} = w_{ij}^3 B_j \left[F_j \frac{\partial E_j^S}{\partial s_{ij}}(S_j)\frac{\partial E_j^Q}{\partial q_{ij}}(Q_j) - T_j F_j \frac{\partial^2 E_j^S}{\partial s_{ij}^2}(-S_j)\right] - \frac{\partial^2 c_{ij}}{\partial s_{ij} q_{ij}}(s_{ij}, q_{ij}) \tag{80}$$

$$\left(\mathbf{H}_{\pi_{ij}}\right)_{22} = w_{ij}^3 B_j F_j E_j^S(S_j)\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) - \frac{\partial^2 c_{ij}}{\partial q_{ij}^2}(s_{ij}, q_{ij}). \tag{81}$$

Therefore, global concavity of π_{ij} and therefore uniqueness of maximizing values of s_{ij} and q_{ij} is established if

$$\left(\mathbf{H}_{\pi_{ij}}\right)_{11}\big|_{S_j = S_j^*, Q_j = Q_j^*} < 0 \qquad \text{or} \qquad \left(\mathbf{H}_{\pi_{ij}}\right)_{22}\big|_{S_j = S_j^*, Q_j = Q_j^*} < 0 \tag{82}$$

$$\text{and} \tag{83}$$

$$\det \mathbf{H}_{\pi_{ij}} > 0\,. \tag{84}$$

The values of each Hessian entry are merely linear combinations of variables and the functions E_j^S, E_j^Q, and the derivatives of these functions. I assume in estimation that the second order derivatives of c_{ij} are constants: these are the estimated parameters δ_S, δ_Q, and κ.

First, it is straightforward to show that the second order derivative of π_{ij} with respect to q_{ij} is always negative. I will show

$$w_{ij}^3 B_j F_j E_j^S(S_j)\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) < \delta_Q\,. \tag{85}$$

Note that E_j^S and E_j^Q are the "smoothed" versions of the perfect certainty ($\sigma_S, \sigma_Q \to 0$) functions $S_j \cdot \mathbb{1}\{S_j \geq \underline{S}_j\}$ and $Q_j \cdot \mathbb{1}\{Q_j \geq \underline{Q}\}$. Values of $E_j^S(S_j)$ when S_j is close to $\underline{S}_j$ are thus biased towards the middle of the range $[0, S_j]$, and we are left with the conditions

$$E_j^S(S_j) > 0 \text{ for all } S_j \in [-1, 1] \tag{86}$$

$$E_j^S(S_j) < S_j \text{ if } S_j \geq \underline{S}_j \tag{87}$$

$$E_j^S(S_j) < \underline{S}_j \text{ if } S_j < \underline{S}_j \tag{88}$$

for S_j (similar conditions for Q_j apply as well). Then I can write

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) < \frac{\delta_Q}{w_{ij}^3 B_j F_j E_j^S(S_j)}\,. \tag{89}$$

$$\tag{90}$$

The largest value taken by F_j is 1, and $w_{ij}^3 < W_j^{(3)} < W_j$, so,

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) < \frac{\delta_Q}{B_j W_j E_j^S(S_j)}\,. \tag{91}$$

$$\tag{92}$$

Finally, we can apply the upper bounds on $E_j^S(S_j)$:

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) < \begin{cases} \frac{\delta_Q}{B_j W_j \underline{S}_j} & \text{if } S_j \in [-1, \underline{S}_j) \\ \frac{\delta_Q}{B_j W_j S_j} & \text{if } S_j \in [\underline{S}_j, 1] \end{cases}\,. \tag{93}$$

Now, I must find the maximum value $\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j)$ can take. Note that

$$\frac{\partial E_j^Q}{\partial q_{ij}}(Q_j^*) = \Phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) + \frac{\underline{Q}}{\sqrt{W_j}\sigma_Q}\phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) \tag{94}$$

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) = \frac{1}{\sqrt{W_j}\sigma_Q}\phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) + \frac{\underline{Q}}{W_j\sigma_Q^2}\phi'\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) \tag{95}$$

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) = \frac{1}{\sqrt{W_j}\sigma_Q}\phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) - \frac{\underline{Q}}{W_j\sigma_Q^2}\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right)\phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) \tag{96}$$

where the last line follows from the property of normal variables that $\phi'(z) = -z\phi(z)$. Then,

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) = \left[\frac{1}{\sqrt{W_j}\sigma_Q} - \frac{\underline{Q}}{W_j\sigma_Q^2}\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right)\right]\phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) \tag{97}$$

The term in brackets is strictly decreasing in Q_j, so $\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j)$ is maximized at a value $Q_j < \underline{Q}$. Moreover, since the term in brackets takes its lowest value when $Q_j = 0$ and $\underline{Q} = 1$, then

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) \leq \left[\frac{1}{\sqrt{W_j}\sigma_Q} + \frac{1}{W_j^{3/2}\sigma_Q^3}\right] \phi\left(\frac{Q_j^* - \underline{Q}}{\sqrt{W_j}\sigma_Q}\right) \tag{98}$$

Finally, note that the maximum value of $\phi(\cdot)$ is when the input is zero, so

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) \leq \left[\frac{1}{\sqrt{W_j}\sigma_Q} + \frac{1}{W_j^{3/2}\sigma_Q^3}\right] \frac{1}{\sqrt{2\pi}\sqrt{W_j}\sigma_Q}. \tag{99}$$

Equivalently,

$$\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j) \leq \left[1 + \frac{1}{W_j\sigma_Q^2}\right] \frac{1}{\sqrt{2\pi}W_j\sigma_Q^2}. \tag{100}$$

Combining the upper bound on $\frac{\partial^2 E_j^Q}{\partial q_{ij}^2}(Q_j)$ with our condition for concavity we get

$$\left(1 + \frac{1}{W_j\sigma_Q^2}\right) \frac{1}{\sqrt{2\pi}\sigma_Q^2} \leq \begin{cases} \frac{\delta_Q}{B_j\underline{S}_j} & \text{if } S_j \in [-1, \underline{S}_j) \\ \frac{\delta_Q}{B_j S_j} & \text{if } S_j \in [\underline{S}_j, 1] \end{cases}. \tag{101}$$

This condition is rather intuitive: larger values of the variance of $\mathring{Q}_j$, $W_j\sigma_Q^2$, as well as values of $\underline{Q}$ close to zero, are necessary for uniqueness. The intuition is that the game with $\sigma_Q \to 0$ would have a discontinuous payoff at $\underline{Q}$, and so global concavity could would not be possible by definition.

Figure 6 is a histogram of the inverse hyperbolic sine[22] of the second order condition lower-bound

$$H_{22}^{LB} = \frac{\hat{\delta}_Q}{B_j\underline{S}_j}\mathbb{1}\left\{S_j^* < \underline{S}_j\right\} + \frac{\hat{\delta}_Q}{B_j S_j^*}\mathbb{1}\left\{S_j^* \geq \underline{S}_j\right\} - \left(1 + \frac{1}{W_j\hat{\sigma}_Q^2}\right)\frac{1}{\sqrt{2\pi}\hat{\sigma}_Q^2}. \tag{102}$$

where $\hat{\cdot}$ is the estimate of $\cdot$. For all observations, $H_{22}^{LB} > 0$, so we can conclude $\left(\mathbf{H}_{\pi_{ij}}\right)_{22}\big|_{S_j = S_j^*, Q_j = Q_j^*} < 0$.

The other second order condition to confirm is that the determinant the the Hessian of π_{ij} satisfies

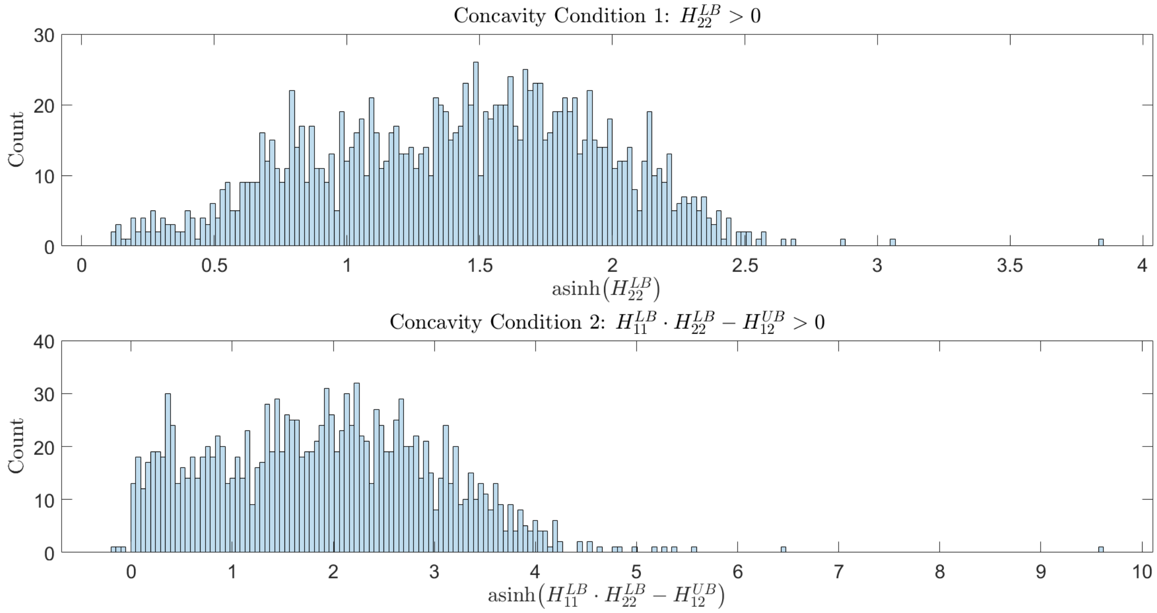
$$\det \mathbf{H}_{\pi_{ij}} = \left(\mathbf{H}_{\pi_{ij}}\right)_{11} \cdot \left(\mathbf{H}_{\pi_{ij}}\right)_{22} - \left(\mathbf{H}_{\pi_{ij}}\right)_{12}^2 > 0 \tag{103}$$

Note that we can similarly find lower bound on the second order derivative of π_{ij} with respect to s_{ij} as

$$H_{11}^{LB} = \frac{\hat{\delta}_S}{B_j\underline{Q}}\mathbb{1}\left\{Q_j^* < \underline{Q}\right\} + \frac{\hat{\delta}_S}{B_j Q_j^*}\mathbb{1}\left\{Q_j^* \geq \underline{Q}\right\} - \left(1 + \frac{\underline{S}_j(S_j^* - \underline{S}_j)}{W_j\hat{\sigma}_S^2}\right)\frac{1}{\sqrt{2\pi}\hat{\sigma}_S^2}. \tag{104}$$

[22] The inverse hyperbolic sine function is defined as $\text{asinh}(z) = \ln\left(z + \sqrt{z^2+1}\right)$. Like other concave-increasing functions such as $\ln(z)$ and $\sqrt{z}$, $\text{asinh}(z)$ decreases the skew of z (in this case, only for visualization purposes). It's advantage is that it has a domain over the entire real line, and $\text{asinh}(0) = 0$.

Figure 6: **Histograms of Second Order Condition Bounds for Participant Payoff Maximization Problem**



Note: This figure contains histograms displaying the concavity conditions for the multitasking game defined in Appendix B. If an observation for ACO j satisfies the conditions of both panels, then a unique equilibrium exists.

Now, we must find an upper bound on $\left(\mathbf{H}_{\pi_{ij}}\right)_{12}^{2}$. This value is zero when κ is close to the cross partial derivative of $w_{ij}E_j^P(\cdot)$. Other extreme values are when the same cross partial is zero, and $\left(\mathbf{H}_{\pi_{ij}}\right)_{12}^{2} = \kappa^2$, or when the cross partial takes its maximum value. In other words,

$$\left(\mathbf{H}_{\pi_{ij}}\right)_{12}^{2} = \left[w_{ij}^3 B_j \left[F_j \frac{\partial E_j^S}{\partial s_{ij}}(S_j) \frac{\partial E_j^Q}{\partial q_{ij}}(Q_j) - T_j F_j \frac{\partial^2 E_j^S}{\partial s_{ij}^2}(-S_j)\right] - \kappa\right]^2 \tag{105}$$

$$\leq \max\left\{\left(W_j B_j S_j Q_j - \kappa\right)^2, \left(-W_j B_j - \kappa\right)^2, \kappa^2\right\} \tag{106}$$

Defining

$$H_{12}^{UB} = \max\left\{\left(W_j B_j S_j^* Q_j^* \hat{\kappa}\right)^2, \left(-W_j B_j - \hat{\kappa}\right)^2, \hat{\kappa}^2\right\} \tag{107}$$

I can finally confirm concavity by showing the histogram of the inverse hyperbolic sine of

$$H_{11}^{LB} \cdot H_{22}^{LB} - H_{12}^{UB} \tag{108}$$

also in in Figure 6. Virtually all ($\sim 99.8\%$ of observations) satisfy this condition as well, and the those that do not are very close to the bound of 0. Because the derived bounds are only sufficient and not necessary, those observations could very well still satisfy a less strict sufficient condition.

C Data Appendix

C.1 Computing Influence Weights

I've defined influence weights $\{w_{ij}\}_{i \in I_j}$ such that

$$\sum_{i \in I_j} w_{ij} s_{ij} = S_j \qquad \sum_{i \in I_j} w_{ij} q_{ij} = Q_j \tag{109}$$

where $\sum_{i \in I_j} w_{ij} = 1$. Note that for participant savings efforts s_{ij} have to have a definition analogous to that of S_j, we would have

$$S_j = \frac{BE_j - E_j}{BE_j} = \frac{\sum_{i \in I_j} BE_{ij} - \sum_{i \in I_j} E_{ij}}{\sum_{i \in I_j} BE_{ij}} = \sum_{i \in I_j} w_{ij} \frac{BE_{ij} - E_{ij}}{BE_{ij}} = \sum_{i \in I_j} w_{ij} s_{ij} \tag{110}$$

where BE_j and E_j are the benchmark expenditure and expenditure of ACO j (both real quantities observed in data) and BE_{ij} and E_{ij} are the benchmark expenditure and expenditure of participant i in ACO j (both theoretical quantities). Thus, a definition of w_{ij} consistent with the above is $w_{ij} = \frac{BE_{ij}}{BE_j}$, or simply participant i's share of ACO benchmark expenditure. Intuitively, this means that a very influential participant i in ACO j will have a relatively large share of expected expenditure on assigned beneficiaries.

In data, I measure w_{ij} as shares of expenditure for each *type* of provider within an ACO. To be specific, suppose provider i has type τ. Then,

$$w_{\{i=\tau\}j} = \frac{\text{Total Spending by type } \tau}{(\text{Total \# of } i \text{ with type } \tau) \cdot E_j}. \tag{111}$$

This measure of w_{ij} has two implicit assumptions. First, it assumes that providers of the same type have similar shares of overall expenditure within an ACO. This is likely the case, since ACOs tend to be predominantly hospital based or group practice based. Second, this measure requires that the *ratio* BE_{ij}/BE_j is close to the ratio E_{ij}/E_j, since w_{ij} as defined in Equation 111 is actually the latter ratio.

C.2 Elements of X_j

Table 5 lists all the variables used for estimation in X_j.

C.3 Model Estimation

To estimate $\boldsymbol{\theta}$, σ_S, and σ_Q, I apply Assumptions 4.1, 4.2, and 4.3 and minimize a GMM objective function with X_j and lagged savings rates and quality scores as instruments. While the function $\boldsymbol{\Xi}$ is nonlinear in parameters, the marginal cost parts MC_j^S and MC_j^Q are linear in $\boldsymbol{\theta}$. This means I can accelerate estimation by nesting the estimation of $\boldsymbol{\theta}$ for a given σ_S and σ_Q within the estimation of σ_S and σ_Q. The drawback of

Table 5: **Elements of X_j**

Abbreviated Variable Name	**Description**
# states	Number of states where beneficiaries assigned to the ACO reside.
# beneficiaries	Number of beneficiaries assigned to the ACO in thousands.
average risk score	Average CMS HCC risk score of aged, non-dual beneficiaries assigned to the ACO.
% over 75	Percent of assigned beneficiaries over age 75.
% male	Percent of assigned beneficiaries that are male.
% nonwhite	Percent of assigned beneficaries that are non-white.
# providers	Total number of individual providers in an ACO in thousands.
fraction PCP	Proportion of individual providers that are primary care physicians.
all group	Indicates every participant in ACO is a group practice or hospital.
fraction inpatient	Proportion of expenditures that are inpatient expenditures (includes short term, long term, rehabilitation, and psychiatric).
fraction outpatient	Proportion of expenditures that are outpatient expenditures.
# PC services	Total number of primary care services in thousands.
# admissions	Total number of inpatient hospital discharges in thousands.
fraction PC served by PCP	Proportion of primary care services provided by primary care physician.
age	Number of performance-years in MSSP of ACO.

This table shows control variables used for estimation of marginal cost parameters. Not listed: constant term, fixed-effects for year and census division.

nesting estimation in this manner is that I must use bootstrapping to obtain standard errors for parameter estimates.

The following steps explain the estimation procedure:

1. Define the matrices

$$\mathbf{X} = \begin{bmatrix} S_j^* & Q_j^* & 0 & X_j & \mathbf{0} \\ 0 & S_j^* & Q_j^* & \mathbf{0} & X_j \end{bmatrix}_{j \in \mathcal{J}} \tag{112}$$

$$\mathbf{Z} = \begin{bmatrix} S_{j,PY-1}^* & Q_{j,PY-1}^* & 0 & X_j & \mathbf{0} \\ 0 & S_{j,PY-1}^* & Q_{j,PY-1}^* & \mathbf{0} & X_j \end{bmatrix}_{j \in \mathcal{J}} . \tag{113}$$

Note that each matrix has dimension $2|\mathcal{J}| \times (2k+3)$.

2. For candidate values of $\hat{\sigma}_S$ and $\hat{\sigma}_Q$, the unexplained variation in marginal cost is given by the parameterized vector-valued function $\mathbf{\Xi}\left(\hat{\sigma}_S, \hat{\sigma}_Q; \hat{\boldsymbol{\theta}}\right)$. This function is defined as follows:

    (a) Using Equations 33 and 34, define the $2|\mathcal{J}| \times 1$ vector

    $$\mathbf{MB}\left(\hat{\sigma}_S, \hat{\sigma}_Q\right) = \begin{bmatrix} MB_j^S\left(\hat{\sigma}_S, \hat{\sigma}_Q\right) \\ MB_j^Q\left(\hat{\sigma}_S, \hat{\sigma}_Q\right) \end{bmatrix}_{j \in \mathcal{J}} . \tag{114}$$

    (b) Solve the linear simultaneous equation model:

    $$\hat{\boldsymbol{\theta}}(\hat{\sigma}_S, \hat{\sigma}_Q) = \left[\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}\left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{MB}\left(\hat{\sigma}_S, \hat{\sigma}_Q\right)\right) . \tag{115}$$

    (c) Define the $2|\mathcal{J}| \times 1$ estimate error vector

    $$\mathbf{\Xi}\left(\hat{\sigma}_S, \hat{\sigma}_Q; \hat{\boldsymbol{\theta}}\right) = \begin{bmatrix} MB_j^S\left(\hat{\sigma}_S, \hat{\sigma}_Q\right) - MC_j^S\left(\hat{\boldsymbol{\theta}}\right) \\ MB_j^Q\left(\hat{\sigma}_S, \hat{\sigma}_Q\right) - MC_j^Q\left(\hat{\boldsymbol{\theta}}\right) \end{bmatrix}_{j \in \mathcal{J}} . \tag{116}$$

3. Find the $\hat{\sigma}_S^*$ and $\hat{\sigma}_Q^*$ that solve the non-linear optimization problem

$$\left(\hat{\sigma}_S^*, \hat{\sigma}_Q^*\right) = \arg\min_{\sigma_S, \sigma_Q} \mathbf{\Xi}\left(\sigma_S, \sigma_Q; \hat{\boldsymbol{\theta}}\left(\sigma_S, \sigma_Q\right)\right)' \mathbf{\Xi}\left(\sigma_S, \sigma_Q; \hat{\boldsymbol{\theta}}\left(\sigma_S, \sigma_Q\right)\right) . \tag{117}$$

4. The final estimates are $\hat{\sigma}_S^*$, $\hat{\sigma}_Q^*$, and $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}\left(\hat{\sigma}_S^*, \hat{\sigma}_Q^*\right)$.

D Additional Estimation and Simulation Results

Recall that the objects of estimation are the cost function parameters δ_S, δ_Q, κ, $\boldsymbol{\gamma}_S$, and $\boldsymbol{\gamma}_Q$ and the uncertainty parameters σ_S and σ_Q where

$$c_{ij}(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \left(\boldsymbol{\gamma}_S' x_{ij} + \xi_{ij}^S\right)s + \left(\boldsymbol{\gamma}_Q' x_{ij} + \xi_{ij}^Q\right)q + \kappa sq \tag{118}$$

$$\mathring{S}_j \sim N\left(S_j, W_j\sigma_S^2\right) \tag{119}$$

$$\mathring{Q}_j \sim N\left(Q_j, W_j\sigma_Q^2\right). \tag{120}$$

Tables 6, 7, and 8 show the estimation results, where Table 6 contains the nonlinear parameters, and Tables 7 and 8 contain parameters that form a linear map from observed ACO characteristics and marginal cost.[23]

Table 6: **Cost Function and Uncertainty Parameter Estimates**

Parameter	Estimate	Standard Error	P-value	95% CI Lower Bound	95% CI Upper Bound
δ_S	185.400	71.225	0.009	70.289	345.420
δ_Q	8.820	1.637	0.000	6.062	12.513
κ	18.037	8.318	0.030	3.157	35.403
σ_S	0.160	0.010	0.000	0.150	0.190
σ_Q	0.420	0.030	0.000	0.340	0.430

Note: $|\mathcal{J}| = 1486$ observations. This table displays parameter estimates for the effort cost function $c(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \boldsymbol{\gamma}_S s + \boldsymbol{\gamma}_Q q + \kappa sq$ and for the standard deviations of the effort-shocks η_{ij}^S and η_{ij}^S. Standard errors, p-values, and confidence interval (CI) bounds are obtained by bootstrapping with 42,327 rep. Parameter estimates not shown here, including year and Census Division fixed-effects, are available in Tables 7 and 8 in Appendix C.

As a robustness check, I also estimate a restricted version of the model where $\sigma_S = \sigma_Q = 0$. The expected payoff function in this environment is then simply the same as the actual payoff function $P_j(S_j, Q_j; T_j; F_j)$. Estimation is the same as is for the unrestricted model, however weighted-average marginal benefit of savings and quality, MB_j^S and MB_j^Q, now have the simple multiplicative forms

$$MB_j^S = W_j^{(3)} B_j \left(F_j Q_j^* \mathbb{1}\{S_j^* \geq \underline{S}_j\}\mathbb{1}\{Q_j^* \geq \underline{Q}\} + T_j(1 - F_j Q_j^*)\mathbb{1}\{S_j^* \leq -\underline{S}_j\}\right) \tag{121}$$

$$MB_j^Q = W_j^{(3)} B_j \left(F_j S_j^* \mathbb{1}\{S_j^* \geq \underline{S}_j\}\mathbb{1}\{Q_j^* \geq \underline{Q}\} + T_j F_j S_j^* \mathbb{1}\{S_j^* \leq -\underline{S}_j\}\right) \tag{122}$$

and the cost function stays the same. Note that while I can estimate parameters of this model and compare them with the unrestricted model, we cannot easily compare (or even compute) differences in counterfactual simulations. The reason is that the multitasking game played by ACO participants can have several equilibria when there is no uncertainty, even if payoff functions are strictly concave. The problem lays with discontinuity

[23] Table 2 in Section 5 is the same as Table 6 in Appendix D.

Table 7: **Cost Function Parameter Estimates: Marginal Cost of Savings**

	Parameter	Estimate	Standard Error	P-value	95% CI Lower Bound	95% CI Upper Bound
γ_S	Constant	-32.836	49.478	0.507	-132.230	61.895
	# states	3.728	2.096	0.075	-1.006	7.271
	# beneficiaries	0.655	0.190	0.001	0.344	1.087
	risk	119.520	45.133	0.008	15.679	192.320
	% over 75	-0.108	0.296	0.715	-0.663	0.498
	% nonwhite	-0.034	0.091	0.711	-0.196	0.159
	% pctmale	1.085	1.205	0.368	-0.979	3.723
	# providers	-3.439	2.246	0.126	-8.388	0.457
	fraction PCP	-9.894	8.286	0.232	-26.097	6.512
	fraction inpatient	-237.200	76.100	0.002	-384.830	-87.364
	fraction outpatient	-191.600	45.121	0.000	-284.320	-107.850
	PC services	-3.414	1.731	0.049	-6.656	0.121
	# admissions	-37.071	49.581	0.455	-138.190	52.012
	fraction PC served by PCP	34.555	14.154	0.015	5.779	61.537
	age	-0.199	0.044	0.000	-0.306	-0.134
	all group	50.435	5.706	0.000	39.414	61.739
	2015	-1.168	4.314	0.787	-9.944	6.981
	2016	3.256	4.462	0.466	-6.172	11.229
	2017	19.323	5.756	0.001	6.759	29.320
	division 2	-3.850	8.586	0.654	-21.624	11.999
	division 3	-10.038	7.848	0.201	-26.645	4.088
	division 4	-27.593	7.898	0.000	-43.925	-13.206
	division 5	-10.475	8.038	0.192	-27.204	4.193
	division 6	-28.107	7.966	0.000	-44.745	-13.611
	division 7	-15.302	7.745	0.048	-31.549	-1.132
	division 8	-4.244	9.625	0.659	-24.349	13.365
	division 9	-8.652	7.828	0.269	-25.233	5.301

Note: $|\mathcal{J}| = 1486$ observations. This table displays parameter estimates for the effort cost function $c(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \gamma_S s + \gamma_Q q + \kappa sq$. Standard errors, p-values, and confidence interval (CI) bounds are obtained by bootstrapping with 42,327 rep.

Table 8: **Cost Function Parameter Estimates: Marginal Cost of Savings**

	Parameter	Estimate	Standard Error	P-value	95% CI Lower Bound	95% CI Upper Bound
	Constant	-11.222	3.554	0.002	-18.93	-5.072
	# states	0.299	0.191	0.119	-0.082	0.646
	# beneficiaries	0.018	0.013	0.176	-0.001	0.048
	risk	7.922	3.909	0.043	0.245	15.091
	% over 75	0.012	0.019	0.518	-0.021	0.052
	% nonwhite	0.009	0.006	0.104	-0.001	0.021
	% pctmale	0.078	0.083	0.352	-0.062	0.264
	# providers	-0.123	0.141	0.382	-0.457	0.089
	fraction PCP	0.172	0.49	0.726	-0.752	1.167
	fraction inpatient	-9.995	4.777	0.036	-20.4	-1.726
	fraction outpatient	-8.631	2.242	0	-13.385	-4.636
	PC services	-0.313	0.101	0.002	-0.511	-0.119
	# admissions	-2.643	2.605	0.31	-7.362	2.726
γ_Q	fraction PC served by PCP	0.708	0.765	0.355	-0.827	2.179
	age	-0.007	0.003	0.012	-0.013	-0.003
	all group	1.551	0.314	0	1.01	2.235
	2015	-0.613	0.279	0.028	-1.171	-0.068
	2016	-0.811	0.27	0.003	-1.405	-0.349
	2017	0.194	0.391	0.619	-0.623	0.906
	division 2	-0.14	0.361	0.699	-0.855	0.559
	division 3	-0.543	0.342	0.113	-1.275	0.07
	division 4	-1.075	0.332	0.001	-1.778	-0.481
	division 5	-0.429	0.376	0.253	-1.18	0.3
	division 6	-0.897	0.369	0.015	-1.666	-0.218
	division 7	-0.475	0.354	0.18	-1.223	0.168
	division 8	0.399	0.479	0.405	-0.538	1.34
	division 9	-0.167	0.325	0.607	-0.829	0.45

Note: $|\mathcal{J}| = 1486$ observations. This table displays parameter estimates for the effort cost function $c(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \gamma_S s + \gamma_Q q + \kappa sq$. Standard errors, p-values, and confidence interval (CI) bounds are obtained by bootstrapping with 42,327 rep.

of payment at the MSR/MLR: because of this discontinuity, the game turns into one in which at least two intersections of marginal payment and marginal cost exist for most ACO participants.

Tables 9, 10, and 11 contain estimates of model parameters in the restricted model.

Table 9: **Cost Function Parameter Estimates (No Uncertainty Model)**

Parameter	Estimate	Standard Error	P-value	95% CI Lower Bound	95% CI Upper Bound
δ_S	224.090	62.528	0.000	116.39	362.990
δ_Q	8.383	1.563	0.000	5.665	11.795
κ	11.214	6.444	0.082	1.321	26.011

Note: $|\mathcal{J}| = 1486$ observations. This table displays parameter estimates for the effort cost function $c(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \gamma_S s + \gamma_Q q + \kappa sq$. Standard errors, p-values, and confidence interval (CI) bounds are obtained by bootstrapping with 50,000 rep. Parameter estimates not shown here, including year and Census Division fixed-effects, are available in Tables 10 and 11 in Appendix C.

Table 10: **Cost Function Parameter Estimates: Marginal Cost of Savings (No Uncertainty Model)**

	Parameter	Estimate	Standard Error	P-value	95% CI Lower Bound	95% CI Upper Bound
	Constant	-74.333	41.071	0.07	-160.81	0.712
	# states	4.363	1.965	0.026	0.059	7.798
	# beneficiaries	0.368	0.173	0.033	0.093	0.763
	risk	125.82	38.383	0.001	30.978	182.99
	% over 75	0.159	0.245	0.516	-0.286	0.677
	% nonwhite	0.004	0.073	0.953	-0.132	0.156
	% pctmale	1.288	1.112	0.247	-0.596	3.755
	# providers	-1.272	1.883	0.499	-5.513	1.889
	fraction PCP	5.338	6.904	0.439	-7.668	19.277
	fraction inpatient	-149.98	62.677	0.017	-279.6	-34.49
	fraction outpatient	-134.95	32.214	0	-199.91	-74.182
	PC services	-4.066	1.309	0.002	-6.428	-1.275
	# admissions	-71.603	27.872	0.01	-117.28	-8.186
γ_S	fraction PC served by PCP	9.904	10.281	0.335	-11.049	29.466
	age	-0.129	0.032	0	-0.204	-0.08
	all group	25.863	4.743	0	17.183	35.818
	2015	-0.309	3.863	0.936	-8.27	7.027
	2016	-1.121	3.644	0.758	-8.958	5.283
	2017	14.138	4.819	0.003	3.478	22.353
	division 2	-1.766	6.911	0.798	-16.117	11.031
	division 3	-7.475	6.32	0.237	-21.161	3.288
	division 4	-16.727	6.357	0.009	-30.522	-5.945
	division 5	-7.322	6.61	0.268	-21.327	4.395
	division 6	-15.794	6.462	0.015	-29.694	-4.584
	division 7	-6.628	6.321	0.294	-20.381	4.274
	division 8	1.395	8.058	0.863	-15.224	16.468
	division 9	-5.888	6.163	0.339	-19.381	4.572

Note: $|\mathcal{J}| = 1486$ observations. This table displays parameter estimates for the effort cost function $c(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \gamma_S s + \gamma_Q q + \kappa sq$. Standard errors, p-values, and confidence interval (CI) bounds are obtained by bootstrapping with 50,000 rep.

Table 11: **Cost Function Parameter Estimates: Marginal Cost of Quality (No Uncertainty Model)**

	Parameter	Estimate	Standard Error	P-value	95% CI Lower Bound	95% CI Upper Bound
	Constant	-11.284	3.564	0.002	-19.034	-5.088
	# states	0.318	0.191	0.097	-0.062	0.669
	# beneficiaries	0.014	0.013	0.261	-0.004	0.045
	risk	8.806	3.819	0.021	1.346	15.831
	% over 75	0.009	0.018	0.621	-0.023	0.048
	% nonwhite	0.011	0.005	0.033	0.002	0.022
	% pctmale	0.078	0.084	0.355	-0.064	0.266
	# providers	-0.101	0.141	0.476	-0.436	0.112
	fraction PCP	0.396	0.488	0.417	-0.513	1.395
	fraction inpatient	-11.077	4.785	0.021	-21.514	-2.699
	fraction outpatient	-8.42	2.253	0	-13.155	-4.324
	PC services	-0.317	0.101	0.002	-0.515	-0.123
	# admissions	-3.381	2.492	0.175	-7.964	1.695
γ_Q	fraction PC served by PCP	0.726	0.759	0.339	-0.811	2.177
	age	-0.006	0.003	0.018	-0.012	-0.002
	all group	1.348	0.318	0	0.792	2.028
	2015	-0.587	0.278	0.035	-1.134	-0.041
	2016	-0.788	0.27	0.004	-1.378	-0.32
	2017	0.269	0.388	0.488	-0.537	0.977
	division 2	-0.181	0.368	0.622	-0.911	0.534
	division 3	-0.494	0.357	0.166	-1.256	0.132
	division 4	-0.982	0.345	0.004	-1.71	-0.369
	division 5	-0.338	0.39	0.387	-1.131	0.403
	division 6	-0.708	0.385	0.066	-1.512	-0.002
	division 7	-0.326	0.367	0.373	-1.104	0.337
	division 8	0.415	0.495	0.401	-0.558	1.391
	division 9	-0.147	0.337	0.662	-0.846	0.477

Note: $|\mathcal{J}| = 1486$ observations. This table displays parameter estimates for the effort cost function $c(s,q) = \frac{\delta_S}{2}s^2 + \frac{\delta_Q}{2}q^2 + \gamma_S s + \gamma_Q q + \kappa sq$. Standard errors, p-values, and confidence interval (CI) bounds are obtained by bootstrapping with 50,000 rep.

References

Agha, L., Frandsen, B., & Rebitzer, J. B. (2019). Fragmented division of labor and healthcare costs: Evidence from moves across regions. *Journal of Public Economics*, *169*, 144-159.

Aswani, A., Shen, Z.-J. M., & Siddiq, A. (2019). Data-driven incentive design in the medicare shared savings program. *Operations Research*, *67*(4), 1002-1026.

Bai, C.-E., & Xu, L. C. (2005). Incentives for ceos with multitasks: Evidence from chinese state-owned enterprises. *Journal of Comparative Economics*, *33*(3), 517 - 539. (Symposium: Monetary and Exchange Rate Policies in the CIS Countries: Between the EU and Russia)

Chandra, A., Cutler, D., & Song, Z. (2011). Chapter six - who ordered that? the economics of treatment choices in medical care. In M. V. Pauly, T. G. Mcguire, & P. P. Barros (Eds.), *Handbook of health economics* (Vol. 2, p. 397 - 432). Elsevier.

Choné, P., & Ma, A. (2011). Optimal health care contract under physician agency. *Annals of Economics and Statistics*(101/102), 229–256.

Dumont, E., Fortin, B., Jacquemet, N., & Shearer, B. (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics*, *27*(6), 1436 - 1450.

Einav, L., Finkelstein, A., & Mahoney, N. (2018). Provider incentives and healthcare costs: Evidence from long-term care hospitals. *Econometrica*, *86*(6), 2161-2219.

Eliason, P. J., Grieco, P. L. E., McDevitt, R. C., & Roberts, J. W. (2018, November). Strategic patient discharge: The case of long-term care hospitals. *American Economic Review*, *108*(11), 3232-65.

Encinosa, W. E., Gaynor, M., & Rebitzer, J. B. (2007). The sociology of groups and the economics of incentives: Theory and evidence on compensation systems. *Journal of Economic Behavior & Organization*, *62*(2), 187 - 214.

Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy*, *21*(3), 673-705.

Foo, P. K., Lee, R. S., & Fong, K. (2017). Physician prices, hospital prices, and treatment choice in labor and delivery. *American Journal of Health Economics*, *3*(3), 422-453.

Frandsen, B., Powell, M., & Rebitzer, J. B. (2019). Sticking points: common-agency problems and contracting in the us healthcare system. *The RAND Journal of Economics*, *50*(2), 251-285.

Frandsen, B., & Rebitzer, J. B. (2015). Structuring incentives within accountable care organizations. *Journal of Law, Economics, and Organization*, *31*(suppl. 1), 77-103.

Frech, H. E., Whaley, C., Handel, B. R., Bowers, L., Simon, C. J., & Scheffler, R. M. (2015, Sep 01). Market power, transactions costs, and the entry of accountable care organizations in health care. *Review of Industrial Organization*, *47*(2), 167–193.

Gaynor, M. (2006, June). *What do we know about competition and quality in health care markets?* (Working Paper No. 12301). National Bureau of Economic Research.

Gaynor, M., Ho, K., & Town, R. J. (2015, June). The Industrial Organization of Health-Care Markets. *Journal of Economic Literature*, *53*(2), 235-284.

Gaynor, M., Mehta, N., & Richards-Shubik, S. (2020, May). *Optimal contracting with altruistic agents: A structural model of medicare payments for dialysis drugs* (Working Paper No. 27172). National Bureau of Economic Research.

Gaynor, M., Rebitzer, J., & Taylor, L. (2004). Physician incentives in health maintenance organizations. *Journal of Political Economy*, *112*(4), 915–931.

Glied, S., & Hong, K. (2018). Health care in a multi-payer system: Spillovers of health care service demand among adults under 65 on utilization and outcomes in medicare. *Journal of Health Economics*, *60*, 165 - 176.

Grassi, S., & Ma, A. (2016). Information acquisition, referral, and organization. *The RAND Journal of Economics*, *47*(4), 935-960.

Hackmann, M. B. (2019, May). Incentivizing better quality of care: The role of medicaid and competition in the nursing home industry. *American Economic Review*, *109*(5), 1684-1716.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.

Ho, K., & Pakes, A. (2014). Hospital choices, hospital prices, and financial incentives to physicians. *The American Economic Review*, *104*(12), 3841–3884.

Hong, F., Hossain, T., List, J. A., & Tanaka, M. (2018). Testing the theory of multitasking: Evidence from a natural field experiment in chinese factories. *International Economic Review*, *59*(2), 511-536.

Kim, M., Sudhir, K., & Uetake, K. (2022). A structural model of a multitasking salesforce: Incentives, private information, and job design. *Manag. Sci.*, *68*(6), 4602–4630.

McWilliams, J. M. (2014, Oct 01). Aco payment models and the path to accountability. *Journal of General Internal Medicine*, *29*(10), 1328–1330.

McWilliams, J. M., Hatfield, L. A., Landon, B. E., Hamed, P., & Chernew, M. E. (2018). Medicare spending after 3 years of the medicare shared savings program. *New England Journal of Medicine*, *379*(12), 1139-1149.

Mullen, K. J., Frank, R. G., & Rosenthal, M. B. (2010). Can you get what you pay for? pay-for-performance and the quality of healthcare providers. *The RAND Journal of Economics*, *41*(1), 64-91.

Rebitzer, J. B., & Votruba, M. E. (2011, October). *Organizational economics and physician practices* (Working Paper No. 17535). National Bureau of Economic Research.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, *33*(3), 520–534.

Slade, M. E. (1996). Multitask agency and contract choice: An empirical exploration. *International Economic Review*, *37*(2), 465–486.

Wilson, M., Guta, A., Waddell, K., Lavis, J., Reid, R., & Evans, C. (2020). The impacts of accountable care organizations on patient experience, health outcomes and costs: a rapid review. *Journal of Health Services Research & Policy*, *25*(2), 130-138.