The Commission Paradox: Evidence from Prison Telecommunications*

Marleen Marra† Nathan H. Miller‡ Gretchen Sileo§

March 26, 2026

Abstract

When an intermediary procures services for end users and receives a commission from the provider, a tension arises: procurement designs that more effectively extract provider rents can simultaneously raise prices and reduce total surplus. We formalize this *commission paradox* in a first-score auction model and estimate the model using procurement data from the prison telecommunications industry. The paradox emerges empirically, as commissions reduce provider rents but raise call rates. We also find that when commissions enter the scoring rule, increased competition among providers raises payments to the correctional authority but does not lower rates; banning commissions restores the link. Our estimates indicate that recent federal regulations banning commissions and capping rates would lower costs for incarcerated individuals while preserving provider profitability.

JEL Codes: D43, D44, H57, L13, L51, L96

Keywords: commissions, public procurement, scoring auctions, regulation, competition policy, telecommunications, prison systems, incarcerated individuals

*We thank David Byrne, Francisco Garrido, Brent Hickman, Andres Gonzalez-Lira, John Mayo, Bill Pope, Stephen Raher, Diwakar Raisingh, Eric Richert, Ben Rosa, Gloria Sheu, Charles Taragin, Peter Wagner, Jonathan Williams, and numerous seminar and conference participants for helpful comments. Matthew Parayil and Maria Abissi provided research assistance. The research was funded by the National Science Foundation (SES 2117197). An earlier working paper circulated under the title "Phoning Home: The Procurement of Telecommunications for Incarcerated Individuals in the United States."

†KU Leuven, Department of Economics, and CEPR. Email: marleen.marra@kuleuven.be.

‡Georgetown University and NBER. Email: nathan.miller@georgetown.edu.

§Temple University, Department of Economics. Email: gretchen.sileo@temple.edu.

1 Introduction

When an intermediary procures services on behalf of end users, procurement design shapes the distribution of surplus among the provider, the intermediary, and end users (Laffont and Tirole, 1993). A common arrangement is for the provider to pay a commission to the intermediary, funded by revenue from end users. Examples include prescription drugs, where pharmacy benefit managers receive rebates from manufacturers (e.g., Ho and Lee, 2024), employer-sponsored retirement plans, where employers select fund providers that make revenue-sharing payments (e.g., Bhattacharya and Illanes, 2025), and grocery retail, where chains receive slotting fees from brand manufacturers (e.g., Hristakeva, 2022). In each case, though details of price-setting vary, the intermediary both selects the provider and receives a financial stake in the contract, raising the question of whether procurement mechanisms that serve the intermediary's interest also benefit end users.

We identify that, in such settings, procurement designs that more effectively reduce provider rents can also raise prices on end users and reduce total surplus. In competitive settings, reducing rents benefits consumers as prices fall toward costs and gains from trade grow. With commissions, the mechanism works differently. The intermediary extracts rents not by demanding lower prices but by claiming a share of revenue, which rewards providers that charge more. Reducing provider rents in this manner shrinks the gains from trade between the provider and end users. Thus, objectives that coincide in standard competitive settings can diverge when the intermediary extracts rents with revenue-sharing rather than price competition. We call this the *commission paradox*.

We study the commission paradox in the market for inmate calling services (ICS), where correctional authorities both select the telecommunications provider and receive commission payments. The setting is well-suited to our analysis. First, correctional authorities award contracts through scoring auctions in which bids on rates and commissions are combined according to predetermined weights, thereby creating a transparent mechanism for tracing how commissions interact with rate-setting. Second, incarcerated individuals and their social contacts, who pay for calls, are external to the procurement process; they neither participate in procurement nor discipline the intermediary by choosing the provider. This absence of consumer choice makes the paradox empirically sharp.

There are more than two million incarcerated individuals in the United States, and aggregate spending on ICS has exceeded $1 billion annually (Shields, 2012). The industry has attracted sustained policy attention, culminating in the FCC's 2024 and 2025 Orders banning commissions and capping rates in telecommunications; the Orders cite prior versions of this paper (FCC, 2024, 2025). As providers have shifted toward tablet-based entertainment and other services not covered by the Orders, correctional authorities face the same challenges in adjacent markets. Our results provide practical guidance on how the treatment of commissions in procurement affects both the rates paid by incarcerated individuals and the revenue flowing to the correctional authority.

The empirical foundation of our study is a set of public records requests submitted to all fifty states regarding ICS procurement for state-level prison systems. We obtained and digitized requests

for proposals, winning and losing bids, bid evaluations, and the ensuing contracts. Thus, for a set of procurement events, we have comprehensive data on the procurement lifecycle. Observing the complete set of submitted bids—including those of losing providers—is important because it allows us to estimate how procuring entities translate proposed rates and commissions into scores using within-auction variation across bidders. This lets us accommodate auction-level heterogeneity in how scores are assigned, which would not be possible with winning bids alone.

A salient feature of the data is the relationship between commissions and rates: when commissions enter the scoring rule, they are positively correlated with rates, and rates are substantially higher than in auctions where only rates are subject to bidding. This pattern is consistent with the central mechanism of the model, in which commission-based scoring channels competitive pressure toward commission payments rather than rate reductions. Separately, natural experiments from two states that sharply reduced call prices identify demand elasticity, a key input to the model. New York and New Jersey each implemented large, discrete rate reductions that generated substantial demand responses: call volumes roughly doubled in New York and tripled in New Jersey, while total expenditures fell. These responses inform our welfare calculations.

We model ICS procurement as a first-score auction, following Asker and Cantillon (2008). Each provider privately observes its cost and submits bids comprising a rate and a commission. In equilibrium, the rates maximize the bilateral surplus of the provider and the procuring entity, while competition enters through the commission. Equilibrium can be characterized in terms of providers' *pseudotypes*, defined as the bilateral surplus each provider creates at its equilibrium rate. The expected gap between the best and second-best pseudotype governs provider rents.

Two lemmas characterize how shifting scoring weight toward commissions affects outcomes. Lemma 1 establishes that equilibrium rates rise for all providers, reducing consumer surplus and total surplus. Lemma 2 establishes that the pseudotype distribution compresses, reducing provider rents. The compression arises because a dollar of commission burdens every firm equally, whereas low-cost firms disproportionately retain profit with low rates; greater weight on commissions therefore reduces differentiation. Together, the lemmas establish the commission paradox (Theorem 1). This contrasts with reserve prices, the canonical instrument for reducing provider rents (Myerson, 1981; Riley and Samuelson, 1981), which exclude high-cost firms and reduce allocative efficiency. Commissions extract rents without altering allocation, at the cost of raising prices for end users.

We then extend the model to allow multidimensional heterogeneity, where providers also differ in quality. The extension introduces selection effects—shifting weight toward commissions alters which firm wins, not only how the winner bids—and provides the foundation for our empirical model of ICS. The paradox cannot be established from primitives alone, as it depends on the joint distribution of costs and quality among competing providers. Quantifying whether it holds in practice, and measuring the welfare consequences of alternative procurement designs, requires taking the model to data. We therefore estimate the structural primitives using the ICS procurement record, recovering the cost and quality distributions that govern equilibrium bidding.

We estimate the model using the simulated method of moments, computing the full equilibrium for each candidate parameter vector. The theoretical model features auctions in which commissions are determined through competitive bidding. Our estimation sample also includes auctions in which commissions are predetermined by the procuring entity. These two auction types require different computational approaches: for the former, the Asker and Cantillon (2008) equilibrium can be computed in closed form; for the latter, we approximate equilibrium using an iterative best-response algorithm (Bajari, 2001; Armantier et al., 2008). Combining both types provides variation in the treatment of commissions that aids identification. The recovered cost distribution is consistent with independent estimates that rely on accounting data (Bazelon et al., 2023).

The commission paradox emerges in our quantitative analysis. In numerical simulations, increasing the commission weight in a representative auction compresses the pseudotype distribution, increasing competitive pressure and allowing the procuring entity to extract providers' rents in the form of higher commission payments. Equilibrium rates rise in tandem, so consumer surplus and total surplus fall. We find the converse when we implement a commission ban in the auctions in our sample. Averaging across auctions, the ban reduces expected rates from \$1.34 to \$0.52 per 15-minute call and increases total surplus by \$6.39 per inmate-month, as consumer surplus and profit gains more than offset lost commission revenue. Applied to the 1.2 million individuals in state prisons, these estimates imply total surplus gains on the order of \$92 million annually.

We use the estimated model to evaluate additional policy interventions. First, rate regulation that predetermines rates while allowing competition over commissions shifts the cost burden from consumers to the procuring entity, but increases provider market power, as higher-cost firms are disproportionately disadvantaged. Second, we find that a rate cap of \$1.35 per 15-minute call, as described in the FCC's 2025 Order, would have limited effects on participation in our simulations because the costs we estimate tend to be below the cap. Finally, we examine competition policy. With commissions in the scoring rule, increased competition raises commission payments but does not reduce rates; banning commissions restores the link between competition and rates.

Our paper relates to a growing literature on the role of intermediary commissions in procurement and market design. Commissions appear in the procurement of prescription drugs (Ho and Lee, 2024) and infant formula (Abito et al., 2022; An et al., 2023), as well as in financial and real estate markets where intermediaries receive payments from product providers (Allen et al., 2014; Badoer et al., 2020; Pool et al., 2024; Bhattacharya et al., 2025; Benetton et al., 2025; Grunewald et al., 2023). Among these, Bhattacharya and Illanes (2025) is the closest methodologically, developing and estimating a model of employers' procurement of defined contribution plans to evaluate the efficacy of regulation. We are the first to formalize the result that commissions can simultaneously reduce provider rents and total surplus, creating the commission paradox, and the first to study commissions in the context of public procurement for correctional services.

The paper also contributes to the empirical literature on government procurement auctions. Recent work examines how auction design affects outcomes in oil lease auctions (Bhattacharya et

al., 2022), federal telecommunications procurement (Kang and Miller, 2022), highway construction (Bolotnyy and Vasserman, 2023), bank resolution (Allen et al., 2024), and defense contracts (Carril et al., 2025). A common feature of this literature is the development of structural models tailored to the institutional details of specific procurement settings. We take the same approach, adapting the scoring auction framework to ICS and using the estimated model to evaluate procurement designs and regulatory interventions relevant to ongoing federal policy.

More broadly, the commission paradox can be understood as an incentive misalignment in delegated procurement. The procuring entity acts as an agent for end users but is compensated through a channel—revenue-sharing—that rewards higher prices. This connects our analysis to the literature on moral hazard in agency relationships (Grossman and Hart, 1983; Holmström, 1979). Our contribution is to formalize how this misalignment operates within a competitive procurement mechanism: commissions do not simply transfer surplus but alter equilibrium bidding, reducing provider rents while raising prices on the population the intermediary nominally serves. This also points to consumers' ability to select among intermediaries as a key moderating factor; in ICS, incarcerated individuals cannot, so the only discipline on prices is the demand response.[1]

The remainder of the paper proceeds as follows. Section 2 provides institutional background and describes our data. Section 3 develops the theoretical model. Section 4 presents the empirical implementation. Section 5 uses the estimated model to establish the commission paradox and examine the effects of a commission ban. Section 6 considers additional policy analyses, including rate regulation, rate caps, and competition policy. Section 7 concludes.

2 Inmate Calling Services

2.1 Background

In this section, we describe the ICS industry in greater detail.[2] Three features of the industry are central to our analysis. First, the procuring authority is both the agent that selects the ICS provider and a direct financial beneficiary of the contract through commission payments.[3] Second, incarcerated individuals and their social contacts, who ultimately pay for calls, are excluded from procurement and have no choice of provider. Without outside options, the demand response that would otherwise discipline prices is limited, amplifying the resulting distortion. Third, the scoring auction used to award contracts creates a formal mechanism through which commission payments interact with rate-setting, generating the tradeoffs we model in Section 3. These features are not

[1] The mechanism we examine is distinct from double marginalization (Spengler, 1950), as the intermediary does not set prices to end users; commissions distort equilibrium bidding rather than stacking markups.

[2] We draw on regulatory filings (e.g., FCC, 2012, 2013, 2015, 2021, 2024), an article written by FCC economists (Baker et al., 2020), a small academic literature (e.g., Jackson, 2005; Raher, 2020; Bazelon et al., 2023), and our conversations with industry experts.

[3] For state-level prison systems, the procuring authority is typically a Department of Corrections (DOC). For county jails, it is often a sheriff's office.

unique to ICS; similar structures arise in other industries.

Most incarcerated individuals reside in a county jail before sentencing and in a state prison facility after sentencing.[4] Facility-specific rules determine inmates' access to phones and who they can call, and a number of security measures are in place to ensure compliance. The rules also limit the duration of calls. A typical maximum call length is 15 minutes, but some facilities allow 20 or 30 minute calls. Many calls are collect, so the non-incarcerated recipient pays. Payments from incarcerated individuals and their social contacts to the ICS provider are based on pricing schedules set during the procurement process. The overall expenditure for a call can depend on several factors, including its length, whether it is intrastate or interstate, the payment type, and whether there is a fixed connection charge.

The providers are privately owned telecommunications companies. Two providers own most contracts for prison systems: Securus and GTL. More providers have contracts with county jails, including IC Solutions and smaller providers like NCIC and CPC. The industry's current configuration reflects a consolidation period during which GTL and Securus acquired many of the smaller providers. Most recently, GTL acquired Telmate in 2018, Securus abandoned the acquisition of IC Solutions after the Department of Justice expressed antitrust concerns, and IC Solutions acquired CenturyLink in 2020.[5] GTL recently rebranded as ViaPath.

Correctional authorities use procurement processes that resemble first-price scoring auctions to select their ICS provider. The broad contours are as follows: first, the correctional authority issues a request-for-proposal (RFP) that outlines the technical requirements that providers must meet, many of which relate to security measures. The RFP also specifies a binding scoring rule describing how the contracting authority will evaluate bids. Second, there is a formal question-and-answer period during which providers can gain additional information about the technical requirements and the facilities involved; prospective bidders can sometimes participate in formal visits ("walk-throughs") of the facilities. Providers also learn about their likely competitors through these interactions. Third, providers submit bids that describe their technical capabilities and propose financial terms. Finally, the correctional authority evaluates the bids according to the scoring rule, and the provider with the highest score wins the contract at the terms that they propose. Contracts are three or more years in duration.[6] Figure D.1 provides an illustrative example of the procurement process.

The financial terms in the scoring rule include the pricing schedule and the commission pay-

[4]The Federal Bureau of Prisons and Immigration and Customs Enforcement also operate facilities that house individuals charged or convicted of violating federal laws and immigration laws, respectively.

[5]The ICS industry began in the 1970s as prison systems relaxed rules restricting incarcerated individuals to a single phone call every three months. AT&T had a monopoly until the 1984 Consent Decree authorized its breakup. A number of providers entered the market in the following years, including large telecommunications companies like MCI and Sprint, as well as more specialized providers like GTL. In the 1990s, almost 30 ITS providers competed for prison and jail contracts. For a discussion of this history, see Jackson (2005).

[6]Many procurement processes differ from our description to some extent. For example, some authorities conduct the bidding in two rounds, with a subset of providers being asked to present more information—and possibly better financial terms—in the second round. Furthermore, losing bidders can contest the decision, which occasionally succeeds in changing the outcome and can lead to an entirely new procurement auction.

ment. In many procurement processes, the correctional authority predetermines either the pricing schedule or the commission, allowing the other to be set through the bidding process. In other auctions, both are subject to bidding, in which case they are usually evaluated separately. The commissions that correctional authorities receive are placed in "inmate welfare" funds that, in principle, support non-essential purchases of books, exercise equipment, and other amenities valued by incarcerated individuals. However, a report of the Prison Policy Initiative, a non-profit advocacy organization, documents that oversight is weak and that funds are sometimes under-utilized or misused, for example, to pay for operational expenses that would normally come from the general budget or, in an extreme case, perks for staff (Nam-Sonenstein, 2024). The gap between the stated purpose of commissions and their actual use underscores the potential for misalignment between the procuring authority's financial interest and the welfare of incarcerated individuals.

The large providers have two main revenue streams. The first comes from the prices charged for the phone calls that incarcerated individuals make. We refer to this as "non-fee revenue." The second comes from ancillary fees. The fees are obtained in a variety of ways. For example, providers can require that calls be made using prepaid calling cards and can levy fees when money is placed on those cards. Commissions are not paid on the revenue obtained from ancillary fees. Fee data are not always shared with the procuring authority and, therefore, are not always accessible with public record requests; furthermore, the fees set by some providers may not be state-specific. In our data, we observe fee schedules in some bids (e.g., see Figure D.2) but not others. The available evidence indicates that fee revenue is significant.[7]

There are at least two major costs associated with providing ICS. First, when providers start serving clients, they install their own phones and equipment (the telecommunication lines, though, typically do not need to be replaced).[8] Second, providers operate and pay for data centers that store call recordings and associated metadata. Qualitatively, the cost of an "install" increases with the number and size of the facilities, and data center costs increase with the calls being made. As private equity companies own the large providers, high-fidelity financial information that breaks down the relative magnitude of these costs is not publicly available. However, one study conducts an accounting exercise and estimates the per minute average call cost to be \$0.010-\$0.012, depending on the size of the facility (Bazelon et al., 2023).

Facilities also incur costs due to ICS. However, these appear to be small. The FCC's 2024 Order accounts for \$0.02 per minute in facility costs for prisons (FCC, 2024).

[7]The FCC determined that ancillary fees can raise the costs to incarcerated individuals and their social contacts by as much as 40%. See FCC Press Release, October 22, 2015, DOC-335984A1. Also notable is that one provider recently settled a class action lawsuit that alleged it had seized over \$100 million from prepaid accounts following periods of inactivity. See Githieya v. Glob. Tel*Link Corp., USDC (N.D. Ga.), Case No. 1:15-CV-00986.

[8]Incumbents are sometimes thought to have an advantage in the procurement process for this reason; correctional authorities may also perceive them as being more likely to have high-quality technical capabilities.

2.2 Regulatory History

The ICS industry has received considerable scrutiny at the federal and state levels. At the federal level, the FCC adopted interim rate caps on interstate calls in 2013 and permanent caps on both intrastate and interstate calls in 2015, but the latter were vacated in a 2017 court decision (*GTL vs. FCC*) partly because the FCC lacked statutory authority over intrastate calls. The Martha Wright-Reed Act of 2023 granted that authority and led to the FCC's 2024 and 2025 Orders (FCC, 2024, 2025), which caps rates and bans both commissions and ancillary fees.[9] At the state level, at least ten states have passed laws eliminating commissions in ICS contracts for state-level facilities, and several have made or scheduled free calls; most of these changes occurred after 2019, when our sample ends.[10] Our counterfactual analyses evaluate relevant policy interventions, including commission bans, rate regulation, and rate caps. We also explore stricter competition policy.

2.3 Data Collection and Summary Statistics

Our data come from requests for public records that we submitted to all 50 states in the 2020-2021 academic year. Thus, data pertain to state prison systems rather than county jails. We targeted documents and data on ICS that span the previous two decades. In particular, we asked for the RFPs, all of the bids submitted by providers, how those bids were evaluated and scored, and the contracts with the winning providers. We also requested monthly data on call volumes and the average daily population (ADP) in each prison system.[11] Our empirical analysis is based on the procurement events ("auctions") for which we received complete information (Table D.1).

Table 1 shows selected summary statistics at the auction level, on realized financial terms, and the scoring rule weights placed on the rate, the commission, and the providers' technical capabilities. We measure the "rate" as the price of a 15-minute local collect call. For the commission, we use a measure of the payment from the provider per inmate-month that we describe later in this section. We use "technical capabilities" as a composite of the various considerations that relate to a bidder's ability to meet the buyer's technical requirements.[12] Across all auctions, the average rate is $1.65, and the average commission payment per inmate-month is $11.34. The average rate, commission, and technical weights are 0.22, 0.12, and 0.66, respectively. The rate and commission weights can be zero, reflecting that they are predetermined in some auctions. The average auction

[9] The 2024 and 2025 Orders cite prior versions of this paper to support demand elasticity estimates and an economic analysis of commissions. The history of federal regulation is detailed in FCC (2024).

[10] The states that eliminated commissions are Colorado, Illinois, Maryland, Nebraska, New Jersey, New Mexico, New York, Ohio, Rhode Island, and South Carolina. States that have implemented or scheduled free calls include Connecticut, California, Colorado, Minnesota, and Massachusetts.

[11] We submitted public records requests to all 50 states and received at least some information from 43. We obtained a complete set of documents and data on at least one procurement event from 26 states; nine states provided complete records for multiple events. The requests required extensive correspondence over two academic years, as responses were shaped by each state's compliance requirements, document retention practices, and willingness to engage.

[12] For example, in 2016, North Dakota considered information technology, experience, qualifications, financial strength, and the presentation of each bidder. We treat all of these as part of a bidder's technical capabilities.

Table 1: Selected Auction-Level Summary Statistics

	Mean	St. Dev.	10%	25%	50%	75%	90%
All Auctions							
Rate	1.62	1.32	0.27	0.58	1.65	2.25	3.12
Commission	11.34	10.53	0.00	0.05	10.74	19.61	28.27
Rate Weight	0.22	0.23	0.00	0.00	0.20	0.33	0.43
Commission Weight	0.12	0.19	0.00	0.00	0.00	0.15	0.32
Technical Weight	0.66	0.23	0.39	0.60	0.70	0.80	0.88
Number of Bidders	3.92	1.14	3.00	3.00	4.00	5.00	5.00
Rate and Commission in the Scoring Rule							
Rate	2.55	1.98	0.68	0.97	1.65	4.07	4.68
Commission	10.11	7.52	3.15	5.46	10.58	13.18	18.52
Rate Weight	0.19	0.14	0.04	0.10	0.20	0.22	0.32
Commission Weight	0.15	0.18	0.04	0.07	0.10	0.13	0.31
Technical Weight	0.66	0.31	0.40	0.68	0.70	0.80	0.92
Number of Bidders	3.29	0.95	2.60	3.00	3.00	3.50	4.40
Rate Predetermined, Commission in the Scoring Rule							
Rate	1.91	0.39	1.59	1.76	1.90	2.21	2.31
Commission	23.1	8.16	13.99	15.68	24.10	28.57	29.56
Commission Weight	0.33	0.20	0.14	0.20	0.30	0.34	0.48
Technical Weight	0.68	0.20	0.52	0.66	0.70	0.81	0.86
Number of Bidders	4.00	0.67	3.00	4.00	4.00	4.00	5.00
Rate in the Scoring Rule, Commission Predetermined							
Rate	0.88	0.77	0.19	0.33	0.60	1.24	2.06
Commission	4.58	6.34	0.00	0.00	0.62	8.49	13.11
Rate Weight	0.38	0.21	0.20	0.29	0.32	0.41	0.59
Technical Weight	0.62	0.21	0.41	0.59	0.69	0.72	0.80
Number of Bidders	4.17	1.38	2.70	3.00	4.00	5.00	6.00

Notes: The table provides selected summary statistics at the auction level for the 37 distinct auctions in the data. The rate is the cost of a 15-minute local collect call, and the commission is in dollars per inmate-month. Statistics are shown for the full sample and subsamples constructed based on which financial terms are subject to bidding. There are 37 auctions in the full sample and seven, ten, and 18 in the subsamples, respectively. Two auctions do not place any weight on financial terms in the scoring rule and are therefore omitted from the subsamples.

has 3.92 bidders.[13]

We allocate the auctions into subsamples based on whether both the rate and the commission are subject to bidding, whether only the commission is subject to bidding, and whether only the rate is subject to bidding. The average rate in these subsamples is \$2.55, \$1.91, and \$0.88, respectively. Two stylized facts are that the rate tends to be the lowest in auctions for which it is the only financial term subject to bidding, and the commission tends to be the largest in auctions for which it is the only financial term subject to bidding. We also observe that the rate and commission receive similar weights in the scoring rule when both are subject to bidding, and the combined weight of the financial terms is similar across the subsamples.

Table 2 shows selected summary statistics at the bid level, based on the 155 bids that were

[13]All prospective providers submitted a single bid in 34 of the 37 auctions, so the average number of bids (4.19) is close to the average number of bidders. Our empirical model assumes that prospective providers submit a single bid. The conditions under which firms submit multiple bids in scoring auctions are explored in Allen et al. (2024).

Table 2: Selected Bid-Level Summary Statistics

	Mean	St. Dev.	10%	25%	50%	75%	90%
Rate	1.33	1.11	0.27	0.50	0.90	1.73	2.72
Commission	17.52	8.35	5.62	11.35	16.67	23.65	28.28
Rate Score	0.71	0.28	0.32	0.54	0.76	0.99	1.00
Commission Score	0.75	0.25	0.42	0.66	0.83	0.92	1.00
Technical Score	0.76	0.18	0.50	0.64	0.80	0.92	0.98

Notes: The table provides selected summary statistics at the bid level for the 155 bids in the data. The rate is the cost of a 15-minute local collect call, and the commission is in dollars per inmate-month. The scores are the fraction of the total points awarded to the bid. The statistics for rates and commissions are conditional, in that we restrict the sample for rates and rate scores to bids in auctions that place a positive weight on rates ($N = 106$), and analogously for commissions and commission scores ($N = 65$).

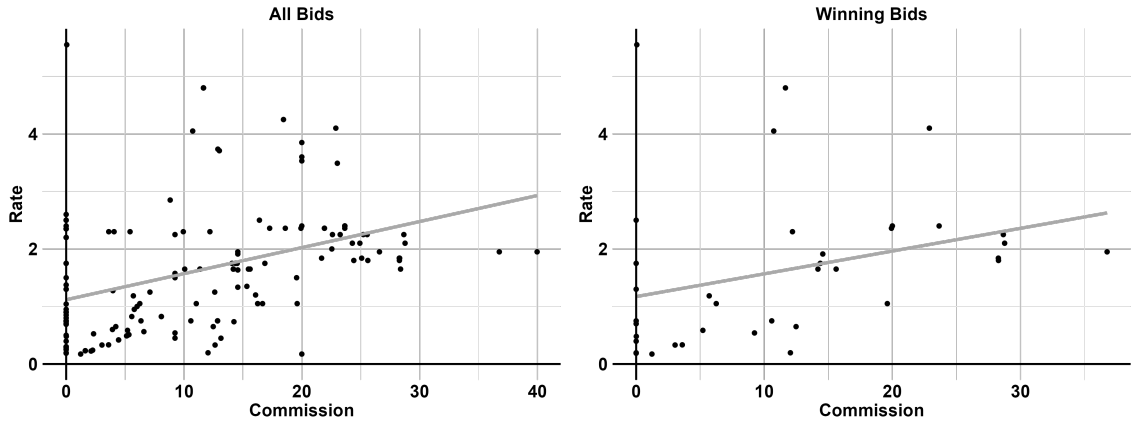


Figure 1: Empirical Relationship Between Rates and Commissions

Notes: The figure provides scatter plots of the commission and the rate in the bid data, using the full sample of 155 bids (left panel) and the 37 winning bids (right panel). Lines of best fit are also shown. The rate is in dollars per 15-minute local collect call. The commission is in dollars per inmate-month.

submitted across all auctions. We examine the proposed rates and commissions and the assigned scores. The statistics for rates and commissions are conditional in that we restrict the sample for rates and rate scores to bids in auctions that place a positive weight on rates, and analogously for commissions and commission scores. The mean proposed rate is $1.33, and the mean proposed commission is $17.52. The scores have means of 0.71, 0.75, and 0.76 for rates, commissions, and technical capabilities, respectively, measured as a share of the maximum possible points.

Figure 1 explores the correlation between rates and commissions in the bid-level data. The left panel uses the full sample of 155 bids; the right panel focuses on the 37 winning bids. Both provide a scatter plot and a line of best fit. A positive empirical relationship between commissions and rates is evident, and we interpret that correlation as a third stylized fact. The slope coefficients in the lines of best fit are statistically significant at the one and ten percent levels, respectively, and the bivariate correlation coefficients are 0.386 and 0.316.

We now return to our measure of commissions. We observe that commission payments are most

commonly specified as a percentage of the provider's non-fee revenue. Among auctions with such terms, the average payment is 55% of non-fee revenue. Six auctions result in fixed commission payments that do not depend on revenue. We allocate these payments to the inmate-month level using data on ADP and contract duration. The average such payment is $13.57 per inmate-month. The commission measure reported in Tables 1 and 2 combines percentage and fixed commissions. We first calculate non-fee revenue using rate data and a demand function that yields the number of calls per inmate-month (Section 4.2). We then apply the commission percentage to non-fee revenue and adjust for any fixed commission payments, allocated to the inmate-month level.

3 Theoretical Model

We establish the commission paradox in a model of procurement. Commissions mitigate the market power of upstream providers, but decrease consumer surplus and total surplus, with the connection being that commissions are funded by higher prices to end users. We first present the model, then characterize the paradox, and finally extend the model to multidimensional heterogeneity.

3.1 Setup and Equilibrium

Procuring entities ("buyers") use sealed-bid first-score auctions to procure a good or service from prospective providers ("firms") on behalf of end users. We index buyers with i and firms with j. Bids comprise the per-unit rate to be paid by end users (r) and a commission payment (k) that is a transfer to be paid between the buyer and firm. The buyer specifies a scoring rule, $s_i(\cdot)$, that converts the bids into a single number (a "score"). The firm with the highest score wins the auction at the terms proposed in its bid. The buyers and firms are risk-neutral and the number of firms that participate in the auction, J_i, is exogenously determined.

We require the scoring rules to be differentiable, additively separable, and quasi-linear in the commission:

ASSUMPTION 1 (Scoring Rule). *For each auction i, the score is given by*

$$s_i(r,k) = \omega_i^r \, s_i^r(r) + \omega_i^k k,$$

with weights $\omega_i^r, \omega_i^k \geq 0$ ($\omega_i^r + \omega_i^k = 1$). The contribution function, $s_i^r(\cdot)$, is continuously differentiable, decreasing, and weakly concave.

The scoring rules we observe in our data also incorporate quality, but otherwise are consistent with our assumption.[14] Scores decrease in the rate and increase in the commission. The concavity

[14] We consider quality, and thus multidimensional bidder heterogeneity, in Section 3.3. Also, our notation is analogous to that of Che (1993) and Asker and Cantillon (2008), where the score is additive in price (a transfer) and a function of quality; in our model "price" is the commission and "quality" is the rate.

of $s^r(\cdot)$ helps ensure a unique equilibrium. The commission is a lump-sum transfer. This nests commissions specified as a percentage of revenue because, when the rate and commission are determined simultaneously, the firm can achieve any dollar payment by adjusting the revenue share; thus, the two forms of commission are equivalent. The distinction matters only if the commission is set before the rate, because then, with a revenue share, the firm would internalize that a higher rate increases the commission payment. We assume the weights, the contribution function, and the number of bidders are common knowledge.

Each firm knows its cost, c_{ij}, but not the costs of its competitors. The cost determines the equilibrium strategy, so it characterizes the firm's type. Formally, we make the following assumption on information and the distribution of firm types:

ASSUMPTION 2 (Information and Distributions). *The cost, c_{ij}, is the private information of firm j, with $c_{ij} \sim^{i.i.d.} F_C$. The corresponding probability distribution function, f_C, is absolutely continuous with support over $\mathcal{C} = [\underline{c}, \bar{c}] \in \mathbb{R}$. The distribution F_C is common knowledge.*

The winning firm's profit is the contract value less the commission. The commission is a pure transfer and does not affect contract value. We make the following assumption on contract value, with subscript notation indicating derivatives applying to the relevant buyer-specific functions:

ASSUMPTION 3 (Contract Value). *A firm of type c that wins with bid (r, k) earns the contract value, $\Pi(r; c)$, less the commission:*

$$\pi(r, k; c) = \Pi(r; c) - k,$$

with $\Pi \in C^2$ and $\Pi_c(r; c) < 0$. Furthermore, $\Pi_{cr}(r; c) > 0$ and $\Pi_{rr}(r; c) < 0$.

Thus, contract value decreases in cost and is concave in rate. With downward-sloping demand, $q(r)$, and a profit function of the form $\Pi(r; c) = (r - c)q(r)$, as we impose shortly, the cross-partial condition $\Pi_{cr} > 0$ follows immediately because a higher rate reduces quantity demanded, making marginal profit less sensitive to costs.

Auction outcomes also affect end users (consumers), who must pay the rate of the winning provider to purchase the good or service. Higher rates reduce quantity-demanded and consumer surplus. To incorporate this into the model, we refine our assumption on contract value:

ASSUMPTION 4 (Contract Value Refinement). *Contract value takes the form:*

$$\Pi(r; c) = (r - c)q(r)$$

where $q : [\underline{r}, \bar{r}] \to \mathbb{R}_+$, the demand for calls, is twice continuously differentiable with $q'(r) < 0$ for all $r \in [\underline{r}, \bar{r})$ and $q(\bar{r}) = 0$, where $\bar{r} < \infty$ is the choke rate.

The conditions $\Pi_c < 0$ and $\Pi_{cr} > 0$ from Assumption 3 are satisfied under Assumption 4. The condition $\Pi_{rr} < 0$ requires a joint restriction on demand curvature and markups; it is satisfied by log-concave demand on the domain of equilibrium rates, and holds globally with linear demand.

The equilibrium concept is symmetric Bayes-Nash equilibrium. An equilibrium strategy is a measurable map $\sigma_i : \mathcal{C} \to \mathbb{R}^2$, with $\sigma_i(c) = \big(r_i(c), k_i(c)\big)$, that maximizes the expected profit of a bidder conditional on other bidders using the same strategy. It satisfies the following condition:

$$\sigma_i(c) = \underset{(r,k)}{\arg\max} \big(\Pi(r;c) - k\big) P_i(r,k) \tag{1}$$

where $P_i(r,k)$ denotes the probability that a firm receives the highest score in auction i when bidding (r,k). Specifically, it equals the probability that $J_i - 1$ competing bidders receive lower scores given i.i.d. costs as in:

$$P_i(s(r,k)) = H_i(s)^{J_i - 1} \tag{2}$$

where $H_i(\cdot)$ denotes the cumulative distribution function of each rival bidder's score.

Finally, we define three notions of surplus: consumer surplus, public surplus (end users and procuring entity), and total surplus. Letting r^* denote the equilibrium rate, these are:

$$\begin{aligned} \text{Consumer Surplus } (CS) &\equiv \int_{r^*}^{\bar{r}} q(p)\, dp \\ \text{Public Surplus } (PS) &\equiv CS + k \\ \text{Total Surplus } (TS) &\equiv CS + k + \pi = CS + \Pi \end{aligned}$$

We focus on these surplus measures in our results below.

3.2 Commissions and Economic Surplus

We now examine how auction design affects outcomes and surplus. To do so, we solve for the equilibrium following Asker and Cantillon (2008). Here, we introduce the two key properties of equilibrium; the full derivation is in Appendix A.1. First, the equilibrium rate of each firm maximizes the "bilateral surplus" of the provider and the buyer, defined as the sum of the contract value and the non-commission components of the score, expressed in currency units. Bilateral surplus depends on the scoring rule weights; we use it to characterize equilibrium but not to evaluate welfare. Formally, it equals

$$\Pi(r;c) + \psi_i(r) \qquad \text{where} \qquad \psi_i(r) \equiv \frac{\omega_i^r}{\omega_i^k}\, s_i^r(r) \tag{3}$$

Equilibrium rate bids, r_{ij}^*, are unique and increasing in costs under the concavity conditions in Assumptions 1 and 3. They are also non-strategic because they are independent of the distribution of competing bids. To characterize the downward pressure on rate from the scoring rule, we define the *monopoly rate*, $r_i^M(c)$, as the rate that maximizes contract value: $\Pi_r(r_i^M; c) = 0$. Absent a positive weight on rate in the scoring rule, providers' equilibrium bids would feature their monopoly rates, which maximize profit and bilateral surplus.

Second, each bidder's competitive position is determined by its *pseudotype*, defined as the bilateral surplus it creates at its equilibrium rate bid:

$$x_{ij} = x_i(c_{ij}) \equiv \Pi(r^*_{ij}; c_{ij}) + \psi_i(r^*_{ij}) \tag{4}$$

In equilibrium, the firm with the highest pseudotype wins the auction. In the second-score version of the auction, the winner's realized profit equals the gap between its pseudotype and the second-best pseudotype: $x_i^{(1)} - x_i^{(2)}$, and revenue equivalence ensures that expected profits across the two auction formats coincide conditional on a firm's own type. More generally, the expected gap between the best and second-best pseudotypes governs the winning provider's profit. Residually, this determines the commission paid to the procuring entity.

We now characterize how shifting scoring weight from rate to commissions affects the rate that each provider offers and the corresponding surplus measures:

Lemma 1 (Comparative Statics)**.** *Under Assumptions 1–4, if the commission weight, ω^k, increases with an offsetting decrease in the rate weight, ω^r, then for any cost realization c,*

(i) *The equilibrium rate bid increases: $\frac{dr^*}{d\omega^k} > 0$, with $r^*(c) \to r^M(c)$ as $\omega^r/\omega^k \to 0$.*

(ii) *Consumer surplus decreases: $\frac{dCS(r^*)}{d\omega^k} < 0$.*

(iii) *Contract value increases: $\frac{d\Pi(r^*;c)}{d\omega^k} > 0$.*

(iv) *Total surplus decreases: $\frac{dTS(r^*;c)}{d\omega^k} < 0$.*

All proofs are in Appendix A.2. As auction weights shift toward commissions, rate rises because the scoring penalty on high rates weakens and, simultaneously, the revenue that high rates generate enable providers to offer higher commission payments that matter more in the scoring rule. In the limit, the rate that maximizes bilateral surplus converges to the monopoly rate. Parts (ii) and (iii) follow directly: consumer surplus falls and contract value increases as rate rises. Part (iv) follows because higher rates eliminate efficient calls: the consumer surplus loss exceeds the contract value gain by the resulting deadweight loss. These results hold for any cost realization and therefore for every bidder.

We have established that shifting auction weights toward commissions raises each provider's equilibrium rate and reduces the total surplus it offers. This shift also compresses the pseudotype distribution:

Lemma 2 (Pseudotype Compression)**.** *Under Assumptions 1–4, increasing the commission weight, ω^k, with an offsetting decrease in the rate weight, ω^r, decreases the variance of the pseudotypes:*

$$\frac{d\,\mathrm{Var}(x)}{d\omega^k} < 0$$

Shifting the scoring rule toward commissions compresses the pseudotype distribution because it moves competition toward a dimension where the advantage of low-cost firms is weaker. To see why, note that lower rates reduce the profit of all firms, but especially the profit of high-cost firms. The reason is that rates interact with demand: for high-cost firms, the margin per call is smaller, so the additional volume generated by a lower rate does less to offset per-unit revenue loss.[15] Commissions, by contrast, are a transfer that costs every firm equally regardless of type. Therefore, as weight shifts from rates to commissions, the surplus differences among firms narrow, the pseudotype distribution compresses, and competition intensifies.

The compression is visible in the pseudotype expression because, as the commission weight increases, pseudotypes become less sensitive to cost. Substituting (3) into (4), we have:

$$x_i(c) = \frac{\omega_i^r}{\omega_i^k}\, s_i^r(r^*(c)) + \Pi(r^*(c); c) \tag{5}$$

Increasing ω_i^k reduces sensitivity to cost: the direct effect operates through the ratio ω_i^r/ω_i^k that scales the rate score, and the indirect effect through r^* vanishes by the envelope theorem. Tracking the discussion above, we establish in the proof that cost sensitivity equals $q(r^*(c))$. That is, a low-cost firm's advantage is proportional to the number of calls, as each call generates one dollar of additional profit per dollar of cost reduction. As ω^k rises, the equilibrium rate increases, reducing demand and thus the marginal value of cost differences. Thus, given the cost distribution, the pseudotype distribution compresses with the commission weight.

Lemma 2 establishes that the variance of the pseudotype distribution falls as the commission weight increases. To translate this into a statement about expected provider profit, the variance reduction must imply a reduction in $E[x^{(1)} - x^{(2)}]$, the expected gap between the best and second-best pseudotypes. This is a weak condition, however, that holds for a wide class of distributions, including any distribution in the location-scale family. We impose it directly as an assumption:

ASSUMPTION 5 (Dispersive Order). *As ω^k increases, the pseudotype distribution contracts in dispersive order: for any $\alpha \in (0, 1)$, the difference $F_x^{-1}(\alpha) - F_x^{-1}(1-\alpha)$ decreases.*

We now state the commission paradox as a theorem:

Theorem 1 (Commission Paradox). *Under Assumptions 1–5, increasing the commission weight, ω^k, with an offsetting decrease in the rate weight, ω^r:*

(i) *increases the expected rate and reduces expected consumer surplus and total surplus;*

(ii) *reduces expected provider profit and increases expected commission, with the increase in the expected commission being larger than the reduction in expected profit.*

[15] Mathematically, we have $\partial\Pi/\partial r = q(r) + (r-c)q'(r)$ and $\partial^2\Pi/\partial r\,\partial c = -q'(r) > 0$, so profit is more sensitive to rate reductions for higher-cost firms. This is the cross-partial condition $\Pi_{cr} > 0$ imposed in Assumption 3.

Part (i) follows from Lemma 1 and the fact that auction weights do not affect winner identity—the lowest-cost firm wins regardless. The profit result in part (ii) follows from Lemma 2 and Assumption 5. The commission result follows because the winner's commission equals contract value minus profit. As contract value increases and profit falls, the commission rises by more than profit falls, with the excess equal to the increase in contract value.

Whether public surplus increases with the commission weight depends on whether the rent extraction from providers exceeds the deadweight loss imposed on consumers. These are determined by different primitives—the pseudotype distribution on one side and demand elasticity on the other. However, given a cost distribution, public surplus increases with the commission weight if demand is sufficiently inelastic, so deadweight loss is small. In this case, there is scope for Pareto improvements among the procuring entity and end users if commissions are redistributed.

3.3 Multidimensional Heterogeneity

We now extend the model to incorporate quality as a third bid dimension. Firms differ in both cost and technical capability, and the scoring rule assigns weight to all three components—rate, commission, and quality. The procuring entity evaluates each firm's technical capability, v^*_{ij}, and assigns a technical score, $v_{ij} \equiv s^v_i(v^*_{ij})$, that is a strictly increasing transformation of v^*_{ij}. Each firm knows its technical capability when it bids, and thus also its technical score. Anticipating the empirical application, we assume that technical capability does not affect contract value. We replace Assumptions 1 and 2 with the following:

ASSUMPTION 6 (Scoring Rule). *For each auction i, the score is given by*

$$s_i(r,k,v) = \omega^r_i\, s^r_i(r) + \omega^k_i s^k_i(k) + \omega^v_i s^v_i(v^*),$$

with weights $\omega^r_i, \omega^k_i, \omega^v_i \geq 0$ ($\omega^r_i + \omega^k_i + \omega^v_i = 1$). The contribution functions, $s^r_i(\cdot)$, $s^v_i(\cdot)$, and $s^k_i(\cdot)$ are continuously differentiable with $s^{r\prime}_i(r) < 0$, $s^{k\prime}_i(k) > 0$, and $s^{v\prime}_i(v^*) > 0$. In addition, $s^{r\prime\prime}_i(r) \leq 0$ and $s^k_i(\cdot)$ is linear with slope α.

ASSUMPTION 7 (Information and Distributions). *The cost, c_{ij}, and the technical score, v_{ij}, are the private information of firm j, with $c_{ij} \sim^{i.i.d.} F_C$ and $v_{ij} \sim^{i.i.d.} F_V$. The corresponding probability distribution functions, f_C and f_V, are absolutely continuous with support over $\mathcal{C} = [\underline{c}, \bar{c}] \in \mathbb{R}$ and $\mathcal{V} = [\underline{v}, \bar{v}] \in \mathbb{R}$. The distributions F_C and F_V are common knowledge, and $c_{ij} \perp v_{ij}\ \forall ij$.*

Under Assumptions 3, 4, 6, and 7, the Bayes-Nash equilibrium features strategies that are contingent on the multidimensional firm type:

$$\sigma_i(c,v) = \underset{(r,k)}{\arg\max}\, \big(\Pi(r;c) - k\big) P_i(r,k;v) \tag{6}$$

where $P_i(r,k;v)$ denotes the probability that a firm with technical score of v receives the highest

score in auction i when bidding (r, k). Specifically, it equals the probability that $J_i - 1$ competing bidders receive lower scores given i.i.d. costs and technical scores:

$$P_i(r, k; v) = H_i\big(s(r, k, v)\big)^{J_i - 1} \tag{7}$$

where $H_i(\cdot)$ denotes the cumulative distribution function of each rival bidder's score. The Asker and Cantillon (2008) characterization of equilibrium extends to this setting when $\omega_i^k > 0$ (Appendix A.1). In that case, the equilibrium rate again maximizes bilateral surplus, the pseudotype depends on both cost and technical score, and expected profit equals the expected gap between the best and second-best pseudotype. The pseudotype is given by

$$x_i(c, v) = \frac{\omega_i^r}{\omega_i^k}\frac{1}{\alpha}s_i^r(r^*(c)) + \Pi(r^*(c); c) + \frac{\omega_i^v}{\omega_i^k}\frac{1}{\alpha}v \tag{8}$$

The lemmas extend under this treatment of multidimensional heterogeneity because the quality term enters the pseudotype additively and is independent of the commission weight. Indeed, the proofs we provide explicitly cover this more general setting.[16] Nonetheless, while the lemmas point to the commission paradox in this setting, they do not establish it because changes in the auction weights can affect which firm wins the auction. The equilibrium selection effects cannot be signed from primitives alone, but the direction of selection can be partially characterized. As the commission weight increases, the pseudotype becomes less sensitive to both cost and technical score, but the relative change depends on the extent to which equilibrium rates respond to the shift and on the elasticity of demand. We formalize this as our final result:

Theorem 2 (Relative Selection)**.** *Under Assumptions 3, 4, 6, and 7, increase the commission weight ω^k with an offsetting decrease in the rate and/or technical score weights, so that $d\omega^k > 0$, $d\omega^r \leq 0$, $d\omega^v \leq 0$, and $d\omega^k + d\omega^r + d\omega^v = 0$. Let $\varepsilon_q(c) \equiv \frac{|q'(r^*(c))|\, r^*(c)}{q(r^*(c))}$ be the elasticity of demand with respect to the rate and $\varepsilon_r(c) \equiv \frac{\omega^k}{r^*(c)}\frac{dr^*(c)}{d\omega^k}$ be the elasticity of the equilibrium rate with respect to the commission weight. A sufficient condition for the relative sensitivity of the pseudotype to cost versus the technical score to be increasing in ω^k is*

$$\varepsilon_q(c) \cdot \varepsilon_r(c) < 1 + \frac{\omega^k}{\omega^v}\left|\frac{d\omega^v}{d\omega^k}\right| \quad \forall\, c \in \mathcal{C}$$

where the last term is the elasticity of the technical score weight with respect to the commission weight. The condition is also necessary when ω^v is fixed ($d\omega^v = 0$).

The condition has a natural interpretation. From (8), the sensitivity of pseudotypes to the

[16] Moreover, our proof of Lemma 2 requires only $\text{Cov}(c, v) \leq 0$, which is weaker than the independence condition of Assumption 7. Thus, the assumption is that firm capability, broadly construed, drives both lower costs and better quality. By contrast, the lemma can fail if higher quality increases cost. Finally, if quality affects end-user demand, shifting weight toward commissions still compresses the cost-driven component of pseudotype variance, whereas the quality-driven component may not be compressed.

technical score declines mechanically with the commission weight, as it is scaled by ω_i^v/ω_i^k. Cost sensitivity also declines, but through a different channel: the equilibrium rate increases, reducing the number of calls over which cost differences are realized. The product $\varepsilon_q \cdot \varepsilon_r$ measures the strength of this channel—how much quantity, and hence cost sensitivity, is lost as the commission weight rises. When the product is below one, cost sensitivity erodes more slowly than technical score sensitivity, and the auction tilts toward selecting on cost. The bound relaxes when weight is also drawn from the technical score, as cost sensitivity then faces a weaker standard to dominate.

4 Empirical Model and Estimation

We introduce an empirical auction model with multidimensional heterogeneity using Assumptions 3, 4, 6, and 7. The objects to be estimated are the contribution functions, the demand function, the cost distribution, and the technical score distribution. We adopt parametric specifications for demand and the contribution functions that yield closed-form expressions for equilibrium bids, and place only weak assumptions on the distributions of cost and the technical score, which determine the pseudotype distribution that governs competitive intensity, the division of bilateral surplus, and whether the commission paradox emerges. The model primitives are nonparametrically identified (Appendix B.1).

4.1 Contribution Functions

We assume the contribution functions are linear in their arguments and known to the bidders:

$\mathsf{s}_i^r(r_{ij}) = \alpha_{0i}^r + \alpha_{1i}^r r_{ij}$(9)

$\mathsf{s}_i^k(k_{ij}) = \alpha_{0i}^k + \alpha_1^k k_{ij}$(10) Thus, we account for unobserved heterogeneity by including fixed effects in the contribution functions. We also allow for limited heterogeneity in the slope parameter in the contribution function for rates, as discussed next. We estimate the parameters by minimizing the sum of squared distances between the contribution functions and the data, and implement this using OLS. The main identifying assumption is that the contribution functions are, in fact, linear. In support of the assumption, the linear functions explain most of the variation in rate and commission score, as evidenced by R^2 values above 0.90. See also the discussion in Section B.4.

Table 3 shows estimates of the slope parameters. Columns (i)-(iii) focus on rates. Column (i) uses a sample of all bids for which the rate receives a positive weight in the scoring rule; the slope coefficient indicates that if the proposed cost of a 15-minute local collect call is $1.00 greater, then the associated rate score is 0.184 lower, on average. Columns (ii) and (iii) use subsamples that differ based on whether a positive weight is placed on the commission in the auction's scoring rule (in

Table 3: Estimated Contribution Functions

	Rates			Commissions
	(i)	(ii)	(iii)	(iv)
Rate	-0.184	-0.250	-0.040	
	(0.041)	(0.036)	(0.021)	
Commission				0.018
				(0.009)
R^2	0.925	0.910	0.984	0.972
# of Observations	106	83	23	62

Notes: The table summarizes OLS regression results. The unit of observation is a bid. The dependent variable is the rate score in columns (i)-(iii) and the commission score in column (iv). We measure the rate score and the commission score as the fraction of the maximum available points that is awarded to a bid. The independent variables are the rate, which we measure as the price of a 15-minute local collect phone call, and the commission, which we measure in terms of dollars per inmate-month. All regressions include auction fixed effects. The sample in column (i) includes bids in auctions that place a positive weight on rates. The sample in column (ii) includes bids in auctions that place a positive weight on rates but not on commissions. The sample in column (iii) includes bids in auctions that place a positive weight on both rates and commissions. The sample in column (iv) includes bids in auctions that place a positive weight on commissions. Robust standard errors are in parentheses.

addition to the rate). The slope coefficient is more negative for auctions that do not consider commissions, and the difference is statistically significant. We use columns (ii) and (iii) for estimation and counterfactuals. Column (iv) shows our commission results. The slope coefficient indicates that a $1.00 increase in the proposed commission payment per inmate-month is associated with a commission score that is 0.018 higher, on average. This relationship does not depend significantly on the weight that rates receive in the scoring rule.[17]

4.2 Demand Function

We assume that the number of calls per person is a linear function of the rate. Adding time subscripts, t, to reflect monthly data on calls and rates, the function is

$$q_{it} = \beta_0 + \beta_1 r_{it} + u_{it}^d \tag{11}$$

where q_{it} is the number of calls per person/month and u_{it}^d captures seasonal and idiosyncratic factors. We estimate demand using OLS, exploiting variation created by policy-induced price reductions in two states, New York and New Jersey. The identifying assumption is that the rate reductions are orthogonal to changes in the error term. This would be violated if the rate changes

[17] We use our baseline measure of commissions: the payment per inmate-month. We obtain similar results if we use the percentage of revenue to be shared with the procuring entity (and restrict the sample to auctions in which commissions take that form). We omit three bids from the sample that have a commission score of zero. One was in the 2014 Utah auction, and the other two were in the 2019 Utah auction. We suspect the bidders were disqualified for other reasons. If we include these three bids, the slope coefficient increases to 0.025 and remains statistically significant.

Table 4: Demand for Calls

	(i)	(ii)	(iii)	(iv)
Rate	-3.30	-4.37	-4.46	-4.35
	(0.38)	(0.07)	(0.27)	(0.07)
Constant	22.00		19.07	29.86
	(1.35)		(0.61)	(0.29)
NJ Constant		29.90		
		(0.28)		
NY Constant		18.93		
		(0.27)		
R^2	0.53	1.00	0.93	0.99
Sample	NJ/NY	NJ/NY	NJ	NY

Notes: The table summarizes OLS regression results. Observations are at the state-month level. The dependent variable is the average number of calls per inmate-month. The main independent variable is the price of a 15-minute call. The sample for New York includes the months January-December 2009 and April 2010 - March 2011. The sample for New Jersey includes the months January-October 2013 and January-December 2016.

coincide with changes in phone access, but to our knowledge, such access changes did not occur.[18] We run the regression on the two states separately and together, with and without state-specific constants. Table 4 summarizes the results. We average the coefficients from the state-level regressions to construct the demand function used in the model. The function is: $q = 24.47 - 4.41r$. Appendix C provides a more extensive analysis of these price reductions.

Together, the linear contribution functions and the linear demand function yield closed-form expressions that we use in estimation and counterfactuals. Substituting into the generalized version of (3), the non-commission component of the score, in currency units, becomes

$$\psi_i(r_{ij}, v_{ij}) = \frac{\omega_i^r \alpha_1^r}{\omega_i^k \alpha_1^k} r_{ij} + \frac{\omega_i^v}{\omega_i^k \alpha_1^k} v_{ij} + \frac{\omega_i^k \alpha_{0i}^k + \omega_i^r \alpha_{0i}^r}{\omega_i^k \alpha_1^k} \tag{12}$$

and the equilibrium rates can be expressed:

$$r_{ij}^* = \frac{1}{2}\left(-\frac{\beta_0}{\beta_1} + c_{ij} - \frac{1}{\beta_1}\frac{\omega_i^r \alpha_1^r}{\omega_i^k \alpha_1^k}\right) \tag{13}$$

4.3 Distributions of Costs and Technical Scores

We use simulated method of moments (SMM) to estimate the distributions of costs and technical scores, F_C and F_V, following Pakes and Pollard (1989) and McFadden (1989). The identifying assumption is that differences between observed auction outcomes and those implied by the true

[18]The midpoint-adjusted arc elasticities are 0.55 and 0.69 for the two policy changes. The FCC's 2025 Order adopts these as its empirical estimates of demand elasticity (FCC, 2025).

distributions are orthogonal to a set of instruments. To evaluate prediction error, we evaluate equilibrium for each set of candidate distributions. For auctions in which commissions are endogenous ($\omega_i^k > 0$), we compute the full equilibrium of bids following Asker and Cantillon (2008). For auctions in which commissions are exogenous ($\omega_i^k = 0$), we approximate the full equilibrium using an iterative best-response algorithm following Bajari (2001) and Armantier et al. (2008). Details on these treatments of equilibrium are in Appendices A.1 and A.3, respectively.

For the cost distribution, we use a flexible integrated spline (an I-spline) to parameterize the inverse CDF that maps percentiles to costs. This provides a natural representation because I-splines are continuously differentiable and can be made monotonically non-decreasing under sign constraints, ensuring a valid distribution (see Appendix B.2). For the technical scores, we assume that bidder j in auction i receives

$$v_{ij} = \mu_i + \sigma\, \eta_{ij} \tag{14}$$

where μ_i incorporates auction-level heterogeneity, η_{ij} is an iid shock from a standardized distribution, and σ is a scale parameter that gives the standard deviation of that distribution. We pre-estimate μ_i as the auction-level means of technical scores using the full sample of 155 bids. This is computationally tractable compared to joint estimation of the auction-specific means alongside the other structural parameters in SMM.[19]

Let $l = 1, \ldots, L$ denote simulated auctions, each comprising draws on costs and technical scores for the auction-specific number of bidders, and $\boldsymbol{\theta}$ be the vector of parameters to be estimated. The empirical moments take the form:

$$\mathbf{m}_r^{(q)}(\boldsymbol{\theta}) = \frac{1}{I}\sum_i \left(r_i^{(q)} - \frac{1}{L}\sum_{l=1}^{L} \tilde{r}_{il}^{(q)}(\boldsymbol{\theta}) \right) \mathbf{Z}_i^r \tag{15}$$

$$\mathbf{m}_k^{(q)}(\boldsymbol{\theta}) = \frac{1}{I}\sum_i \left(k_i^{(q)} - \frac{1}{L}\sum_{l=1}^{L} \tilde{k}_{il}^{(q)}(\boldsymbol{\theta}) \right) \mathbf{Z}_i^k \tag{16}$$

where I is the number of auctions, the superscript (q) identifies the rank of the firms by their overall score, $\tilde{r}_{il}^{(q)}(\boldsymbol{\theta})$ and $\tilde{k}_{il}^{(q)}(\boldsymbol{\theta})$ are the simulated rates and commissions, respectively, and $\mathbf{Z}_i^r$ and $\mathbf{Z}_i^k$ are matrices of instruments. For the instrument matrices, the former includes a constant and the scoring-rule weight on rates, and the latter includes a constant and the scoring-rule weight on commissions. As nearly all auctions feature at least three firms, we construct moments for each of the top three bids. Thus, in total, there are 12 moments. Auctions with exogenous commissions contribute only to the rate moments. We use a one-step estimator with $L = 1,000$ and a block-diagonal weighting matrix; each block corresponds to one bid rank and instrument set, with weights given by the diagonal elements of $(\mathbf{Z}'\mathbf{Z})^{-1}$, computed separately for $\mathbf{Z}^r$ and $\mathbf{Z}^k$.

[19] Identification of σ runs through equilibrium information rents rather than the sample's score variance, such that the estimate reconciles observed commissions with the model's predictions. Appendix B.1 provides further details, and Figure D.4 shows a histogram of the raw technical score residuals.

Table 5: Results from Simulated Method of Moments

		Estimate	St. Error
Technical Score Scale	σ	0.059	(0.001)
I-Spline Shifter	τ_0	-1.021	(0.077)
I-Spline Weight (First)	τ_1	1.358	(0.025)
I-Spline Weight (Second)	τ_2	0.282	(0.025)
I-Spline Weight (Third)	τ_3	0.118	(0.015)

Notes: The sample comprises all the auctions in our data except Arkansas in 2014, Michigan in 2018, New Hampshire in 2013, and Vermont in 2016, which have extreme scoring-rule weights. Robust standard errors are in parentheses.

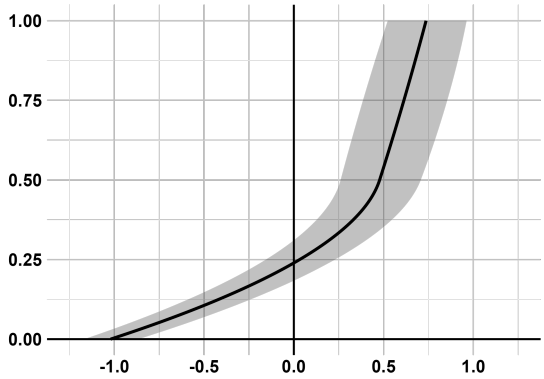


Figure 2: Cumulative Distribution Function of Costs

Notes: The figure plots the CDF of costs (net of ancillary fees) along with a 95% confidence interval. The CDF is obtained by inverting an I-Spline with linear basis functions, a knot at 0.50, and basis function weights that are estimated with simulated method of moments. See Appendix B.2 for details.

Table 5 contains point estimates and standard errors. The within-auction standard deviation of technical scores is 0.059. As the spline parameters are difficult to interpret on their own, we provide the estimated cost distribution in Figure 2, along with a 95% confidence interval. The mean and standard deviation are 0.27 and 0.47, respectively, and the median is 0.48. Costs can be negative, consistent with firms obtaining revenue from ancillary fees. Averaging across auctions and simulation draws, the cost of the winning firm is -$0.07. Thus, our results are consistent with Bazelon et al. (2023), which places explicit costs in the $0.15-$0.18 range for a 15-minute call, provided the winning firms obtain some revenue from fees.

For auctions with endogenous commissions, the model implies a mean rate of $2.10 per 15-minute call and a mean commission of $17.38 per inmate-month, close to the empirical means of $2.27 and $18.54, respectively. The equilibrium rates imply average non-fee revenue of $31.28 per inmate-month, of which the commission accounts for 56%, leaving the winning firm an average profit of $14.40 per inmate-month. For auctions with exogenous commissions, the mean model-predicted rate is $0.71, compared to an empirical mean of $0.76.

Appendix B.4 examines the sensitivity of these results to the demand specification and the

linearity of the contribution functions.

5 Auction Design and the Commission Paradox

We use the empirical model to quantify the commission paradox in the ICS industry. Section 5.1 uses numerical simulations to establish that the paradox emerges under the estimated primitives and to examine how its magnitude varies with market structure and provider heterogeneity. Section 5.2 applies the model to the auctions in our sample, using counterfactual commission bans and maximum commission weights to bracket the welfare consequences of procurement design in ICS.

5.1 Numerical Simulations

We vary the relative weight on commissions and rates in the scoring rule, holding the technical score weight fixed, to trace how expected equilibrium outcomes change as the procurement design shifts toward commissions. Specifically, we set $\omega^v = 0.5$ and let ω^k range from 0.10 to 0.40 with ω^r adjusting one-for-one. We separately consider a scenario with $\omega^k = 0$, which requires us to employ equilibrium approximation. We assume four bidders throughout, corresponding to the median number of bidders in the data.

Figure 3 summarizes the results for $\omega^k \in [0.10, 0.40]$. The top two panels establish the commission paradox. The top-left panel shows that as the commission weight increases, the commission rises, whereas expected consumer surplus and profit both fall. The top-right panel shows that total surplus falls. This is the commission paradox: the auction design that more effectively reduces providers' rents simultaneously decreases total surplus. We also find that the sum of the commission and consumer surplus—what we call public surplus—increases with the commission weight at low values of ω^k, as the commission gains initially offset the consumer surplus losses. These gains, where they exist, accrue entirely to the procuring entity; consumers are strictly worse off. If the procuring entity transfers surplus to incarcerated individuals (e.g., through inmate welfare funds), then modest commission weights may admit Pareto improvements among non-providers. In our simulations, however, the public surplus effects are small.

The bottom-left panel illustrates the mechanism behind these results, showing the empirical density of demeaned pseudotypes under commission weights of 0.10 (solid line) and 0.40 (dashed line). Placing greater weight on the commission reduces providers' rents by compressing the pseudotype distribution, thereby increasing substitutability and intensifying competition (Section 3.2). Indeed, we find that the distribution is visibly tighter with the higher commission weight; the corresponding variances are 177.69 and 38.34, and the expected gap between the best and second-best pseudotypes falls from \$13.13 to \$6.23. This compression reflects the auction mechanism alone, as the underlying cost and technical score distributions are held fixed—the pseudotypes simply become less responsive to firm characteristics as the commission weight increases.

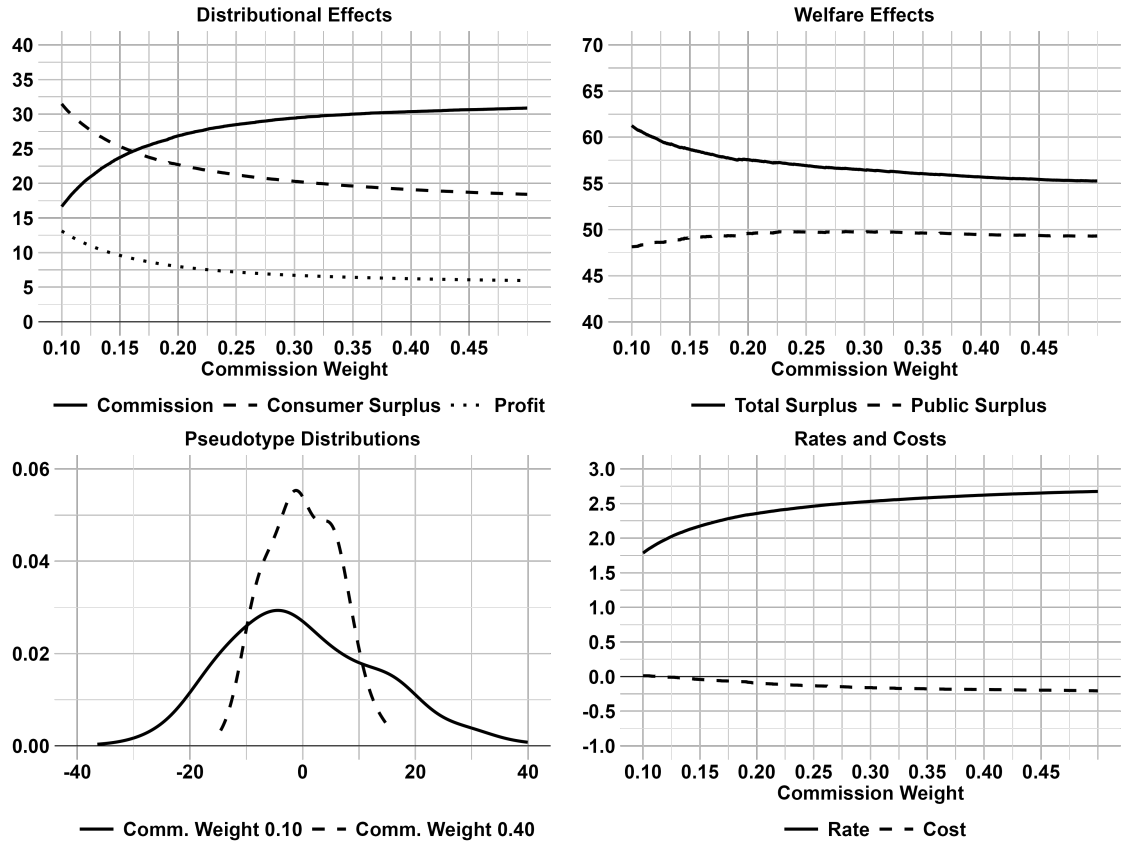


Figure 3: The Commissions Paradox

Notes: The panels summarize counterfactual simulations in which the commission weights vary from 0.10 to 0.40 and rate weights that shift one-for-one from 0.40 to 0.10; the technical score weight is 0.50 throughout.

The bottom-right panel shows that as the commission weight increases, expected rates of the winner rise (solid line), and the expected costs fall (dashed line). The latter effect points to higher commission weights tilting selection toward lower-cost providers (or high-fee providers, as cost is net of fees), which, by Lemma 1, generate higher contract value and can therefore afford larger commissions. Thus, the increase in rates occurs despite a shift toward low-cost providers.

If we remove commissions from the scoring rule entirely, using the iterative best-response algorithm to approximate equilibrium, the outcomes extend the patterns in Figure 3: rates fall further, and total surplus rises. Table D.2 summarizes the results.

Figure 4 illustrates how the primitives that govern the commission paradox interact. The paradox itself is robust across specifications. The top row features eight bidders rather than four. With more competitors, providers earn lower rents, attenuating the commission's effect on total surplus. The bottom row features four competitors, but scales the cost and technical score distributions toward their means by 75%, which significantly reduces provider heterogeneity. With fewer rents to extract, the deadweight loss from higher rates is larger relative to rent extraction. This is visible in

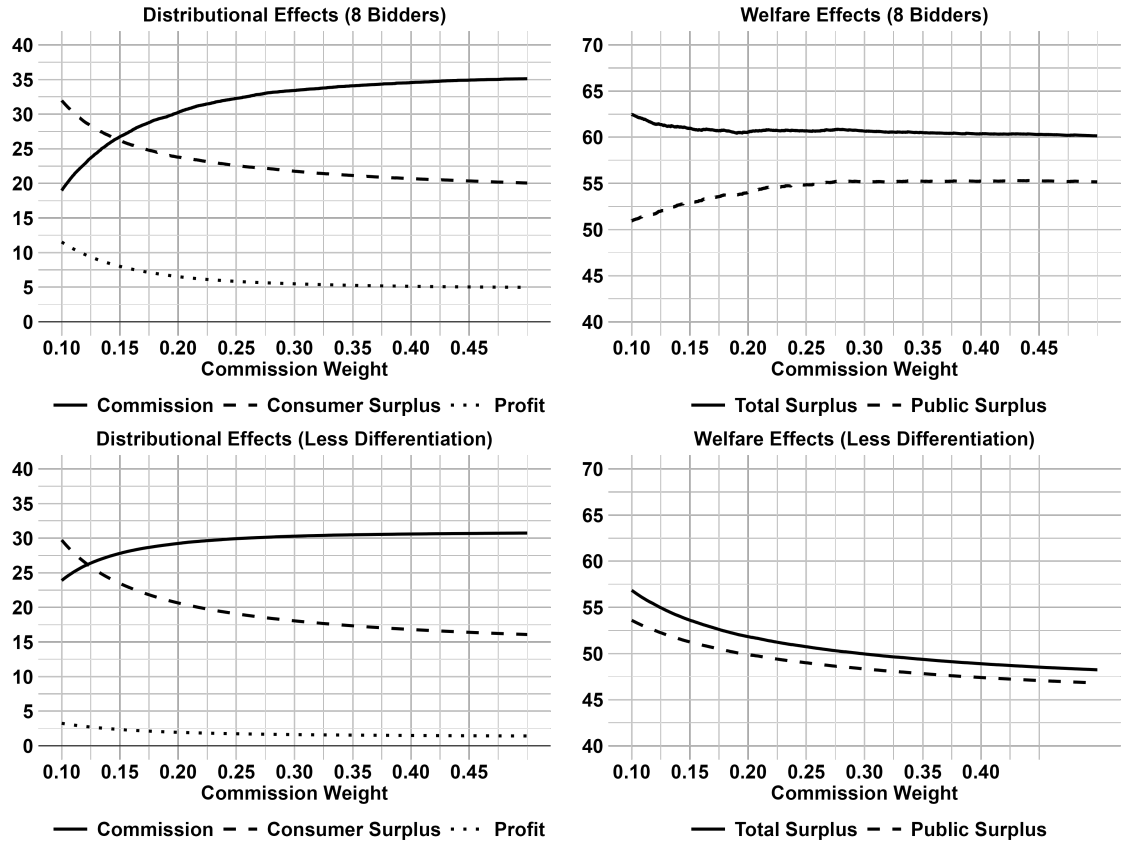


Figure 4: The Commissions Paradox: Sensitivity

Notes: The panels summarize counterfactual simulations in which the commission weights vary from 0.10 to 0.40 and rate weights that shift one-for-one from 0.40 to 0.10; the technical score weight is 0.50 throughout. Relative to our baseline simulations (four bidders with the estimated cost and technical score distributions), the top panels feature eight bidders and the bottom panels reduce the scale of the cost and technical score distributions by 75%.

public surplus, which rises with the commission weight in the top row but falls in the bottom row. More generally, the effect on public surplus depends on the balance between rent extraction and deadweight loss; its sign is an empirical question.

5.2 Commissions in the ICS Industry

Our primary counterfactual is a commission ban. For auctions in which commissions enter the scoring rule, we reallocate the commission weight to rates and incorporate the steeper rate contribution function estimated from auctions without commissions. For auctions with predetermined commissions, we simply eliminate the payment. In both cases, we apply the iterative best-response algorithm to obtain equilibrium strategies under the ban. We also simulate a "maximum commission" counterfactual, in which all non-quality weight is reallocated to commissions, to characterize the range of feasible procurement designs. Together, the two counterfactuals span the welfare

Table 6: Effects of a Commission Ban in the ICS Industry

	All Auctions			*Scored Commissions*			*Predetermined Commissions*		
	Before	After	Change	Before	After	Change	Before	After	Change
Commission	11.16	0.00	-11.16	18.64	0.00	-18.64	4.12	0.00	-4.12
Consumer Surplus	41.12	56.97	15.85	27.07	55.63	28.56	54.34	58.23	3.89
Profit	17.67	19.36	1.70	17.29	19.8	2.51	18.02	18.95	0.93
Total Surplus	69.94	76.33	6.39	63.00	75.43	12.43	76.48	77.18	0.70
Public Surplus	52.28	56.97	4.69	45.71	55.63	9.92	58.46	58.23	-0.23
Rate	1.34	0.52	-0.82	2.08	0.58	-1.50	0.64	0.46	-0.18

Notes: The table reports simulated outcomes before and after a commission ban, averaged across auctions within each subsample. For auctions with scored commissions, the ban reallocates commission weight to rates and incorporates the steeper rate contribution function from Table 3. For auctions with predetermined commissions, the ban eliminates the predetermined payment. Both "before" and "after" results are based on model simulations. The iterative best-response algorithm is used to approximate equilibrium in all "after" columns and in the "before" period for predetermined commissions. We assume the rate is subject to competitive bidding. The sample comprises 33 auctions: 16 with scored commissions and 17 with predetermined commissions. All values are in dollars per inmate-month except the rate, which is the price of a 15-minute local collect call.

consequences of commission weight, from its elimination to its maximum.

Table 6 summarizes the simulated outcomes before and after the commission ban. Averaging across all auctions, the ban reduces expected rates from \$1.34 to \$0.52 and eliminates commission payments of \$11.16 per inmate-month. Consumer surplus rises by \$15.85 and profit rises by \$1.70 per inmate-month, so total surplus increases by \$6.39 per inmate-month, as lower rates restore efficient trades that were suppressed in the equilibria with commissions. Public surplus also increases, as consumer gains exceed the lost commission revenue.

The effects are concentrated in auctions where commissions enter the scoring rule ("scored commissions"). In those auctions, expected rates fall from \$2.08 to \$0.58, consumer surplus doubles, and total surplus rises by \$12.43 per inmate-month. This combination—higher profit, lower rates, higher total surplus—is the commission paradox. The ban also improves outcomes in auctions with predetermined commissions, where rates fall from \$0.64 to \$0.46 and total surplus rises by \$0.70 per inmate-month. The smaller magnitudes reflect that predetermined commissions are generally lower, and many are zero—for those auctions, the ban has no effect, and their inclusion in the average attenuates the average changes reported.

Our second counterfactual maximizes the commission weight, setting $\omega_i^k = 1 - \omega_i^v$ and replacing any predetermined commission payments with scored commissions. The rate weight is set to zero; although rates remain subject to competitive bidding, the equilibrium features bids at the monopoly rate. Table 7 reports the results. Averaging across all auctions, rates rise from \$1.34 to \$2.58 as they move uniformly to the monopoly level, commissions rise from \$11.16 to \$27.78, and total surplus falls by \$10.34 per inmate-month. Public surplus falls by \$4.62 per inmate-month. Relative to the commission ban scenario, the difference in total surplus is \$16.73 per inmate-month, illustrating the range of outcomes available to a procurement authority through the choice of commission weight alone. The status quo captures approximately sixty percent of the total surplus

Table 7: Effects of Maximum Commission Weight in the ICS Industry

	All Auctions			*Scored Commissions*			*Predetermined Commissions*		
	Before	After	Change	Before	After	Change	Before	After	Change
Commission	11.16	27.78	16.63	18.64	26.51	7.87	4.12	28.98	24.87
Consumer Surplus	41.12	19.87	-21.25	27.07	19.62	-7.45	54.34	20.10	-34.23
Profit	17.67	11.95	-5.72	17.29	12.72	-4.57	18.02	11.23	-6.80
Total Surplus	69.94	59.60	-10.34	63.00	58.85	-4.16	76.48	60.31	-16.16
Public Surplus	52.28	47.65	-4.62	45.71	46.13	0.41	58.46	49.09	-9.37
Rate	1.34	2.58	1.24	2.08	2.60	0.51	0.64	2.56	1.92

Notes: The table reports simulated outcomes, averaged across auctions, before and after maximizing the commission weight. For auctions with scored commissions, the commission weight is set to $1 - \omega^v$ and the rate weight to zero. For auctions with predetermined commissions, the predetermined payment is replaced by a scored commission with weight $1 - \omega^v$. Both "before" and "after" results are based on model simulations. The iterative best-response algorithm is used to approximate equilibrium in the "before" period with predetermined commissions. We assume the rate is subject to competitive bidding. The sample comprises 33 auctions: 16 with scored commissions and 17 with predetermined commissions. All values are in dollars per inmate-month except the rate, which is the price of a 15-minute local collect call.

gains available from a ban, with the remainder reflecting the additional total surplus losses that heavier reliance on commissions could impose.

To put these numbers in context, approximately 1.2 million individuals are housed in state-level prison systems. The commission ban generates an average total surplus gain of $6.39 per inmate-month, implying aggregate gains of roughly $92 million per year from state prisons alone. County jails, which house an additional 700,000 individuals and often have higher commission rates, are outside our sample but would increase this figure. These calculations assume that estimated cost and demand parameters are representative across state systems; to the extent systems differ in scale or competitive conditions, the extrapolation is approximate.

6 Additional Policy Analyses

We consider three additional policy analyses: First, we examine a regulation that predetermines rates but allows competition to determine the magnitude of commission payments. Second, we examine regulation that caps competitively determined rates. Third, to inform competition policy, we examine how outcomes vary with the number of competitors.

6.1 Rate Regulation

We first explore regulation that determines rates; the procuring entity still conducts a scoring auction that determines the commission payment. We examine how outcomes vary with the regulated rate. The analysis can be motivated by recently passed state laws in Connecticut, California, Colorado, Minnesota, and Massachusetts that make calls free for incarcerated individuals; these states

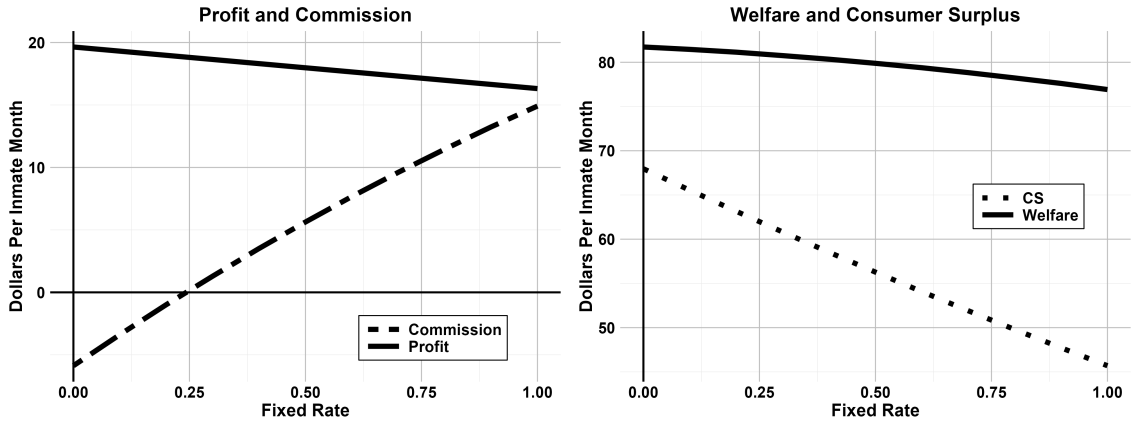


Figure 5: The Economic Effects of Rate Regulation

Notes: The figure is based on 1,000 simulations of an auction with four bidders, scoring rule weights of $\omega^k = 0.33$ and $\omega^v = 0.67$, and regulated rates between \$0.00 and \$1.00 per 15-minute call (horizontal axis). The left panel shows the expected equilibrium commission payment and profit. The right panel shows the expected consumer surplus and welfare generated by the winning firm.

pay or will pay for ICS, which could be interpreted as a negative commission.[20] Our starting point is an auction with four bidders and scoring rule weights of $\omega^k = 0.33$ and $\omega^v = 0.67$. We compute equilibrium for rates from \$0.00 to \$1.00 per 15-minute call.

Figure 5 presents the results. The left panel plots the expected equilibrium commission and profit. With a \$1.00 regulated rate, the commission payment is \$14.91 per inmate-month. As the regulated rate falls (from right to left), the commission decreases because the winning firm earns less revenue from calls and can therefore pay less in commission. For regulated rates below \$0.25, the commission is negative, so the flow of funds reverses and the buyer pays the provider. With free calls, the procuring entity pays the provider \$5.87 per inmate-month. We do not observe payments from the procuring entity to the provider in our data, but they are standard in other settings.

Profit increases as the regulated rate falls. Mechanically, the loss of revenue from rates and fees is more than offset by smaller commission payments. This occurs because rate regulation increases the winning provider's market power by disadvantaging higher-cost providers in the bidding process. Although all firms lose profit from calls as regulated rates fall, this effect is amplified for high-cost firms. These disadvantaged firms bid less aggressively, enabling the winning provider to obtain better financial terms. As regulated rates fall, the provider revenue increasingly comes from the buyer via a fixed payment rather than from incarcerated consumers via the rate.

The right panel illustrates the welfare effects of rate regulation. As the regulated rate decreases from right to left, the expected consumer surplus generated by the ICS contract increases (dotted line). Welfare also increases, but to a more modest extent (solid line). This reflects that, as rates fall, commission payments decline, but this is more than offset by higher consumer surplus and

[20]For example, Connecticut Senate Bill 972 (2021) states: "The annualized cost for paying the [ICS] vendor for telephone services is approximately \$4.5-\$5.5 million per year...."

a moderate increase in profit. Thus, in the model, rate regulation shifts the cost of ICS from consumers to the procuring entity. Enacted in isolation, it also increases the market power and profit of the winning firms by placing higher-cost providers at a disadvantage in the auction.

6.2 Rate Caps

We now consider regulation that caps competitively determined rates at $1.35 per 15-minute call, a level we select to be consistent with the FCC's 2025 Order. Such a cap directly affects equilibrium rates whenever those rates exceed the cap; it can also have indirect effects if it screens out firms with costs above the cap. In this section, we analyze the indirect effects as auction participation is exogenous in our model. Also in keeping with the Order, we examine the distribution of costs in an auction without commissions and determine the frequency with which costs exceed the cap.

We use an auction with four bidders and scoring rule weights of $\omega^v = 0.67$ and $\omega^r = 0.33$. The results are summarized in the top panels of Figure 6, which provide histograms of the cost distributions of the first-best firm (the winner) and each of the other firms. We obtain the distributions using 1,000 simulations. The vertical red line in each panel is the rate cap. Taking the firms in turn, the fractions of simulations in which their costs are below the cap are 98.6%, 90.2%, 65.8%, and 25.4%, respectively. As the third- and fourth-best firms are the least competitively significant, the exercise indicates that the rate cap may have modest indirect effects through auction participation, assuming that bidders observe their costs before entry (e.g., as in the canonical model of Samuelson, 1985).

However, the FCC Order also bans ancillary fees. In our model, where costs are net of fee revenue, eliminating fees would raise costs for all providers. To examine the sensitivity of our participation results to this channel, we shift the cost distribution upward by $0.50 and recompute the equilibrium in 1,000 simulations. The histograms of the resulting cost distributions are in the lower panels of Figure 6. Taking the firms in turn, the fractions of simulations in which their costs fall below the cap are 96.7%, 78.6%, 46.4%, and 12.5%, respectively. Only 45.6% of auctions would have at least three bidders if firms bid when their costs are below the cap. Thus, when paired with a fee ban, the rate cap could moderately affect participation. These results should be interpreted with some caution, as we cannot separately identify fee revenue from the cost of service. If fee revenue varies across providers, a uniform cost shift may not fully capture the effects of a fee ban. Provider-level financial data, which are not publicly available, could help resolve this.

6.3 Competition Policy

Our final analysis examines how outcomes vary with the number of competitors. We motivate the exercise by the industry's extended period of consolidation and the Department of Justice's recent decision to block the Securus and ICS merger. In our empirical model, adding a firm affects outcomes for two reasons. First, it reduces the probability that any given firm wins, thereby inducing

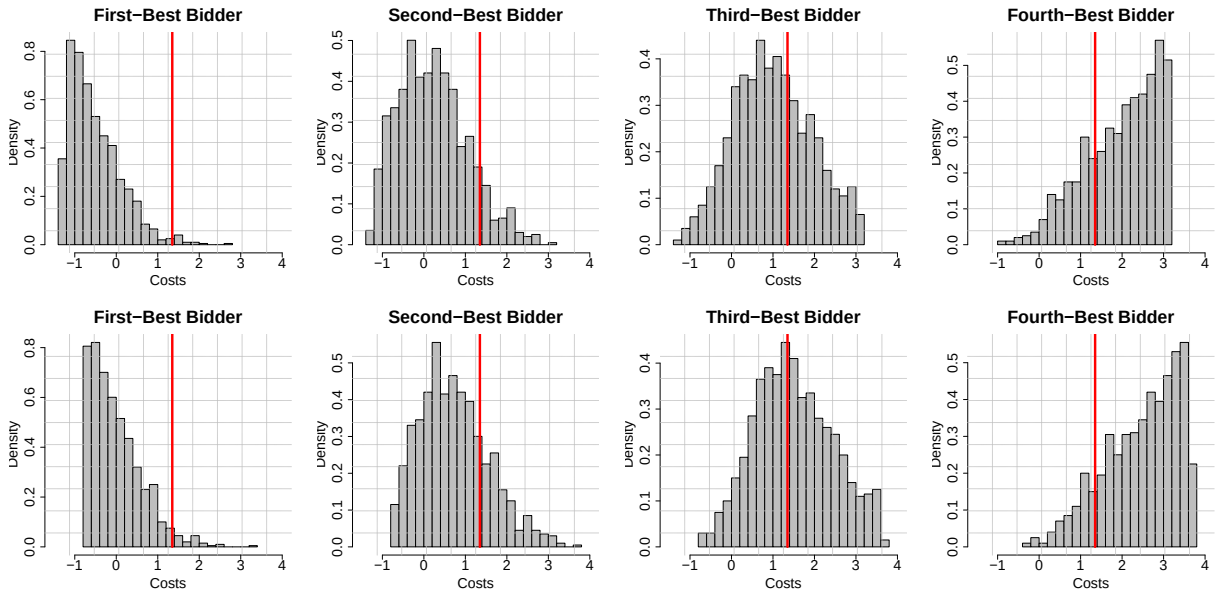


Figure 6: The Cost Distribution and the Rate Cap by Bidder

Notes: From left to right, the figure provides histograms of the cost distribution for the winning firm, the second-best firm, the third-best firm, and the fourth-best firm. The vertical line plots the 2025 imposed rate caps by the FCC of $1.35 per 15-minute call. The top panels are based on 1,000 simulations of an auction with four bidders and scoring rule weights of $\omega^r = 0.33$ and $\omega^v = 0.67$. The bottom panels incorporate a $0.50 rightward shift in the cost distribution but otherwise are based on the same assumptions.

more aggressive bidding. Second, it improves the expected characteristics of the winner (in our model, this implies lower cost and better technical capabilities). We refer to these as a *competition effect* and a *composition effect*, respectively.

We first consider an auction with commissions in the scoring rule. In that context, our description of the composition and competition effects can be sharpened. The composition effect occurs because the expected first-best pseudotype increases with the number of firms, whereas the competition effect occurs because the expected second-best pseudotype increases. We consider an auction with a technical weight of 0.67 and rate and commission weights of 0.165, and compute expected equilibrium outcomes with different numbers of competitors. We report "unadjusted" results that combine the competition and composition effects. We also report "adjusted" results obtained by rescaling the variances of the cost and technical score distributions such that their expected minimums and maximums coincide with those from duopoly. Thus, the expected characteristics of the winner do not vary with the number of firms, allowing us to isolate the competition effect.

Figure 7 shows the results. In the top panels, both the rate and the commission are subject to competitive bidding. The top-left panel focuses on the expected equilibrium rate. As we increase the number of firms from two to six, the unadjusted rate decreases from $2.39 to $1.92, whereas the adjusted rate remains flat. Thus, any changes in the rate are due to compositional effects. This tracks our theoretical results, which establish that the winning firm chooses rates to maximize

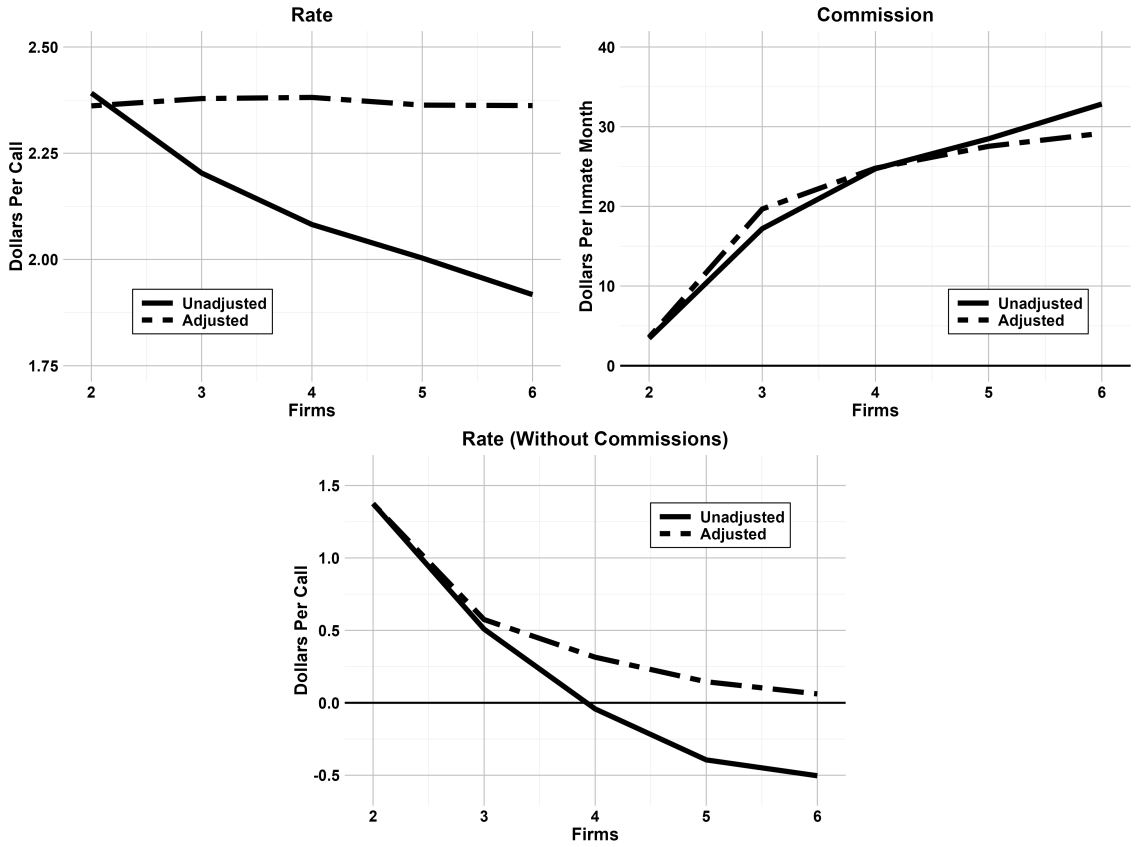


Figure 7: The Effect of Competition on Rates and Commissions

Notes: The panels depict simulated rates and commissions as the number of firms varies from two to six. In the top panels, both the rate and the commission are subject to competitive bidding. In the bottom panel, on the rate is subject to competitive bidding; there is not commission. The unadjusted trends combine two effects: a competition effect (more aggressive bidding) and a composition effect (more firms lower the expected cost of the winner). Adjusted trends rescale the cost and technical score distributions so that the expected best cost and quality remain constant, isolating the competition effect.

the bilateral surplus it creates with the procuring entity, not based on the identity or number of competitors. The top-right panel, in contrast, shows that the expected equilibrium commission increases with the number of firms, in both the unadjusted and adjusted cases. These results indicate that adding more firms benefits the procuring entity when commissions are subject to competitive bidding, but whether consumers gain depends on the characteristics of the new firms.

The bottom panel considers auctions in which commissions are not in the scoring rule. In that setting, the rate is the only financial term subject to competition, and adding firms unambiguously lowers rates. In our experiments, we assume a technical weight of 0.67 and a rate weight of 0.33. As shown, both the unadjusted and adjusted expected equilibrium rates decrease. Unadjusted rates decline from \$1.37 to about -\$0.50, whereas adjusted rates decrease to \$0.06 as we increase the number of firms from two to six. (Recall that costs in our model are net of fees and so can be negative.) When commissions are absent, the link between competition and rates is robust.

7 Conclusion

Procurement designs that more effectively reduce provider rents do not necessarily improve outcomes for end users. In the ICS industry, we find that shifting scoring weight toward commissions reduces provider rents, but raises rates and reduces total surplus—the commission paradox. We connect this to economic primitives using a first-score auction model in the spirit of Asker and Cantillon (2008). As scoring weight shifts toward commissions, providers internalize less downward pressure on rates, and equilibrium rates rise toward monopoly levels. Simultaneously, providers' competitive positions become less differentiated, eroding rents. These two forces—higher rates and lower rents—are driven by the same shift in the scoring rule. Thus, a primary tool available to intermediaries for extracting upstream rents can simultaneously harm downstream consumers.

The trade-off we identify between rent extraction through commissions and end-user welfare may be relevant in other intermediated procurement settings, though institutional differences will determine the extent to which it operates. A key determinant is likely consumers' ability to select among intermediaries. When intermediaries compete for consumers, competition may discipline commissions, attenuating the paradox. However, if commissions can be shared with consumers—through lower fees, better service, or direct transfers—competition could instead sustain or amplify commissions in equilibrium. In this light, ICS represents one pole: incarcerated individuals cannot switch providers and receive no such transfers, so the only discipline on prices is the demand response. Whether the paradox attenuates or amplifies with competition among intermediaries is an empirical question that we leave open, and a natural direction for future research.

Within ICS specifically, recent state-level commission bans and the FCC's 2024 and 2025 Orders create opportunities to test the commission paradox with reduced-form methods, complementing our model-based approach. More broadly, Raher (2020) highlights that prices are inflated for incarcerated individuals across many goods and services beyond telecommunications, suggesting that procurement distortions in correctional settings could extend well beyond ICS. Future research could examine how call prices affect broader outcomes such as recidivism and post-release employment, or explore the effects of visitation and social connectedness more broadly.

References

Abito, Jose Miguel, Kathleen Hui, Yuval Salant, and Kosuke Uetake, "Demand Spillover and Inequality in the WIC Program," 2022. Working Paper.

Allen, Jason, Robert Clark, and Jean-François Houde, "Price Dispersion in Mortgage Markets," *The Journal of Industrial Economics*, 2014, *62* (3), 377–416.

_ , _ , Brent Hickman, and Eric Richert, "Resolving Failed Banks: Uncertainty, Multiple Bidding, and Auction Design," *Review of Economic Studies*, 2024, *91* (3), 1201–1242.

An, Yonghong, David Davis, Yizao Liu, and Ruli Xiao, "Procurement in Welfare Programs: Evidence and Implications from WIC Infant Formula Contracts," 2023. Working Paper.

Armantier, Olivier, Jean-Pierre Florens, and Jean-Francois Richard, "Approximation of Nash Equilibria in Bayesian Games," *Journal of Applied Econometrics*, 2008, *23*, 965–981.

Asker, John and Estelle Cantillon, "Properties of Scoring Auctions," *RAND Journal of Economics*, 2008, *39* (1), 69–85.

Badoer, Dominique C., Charlie Costello, and Christopher M. James, "I Can See Clearly Now: The Impact of Disclosure Requirements on 401(K) Plans," *Journal of Financial Economics*, 2020, *136*, 471–489.

Bajari, Patrick, "Comparing Competition and Collusion: A Numerical Approach," *Economic Theory*, 2001, *18*, 187–205.

Baker, Allison, Patrick Brogan, Octavian Carare, Nicholas Copeland, Patrick DeGraba, Steven Kauffman, Paul Lafontaine, Catherine Matreves, Jeffrey Prince, Sean Sullivan, Patrick Sun, and Emily Talaga, "Economics at the FCC 2019-2020: Spectrum Policy, Universal Service, Inmate Calling Services, and Telehealth," *Review of Industrial Organization*, 2020, *57*, 827–858.

Bazelon, Coleman, Paroma Sanyal, L. Shea Peretz, Preetul Sen, Natalie Selfe, and Sarah Mather, "Inmate Calling Services Cost Model: A Model Carrier Approach to Estimating the Communications Services Cost for the U.S. Incarcerated Population," 2023. Available at SSRN: `https://ssrn.com/abstract=4530948`.

Benetton, Matteo, Alessandro Gavazza, and Paolo Surico, "Mortgage Pricing and Monetary Policy," *American Economic Review*, 2025, *115* (3), 823–863.

Bhattacharya, Vivek and Gastón Illanes, "The Design of Defined Contribution Plans," 2025. Working Paper.

_ , Andrew Ordin, and James W. Roberts, "Bidding and Drilling Under Uncertainty: An Empirical Analysis of Contingent Payment Auctions," *Journal of Political Economy*, 2022, *130* (5).

_ , **Gastón Illanes, and Manisha Padi**, "Fiduciary Duty and the Market for Financial Advice," *Econometrica*, 2025, *93* (4), 1449–1480.

Bolotnyy, Valentin and Shoshana Vasserman, "Scaling Auctions as Insurance: A Case Study in Infrastructure Procurement," *Econometrica*, 2023, *91* (4), 1205–1259.

Carril, Rodrigo, Andres Gonzalez-Lira, and Michael S. Walker, "Competition Under Incomplete Contracts and the Design of Procurement Policies," *American Economic Review*, 2025. Forthcoming.

Che, Yeon-Koo, "Design Competition through Multidimensional Auctions," *RAND Journal of Economics*, 1993, *24* (4), 668–680.

de Boor, Carl, *A Practical Guide to Splines*, revised ed., Springer, 2001.

FCC, "Rates for Interstate Inmate Calling Services," 2012. WC Docket No. 12-375, Notice of Proposed Rulemaking.

_ , "Rates for Interstate Inmate Calling Services," 2013. WC Docket No. 12-375, Report and Order and Further Notice of Proposed Rulemaking.

_ , "Rates for Interstate Inmate Calling Services," 2015. WC Docket No. 12-375, Second Report and Order and Third Further Notice of Proposed Rulemaking.

_ , "Rates for Interstate Inmate Calling Services," 2021. WC Docket No. 12-375, Third Report and Order, Order on Reconsideration, and Fifth Further Notice of Proposed Rulemaking.

_ , "Incarcerated People's Communications Services; Implementation of the Martha Wright-Reed Act; Rates for Interstate Inmate Calling Services," 2024. WC Docket No. 12-375, Report and Order, Order on Reconsideration, Clarification and Waiver, and Further Notice of Proposed Rulemaking.

_ , "Incarcerated People's Communications Services; Implementation of the Martha Wright-Reed Act; Rates for Interstate Inmate Calling Services," 2025. WC Docket Nos. 23-62 and 12-375, Report and Order and Further Notice of Proposed Rulemaking.

Grossman, Sanford J. and Oliver D. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica*, 1983, *51* (1), 7–45.

Grunewald, Andreas, Jonathan A. Lanning, David C. Low, and Tobias Salz, "Auto Dealer Loan Intermediation: Consumer Behavior and Competitive Effects," 2023.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong, "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 2000, *68* (3), 525–574.

Ho, Kate and Robin S. Lee, "Contracting over Pharmaceutical Formularies and Rebates," 2024. Working Paper.

Holmström, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, 1979, *10* (1), 74–91.

Hristakeva, Sylvia, "Vertical Contracts with Endogenous Product Selections: An Empirical Analysis of Vendor-Allowance Contracts," *Journal of Political Economy*, 2022, *130* (12), 3202–3252.

Jackson, Steven J., "Ex-Communication: Competition and Collusion in the U.S. Prison Telephone Industry," *Critical Studies in Media Communication*, 2005, *22* (4), 263–280.

Kang, Karam and Robert A. Miller, "Winning by Default: Why is There So Little Competition in Government Procurement?," *Review of Economic Studies*, 2022, *89* (3), 1495–1556.

Kong, Yunmi, Isabelle Perrigne, and Quang Vuong, "Multidimensional Auctions of Contracts: An Empirical Analysis," *American Economic Review*, 2022, *112* (5), 1703–1736.

Laffont, Jean-Jacques and Jean Tirole, *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press, 1993.

_ , Herve Ossard, and Quang Vuong, "Econometrics of First-Price Auctions," *Econometrica*, 1995, pp. 953–980.

McFadden, Daniel, "A Method of Simulated Moments for Estimation of Discrete Response Models," *Econometrica*, 1989, *57* (5), 995–1026.

Myerson, Roger B., "Optimal Auction Design," *Mathematics of Operations Research*, 1981, *6* (1), 58–73.

Nam-Sonenstein, Brian, "Shadow Budgets: How Mass Incarceration Steals from the Poor to Give to the Prison," 2024. Report of the Prison Policy Initiative, Available at `https://www.prisonpolicy.org/reports/shadowbudgets.html`.

Pakes, Ariel and David Pollard, "Simulation and Asymptotics of Optimization Estimators," *Econometrica*, 1989, *57* (5), 1027–1057.

Pool, Veronika K., Clemens Sialm, and Irira Stefanescu, "Mutual Fund Revenue Sharing in 401K Plans," *Management Science*, 2024. Forthcoming.

Raher, Stephen, "The Company Store and the Literally Captive Consumers: Consumer Law in Prisons and Jails," *Hastings Race and Poverty Law Journal*, 2020, *17* (1).

Riley, John G. and William F. Samuelson, "Optimal Auctions," *American Economic Review*, 1981, *71* (3), 381–392.

Samuelson, William F., "Competitive Bidding with Entry Costs," *Economics letters*, 1985, *17* (1-2), 53–57.

Schumaker, Larry L., *Spline Functions: Basic Theory*, Wiley, 1981.

Shaked, Moshe and J. George Shanthikumar, *Stochastic Orders*, New York: Springer, 2007.

Shields, Todd, "Prison Phones Prove Captive Market for Private Equity," *Bloomberg BusinessWeek*, 2012. (October 4).

Spengler, Joseph J., "Vertical Integration and Antitrust Policy," *Journal of Political Economy*, 1950, *58* (4), 347–352.

Appendix

For Online Publication

A Theoretical Model

We provide details on the Bayes-Nash equilibrium in auctions with commissions in the scoring rule ($\omega_t^k > 0$), prove the results in Section 3, and finally provide more details about how we approximate Bayes-Nash equilibrium for other auctions ($\omega_t^k = 0$).

A.1 Equilibrium with Commissions in the Scoring Rule

We characterize equilibrium for auctions in which commissions enter the scoring rule ($\omega_t^k > 0$) based on a reformulation of the model developed in Asker and Cantillon (2008). We use Assumptions 6 and 7, which are more general than Assumptions 1 and 2. We also use Assumption 3.

First-Order Conditions and Equilibrium Rates

Our model falls within the class of quasi-linear scoring auctions studied in Asker and Cantillon (2008) when commissions enter the scoring rule ($\omega_i^k > 0$). Differentiating the expected profit function on the right-hand side of (1) with respect to rates and commissions, and rearranging, yields two first-order conditions that characterize optimal bids:

$$\Pi_r(r;c) = -\frac{P_r}{P_k}(r,k;v) \tag{A.1}$$

$$\Pi(r;c) = k + \frac{P}{P_k}(r,k;v) \tag{A.2}$$

where subscripts denote derivatives. The first-order condition in (A.1) equates the marginal profit from raising the rate to the marginal rate of substitution between rate score and commission score in the scoring rule. The first-order condition in (A.2) determines how aggressively the firm bids on commissions, weighing the cost of higher payments against an improved win probability.[21]

A key result is that, under Assumptions 1 and 3, the rate that solves (A.1) is independent of both k and v. First, these terms do not affect the contract value, Π, and thus are excluded from Π_r.

[21] Analogous first-order conditions obtain generally in scoring auction models and can form the basis of nonparametric identification and estimation under a best-response approach (Guerre et al., 2000; Kong et al., 2022). The first-order condition in (A.1) is obtained by plugging (A.2) into the derivative of expected profit with respect to the rate. It thus characterizes profit-maximizing rates conditional on the firm bidding a profit-maximizing commission.

Second, the iso-P slope given by $\frac{P_r}{P_k}(r,k;v)$ is independent of v under additivity and independent of k under linearity of $r_i^k(\cdot)$. Specifically:

$$\frac{P_r}{P_k}(r,k;v) = \frac{P_i'(s)\omega_i^r s_i^{r\prime}(r)}{P_i'(s)\omega_i^k \alpha} = \frac{\omega_i^r s_i^{r\prime}(r)}{\omega_i^k \alpha} \tag{A.3}$$

This delivers a triangular structure where the optimal rate bid, $r^*(c)$, solves (A.1) and the optimal commission bid, $k^*(r^*(c),v)$, solves (A.2) given $r^*(c)$. Furthermore, the optimal rate bid maximizes the *bilateral surplus* of the bidder and the buyer defined by a generalized version of (3). To see this, differentiating bilateral surplus with respect to r yields a first-order condition

$$\frac{\partial}{\partial r}\left[\psi_i(r,v) + \Pi(r;c)\right] = \psi_r(r) + \Pi_r(r;c) = 0 \quad \Rightarrow \quad \Pi_r(r;c) = -\psi_r(r)$$

that exactly matches (A.1) and (A.3), given the expression for $\frac{\partial}{\partial r}\psi_i(r,v)$ implied by the generalized version of (3).

Thus, equilibrium rate bids are non-strategic, as they are independent of the win probability and free from expectations over competing bids. They are also unique and strictly increasing in the bidder's cost because bilateral surplus is strictly concave in r under Assumptions 1 and 3. Strategic considerations enter through the commission.

Pseudotypes and Equilibrium Commissions

We characterize equilibrium using a strategically equivalent reformulation in which firms submit expected scores (in currency units) rather than commissions. This involves only a change of variables. In the original auction, the score is $\psi_i(r,v)+k$, so choosing a score is equivalent to choosing a commission. Consistent with this change of variables, the commission paid by a firm that wins with a score bid of $\hat{s}$ is given by

$$k = \hat{s} - \psi_i(r^*, v) \tag{A.4}$$

The winner of the reformulated auction is the bidder with the highest score. Thus, the reformulation preserves both the allocation rule and the payments.

Applying a change of variables to equation (1), the equilibrium strategy satisfies:

$$\sigma_i(c_{ij}, v_{ij}) = \arg\max_{\tilde{s}} (x_{ij} - \tilde{s}) P_i(\tilde{s}) \tag{A.5}$$

where $P_i(\cdot)$ is the win probability, as used in (1), and x_{ij} is the *pseudotype* of the bidder, defined in (4) as the bilateral surplus it can create with the unique optimal rate bid. Thus, the first term in (A.5) is the firm's profit if it wins: the bilateral surplus it creates minus what it delivers to the buyer. Together with the equilibrium rate from (A.1), the solution to equation (A.5) characterizes equilibrium bidding. Furthermore, firms with the same pseudotype adopt the same equilibrium strategy, i.e., $\sigma_i(c_{ij}, v_{ij}) = \sigma_i(x_{ij})$, and the multidimensional type space (c,v) collapses to a single

index (x), as in Asker and Cantillon (2008). In equilibrium, the score bid is strictly increasing in the pseudotype, so firms with higher pseudotypes are more likely to win.

The win probability is a function of the distribution of pseudotypes (denoted by $T_i()$) and equals the probability that $J_i - 1$ bidders have a lower pseudo-type:

$$P_i(win|s_{ij}) = T_i\big(\sigma_i^{-1}(s_{ij})\big)^{J_i-1} \tag{A.6}$$

The distribution of pseudo-types is given by

$$T_i(x) = \int\int \mathbb{1}\big(x_i^*(c,v) \leq x\big) dF_C(c) dF_V(v) \tag{A.7}$$

where $x_i^*(c,v)$ is the pseudotype of a bidder with cost c and match value v (equation (4)). A closed-form solution can now be derived. Denoting the support of $T_i(\cdot)$ by $[\underline{t}, \bar{t}]$ and adding the boundary condition that $\sigma_i(\underline{t}) = \underline{t}$, the equilibrium (score) bidding strategy is given by

$$s_{ij}^* = \sigma_i(x_{ij}^*) = x_{ij}^* - \frac{\int_{\underline{t}}^{x_{ij}^*} T_i(u)^{J_i-1} du}{T_i(x_{ij}^*)^{J_i-1}} \tag{A.8}$$

Finally, the equilibrium commission is given by

$$k_{ij}^* = s_{ij}^* - \psi_i(r_{ij}^*, v_{ij}) \tag{A.9}$$

Equations (A.1) and (A.9) characterize the unique symmetric Bayes-Nash equilibrium for any auction in which rates and commissions both receive weight in the scoring rule (Che (1993), Asker and Cantillon (2008)). If commissions receive weight in the scoring rule but rates do not, this reduces to an independent private values first price-auction for which equation (A.9) alone characterizes the unique symmetric Bayes-Nash equilibrium.

In estimation, we compute the full equilibrium for every candidate parameter vector by evaluating (A.8) directly, estimating $T_i(\cdot)$ with a Gaussian kernel (Sheather-Jones bandwidth) and computing the integral via cumulative trapezoidal integration on a fine grid. Equilibrium commissions then follow from (A.9). In many of our counterfactuals, for which we need only the commission of the winning bidder, we follow Laffont et al. (1995) and use revenue equivalence to circumvent solving the full equilibrium. Revenue equivalence holds in our setting under quasi-linearity, risk neutrality, and i.i.d. types (Myerson, 1981). As the allocation depends only on pseudotypes, equilibrium payments can be characterized using a second-score construction: the winner's score (in currency units) equals the second-highest pseudotype among bidders. From (A.4), this determines the commission paid by the winner. In this construction, the profit of the winner equals $x^{(1)} - x^{(2)}$ where the superscripts refer to the rank of pseudotypes.

A.2 Proofs

Lemma 1

Proof. We prove the lemma under Assumptions 6 and 7, which are more general than Assumptions 1 and 2. Let $\lambda \equiv \omega^k/\omega^r$. Under the conditions of the proposition ($d\omega^k > 0$, $d\omega^r \leq 0$), λ is increasing whenever $\omega^r > 0$. Since the equilibrium rate depends on the scoring rule weights only through λ, it suffices to show the results with respect to λ. If $\omega^r = 0$, the equilibrium rate equals $r^M(c)$ regardless of ω^k, and the comparative statics hold trivially at the boundary.

Part (i). From Appendix A.1, the equilibrium rate solves the first-order condition for bilateral surplus maximization:

$$\Pi_r(r; c) + \frac{1}{\lambda\alpha}\, s^{r\prime}(r) = 0$$

Differentiating implicitly:

$$\frac{dr^*}{d\lambda} = \frac{s^{r\prime}(r^*)/(\lambda^2\alpha)}{\Pi_{rr}(r^*; c) + \frac{1}{\lambda\alpha}\, s^{r\prime\prime}(r^*)}$$

The numerator is negative ($s^{r\prime} < 0$, $\alpha > 0$). The denominator is strictly negative: $\Pi_{rr} < 0$ by Assumption 3 and $s^{r\prime\prime} < 0$ by Assumption 1. Hence $dr^*/d\lambda > 0$. As $\lambda \to \infty$, the first-order condition becomes $\Pi_r(r; c) = 0$, whose unique solution is $r^M(c)$.

Part (ii). $\frac{dCS}{d\lambda} = \frac{d}{d\lambda}\int_{r^*}^{\bar{r}} q(p)\, dp = -q(r^*)\, \frac{dr^*}{d\lambda} < 0$, since $q(r^*) > 0$ and $dr^*/d\lambda > 0$.

Part (iii). $\frac{d\Pi}{d\lambda} = \Pi_r(r^*; c)\, \frac{dr^*}{d\lambda} > 0$, since $r^* < r^M$ implies $\Pi_r(r^*; c) > 0$ by concavity.

Part (iv). $\frac{dTS}{d\lambda} = [-q(r^*) + \Pi_r(r^*; c)]\, \frac{dr^*}{d\lambda}$. Using $\Pi(r; c) = (r - c)q(r)$, the bracketed term equals $(r^* - c)\, q'(r^*) < 0$ since $r^* > c$ and $q' < 0$. Combined with $dr^*/d\lambda > 0$, we have $dTS/d\lambda < 0$. □

Lemma 2

Proof. We prove the lemma under Assumptions 6 and 7, which are more general than Assumptions 1 and 2. Furthermore, we use $Cov(c, v) \leq 0$, which is weaker than was is stated in Assumption 7. We work directly with ω^k as the varying parameter, with $d\omega^r/d\omega^k \leq 0$, $d\omega^v/d\omega^k \leq 0$, and $d\omega^r/d\omega^k + d\omega^v/d\omega^k = -1$.

Variance Result. Write the pseudotype as

$$x(c, v) = x^c(c; \omega^k) + \frac{\omega^v}{\omega^k\alpha}\, v$$

where $x^c(c; \omega^k) \equiv \Pi(r^*(c); c) + \frac{\omega^r}{\omega^k\alpha}\, s^r(r^*(c))$ collects the cost-driven terms. The variance decomposes as

$$\mathrm{Var}(x) = \mathrm{Var}[x^c] + \left(\frac{\omega^v}{\omega^k\alpha}\right)^2 \mathrm{Var}(v) + 2\,\frac{\omega^v}{\omega^k\alpha}\,\mathrm{Cov}(x^c,\, v)$$

We show that each term decreases with ω^k.

Term 1: $\mathrm{Var}[x^c]$. Using the identity $\frac{\partial \mathrm{Var}[g]}{\partial \omega^k} = 2\,\mathrm{Cov}\Big[g,\, \frac{\partial g}{\partial \omega^k}\Big]$, we compute $\partial x^c/\partial \omega^k$. Since r^* maximizes $\Pi(r;c) + \frac{\omega^r}{\omega^k \alpha}\, s^r(r)$, the envelope theorem eliminates indirect effects through r^*, leaving

$$\frac{\partial x^c}{\partial \omega^k} = \frac{d}{d\omega^k}\left(\frac{\omega^r}{\omega^k}\right)\frac{1}{\alpha}\, s^r(r^*(c))$$

Since $d\omega^r/d\omega^k \leq 0$ and $\omega^r \geq 0$, we have $\frac{d}{d\omega^k}\big(\frac{\omega^r}{\omega^k}\big) = \frac{\omega^k\, d\omega^r/d\omega^k - \omega^r}{\omega^{k\,2}} \leq 0$. Because $r^*(c)$ is increasing in c and $s^{r\prime} < 0$, the function $s^r(r^*(c))$ is decreasing in c, so $\partial x^c/\partial \omega^k$ is increasing in c. Meanwhile, x^c is decreasing in c (since $\Pi_c < 0$ and $s^r(r^*(c))$ is decreasing in c). A decreasing function and an increasing function have negative covariance, so

$$\frac{\partial \mathrm{Var}[x^c]}{\partial \omega^k} = 2\,\mathrm{Cov}\left[x^c,\, \frac{\partial x^c}{\partial \omega^k}\right] < 0$$

Term 2: $\big(\frac{\omega^v}{\omega^k \alpha}\big)^2 \mathrm{Var}(v)$. Since $d\omega^v/d\omega^k \leq 0$ and ω^k is increasing, the ratio ω^v/ω^k is decreasing:

$$\frac{d}{d\omega^k}\left(\frac{\omega^v}{\omega^k}\right) = \frac{\omega^k\, d\omega^v/d\omega^k - \omega^v}{\omega^{k\,2}} \leq 0$$

Hence the coefficient $(\omega^v/(\omega^k\alpha))^2$ decreases, and $\mathrm{Var}(v)$ is fixed, so this term decreases.

Term 3: $2\,\frac{\omega^v}{\omega^k \alpha}\,\mathrm{Cov}(x^c,\, v)$. We first sign the covariance. Since x^c is decreasing in c and $\mathrm{Cov}(c, v) \leq 0$, higher v is associated with lower c and hence higher x^c, so $\mathrm{Cov}(x^c, v) \geq 0$.[22] Thus the cross term is non-negative. Differentiating:

$$\frac{d}{d\omega^k}\left[\frac{\omega^v}{\omega^k \alpha}\,\mathrm{Cov}(x^c, v)\right] = \frac{d}{d\omega^k}\left(\frac{\omega^v}{\omega^k \alpha}\right)\mathrm{Cov}(x^c, v) + \frac{\omega^v}{\omega^k \alpha}\,\mathrm{Cov}\left(\frac{\partial x^c}{\partial \omega^k},\, v\right)$$

The first product is non-positive: $\frac{d}{d\omega^k}(\omega^v/\omega^k) \leq 0$ (shown above) and $\mathrm{Cov}(x^c, v) \geq 0$. For the second product, $\partial x^c/\partial \omega^k$ is increasing in c (shown in Term 1) and $\mathrm{Cov}(c, v) \leq 0$, so $\mathrm{Cov}(\partial x^c/\partial \omega^k,\, v) \leq 0$. Since $\omega^v/(\omega^k\alpha) \geq 0$, the second product is also non-positive. Hence the cross term decreases.

All three terms decrease with ω^k, establishing $\frac{d\,\mathrm{Var}(x)}{d\omega^k} < 0$.

□

Theorem 1

Proof. We work under Assumptions 1–5, which abstract from quality heterogeneity. The pseudotype is $x(c) \equiv \Pi(r^*(c); c) + \frac{\omega^r}{\omega^k}\, s^r(r^*(c))$, which is strictly decreasing in c. Consequently, the firm with the lowest cost draws the highest pseudotype and wins the auction regardless of ω^k; there are no selection effects.

[22] Formally, $x^c(c)$ is decreasing in c, so $\mathrm{Cov}(x^c(c), v) = -\mathrm{Cov}(\tilde{x}^c(c), v)$ where $\tilde{x}^c = -x^c$ is increasing in c. Since $\tilde{x}^c$ is increasing in c and $\mathrm{Cov}(c, v) \leq 0$, we have $\mathrm{Cov}(\tilde{x}^c, v) \leq 0$, and therefore $\mathrm{Cov}(x^c, v) \geq 0$.

Part (i). Since the winner is always the lowest-cost firm, the expected rate, consumer surplus, and total surplus are determined entirely by the winning firm's equilibrium rate bid. Lemma 1 establishes that $dr^*(c)/d\omega^k > 0$ for every cost realization c, and parts (ii) and (iv) of the same lemma establish that consumer surplus and total surplus are strictly decreasing in ω^k for any cost realization. Taking expectations over the winning firm's cost, all three results pass directly to expected outcomes.

Part (ii). The winner's payoff satisfies the identity

$$\Pi(r^*; c_{\text{winner}}) = k^* + (x^{(1)} - x^{(2)}) \tag{A.10}$$

where k^* is the commission paid to the procuring entity and $x^{(1)} - x^{(2)}$ is the winner's retained profit in the revenue-equivalent second-score auction. We address expected profit and expected commission in turn.

Expected profit. Applying revenue equivalence, expected profits in the first-score and second-score auctions coincide. In the second-score auction, the winner's profit equals $x^{(1)} - x^{(2)}$. By Lemma 2, $\mathrm{Var}(x)$ decreases with ω^k. Assumption 5 strengthens this to contraction in dispersive order, under which all expected order-statistic spacings decrease (Shaked and Shanthikumar, 2007, Theorem 3.B.13). Therefore, $E[x^{(1)} - x^{(2)}]$ decreases, so expected provider profit falls.

Expected commission. Taking expectations of (A.10) and differentiating with respect to ω^k:

$$\frac{d\,E[k^*]}{d\omega^k} = \frac{d\,E[\Pi(r^*; c_{\text{winner}})]}{d\omega^k} - \frac{d\,E[x^{(1)} - x^{(2)}]}{d\omega^k}$$

By Lemma 1(iii), contract value increases for every cost realization, so $d\,E[\Pi]/d\omega^k > 0$. By the profit result above, $d\,E[x^{(1)} - x^{(2)}]/d\omega^k < 0$. Both terms therefore contribute positively to $d\,E[k^*]/d\omega^k$, establishing that expected commission rises. Furthermore, the increase in expected commission exceeds the reduction in expected profit: the commission gains both from the compression of pseudotype rents and from the increase in contract value that higher rates generate.

□

Theorem 2

Proof. By the envelope theorem applied to (4), since r^* maximizes $\Pi(r; c) + \psi(r, v)$, the sensitivity of the pseudotype to cost is $\frac{\partial x}{\partial c} = \Pi_c(r^*; c)$. Under Assumption 4, $\Pi_c(r; c) = -q(r)$, so $|\partial x/\partial c| = q(r^*(c))$. From (5), the sensitivity to the technical score is $\frac{\partial x}{\partial v} = \frac{\omega^v}{\omega^k \alpha}$. The ratio of cost to technical score sensitivity is therefore

$$R(c; \omega^k) \equiv \frac{|\partial x/\partial c|}{\partial x/\partial v} = \frac{q(r^*(c))\,\omega^k\,\alpha}{\omega^v}.$$

Differentiating with respect to ω^k:

$$\frac{dR}{d\omega^k} = \frac{\alpha}{\omega^v}\left[q(r^*) + \omega^k\, q'(r^*)\frac{dr^*}{d\omega^k}\right] - \frac{q(r^*)\,\omega^k\,\alpha}{\omega^{v\,2}}\frac{d\omega^v}{d\omega^k}$$

Since $\frac{d\omega^v}{d\omega^k} \leq 0$, the second term is non-negative. The expression $dR/d\omega^k > 0$ if

$$q(r^*) + \omega^k\, q'(r^*)\frac{dr^*}{d\omega^k} + \frac{q(r^*)\,\omega^k}{\omega^v}\left|\frac{d\omega^v}{d\omega^k}\right| > 0.$$

Dividing through by $q(r^*) > 0$ and rearranging:

$$1 - \frac{|q'(r^*)|\,\omega^k}{q(r^*)}\frac{dr^*}{d\omega^k} + \frac{\omega^k}{\omega^v}\left|\frac{d\omega^v}{d\omega^k}\right| > 0$$

Multiplying and dividing the middle term by r^* yields the equivalent condition

$$\varepsilon_q(c)\cdot\varepsilon_r(c) < 1 + \frac{\omega^k}{\omega^v}\left|\frac{d\omega^v}{d\omega^k}\right|$$

which is sufficient for all $c \in \mathcal{C}$. When $d\omega^v = 0$, the right-hand side equals one and the derivation is an if-and-only-if, establishing necessity. □

A.3 Equilibrium without Commissions in the Scoring Rule

Under Assumptions 3, 4, 6, and 7, the symmetric Bayes-Nash equilibrium strategies are characterized by (6) and (7) from Section 3.3. Analytical solutions, however, are unavailable when commissions are exogenously determined ($\omega_i^k = 0$). Therefore, we use an iterative best-response algorithm to approximate equilibrium. This involves parameterizing a "belief function" that determines how firms expect their competitors to bid, computing the best responses to those beliefs, updating the belief function based on the best responses, and iterating until the beliefs converge to best responses (e.g., Bajari, 2001; Armantier et al., 2008).[23] The converged belief function is the approximate equilibrium strategy. In ICS, we find a tight relationship between the belief functions and the best-response functions, indicating that the approximation is accurate in practice.

To implement the approximation, we assume that in each iteration, $h = 0, 1, \ldots, H$, all firms believe rival j in auction i will bid according to

$$\hat{s}_{ij}^{(h)} = g_i\left(\mathbf{x}_{ij}; \gamma_i^{(h)}\right) \tag{A.11}$$

where $g_i()$ is the belief function, the auction-specific vector $\gamma_i^{(h)}$ contains "belief coefficients," and the vector $\mathbf{x}_{ij}$ includes exogenous factors. The factors we use are the rival's cost and technical

[23] Under Assumption 3, the profit functions are concave conditional on beliefs, and this holds with the linear demand and contribution functions we use in the empirical model. Therefore, the best-response functions are unique.

score, and a constant; as $g_i()$ is auction-specific, the constant absorbs factors that do not vary across bidders (e.g., the auction weights). Given these beliefs, we compute the best responses, $\hat{s}_{if}^{(h)BR}$, of $f \in \{1, \ldots, F\}$ simulated "focal" firms by solving their profit maximization problem. For each auction, we regress the best responses of these focal firms on the function of their exogenous factors, obtaining the regression coefficients, $\hat{\gamma}_i^{(h)}$. We then update the belief coefficients using

$$\gamma_i^{(h+1)} = \rho^{(h)} \hat{\gamma}_i^{(h)} + \left(1 - \rho^{(h)}\right) \gamma_i^{(h)} \tag{A.12}$$

where $\rho^{(h)} \in [0, 1]$ controls the step size; we reduce the step size with the number of iterations to promote convergence. We run 21 iterations using linear belief functions and then 21 iterations in which $g_i()$ is a second-order polynomial in the exogenous factors, warm-started from the converged linear coefficients, with zeros appended for the higher-order terms.

Evaluated at our final estimates of the structural parameters, the average mean squared error between the belief coefficients and the regression coefficients is only 0.0008 after the 42 iterations. Across auctions, the average R^2 in the final regression is 0.9566. Therefore, there is alignment between beliefs and best responses, and the belief specification captures best-response curvature. We provide more details on implementation and results in Appendix B.3.

B Empirical Model and Estimation

B.1 Nonparametric Identification

B.1.1 Cost Distribution, $F_C(\cdot)$

As established in Appendix A.1, observed equilibrium rates are set to maximize the bilateral surplus. Given the contract value form, $\Pi_r(r_{ij}; c_{ij}) = q(r_{ij}) + (r_{ij} - c_{ij})q\prime(r_{ij})$, so that the rate first-order condition (A.2) obtains:

$$c_{ij} = r_{ij} + \left(q(r_{ij}) + \frac{1}{\lambda\alpha} s_i^r\prime(r_{ij})\right)\left(q\prime(r_{ij})\right)^{-1} \tag{B.1}$$

All objects on the right-hand side of this equation are observed or estimable in the data. Given strictly decreasing $q(r)$, differentiability, and an interior solution, the first-order condition is necessary and sufficient and defines a unique mapping from the observed rate r_{ij} to c_{ij}. Hence, bidder-level costs can be recovered from observed rates, and pooling these across auctions (with $\omega_k > 0$, $\omega_r > 0$, and where rates are not predetermined by the buyer) identifies the unconditional cost distribution $F_C(\cdot)$ nonparametrically.

B.1.2 Scale of Technical Scores, σ

Combining equations (A.8) and (A.9), the equilibrium commission satisfies

$$k_{ij}^* = \Pi(r_{ij}^*; c_{ij}) - \mathrm{IR}_i(x_{ij}^*; \sigma), \tag{B.2}$$

where we use

$$\mathrm{IR}_i(x; \sigma) \equiv \frac{\int_{\underline{t}}^{x} T_i(u; \sigma)^{J_i - 1}\, du}{T_i(x; \sigma)^{J_i - 1}} \tag{B.3}$$

to denote the information rent. Equation (B.2) follows because $x_{ij} - \psi_i(r_{ij}^*, v_{ij}) = \Pi(r_{ij}^*; c_{ij})$ by (8), and intuitively the technical score enters both the pseudotype and ψ_i with the same multiplier $\omega_i^v/(\omega_i^k \alpha)$, so its contribution cancels and the commission is independent of v_{ij} conditional on c_{ij}. Since k_{ij}^* is observed and $\Pi(r_{ij}^*; c_{ij})$ is identified in Step 1, rearranging equation (B.2) recovers $\mathrm{IR}_i(x_{ij}^*; \sigma)$ directly from the data.

The pseudotype distribution T_i as defined in equation (A.7), replacing $v = \mu_i + \sigma\,\eta$ is given by:

$$T_i(x; \sigma) = \int\int \mathbf{1}(x_i^*(c,\, \mu_i + \sigma\eta) \le x)\, dF_C(c)\, dF_\eta(\eta). \tag{B.4}$$

Since $c \perp \eta$ by Assumption 7, the integral in (B.4) is well-defined as a product integral over independent marginals. All objects entering (B.4) (namely F_C, F_η, μ_i, auction weights, and contribution functions) are identified or pre-estimated, so $T_i(\cdot\,; \sigma)$ is a known functional of σ alone.

The information rents uniquely pin down σ provided the mapping $\sigma \mapsto \mathrm{IR}_i(x; \sigma)$ is injective. Since $x_i^*(c, \eta; \sigma) = f_i(c) + \frac{\omega_i^v}{\omega_i^k \alpha}(\mu_i + \sigma\eta)$ for a known function $f_i(c)$ the variance of the pseudotype is

$$\mathrm{Var}(x_i^*; \sigma) = \mathrm{Var}(f_i(c)) + \left(\frac{\omega_i^v}{\omega_i^k \alpha}\right)^2 \sigma^2, \tag{B.5}$$

which is strictly increasing in σ whenever $\omega_i^v > 0$. It follows that $T_i(\cdot\,; \sigma) \neq T_i(\cdot\,; \sigma')$ for any $\sigma \neq \sigma'$, and the mapping $\sigma \mapsto \mathrm{IR}_i(x; \sigma)$ is therefore injective.[24] The observed information rents therefore uniquely identify σ. Intuitively, greater within-auction variance in technical scores spreads the pseudotype distribution, increasing the expected gap between the winner's pseudotype and that of the highest rival, and hence inflating equilibrium information rents.

Cross-auction variation in observables is useful but not necessary. Cross-auction variation in J_i helps separate the contribution of σ to information rents from the mechanical effect of competition: more bidders compress the gap between first- and second-order pseudotype statistics independently of σ. Variation in $\omega_i^v/(\omega_i^k \alpha)$ means that σ has heterogeneous effects on $\mathrm{Var}(x_i^*; \sigma)$ across auctions, providing additional variation in the relationship between σ and information rents.

[24] Injectivity of $\sigma \mapsto \mathrm{IR}_i(x; \sigma)$ is thus based on the observation that the variance of the pseudotype distribution depends strictly on σ via equation (B.5), so that two distinct values of σ cannot generate the same distribution $T_i(\cdot\,; \sigma)$, and hence cannot generate the same information rent.

B.2 Integrated Splines and the Cost Distribution

An integrated spline (I-spline) is a linear combination of basis functions. Each basis function is continuously differentiable and non-decreasing, mapping from [0,1] to [0,1]. Recycling Q and p from a notational perspective, the I-spline takes the form

$$Q(p) = \tau_0 + \sum_{j=1}^{K} \tau_j I_{j,d}(p) \tag{B.6}$$

where $\tau_0, \ldots, \tau_K$ are the parameters we estimate and $I_{j,d}(p)$ are the basis functions. Thus, the I-spline is continuously differentiable and maps from [0,1] to an unrestricted range on the real line. We restrict the weights $(\tau_1, \ldots, \tau_K)$ to be nonnegative, which ensures that the I-spine is non-decreasing. Therefore, it satisfies the properties of a *quantile function*, or inverse CDF. In our application, p is a probability level and $Q(p)$ is the corresponding cost quantile. Figure D.7 shows the quantile function and the implied CDF that we estimate. Drawing $p \sim \text{Uniform}(0, 1)$ and computing $c = Q(p)$ generates a random cost draw from $F_C(c) = Q^{-1}(c)$.

The flexibility of the I-spline depends on the knot placement (which controls where curvature changes) and the polynomial degree d (which controls smoothness). We use $d = 2$ with one interior knot at $c = 0.50$, which yields $K = 3$ basis functions and four parameters to estimate: $\{\tau_0, \tau_1, \tau_2, \tau_3\}$. Our specification allows the quantile function (and the CDF) to be globally convex, globally concave, locally concave then locally convex, or locally convex then locally concave. Figure D.8 illustrates this flexibility using different parameterizations. The I-spline can be made more or less flexible; the more flexibility, the more parameters are required.

The I-spline basis functions $I_{j,d}(p)$ are constructed through a three-step process. First, B-splines of degree d are defined recursively over a knot sequence $\{\kappa_1, \ldots, \kappa_{K+d+1}\}$. These are piecewise polynomials, each with compact support. Second, M-splines are obtained by normalizing B-splines so they integrate to one over their support. Because B-splines are nonnegative, M-splines are also nonnegative. Third, I-spline basis functions are obtained by integrating M-splines:

$$I_{j,d}(p) = \int_{\kappa_1}^{p} M_{j,d}(t)\, dt. \tag{B.7}$$

Since M-splines are nonnegative, each I-spline basis function is nondecreasing. Integration increases the polynomial degree by one, so with $d = 2$ (quadratic M-splines), the I-spline basis functions are piecewise cubic. For more details, see de Boor (2001) and Schumaker (1981).

B.3 Equilibrium Approximation

For each candidate parameter vector evaluated by the optimizer, we approximate equilibrium in the auctions for which the commission does not enter the scoring rule, as described in Appendix A.3. We use 21 iterations with a linear specification of the belief function and use the results to

warm-start 21 iterations with a second-order polynomial specification. Here, we provide details on implementation and then summarize the results, first for the intermediate linear beliefs and then for the final polynomial beliefs.

Details on Implementation

We take ten evenly-spaced draws from both the cost distribution and the within-auction technical score distribution.[25] The combination of these draws provides 100 representative "focal firms" for which we obtain best responses given a set of beliefs. We compute the best responses of these focal firms numerically. For a candidate score bid from a focal firm with a known cost and technical score, we compute the implied rate and profit (conditional on winning) for 999 evenly-spaced draws on the evaluation noise; this allows us to calculate the expected profit conditional on winning. We obtain the probability of winning by determining the fraction of times that the focal firm's candidate score bid exceeds the bids of all its competitors, using 1000 simulated sets of competitors (each defined by cost and technical score draws) and obtaining competitor bids from the beliefs. We draw the simulated sets of competitors using 999 evenly-spaced cost draws and 999 evenly-spaced technical score draws. Thus, we consider a full range of possible competitors; the focal firms are not restricted to competing only with each other. Using the conditional profit and the probability of winning, we compute the expected profit for each candidate bid. The candidate bid that maximizes expected profit is the best response.[26] We update beliefs as described above. We set $\rho^{(h)} = 0.50$ for the initial iterations, $\rho^{(h)} = 0.10$ for intermediate iterations, and $\rho^{(h)} = 0.01$ for the final iterations, both for the initial linear belief function and for the subsequent second-order polynomial. The process is not a contraction mapping, but convergence is reliable in practice.

Detailed Results

Table D.4 shows the regression results obtained in the final iteration with linear beliefs. The dependent variable is the best response of the focal firms; the regressors are the cost and technical score of the focal firms. Each row shows the results for a single auction. In each auction, the score bids of the focal firms decrease with their costs and increase with their technical score, which is consistent with profit-maximizing behavior. The results show that the best responses are well approximated with a linear function: the average R^2 value is 0.9198. The polynomial specification improves on this, as we discuss below. The mean squared error between belief and regression coefficients, averaged across auctions, is 0.0007. This is small relative to the magnitude of the coefficients, indicating alignment between the best responses and the beliefs.

[25] That is, we take the 5th percentile, 15th percentile, 25th percentile, and so on, up to the 95th percentile. Thus, the draws are evenly distributed in probability space.

[26] We use the optimize function in R to find the best responses. The function requires us to specify lower and upper bounds. As the lower bound, we use the score corresponding to the monopoly price with the lowest possible technical score for the focal firm. As the upper bound, we use the score corresponding to at-cost rates with the highest possible technical score.

The heterogeneity in belief coefficients across auctions reflects differences in auction design and competitive intensity. To illustrate, Table D.5 regresses each belief coefficient on the auction weights and the number of bidders. The cost coefficient γ_1 becomes more negative when the rate weight is larger and when there are more competitors, meaning that best responses are more sensitive to costs in auctions where rates matter more and competition is fiercer. The technical score coefficient γ_2 increases with the technical weight and the number of bidders, as expected. The constant γ_0 is not strongly explained by these covariates ($R^2 = 0.06$), consistent with auction-specific factors such as the scoring rule intercepts absorbing much of the level variation. These patterns align with the model's comparative statics and are captured by our auction-specific belief functions.

Table D.6 shows the belief coefficients, R^2 values, and MSE values after all 42 iterations. The polynomial specification improves the fit: the average R^2 increases from 0.9198 to 0.9566. The average MSE between belief and regression coefficients remains small at 0.0008, confirming that the converged beliefs closely approximate best responses. As these are the final beliefs, we treat them as the approximate equilibrium strategies and use them to generate the simulated bids that enter the moment conditions.

B.4 Sensitivity of Estimation Results

We examine the sensitivity of results to the demand specification and the linearity of the contribution functions.

Demand for Calls

We explore the sensitivity of the cost distribution to demand estimation by shifting the demand intercept upward to reflect a $1.00 increase in the willingness-to-pay for calls. From an econometric standpoint, the implications of higher demand are ambiguous for the cost distribution estimates. On one hand, higher demand requires higher costs to rationalize observed commissions. On the other hand, lower costs are required to rationalize observed rates. Figure D.5 plots the cost distribution from the baseline model and the alternative with higher demand. We find that the cost distribution shifts toward higher costs with the alternative demand function. Thus, the commission channel dominates empirically. The mean of costs with the alternative CDF is $0.88, and the average cost of the winning bidder is $0.41.

Contribution Functions

We examine the linearity of the contribution functions by regressing the contribution function residuals from columns (ii), (iii), and (iv) of Table 3 on the square of the proposed contract term. Econometrically, this is equivalent to estimating quadratic contribution functions. We plot the fit of the residuals and the quadratic terms in Figure D.6. An initial observation is that the residuals are small in magnitude relative to the mean of the average rate and commission, implying limited scope for

improvement over the linear specification. Furthermore, the relationship between the residuals and the quadratic term is weak, so the quadratic term has little explanatory power. If the quadratic term is added directly into the contribution functions, the R^2 values increase only trivially: from 0.910 to 0.919, 0.984 to 0.985, and 0.973 to 0.974. Therefore, the linearity assumption is appropriate in our setting.

C Empirical Evidence on the Effect of Rates on Call Activity

We exploit two natural experiments to estimate the demand for calls. Both are based on state-level policy changes. In 2010, New York eliminated a $1.28 connection charge and reduced the per-minute price from $0.068 to $0.048. The implied rate of a 15-minute call fell from $2.30 to $0.72. New Jersey implemented a series of price reductions from January 2014 to May 2015 that lowered the per minute price from $0.33 to $0.044 (from $4.95 to $0.66 for a 15-minute call).[27]

To estimate demand, we regress the number of calls on the price of a 15-minute call, using data from New Jersey and New York before and after the price changes. As prices change only due to the policy changes, our OLS results are identical to 2SLS estimates that use the policy changes as instruments. Table 4 summarizes the results. We consider specifications that pool the two states and specifications that treat them separately. As we interpret the latter as providing the better summary of the empirical variation, we use an average of the subsample coefficients in the empirical model. The coefficients are those in columns (iii) and (iv) in the empirical model.

The effects of the rate reductions are of broader interest, so we examine in the remainder of this appendix how four variables respond: the number of calls per person/month, the minutes per call, the calling minutes per person/day, and the expenditure per person/month. Each variable is computed as the average across all incarcerated individuals in the state prison system and is observed monthly. A visual inspection of the data indicates that each variable is reasonably stable before the price changes (Figure D.3), which also supports the validity of our demand estimation. Afterwards, the number of calls and calling minutes increases, yet expenditures decrease. In New York, the minutes per call decreased slightly, consistent with incarcerated individuals substituting to shorter calls in response to the elimination of connection charges. In New Jersey, where connection charges did not exist, the number of minutes per call changed much less.

Table D.3 summarizes the changes quantitatively, using "before" and "after" periods that we select based on our visual inspection of the data. The average number of calls per person/month increases from 8.82 to 15.86 in New York and from 8.32 to 27.00 in New Jersey. Sample means tests indicate that these changes differ from zero at the 1% level. Similar patterns are observed for

[27] In both states, the same prices were charged for local, intrastate, and interstate calls, and for different payment methods (e.g., debit and collect). In New Jersey, the first change, reducing the per-minute price from $0.33 to $0.19, occurred in February 2014 at the direction of the DOC. We observe subsequent reductions to $0.17 in March 2014, $0.15 in September 2014, and $0.044 in May 2015. The last change was due to a new state law. There was no connection charge in New Jersey before or after these changes.

the average minutes per inmate-day spent on the phone. Average expenditures per inmate-month decrease from $23.77 to $14.35 in New York and from $30.36 to $13.24 in New Jersey, and the changes again are statistically significant. Thus, the raw data are consistent with the hypothesis that price is a meaningful determinant of phone usage.

D Additional Figures and Tables

4.5.1 After determining that a proposal satisfies the mandatory requirements, the evaluator(s) shall use both objective and subjective judgement in conducting a comparative assessment of the proposal in accordance with the evaluation criteria stated below:

Call Expense to Called Parties and Inmates	50%
Experience & Reliability	20%
Proposed method of Performance	30%

(a) Scoring Rule from the RFP

					For DPMM Use Only
NAME OF OFFEROR	Experience And Reliability (Max. 20 Pts.)	Proposed Method of Performance (Max. 30 Pts.)	TOTAL SUBJECTIVE POINTS (Max. 50 Pts.)	COST POINTS INSERTED BY PURCHASING (Max. 50 Pts.)	TOTAL POINTS (Max. 100 Pts.)
1. MCI Worldcom	20	30	= 50 +	50	TOTAL = 100
2. Public Communications Service	15	25	= 40 +	47.75	TOTAL = 87.75

(b) Scoring Sheet for Submitted Bids

Application #1 - Inmate Telephone System

001	$ 0.00	firm, fixed price per minute, local calls;
002	$ 1.30	firm, fixed price per call, set-up charge, local calls
003	$ .14	firm, fixed price per minute, intralata calls;
004	$ 1.00	firm, fixed price per call, set-up charge, intralata calls
005	$ .14	firm, fixed price per minute, interlata calls;
006	$ 1.00	firm, fixed price per call, set-up charge, interlata calls
007	$ .45	firm, fixed price per minute, interstate calls;
008	$ 2.45	firm, fixed price per call, set-up charge, interstate calls

(c) Finalized Prices in the Contract

Figure D.1: Illustrative Example of a Procurement Process

Notes: The figure displays screenshots illustrating the procurement process in Missouri in 2000. On March 7, the Department of Corrections held a "Pre-Proposal Conference" for prospective bidders. A formal RFP was issued on March 29; it detailed the scoring criteria and was made available to all prospective bidders (top panel). Proposals were due April 12. From April 20 to July 18, a committee evaluated the bids and assigned scores to MCI Worldcom and Public Communications Service, the two bidders (middle panel). MCI received the highest score and was awarded the contract at their proposed rates (bottom panel).

On-site Cash Transaction Process Fees		
Low End	Upper End	Cost
$0.00	$300.00	$3.00

On-Site or Over The Internet		
Non-Cash Deposit		
Low End	Upper End	Cost
$0.00	$19.99	$2.95
$20.00	$50.00	$4.25
$50.01	$100.00	$5.50
$100.01	$200.00	$8.50
$200.01	$300.00	$11.50

(a) GTL Account Funding Fees

Fees:

Funding Fee:

Amount Deposited	$5.00 - $25.00	$25.01 - $50.00	$50.01 - $100.00
Funding Fee	$1.99	$2.99	$3.99

Note: Refund Fee $2.00 applies when closing an account and requesting a refund of remaining balance

Policies:

Minimum Deposit Amount: $5
Maximum Deposit Amount: $100

(b) Securus Account Funding Fees

Figure D.2: Illustrative Examples of Ancillary Fee Schedules

Notes: Some bids in our data contain information about the ancillary fees set by providers. The figure shows the ancillary fees of Securus and GTL as described in bids submitted to New Jersey in 2014 (and as captured with screenshots). GTL proposed fees that users would pay when depositing money into an account, and Securus proposed fees that users would pay when closing an account or requesting refunds. CenturyLink also submitted a bid in the 2014 New Jersey auction, but it did not include ancillary fees.

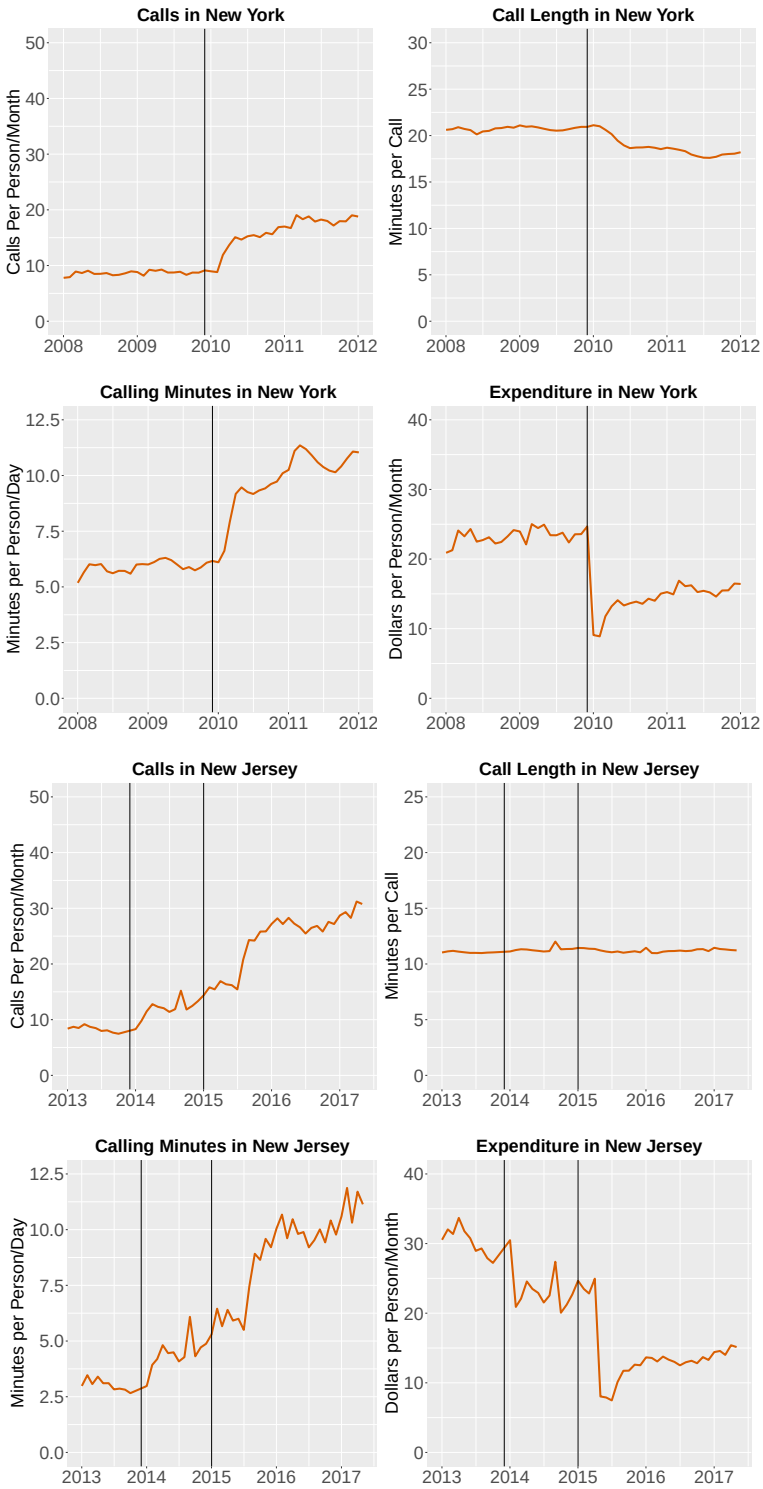


Figure D.3: Calling Patterns Before and After Price Reductions

Notes: The figure plots calls per person/month, minutes per call, minutes per person/day, and expenditure per person/month over time in New York (top four panels) and New Jersey (bottom four panels). The data points are monthly averages. The vertical black lines show the timing of price changes.

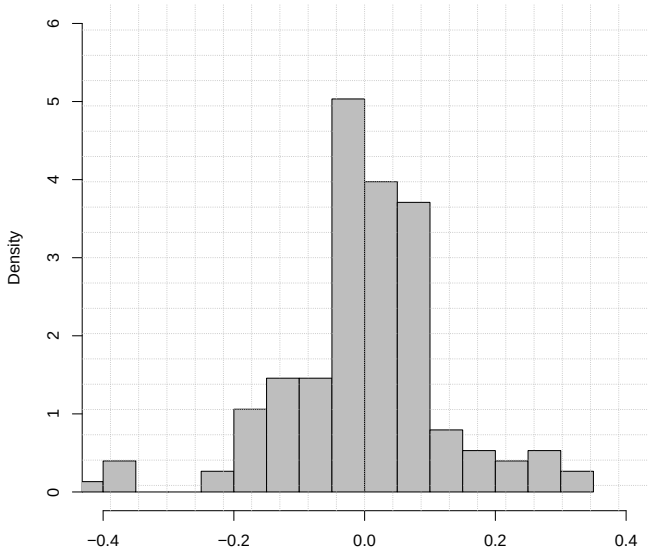


Figure D.4: Distribution of Technical Score Residuals

Notes: The figure provides a histogram of the residuals we obtain from a regression of technical scores on auction fixed effects. We measure the technical score as the score of the bidder divided by the maximum possible score (for all elements of the bid related to technical capabilities or subjective assessments of the provider).

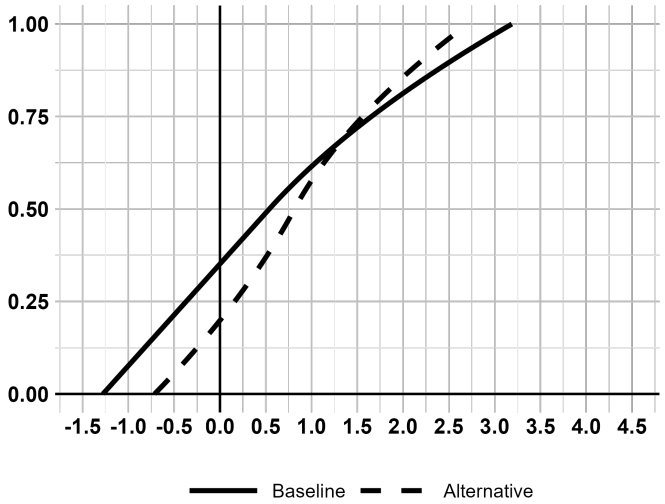


Figure D.5: Sensitivity of Estimated Cost Distribution to Greater Demand

Notes: The figure plots the estimated CDFs of cost (net of fees) obtained with the baseline demand function and an alternative version in which the demand intercept is shifted to produce a one-dollar increase in willingness-to-pay.

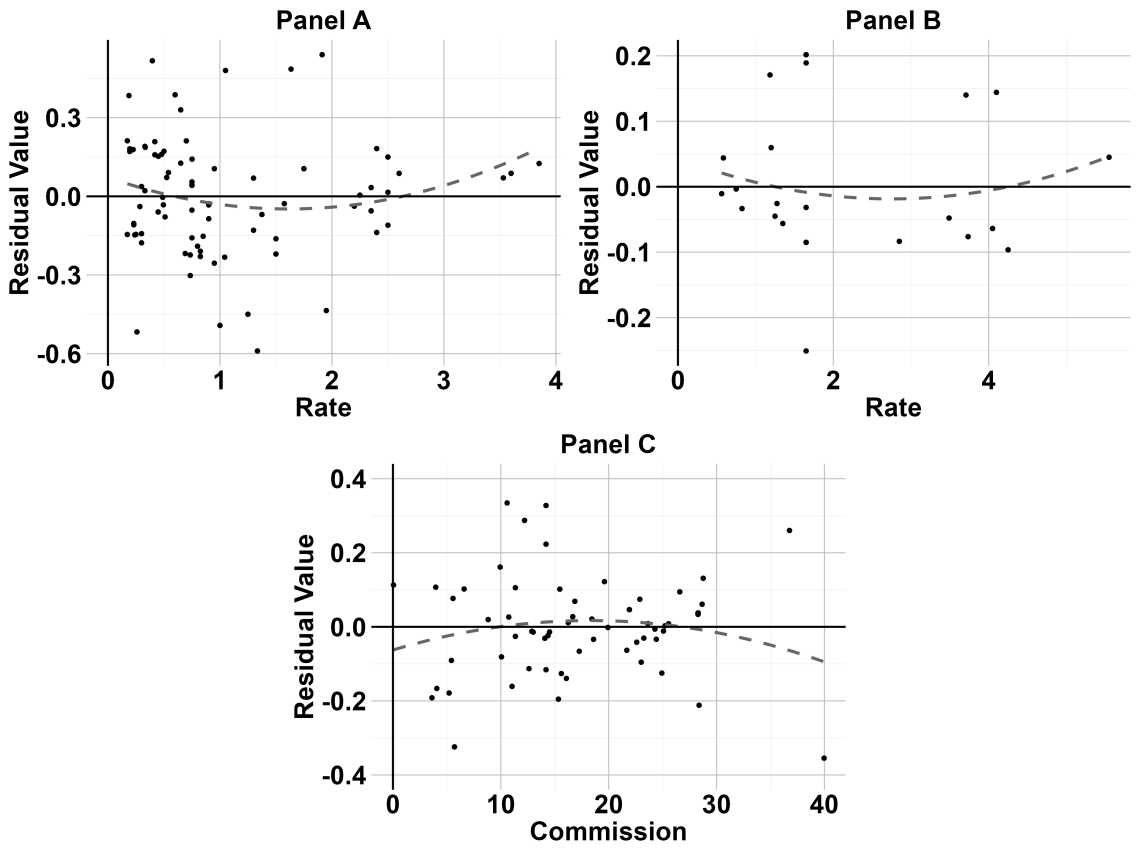


Figure D.6: Assessing Linearity of the Contribution Functions

Notes: The figure plots the residuals obtained from columns (ii), (iii), and (iv) of Table 3 against the rate (Panels A and B) and the commission (Panel C). The dashed gray lines provide quadratic fits. The coefficients on the squared terms (not shown) are small relative to the slope coefficients of Table 3. For Panels A and C, but not Panel B, they are statistically different from zero.

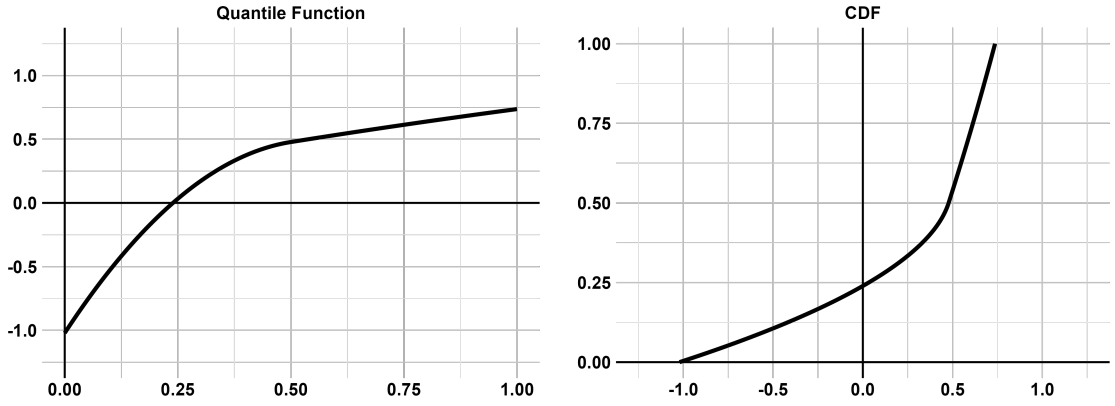


Figure D.7: The Quantile Function and the Implied CDF

Notes: We estimate the parameters of an I-spline that satisfies the requirements to be a *quantile function*, or inverse CDF. We then invert the quantile function to obtain the implied CDF. The left panel shows the quantile function that we estimate with the baseline specification, and the right panel shows the corresponding CDF (see also Figure 2).

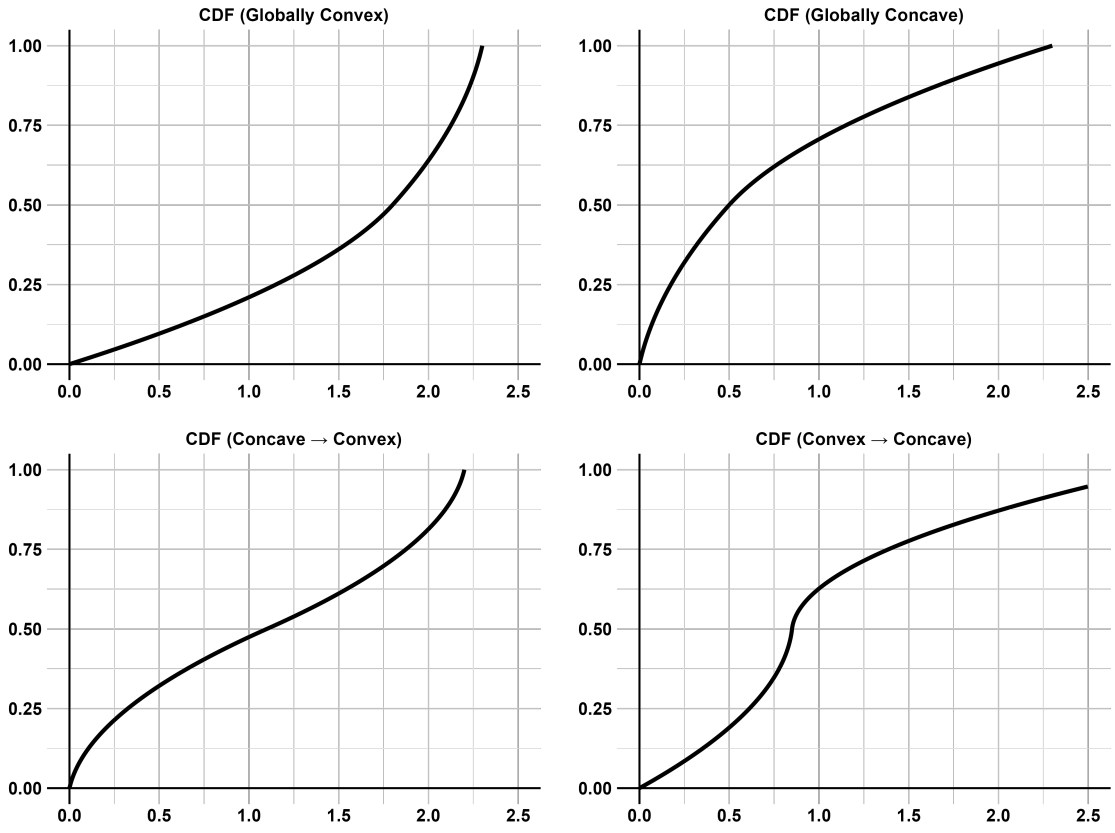


Figure D.8: Illustrating the Flexibility of the I-Spline for CDFs

Notes: The figure illustrates the flexibility of the I-spline in our baseline specification. We set the constant to zero for each panel ($\tau_0 = 0$). In the top left, the CDF is globally convex because the spline weights decrease magnitude ($\tau_1 = 1.4, \tau_2 = 0.8, \tau_3 = 0.1$). In the top right, the CDF is globally concave because the spline weights increase in magnitude ($\tau_1 = 0.1, \tau_2 = 0.8, \tau_3 = 1.4$). In the bottom left, the CDF is locally concave then locally convex, which obtains with a large middle weight ($\tau_1 = 0.1, \tau_2 = 2.0, \tau_3 = 0.1$). In the bottom right, the CDF is locally convex then locally concave, which obtains with a small middle weight ($\tau_1 = 0.8, \tau_2 = 0.1, \tau_3 = 2.0$).

Table D.1: Summary of the Bid Data

	State	Year	Rate Weight	Quality Weight	Comm. Weight	# of Bids	# of Bidders	Winning Bidder	Contract Rate	Contract Commission		
										Percentage	Fixed	Total
1	Alaska	2016	0	0.56	0.44	4	4	Securus	1.05	0.94	0	19.61
2	Arizona	2014	0	0.17	0.83	4	4	CenturyLink	1.84	0.94	0	28.27
3	Arkansas	2014	0	1	0	4	4	Securus	4.80	0.73	0	11.65
4	Florida	2013	0.05	0.90	0.05	3	3	GTL	0.75	0.67	0	10.58
5	Florida	2017	0.17	0.83	0	3	3	Securus	0.19	0	0	0
6	Georgia	2015	0	0.65	0.35	3	3	Securus	1.95	0.97	6.73	36.76
7	Idaho	2014	0.20	0.80	0	5	5	CenturyLink	2.40	0	20.00	20.00
8	Illinois	2012	0.45	0	0.55	3	3	Securus	4.10	0.87	0	22.89
9	Indiana	2010	0.22	0.67	0.11	3	3	PCS	4.05	0.40	0	10.74
10	Kentucky	2012	0	0.74	0.30	5	5	GTL	2.25	0.88	0	28.66
11	Kentucky	2017	0	0.86	0.14	3	3	GTL	1.65	0.50	0	14.19
12	Maine	2015	0.15	0.70	0.15	4	4	Legacy	1.65	0.55	0	15.61
13	Massachusetts	2013	0	0.86	0.14	6	4	GTL	2.36	0.60	0	19.93
14	Michigan	2018	0	1	0	3	3	GTL	2.40	0	23.65	23.65
15	Minnesota	2005	0	0.84	0.16	4	4	MCI	1.75	0.49	0	14.37
16	Minnesota	2016	0.40	0.60	0	8	4	GTL	0.33	0.47	0	3.60
17	Minnesota	2019	0.30	0.70	0	4	4	GTL	0.33	0.40	0	3.04
18	Missouri	2000	0.75	0.25	0	2	2	MCI	1.30	0	0	0
19	Missouri	2006	0.52	0.48	0	5	5	PCS	2.50	0	0	0
20	Missouri	2011	0.42	0.58	0	10	7	Securus	1.75	0	0	0
21	Missouri	2018	0.28	0.72	0	3	3	Securus	0.75	0	0	0
22	Montana	2017	0.20	0.80	0	5	5	CenturyLink	0.54	0	9.24	9.24
23	Nebraska	2008	0.33	0.67	0	6	6	PCS	0.70	0	0	0
24	Nebraska	2016	0.41	0.59	0	6	5	GTL	0.19	0	0	0
25	New Hampshire	2013	1	0	0	4	4	IC Solutions	0.65	0.20	9.67	12.48
26	New Hampshire	2018	0.35	0.65	0	4	4	GTL	0.19	0.20	11.12	12.04
27	New Jersey	2014	0.40	0.60	0	2	2	GTL	0.40	0	0	0
28	North Dakota	2016	0.23	0.68	0.09	5	5	Securus	1.19	0.25	0	5.70
29	Oklahoma	2018	0.30	0.70	0	4	4	Securus	1.91	0	14.57	14.57
30	Utah	2014	0	0.70	0.30	4	4	CenturyLink	2.10	0.90	0	28.76
31	Utah	2019	0	0.70	0.30	4	4	GTL	1.80	0.95	0	28.28
32	Vermont	2010	0	0.70	0.30	5	5	PCS	2.30	0.37	0	12.20
33	Vermont	2016	0.03	0.95	0.02	3	3	GTL	0.58	0.41	0	5.21
34	Virginia	2005	0.20	0.70	0.10	2	2	MCI	5.55	0.41	0	0.05
35	West Virginia	2014	0.30	0.70	0	3	3	CenturyLink	0.48	0	0	0.01
36	Wisconsin	2008	0.30	0.70	0	6	6	Embarq	1.05	0.30	0	6.25
37	Wisconsin	2018	0.20	0.80	0	3	3	CenturyLink	0.17	0.30	0	1.23

Notes: The table summarizes the auction-level data. Rate is the cost of a 15-minute local collect phone call. The commission percentage is the percentage of the non-fee revenue that the provider pays to the state. The fixed commission is a fixed amount of money that the provider pays to the state, converted to be in dollars per inmate-month. The total commission combines these two forms of payments and is in dollars per inmate-month.

Table D.2: Removing Commissions from the Scoring Rule

Expected Outcomes:		$\omega^k = 0.25$	$\omega^k = 0.00$
Commission	a	29.76	—
Consumer Surplus	b	24.01	48.24
Profit	c	11.03	22.10
Total Surplus	$a+b+c$	64.80	70.34
Public Surplus	$a+b$	53.77	48.24
Rate		2.27	0.95

Notes: The first column shows expected equilibrium outcomes in an auction with four bidders and auction weights of $\omega^k = \omega^r = 0.25$ and $\omega^v = 0.50$. The second column shows expected equilibrium outcomes when the weights shift such that $\omega^k = 0.00$ and $\omega^r = \omega^v = 0.50$, based on equilibrium approximation.

Table D.3: Calling Patterns Before and After Price Reductions

	New York				New Jersey			
	Before	After	Change	*p*-value	Before	After	Change	*p*-value
Number of Calls	8.82	15.86	7.04	0.000	8.32	27.00	18.68	0.000
Minutes per Call	20.81	18.87	-1.94	0.000	11.05	11.18	0.13	0.011
Minutes per Day	6.04	9.83	3.80	0.000	3.03	9.91	6.87	0.000
Expenditures	23.77	14.35	-9.43	0.000	30.36	13.24	-17.12	0.000

Notes: The table provides the average number of calls per inmate-month, average minutes per call, average minutes per inmate-day, and average expenditure per inmate-month (in dollars), both before and after price reductions in New York and New Jersey. For each, it also provides the change and the p-value from a sample means test of the null hypothesis that the shift equals zero. We observe the variables at the month-state level; the p-values are computed assuming independence. For New York, we use a "before" period of January-December 2009 and an "after" period of April 2010 - March 2011. For New Jersey, we use a "before" period of January-October 2013, as some data are unavailable for November-December 2013, and an "after" period of January-December 2016.

Table D.4: Regressions with Linear Belief Functions

State	Year	Constant: $\gamma_0^{(31)}$	Cost: $\gamma_1^{(31)}$	Tech Score: $\gamma_2^{(31)}$	R^2	MSE
Florida	2017	0.8077	-0.0256	0.5485	0.9404	0.0002
Indiana	2014	0.7246	-0.0436	0.4523	0.7785	0.0014
Minnesota	2016	0.7457	-0.0597	0.3517	0.9590	0.0001
Minnesota	2019	0.6004	-0.0497	0.4743	0.9689	0.0002
Missouri	2000	1.0718	-0.0517	0.0978	0.9624	0.0000
Missouri	2006	0.8889	-0.0785	0.3086	0.9857	0.0000
Missouri	2011	0.9671	-0.0645	0.3955	0.9793	0.0001
Missouri	2018	0.9535	-0.0305	0.3673	0.9276	0.0003
Montana	2017	0.7886	-0.0358	0.5143	0.9518	0.0001
Nebraska	2008	0.6847	-0.0582	0.4985	0.9768	0.0001
Nebraska	2016	0.7416	-0.0666	0.4101	0.9759	0.0001
New Hampshire	2018	0.7746	-0.0707	0.4048	0.7733	0.0045
New Jersey	2014	0.5702	-0.0469	0.3894	0.8351	0.0008
Oklahoma	2018	0.6823	-0.0563	0.3929	0.7845	0.0034
Wisconsin	2008	0.7423	-0.0512	0.4415	0.9429	0.0000
Wisconsin	2018	0.8197	-0.0299	0.5124	0.9490	0.0002
West Virginia	2014	0.9390	-0.0346	0.3827	0.9452	0.0001

Notes: For auctions in which the commission does not enter the scoring rule, we approximate symmetric Bayes-Nash equilibrium by computing how 100 "focal firms" would bid, given a belief that competitors' bids are a linear function of their costs and technical score. We regress the best responses of the focal firms (e.g., their score bid) on their costs and technical scores, update beliefs, and iterate. The table summarizes the regression results for each auction after 21 iterations. We provide the coefficients, the R^2 of each state-specific regression, and the MSE between the final belief and regression coefficients.

Table D.5: Determinants of Belief Coefficients with Linear Beliefs

	Dependent Variable		
	$\gamma_1^{(31)}$	$\gamma_2^{(31)}$	$\gamma_0^{(31)}$
Constant	-0.0004	-0.0749	0.6728
	(0.0102)	(0.0538)	(0.1357)
Rate Weight	-0.0712		0.4002
	(0.0174)		(0.2311)
Technical Weight		0.6424	
		(0.0724)	
# of Bidders	-0.0061	0.0148	-0.0038
	(0.0017)	(0.0072)	(0.0229)
R^2	0.6494	0.8642	0.1837

Notes: The table summarizes how the belief coefficients correlate with selected auction features. The belief coefficients $(\gamma_1, \gamma_2, \gamma_3)$ correspond to the coefficient on the cost, the technical score, and the constant in the belief regressions (reported in Table D.4). Here, we regress the belief coefficients on the rate weight, the technical weight, and the number of bidders. The unit of observation is an auction, and there are 17 observations. The technical weight equals one less the rate weight, as the commission does not enter the scoring rule for these auctions. Each regression also includes a constant. Standard errors are in parentheses.

Table D.6: Regressions with Flexible Belief Functions

State	Year	Constant	Cost	Tech Score	Cost^2	Tech Score^2	Interaction	R^2	MSE
Florida	2017	0.8113	-0.0269	0.5759	-0.0055	-1.5707	0.1480	0.9784	0.0025
Indiana	2014	0.7129	-0.0413	0.5416	-0.0008	0.4932	-0.0038	0.8497	0.0000
Minnesota	2016	0.7426	-0.0578	0.3180	-0.0090	-0.2359	0.0942	0.9727	0.0000
Minnesota	2019	0.5973	-0.0453	0.4058	-0.0071	-0.3775	0.1359	0.9870	0.0006
Missouri	2000	1.0731	-0.0552	0.1074	-0.0060	-0.3430	0.0507	0.9472	0.0001
Missouri	2006	0.8916	-0.0793	0.2851	-0.0057	-0.4566	0.1125	0.9924	0.0000
Missouri	2011	0.9778	-0.0721	0.3935	-0.0104	-0.5837	0.1443	0.9924	0.0000
Missouri	2018	0.9661	-0.0434	0.4520	-0.0086	-1.0989	0.2393	0.9810	0.0000
Montana	2017	0.7866	-0.0332	0.4714	-0.0046	-0.1696	0.0592	0.9451	0.0000
Nebraska	2008	0.6857	-0.0569	0.4489	-0.0091	-0.6400	0.1634	0.9921	0.0000
Nebraska	2016	0.7433	-0.0667	0.3622	-0.0106	-0.4887	0.1626	0.9913	0.0015
New Hampshire	2018	0.7816	-0.0691	0.5128	-0.0175	-1.4608	0.3261	0.9084	0.0002
New Jersey	2014	0.5720	-0.0564	0.4040	-0.0271	-2.5715	0.4301	0.8341	0.0006
Oklahoma	2018	0.6832	-0.0616	0.5409	-0.0165	-2.1807	0.3247	0.9783	0.0075
Wisconsin	2008	0.7358	-0.0517	0.4142	-0.0054	0.2753	0.0823	0.9501	0.0000
Wisconsin	2018	0.8226	-0.0317	0.5338	-0.0048	-1.4429	0.1584	0.9795	0.0001
West Virginia	2014	0.9497	-0.0451	0.4334	-0.0102	-1.0233	0.2303	0.9826	0.0000

Notes: For auctions in which the commission does not enter the scoring rule, we approximate symmetric Bayes-Nash equilibrium by computing how 100 "focal firms" would bid, given a belief that competitors' bids are a second-order polynomial in their costs and technical score. We regress the best responses of the focal firms (e.g., their score bid) on the corresponding variables, update beliefs, and iterate. The table summarizes the regression results for each auction after 42 iterations (21 with the linear beliefs, 21 with the flexible belief function). We provide the coefficients, the R^2 of each state-specific regression, and the MSE between the final belief and regression coefficients.