

Private Marketplaces and Supply-Side Selection Dynamics in Display Advertising

Simeon Duckworth

Lars Nesheim

Mateusz Mysliwski

2026-04-10

Abstract

Private Marketplaces (PMPs) have grown rapidly in programmatic advertising, yet observed PMP price premia and adoption patterns are difficult to interpret because publishers allocate heterogeneous impressions across marketplaces using information that is not observed by researchers. We develop a supply-side model in which publishers choose whether to participate in RTB and/or PMP (extensive margin) and, conditional on participation in both, allocate impressions between channels (intensive margin). Observed daily average prices are corrected for quality-based selection using inverse Mills ratios, yielding latent, quality-adjusted price indices. We embed these latent indices in a dynamic factor model to capture aggregate shocks and heterogeneous product exposure. The model quantifies the role of selection in observed PMP pricing, estimates marketplace supply elasticities, and provides a decomposition-ready framework for understanding PMP growth under platform and privacy shocks.

1 Introduction

Over the last two decades, Real-Time Bidding (RTB) has often been described as the canonical “free-market” mechanism of digital advertising. By allowing any advertiser, large or small, to compete for impressions in real time, RTB lowers access barriers to premium inventory. A small local advertiser can bid for space on a national news website under the same auction rules as a multinational brand. RTB therefore appears to realise the textbook promise of an open electronic market: broad participation, rapid price discovery, and scalable matching.

Against this backdrop, the rise of invitation-only Private Marketplaces (PMPs) presents a conceptual puzzle. PMPs restrict participation to pre-selected buyers and require bilateral set-up. They are, by design, less open. Yet over the past decade, PMPs have grown to represent a substantial share of transactions, rivalling or exceeding open RTB. PMPs were introduced and widely perceived as a mechanism for signalling higher-quality impressions. Publishers could curate buyer access, bundle inventory with richer audience information, and offer stronger guarantees of brand safety and contextual fit. The expectation, therefore, was that PMPs would

command a price premium relative to open RTB because they reduced adverse selection and improved matching quality.

Empirically, PMPs did command such a premium early on. But that premium has narrowed over time even as PMP participation has expanded. Why does a marketplace designed to signal and monetise quality grow rapidly whilst its apparent premium compresses? If PMPs function primarily as a screening and signalling device, one might expect the premium to persist or even widen as adoption increases. Instead, PMP share rises as the PMP-RTB price gap narrows. That joint evolution is the central puzzle of the paper.

From the perspective of auction theory, the coexistence of open and restricted marketplaces is not inherently paradoxical. When participation is endogenous and entry is costly, restricting access can increase revenue by shaping the composition of bidders and intensifying effective competition. As participation frictions decline, the associated premia can erode (Bhattacharya et al., 2014; Lauermann & Wolinsky, 2025). But these theoretical possibilities leave open a central empirical question. Does the PMP premium reflect intrinsic properties of the marketplace mechanism, or does it reflect endogenous allocation of higher-quality inventory into that mechanism?

This paper argues that the key to the puzzle is dynamic, quality-based selection on the supply side. Publishers control the routing of each impression. They can allocate it to an open RTB marketplace or to invitation-only PMPs. RTB is an open auction with broad buyer participation and minimal ex-ante restrictions. PMPs are curated marketplaces that restrict buyer access and allow publishers to bundle impressions with contractual features or richer signals. These mechanisms compete for the same underlying inventory and coexist within most large publisher stacks. Over time, regulatory and technological shifts, including privacy regulation, browser tracking restrictions, auction-format changes, and increasing automation, have altered both the costs of operating each marketplace and the relative value of open versus restricted trading.

Two descriptive facts motivate our analysis. First, PMP share rises persistently over time. Second, PMPs command a positive price premium early on, but that premium narrows as participation expands, with distinct patterns across display and video. On their own, however, these price differences are not economically interpretable. Publishers observe impression-level signals, including engagement, logged-in status, contextual fit, and audience quality, that the econometrician does not. If higher-quality impressions are disproportionately allocated to PMPs, observed PMP prices mechanically overstate latent PMP price indices. As PMP participation broadens and marginal inventory enters, compositional shifts alone can compress the observed premium even if the underlying revenue properties of the two marketplaces are unchanged. Without modelling selection, we cannot distinguish erosion of a true marketplace premium from changing allocation of quality.

We address this problem with a structural supply-side model of publisher participation and allocation across marketplaces. Publishers first decide whether to operate in RTB only, PMP only, or both marketplaces, and then allocate impressions across active mechanisms. Observed RTB and PMP prices are treated as selected realisations of latent price indices. Selection arises because unobserved quality shocks affect both allocation decisions and prices. We model this explicitly using inverse Mills ratios with product-specific covariance parameters. The latent RTB price, latent PMP price, and allocation index are embedded in a dynamic factor structure

with product fixed effects and a small number of common shocks. This framework allows us to separate persistent heterogeneity, aggregate demand dynamics, endogenous selection, and idiosyncratic noise in a high-dimensional panel with incomplete overlap across marketplaces.

Our empirical contribution differs from most structural work in digital advertising, which has focused primarily on the demand side, including bidder behaviour, auction design, and advertiser information environments, rather than on the publisher’s supply-side allocation problem (e.g., Breitmar et al., 2023; Choi & Carl F. Mela, 2019a; Choi & Carl F. Mela, 2019b). We contribute in three ways. First, we develop and estimate a structural model of publisher participation and allocation between RTB and PMP that matches the institutional dual-marketplace choice problem. Second, we quantify quality-based selection into PMPs and show that it is economically large early in the sample but attenuates as PMP participation expands. Third, we recover supply elasticities and their evolution over time in a framework that supports counterfactual analysis of platform, privacy, and technological shocks. This contribution depends on a substantially longer data horizon than prior empirical work on PMP and RTB. The longer horizon allows us to characterise dynamics in participation, selection, and price evolution rather than treating these mechanisms as static. It is essential because the attenuation of selection and the evolution of elasticities are gradual processes that cannot be identified in short panels.

Three findings emerge. First, effective entry barriers into PMP fell materially over time. Converting utility estimates into dollar-CPM equivalents, effective entry costs decline by roughly \$0.26 per CPM between 2018 Q2 and 2020 Q3, about 4 percent of the mean PMP price, and counterfactual steady-state participation is meaningfully lower when entry costs are held fixed. The extensive margin is therefore a first-order driver of PMP growth. Second, selection on unobserved quality was quantitatively large early in the sample, accounting for roughly 30 percent of video PMP price variance and 14 percent in display, but declines steadily as PMP scales. A substantial share of the narrowing PMP–RTB price gap is explained by this attenuation in selection rather than by a collapse of intrinsic marketplace premia. Third, supply elasticities are large but fall over time. Publishers retain substantial discretion in reallocating inventory across marketplaces, yet responsiveness declines as participation expands and the composition of products changes.

Taken together, the findings point to a supply-side interpretation of PMP growth. The expansion of a restricted-entry marketplace need not reflect persistent exclusivity. In our estimates, it reflects declining entry costs and attenuating selection on unobserved quality. As PMPs scaled, setup frictions fell, the quality composition of routed inventory converged, and informational rents narrowed. The compression of the PMP premium is therefore largely a selection story, driven by changes in who participates and what is allocated, rather than a change in the intrinsic revenue properties of the marketplace mechanism itself.

The rest of the paper is organised as follows. Chapter 3 relates our approach to the economics of dual-channel trade, adverse selection, and supply-side elasticity. Chapter 2 describes programmatic marketplaces, RTB and PMP, and motivates why publishers face a participation and allocation problem. Chapter 4 describes the data, measurement, and stylised facts that discipline the model. Chapter 5 presents the supply model and estimation. Chapter 6 reports the parameter estimates and elasticity calculations. Appendices document Diode data construction and technical derivations.

2 Industry background

This section provides a brief overview of the programmatic advertising marketplace and introduces the terminology we use throughout the paper. We focus on the institutional details that matter for our empirical analysis, so the picture is necessarily simplified. Our setting is the open web, where publishers sell inventory across a decentralised set of websites and intermediaries, rather than within walled gardens, such as social media and video platforms, where inventory is controlled by a single platform. Our focus is the publisher’s choice between Real-Time Bidding (RTB) and Private Marketplaces (PMPs), because these are the two channels through which publishers allocate open-web inventory in our setting.¹ The discussion emphasises the features that matter for pricing, routing, and selection.

2.1 Programmatic advertising

We focus on programmatic advertising on the open web, where ad impressions are matched automatically between publishers and advertisers. An impression is a single opportunity to show an advertisement to a user when they load a web page. In 2022, advertisers spent roughly \$125bn worldwide on programmatic advertising, about a quarter of all digital display spending (GroupM, 2022). The defining characteristic is automation. For millions of impressions, programmatic technology matches buyers with sellers in roughly 200 milliseconds, enabling the winning ad to load before the user’s page renders (IAB, 2014).

Publishers operate websites and own the ad slots. Trade takes place through intermediaries on the sell side and buy side, but we abstract from their strategic role and focus on the allocation problem between publishers and advertisers. The price of inventory is quoted in *CPM* (cost per mille), the cost per thousand impressions.

Programmatic trading takes several forms. We focus on two. RTB is an open-access auction available to many buyers. PMP is an invite-only auction built around Deal IDs. The subsections below set out the institutional features of each.

2.1.1 Real-Time Bidding (RTB)

Real-Time Bidding (RTB) is the archetypal programmatic marketplace. Advertising is sold impression by impression through an open-access auction. For each impression, the publisher or a sell-side intermediary issues a bid request that describes the user and page context. Buyers evaluate the impression and submit bids on behalf of advertisers, and the highest bidder wins (Wang et al., 2017).

Historically, much open-market inventory was sold through a waterfall, in which publishers offered the same impression sequentially across exchanges or networks according to a priority order. This limited direct competition because advertisers and their intermediaries could access the impression through lower-ranked exchanges or networks only if higher-ranked ones declined it. Header bidding changed this architecture by allowing publishers to solicit bids from multiple

¹Publishers can also sell inventory through programmatic guarantee and direct deals (Adshead et al., 2019). These deal types are less common and fall outside our study. We restrict attention to the publisher’s choice between RTB and PMP.

exchanges simultaneously before the final allocation decision.² This increased competition across exchanges and weakened the priority advantage embedded in the waterfall (Despotakis et al., 2021).

Open display also largely transitioned from second-price to first-price auctions during our sample period. Google rolled out first-price in September 2019 (Goke et al., 2022; Google, 2019). Header bidding helped accelerate this shift by making competition across exchanges more direct and weakening the logic of older sequential selling arrangements (Despotakis et al., 2021). Together, header bidding and the move to first-price produced a discrete market-wide change in RTB pricing and auction design that falls within our sample period. They also reduced the operational frictions of allocating inventory across RTB and PMP, which helps explain why dual-market participation became easier over time.

2.1.2 Private Marketplaces (PMPs)

Private marketplaces (PMPs) restrict access to a publisher’s impression inventory to a curated, pre-approved group of advertisers. Access is typically negotiated bilaterally and activated through a Deal ID embedded in RTB bid requests (Media, 2020). Unlike open RTB, PMP participation depends on a commercial agreement between the publisher and the advertiser, often via an intermediary. PMPs therefore operate within the RTB infrastructure, but alter the allocation rules in favour of deal holders.

Historically, the most common PMP structure grants participating advertisers first look at an impression. Before the open RTB auction is conducted, bids bearing a valid Deal ID are evaluated against a negotiated floor price. If the highest qualifying PMP bid clears that floor, the impression is sold privately. Only if no PMP bid clears does the impression pass to the open exchange. This creates a sequential two-tier structure, with a private auction followed, if needed, by an open auction (Media, 2020). Each buyer with a Deal ID therefore holds an option to bid or pass. This option is most valuable to buyers with private user-level signals (Kim et al., 2021).

PMP inventory is typically sold with specified characteristics such as audience, context, and floor prices. Access to publisher-held data is a key differentiator. Publishers can bundle first-party signals that are not available in the open exchange. Set-up and relationship costs are higher than in RTB, including deal creation, buyer selection, and floor-setting, although some adtech costs can be lower (Adshead et al., 2019).

A key economic question is why both channels coexist for the same inventory. Publishers observe impression-level quality when they route each impression. Plausible routing rules favour PMP for high-quality traffic, for impressions that the open auction would underprice, for example cookieless impressions in Safari, or, when monitoring is imperfect, for low-quality impressions. These routing rules create the selection problem that we address in the model.

The empirical challenge is that bid-level logs, which would directly reveal routing and bidding, are typically proprietary and not available across the market and over time. We therefore work with daily aggregates and handle selection explicitly. Our unit of observation is a product,

²Header bidding is a technical arrangement in which the publisher calls multiple exchanges before the ad server makes its final allocation decision. Relative to the older waterfall, it allows exchanges to compete for the same impression in parallel rather than sequentially.

defined as a unique combination of publisher site, creative format, and ad format-size grouping. We describe the data construction and sample in detail in Chapter 4.

Observed PMP prices reflect both a true channel premium and the fact that better impressions are steered into PMP. Observed RTB prices are depressed because RTB receives the residual. We therefore cannot interpret raw price gaps without modelling this allocation.

2.2 PMP growth and market context

PMPs have grown rapidly since 2015. In the United States, PMP spending overtook RTB (Open Exchange) as the dominant programmatic deal type by 2018 (Figure 1). This pattern in a major market illustrates the broader reallocation towards PMP that motivates our analysis.

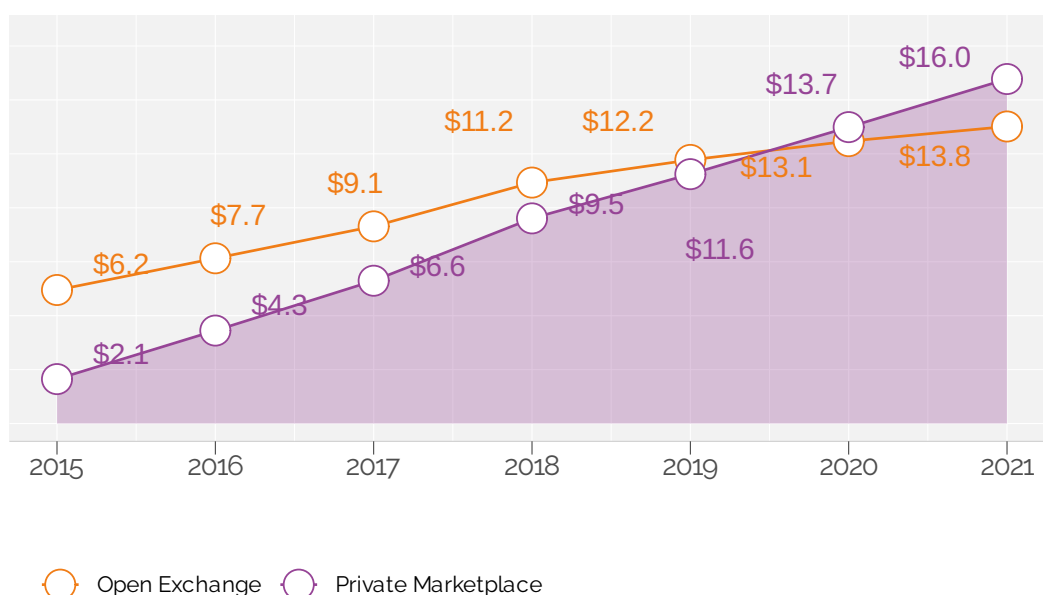


Figure 1: US programmatic advertising spending by deal type, \$billions, 2015–2021. Source: eMarketer.

Several forces are commonly cited for this shift. First, PMP provides a screening mechanism for high-quality inventory (Li et al., 2014). At the same time, there have been repeated negative shocks to the perceived quality of RTB inventory. Ad fraud is pervasive. Around 10% of impressions are fraudulent (CHEQ, 2020). Brand safety also matters. Programmatic ads can appear in unsafe or low-quality contexts, and a sizeable share of impressions runs on Made-For-Advertising sites (ANA, 2023b).

Second, PMP gives publishers stronger incentives to collect and share valuable first-party data. Tracking restrictions have reduced the value of third-party cookies. Apple’s Intelligent Tracking Prevention, GDPR, and browser changes have curtailed user-level targeting. By 2023, fewer than 40% of impressions could be targeted using cookies.

Third, adtech innovation reduced the cost of participating in PMPs. Header bidding is one example. Finally, adverse selection can itself depress the value of RTB inventory, which increases the value of participating in PMP (Balocco et al., 2025).

Our analysis speaks to these aggregate trends through three empirical patterns that we document: PMP share grows substantially over our sample period; PMP price premia decline; and publishers operate in both channels simultaneously. Our model is designed to rationalise these patterns whilst correcting for selection. It does not identify which force dominates. Instead, we treat them as aggregate shocks and focus on how publishers respond to prices when they choose whether and how much to allocate to each channel.

We next discuss how these institutional features connect with economic literatures that motivate our modelling choices.

3 Related Literature

We relate our analysis to three strands of work. First, the economics of dual-channel trade asks why publishers run RTB and PMP simultaneously, and what the revenue consequences of that choice are. Second, theories of adverse selection and the empirical literature on quality sorting motivate our selection correction: if publishers route better inventory to PMP, observed price gaps are not interpretable without modelling that routing decision. Third, the literature on common shocks and price formation motivates our factor structure: market-wide events (auction format changes, privacy regulation, seasonality) move publisher prices together, but with heterogeneous exposure that must be absorbed before supply elasticities can be estimated. We use these three strands to motivate a structural supply-side model of daily participation and allocation decisions.

3.1 Programmatic advertising and dual-channel trade

The central economic question is why guaranteed or semi-guaranteed contracts coexist with RTB for the same inventory, rather than one channel displacing the other. RTB is liquid and operationally cheap, but it gives publishers little control over who buys and limits how they can package and monetise private audience signals. PMP requires upfront costs (deal creation, buyer selection, floor-setting, ongoing maintenance), but in return allows publishers to sell curated inventory to selected buyers with stronger quality controls and a tighter connection to publisher-held data. Theory establishes that this coexistence is an equilibrium outcome rather than a transitional arrangement: publishers maximise revenue by running both channels, not by committing fully to either (Cohen et al., 2022; Kim et al., 2021; Sayedi, 2018). The TV market provides a useful antecedent: PMP corresponds to the upfront (pre-negotiated, first-look) and RTB to scatter (spot auction, residual inventory), and the fundamental allocation problem of how much inventory to commit in advance versus hold for the spot market has the same structure (Araman & Popescu, 2010).

The revenue consequences of running PMP, however, are theoretically ambiguous. Sayedi et al. (2018) and Choi & Sayedi (2022) show that PMP can raise or lower publisher revenue depending on how information is distributed across channels: concentrating informed buyers in PMP can deplete open auction competition, reducing RTB revenue even as PMP revenue rises. A further complication is that most advertisers operate with fixed campaign budgets. (Balseiro et al.,

2015) show that publishers who set RTB floor prices without accounting for budget-constrained bidder behaviour can forgo up to 40% of potential profit. D’Annunzio & Russo (2023) add that the SSP intermediary layer shapes what advertisers know about inventory quality, creating a market context for selection that publishers do not fully control. Zeithammer & Choi (2025) show that this intermediary layer introduces a further distortion: because exchanges shade their bids strategically in the publisher’s header auction, earning the spread between their internal clearing price and their own header bid, the publisher is forced to raise its reserve price above the first-best level to compensate, reducing sale probability and leaving all channel members worse off. Taken together, these results motivate direct estimation from market data rather than relying on theoretical predictions about which channel delivers higher revenue.

3.2 Adverse selection, sorting, and evidence on quality differences

Arnosti et al. (2016) establish the core prediction of the information-asymmetry literature: open auctions expose less-informed advertisers to adverse selection, as performance advertisers who can assess individual impression value cherry-pick the best opportunities, leaving brand advertisers, who cannot, with below-average quality at market prices. The selection problem at the heart of our paper is the supply-side counterpart: the same information asymmetry that disadvantages brand advertisers in the open exchange gives publishers an incentive to route their best inventory away from RTB, making observed PMP premia a confound of genuine channel effects and quality sorting. PMP is a natural institutional response: it curates the buyer pool and lets publishers signal quality through the deal structure rather than exposing every impression to open competition. Bergemann et al. (2022) formalise the information-design rationale: a publisher who controls disclosure can raise more revenue by pooling (not fully revealing) high-value inventory quality, since full disclosure intensifies competition and extracts information rents for bidders rather than the publisher. Levin & Milgrom (2010) show a related result: very fine targeting thins individual auctions by reducing the number of bidders who compete for each impression, so bundling similar impressions into a deal pool can raise revenue by preserving competitive thickness. Abraham et al. (2020) show that the direction of the information asymmetry matters: when private buyer signals identify low-value impressions (“lemons”), open auction revenue is badly depressed; when they identify high-value impressions (“peaches”), the seller is partly protected. All three results point to the same implication: the PMP channel can be understood as an information-management tool, and the price premium it commands reflects both a genuine channel benefit and the fact that better inventory is being routed there.

Selection, this second force, is the econometric problem at the heart of our paper. If publishers systematically steer higher-quality impressions into PMP, then observed PMP prices overstate the true channel premium and observed RTB prices understate it. Balocco et al. (2025) provide the clearest empirical evidence. Combining a game-theoretic model with bid-request data from a European DSP, they show that RTB impressions from dual-channel publishers are approximately 25% less viewable than those from single-channel publishers, a gap invisible to buyers from disclosed signals at auction time and robust across eleven alternative specifications. The adverse selection is structural. Single-channel publishers always lose revenue when PMP exists, because dual-channel publishers degrade the RTB quality pool and force advertisers to shade their bids, a welfare loss that falls entirely on RTB-only competitors. Devaux (2023) offer an important nuance: she finds that CPM contracts reduce viewability by 18% relative to performance

contracts but that PMP does not uniformly crowd out RTB quality. The market-level effect depends on portfolio composition and quality heterogeneity, precisely the cross-product variation our structural model accommodates through publisher-specific selection parameters. PMP prices are only observed when publishers activate PMP on a given day, and that participation decision is itself endogenous to inventory quality. Raw PMP premia and their evolution over time are therefore not interpretable without modelling both the routing and the participation decision, which is what our model does.

3.3 Aggregate shocks and price formation

Market-wide shocks are a first-order feature of our data. The switch from second-price to first-price auctions across all major exchanges 2017–2019, driven by the adoption of header bidding, was a structural break that shifted all publishers’ RTB revenue simultaneously (Despotakis et al., 2021). Critically, demand-side adaptation to this change was slow: for months after the switch, bidders continued to overbid relative to the new rules, mechanically inflating observed RTB prices (Alcobendas & Zeithammer, 2023). Privacy regulation created further common shocks (GDPR, ITP, and CCPA each induced market-wide co-movements), but with heterogeneous channel exposure: Duckworth et al. (2023) find that Apple’s ITP restrictions hit RTB more heavily than PMP, because RTB relies more on third-party tracking. The same policy change therefore moved RTB and PMP prices in different directions for the same publisher products.

Two papers motivate why these common shocks transmit differently across products. Gentzkow et al. (2024) show that a publisher’s pricing power depends not on audience size but on how much exclusive, non-overlapping audience it delivers to advertisers: outlets whose audiences can easily be reached elsewhere face stronger substitution and lower prices. When a common shock reduces the value of targeting, the effect on any one publisher depends on its audience’s substitutability, exactly the cross-sectional heterogeneity in factor loadings our model is designed to recover. Hristakeva & Mortimer (2023) document persistent, relationship-driven price dispersion in TV upfront markets even for equivalent inventory, showing that price variation contains both structural and relational components that aggregate shocks must be separated from. The factor structure we use (common factors with product-specific loadings) is the natural tool for this separation. Breitmar et al. (2023) provide the closest methodological precedent: they show that estimating revenue heterogeneity from grouped platform data (as we have) requires time-varying interactive effects, precisely because aggregate shocks load differently across products.

3.4 Gap and positioning

Most structural work in digital advertising focuses on the demand side: bidder behaviour, auction design, and the information environment facing advertisers. The supply side — how publishers decide whether to activate PMP, and how they allocate impressions across channels conditional on participating — has received far less structural attention, particularly with the kind of grouped, aggregated data that platforms make available (Breitmar et al., 2023; Choi & Carl F. Mela, 2019a; Choi & Carl F. Mela, 2019b).

We contribute in three ways. First, we develop and estimate a supply-side model of publisher participation and allocation between RTB and PMP that matches the institutional choice problem documented in the dual-channel literature. Second, we quantify quality-based selection into PMP, finding that it inflated observed PMP prices by 30–40% in the early period, declining toward

zero by the end of 2020 as PMP scaled and quality differences between channels narrowed. Third, we recover supply elasticities and their evolution over time, finding median own-price elasticities of 3.4 for display and 7.3 for video (selection-corrected), declining substantially as the market matured, in a way that supports counterfactual analysis of platform and privacy shocks.

We turn next to the data and sample construction (Chapter 4).

4 Data

We now describe the data and the main empirical patterns that motivate the structural model. Our dataset derives from Diode, a proprietary panel capturing aggregated daily impression-level advertising transactions from US publisher websites. The observation period spans 1 April 2018 through 30 September 2020, encompassing 913 days and 3,424 distinct products. Each product represents a unique combination of publisher site, creative format (display or video), and size grouping. We measure prices in CPM (cost per mille), the cost of media excluding fees, expressed in dollars per thousand impressions. The construction of our dataset is described in more detail in Appendix A.³

4.1 Descriptive statistics

Table 1 presents summary statistics for our estimation sample, summarising CPMs, impressions, and market structure by product. PMP transactions command a substantial premium over RTB. The median PMP CPM is \$11.31 for video and \$4.64 for display, while the median RTB CPM is \$9.24 for video and \$2.88 for display. Impression volumes are highly skewed, with a handful of large publishers accounting for the bulk of activity.

Heterogeneity is considerable within both channels. Video and display differ in price levels, PMP penetration, and participation patterns. We treat website \times format as the product dimension in part because this format heterogeneity is economically meaningful—advertisers and publishers make distinct choices for video versus display inventory.

³External validation against comScore and Association of National Advertisers (ANA) benchmarks confirms that our sample is broadly representative of the distribution of CPMs and sites. See Appendix A.2 for details.

Table 1: Summary statistics, PMP and RTB transactions.

	Display				Video			
	Mean	P10	P50	P90	Mean	P10	P50	P90
Media CPM (\$)								
All	4.20	1.89	3.57	7.31	10.90	4.77	9.98	18.23
RTB	3.27	1.55	2.88	5.22	9.70	4.70	9.24	14.82
PMP	5.19	2.59	4.64	8.20	12.19	4.86	11.31	21.53
Total Impressions								
All	19.29M	0.01M	1.64M	26.45M	2.43M	0.00M	0.11M	3.45M
RTB	25.68M	0.01M	3.17M	33.65M	2.56M	0.00M	0.16M	4.03M
PMP	12.59M	0.01M	0.92M	20.75M	2.29M	0.00M	0.06M	2.35M
Market Structure								
PMP Share	28.5%	0.5%	19.0%	72.8%	24.0%	0.0%	13.0%	67.6%
% Days Active	73.5%				48.8%			
Sample Dimensions								
Total Observations	2,887,442							
Time Periods (Days)	913							
Sites	515							
Site Categories	22							
Creative Formats	7							
Total Products	3,403							

Note:

CPM in USD; impressions in millions (M) per product-deal over sample period.

% Days Active: share of all 913 sample days on which a product has any price observation (either channel).

% Active Days rows use active days as denominator and sum to 100% across RTB Only, PMP Only, and Both.

PMP = Private Marketplace; RTB = Real-Time Bidding.

4.2 Key market trends

We document four patterns in the data that motivate the structural model. PMP adoption rises over time, the PMP price premium narrows, products increasingly operate in both channels, and heterogeneity across products is substantial.

4.2.1 Rising PMP share

PMP impression share rose from approximately 20% to 60% over the observation period (Figure 2). The increase occurred in both display and video inventory, though the level and trajectory differ by format. Display products adopted PMP earlier and exhibit a steeper climb. Video products began with lower PMP penetration but converged by late 2020. This growth reflects two distinct forces — new publishers joining PMP and existing participants routing more impressions there — which we decompose in Section 4.2.5 below.

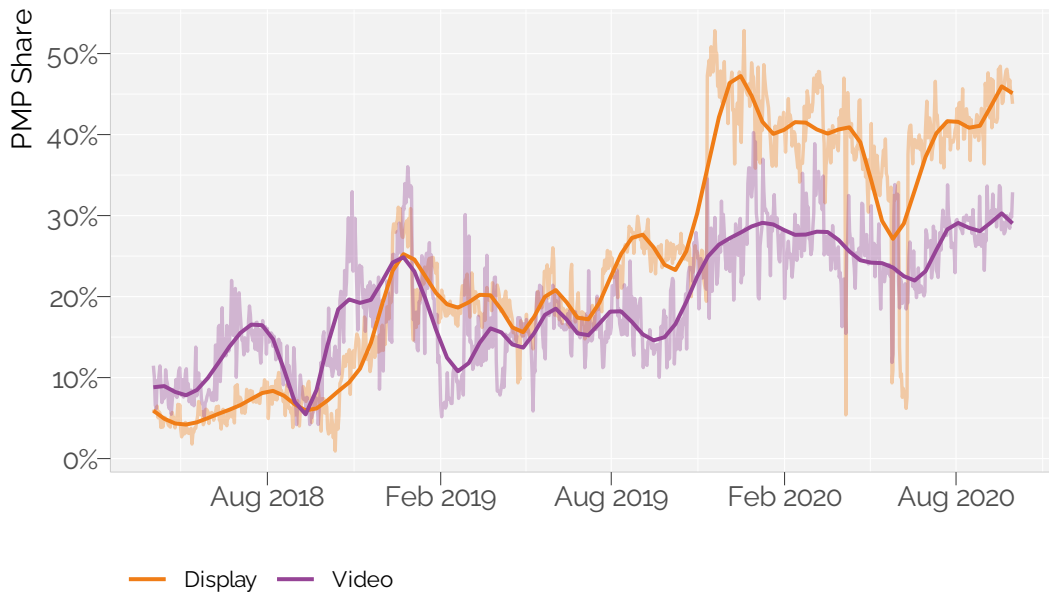


Figure 2: PMP impression share over time, by creative format. Mean share with cubic spline trend.

4.2.2 Declining PMP premium

While PMP prices remain elevated relative to RTB, the gap narrowed over time (Figure 3). The median PMP CPM for display inventory fell from roughly \$8 to \$5, while RTB display remained stable near \$2. Video CPMs exhibit greater volatility but follow a similar convergence pattern. Two mechanisms may explain this narrowing. First, composition matters: if lower-quality sites adopt PMP later, average PMP prices fall mechanically. Second, prices may adjust within product as publishers reduce PMP floors when competition intensifies or RTB demand strengthens. The structural model separates these channels. The variance decomposition in Appendix B shows that product fixed effects account for 35% of price variation, common factors explain 40%, and the Mills correction contributes 8%.

4.2.3 Increasing dual-channel participation

On a given day, a product may transact in RTB only, in PMP only, or in both channels. As Table 1 shows, RTB-only accounts for the bulk of product-days early in the sample, while PMP-only remains rare throughout, consistent with publishers maintaining an RTB backstop. Figure 4 shows a rapid shift towards dual-channel operation over 2018–2020 as PMP adoption diffuses.

Part of this shift is mechanical. Dual-channel operation requires PMP adoption. Its speed and breadth nonetheless point to declining adtech set-up costs. Establishing a PMP requires DSP relationships, ad-server configuration, and negotiated deal structures, and these frictions

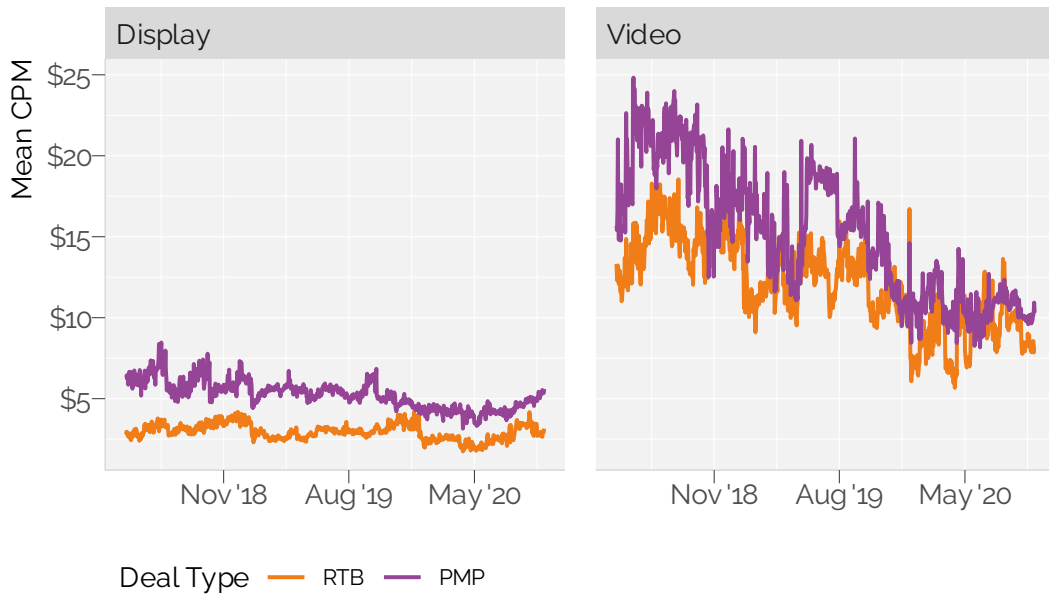


Figure 3: Advertising prices by deal type and creative format, mean cost per thousand impressions (CPM) by product.

plausibly fall over time as programmatic infrastructure matures. We return to this mechanism in the counterfactual analysis (Section 7.1).

4.2.4 Substantial heterogeneity across site categories

We define a product as site (s_j) \times creative format (z_j) because CPM heterogeneity is cross-product as well as temporal. For example, we treat large and small video and display inventory on the New York Times as separate markets. Precise definitions and construction details for each characteristic are in Appendix A. Figure 5 shows the distribution of log CPM by site category, split by PMP and RTB. Technology and business sites command higher CPMs in both channels; news and entertainment account for the largest volumes but exhibit lower median prices. The ridges reveal substantial dispersion both across categories and within them, justifying a product-level modelling approach.⁴

⁴A variance decomposition of log CPM across our 3,403 products and 913 sample days confirms this: roughly half of total variation is between products — persistent cross-sectional differences in price levels — and roughly half is within products over time. This motivates the structural model’s product fixed effects μ_j for cross-sectional levels and common factors $\Lambda_j f_t$ for shared time dynamics.

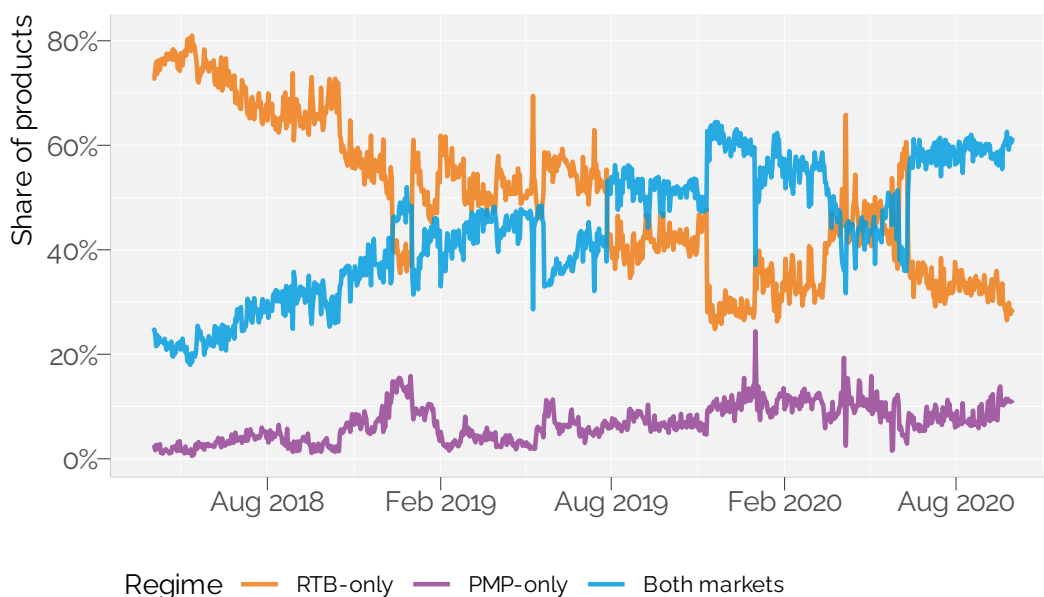


Figure 4: Daily market configurations over time, share of products operating RTB-only, PMP-only, or in both markets.

4.2.5 Two margins of PMP growth

Figure 2 shows that PMP growth occurs along two margins. New publishers adopt PMP over time, and existing participants route a larger share of impressions through PMP.⁵

This decomposition motivates the structure of the model. We estimate a participation model for whether a product operates in PMP and a supply model for how participating products allocate impressions across PMP and RTB. We discuss both models in detail in Chapter 5 and report the estimates in Chapter 6.

5 Model and Estimation

We present a dynamic selection model with high-dimensional heterogeneity. Publisher-products make two distinct decisions, summarised in Figure 7. The extensive-margin (Stage 1) decision is whether to operate RTB-only, PMP-only, or both markets on a given day: a state-dependent switching problem that identifies entry and switching costs and explains the persistence in market configurations documented in Section 4.2. The intensive-margin (Stage 2) decision is how to allocate individual impressions between RTB and PMP, conditional on both markets

⁵Aggregate PMP share can be written as the product of a participation rate w_t and a conditional allocation share \bar{s}_t , $S_t \approx w_t \times \bar{s}_t$. Differencing relative to the first 30 days decomposes ΔS_t into an extensive-margin term ($\bar{s}_0 \cdot \Delta w_t$), an intensive-margin term ($w_0 \cdot \Delta \bar{s}_t$), and their interaction. Both w_t and \bar{s}_t are computed from observed participation choices and observed impression shares on the matched sample; no structural parameters are required.

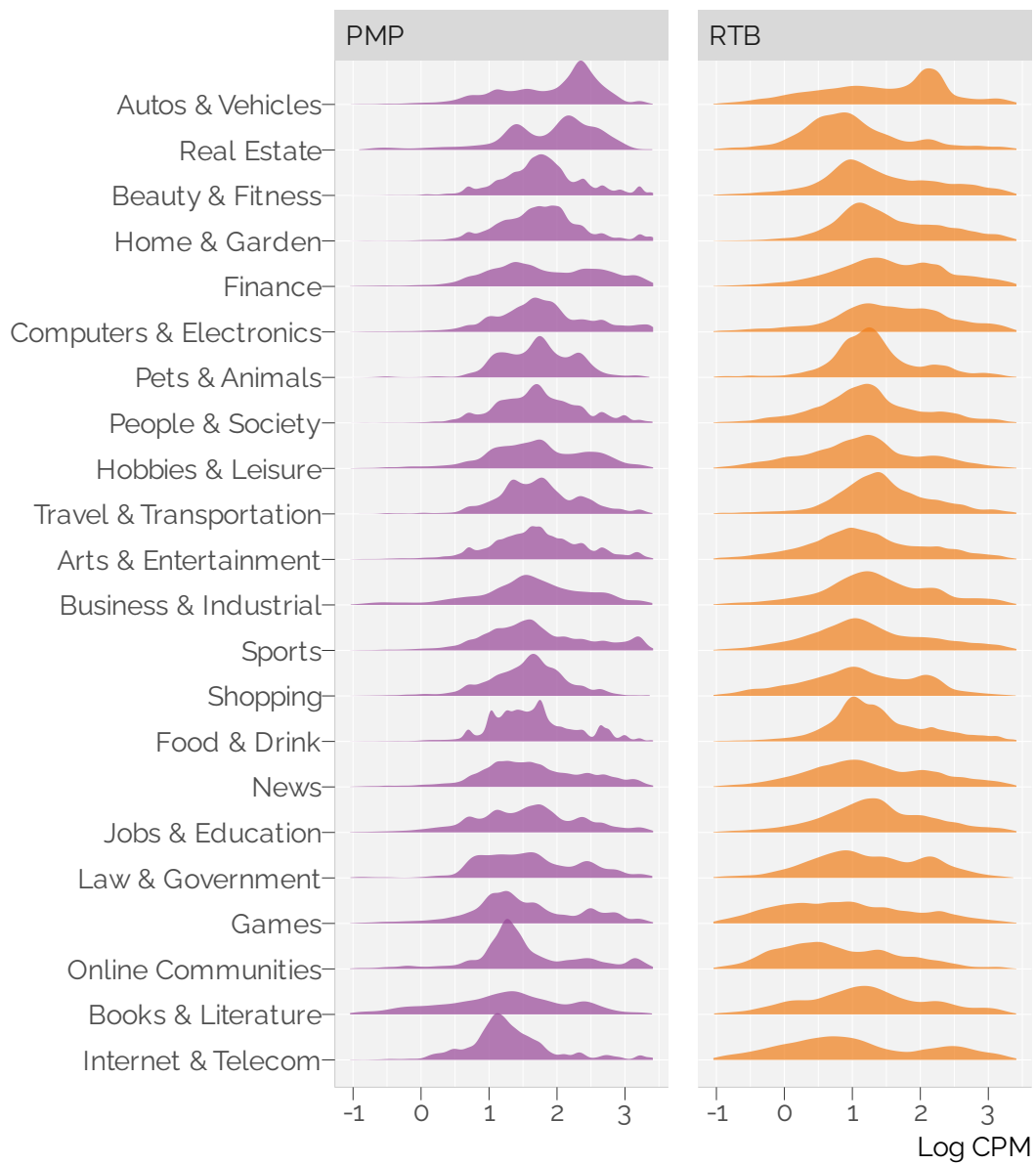


Figure 5: Log CPM distribution by site category and deal type.

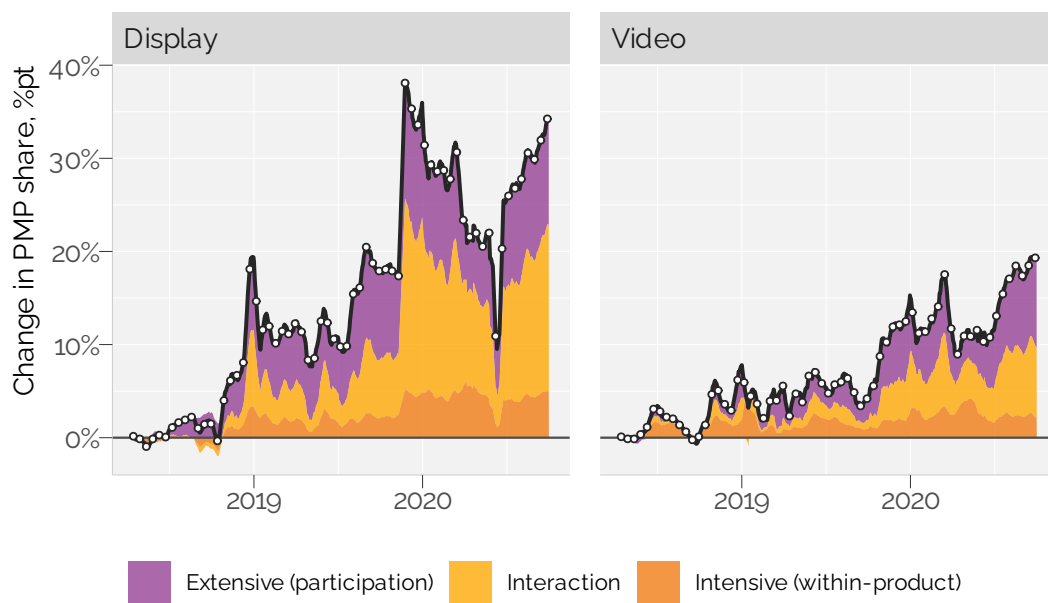


Figure 6: Decomposition of aggregate PMP share growth relative to first 30 days, stacked areas show extensive (participation), intensive (within-product), and interaction components (14-day smoothed). Grey line shows actual ΔS .

being active: this quantifies quality-based selection and underlies the latent price dynamics. Structural estimation is required to recover supply elasticities, separate selection effects from genuine price dynamics, and conduct counterfactual analyses.

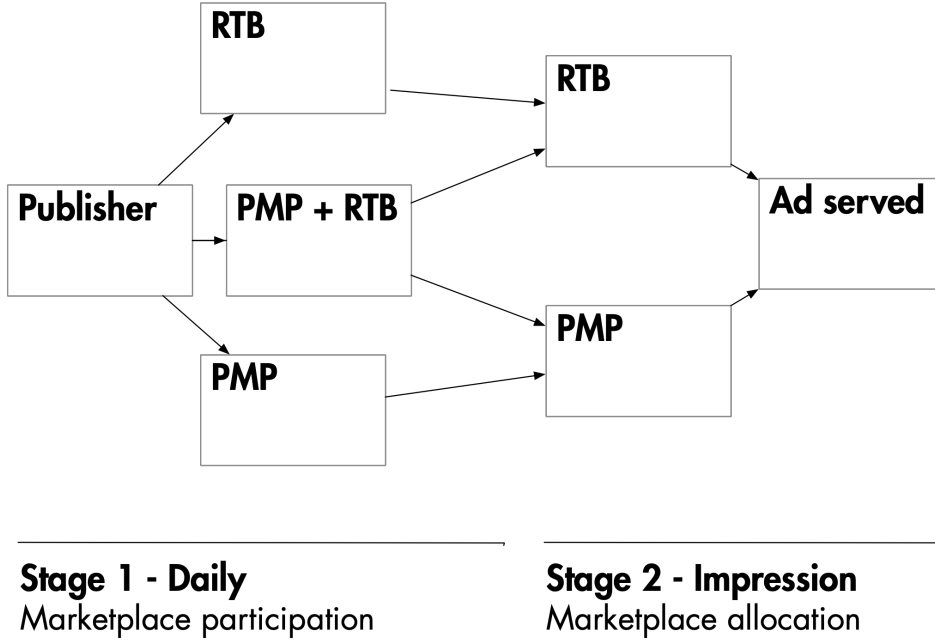


Figure 7: Stylised schematic of the publisher routing problem. Publishers can operate RTB only, PMP only, or both markets. Conditional on joint participation, impressions are routed to RTB or PMP before an ad is served.

We solve the model by backward induction, starting with Stage 2. Stage 2 delivers the expected PMP attractiveness objects \bar{y}_{3tj} and y_{3tj}^+ that enter Stage 1 switching utilities. Without Stage 2 estimates, the Stage 1 switching payoffs are undefined; estimation must therefore follow the same order.

5.1 Stage 2: intensive margin

5.1.1 Supply

Each day, websites (publishers) $s \in \mathcal{S}$ have inventories of impressions that they sell to advertisers. Conditional on PMP being active on a given day, for each impression i a publisher makes an intensive-margin decision: which market to route that impression to. We indicate the RTB market with $m = 1$ and the PMP market with $m = 2$.

Impressions are grouped into a finite number J of product categories based on site (s_j) and creative format ($z_j \in \{1, \dots, 7\}$). The seven formats cover four display sizes (desktop horizontal, desktop vertical, desktop other, mobile) and three video durations (short, long, unknown), so the video/display distinction is embedded in z . Each impression in category j has product characteristics $(s_j, z_j, \varepsilon_{ji})$ where $\varepsilon_{ji} \in \mathbb{R}^3$ is the vector of unobserved quality. All characteristics are observed by the publisher but only z_j is recorded in the Diode dataset. Thus ε_{ji} is the unobserved component of quality. Later, we relate estimated heterogeneity to observable product

characteristics including site category, format, scale, viewability, and advertiser-paid data and technology fees. These variables are used in descriptive plots and regressions on estimated parameters, but they do not enter the structural model directly. Definitions are in Appendix A.

Let $y_{tj} = (y_{1tj}, y_{2tj}, y_{3tj})$ be a vector of aggregate market variables that affect prices and costs for product j in period t . We assume that prices in market $m \in \{1, 2\}$ are given by

$$\begin{aligned} p_{1tji} &= y_{1tj} + \varepsilon_{1tji} \\ p_{2tji} &= y_{2tj} + \varepsilon_{2tji} \end{aligned}$$

Conditional on selling in both markets, the net gain from selling in PMP is

$$\Delta\pi_{tji} = p_{2tji} - p_{1tji} - c_{tji} = y_{3tj} + \varepsilon_{3tji} \quad (1)$$

which implies non-price costs and benefits

$$c_{tji} = y_{2tj} - y_{1tj} - y_{3tj} + \varepsilon_{2tji} - \varepsilon_{1tji} - \varepsilon_{3tji}$$

The net gain $\Delta\pi_{tji} = y_{3tj} + \varepsilon_{3tji}$ determines which market each impression is routed to. These expressions underlie the expected PMP attractiveness objects $(\bar{y}_{3tj}, y_{3tj}^+)$ derived in Stage 1.

We assume $(\varepsilon_{1tji}, \varepsilon_{2tji}, \varepsilon_{3tji}) = \varepsilon_{tji} \sim N(0, \Sigma_j)$ where only the variance of ε_{3tji} is normalised to 1, since this shock determines the discrete choice of market for each impression and its scale is not separately identified from the latent indices.

The intraday variances $(\sigma_{11j}, \sigma_{22j})$ are not identified from daily average data since we cannot recover impression-level variances from daily aggregates. Day-to-day variation in prices and shares is driven by aggregate shocks affecting $(y_{1tj}, y_{2tj}, y_{3tj})$. We leave $(\sigma_{11j}, \sigma_{22j})$ unrestricted as they do not affect the likelihood.⁶

Note that σ_{12j} is not identified. It would not be identified even with impression-level data because we never observe an impression sold in both markets. Bounds on σ_{12j} require estimates of the intraday variances $(\sigma_{11j}, \sigma_{22j})$, which we do not have from daily aggregates - see Identification below.

Interpretation of y_3 The variable y_{3tj} captures factors other than price that affect PMP/RTB market shares. There are times when the relative price of PMP goes up but market share goes down. Several factors explain this. One is pre-existing commitments. A publisher may make longer-term commitments to advertisers to make some types of impressions available on PMP, sacrificing short-term gains from switching to RTB. Alternatively, maintaining a relatively high market share in RTB may offer longer-term gains in attracting future bidders at the cost of short-term PMP gains. Selling additional impressions in RTB could enable learning about new bidder types, improving web content for the future.

⁶As a separate exercise, we show we can bound $(\sigma_{11j}, \sigma_{22j})$ using estimates of intraday variance from a sample of impression-level data - see Appendix C

Aggregate shocks After aggregating the tail of very small websites, there are 590 websites and 7 formats (four display, three video). Not all websites offer all formats. Nevertheless, there are several thousand products. We have daily data for 900+ days. Both cross-sectional and serial correlation are important. We assume price dynamics are driven by a lower-dimensional set of latent aggregate factors. These factors capture exogenous shocks such as Covid-19 and lockdowns, privacy policy changes, and supply-side technology changes. See Appendix C for a detailed discussion of the factors and their economic interpretation.

We assume $(y_{1tj}, y_{2tj}, y_{3tj})$ evolve according to a factor model

$$\begin{aligned} y_{1tj} &= \mu_{1j} + \lambda_{1j}^\top f_t + \xi_{1tj} \\ y_{2tj} &= \mu_{2j} + \lambda_{2j}^\top f_t + \xi_{2tj} \\ y_{3tj} &= \mu_{3j} + \lambda_{3j}^\top f_t + \xi_{3tj} \end{aligned} \quad (2)$$

where $f_t \in \mathbb{R}^r$ is a vector of latent factors and $(\lambda_{1j}, \lambda_{2j}, \lambda_{3j})$ are corresponding factor loadings. We use $\mu = (\mu_{11}, \mu_{21}, \mu_{31}, \dots, \mu_{1J}, \mu_{2J}, \mu_{3J})$ for the $3J$ vector of mean parameters and $\Lambda \in \mathbb{R}^{3J} \times \mathbb{R}^r$ for the stacked factor loadings, where μ_{1j} and μ_{2j} are correlated with average product quality and μ_{3j} is correlated with the selection effect.

Let $\xi_t = (\xi_{1tj}, \xi_{2tj}, \xi_{3tj}, \dots, \xi_{1tJ}, \xi_{2tJ}, \xi_{3tJ})$ and assume $\xi_t \sim N(0, \Sigma_\xi)$. The factors follow

$$\begin{aligned} f_t &= \sum_{s=1}^p A_s f_{t-s} + \eta_t \quad \forall t \geq 1 \\ \eta_t &\sim N(0, \Sigma_\eta) \\ (f_0, \dots, f_{1-p}) &\sim N(\mu_0, \Sigma_0) \end{aligned} \quad (3)$$

5.1.2 Data mapping and selection correction

Conditional on being active in PMP on day t , the market share of PMP for product j is

$$s_{2tj} = \Phi(y_{3tj}) \quad (4)$$

We observe this share only when both markets are active. What is observed therefore depends on the market configuration active at the start of the day. For RTB-only products we observe \bar{p}_{1tj} ; for PMP-only products we observe \bar{p}_{2tj} ; for both we observe $(\bar{p}_{1tj}, \bar{p}_{2tj}, s_{2tj})$. Missing values are pervasive.

The intuition is simple. If publishers steer better impressions into PMP, the observed PMP average price mixes a true price premium with a quality premium. We use a Tobit-style correction to strip out the quality component.

We use the selection index $x_{3tj} = \Phi^{-1}(s_{2tj})$ and inverse Mills ratios M_{1tj} and M_{2tj} , only when both markets are observed on product-day (j, t) . When a product is RTB-only or PMP-only, x_{3tj} , M_{1tj} , and M_{2tj} are undefined, we treat them as missing and drop the corresponding measurement rows for that (j, t) . For product-days with both markets, $M_{1tj} = \phi(x_{3tj})/\Phi(-x_{3tj})$ for RTB and

$M_{2tj} = \phi(-x_{3tj})/\Phi(x_{3tj})$ for PMP, where ϕ and Φ are the standard normal PDF and CDF. The latent indices satisfy

$$\begin{aligned} y_{1tj} &= x_{1tj} + \sigma_{13,j} \cdot M_{1tj} \\ y_{2tj} &= x_{2tj} - \sigma_{23,j} \cdot M_{2tj} \\ y_{3tj} &= x_{3tj} \end{aligned} \tag{5}$$

where $x_{1tj} = \bar{p}_{1tj}$ and $x_{2tj} = \bar{p}_{2tj}$ are observed prices in levels (\$CPM). The selection parameters (σ_{13}, σ_{23}) control how strongly unobserved quality loads into observed prices. RTB receives lower-quality impressions (correction enters positively); PMP receives higher-quality impressions (correction enters negatively).

Missing observations are handled by dropping the corresponding measurement rows on each date. The Kalman filter handles the resulting time-varying pattern automatically.

5.1.3 Estimation

Stage 2 estimates the factor model with selection correction. The DFM summarises a high-dimensional panel into common shocks plus product-specific exposure. We build on the standard large-panel dynamic factor model, estimated by EM using Kalman filtering and smoothing (Bai & Ng, 2002; Doz et al., 2011; Stock & Watson, 2002). Our contribution is to adapt this framework to a structural supply-side setting with endogenous participation and selection correction, following the logic of Heckman (1979).

We estimate the factor model (Equation 2) with dynamics (Equation 3) and selection correction (Equation 5) jointly, setting the factor dimension to $r = 60$ and lag order $p = 2$. The full parameter vector is $\theta = (\mu, \Lambda, \Sigma_\xi, A_1, A_2, \Sigma_\eta, \sigma_{13}, \sigma_{23})$; we also recover the latent factors $\{f_t\}_{t=1}^T$. Estimation proceeds by EM.

The EM algorithm alternates between an E-step and an M-step. In the E-step, we run the Kalman filter and smoother to recover smoothed factors and their second moments. In the M-step, we update parameters. Most updates are closed-form: factor loadings and selection parameters via equation-by-equation OLS; VAR parameters using moments of the smoothed factors. We iterate until the relative change in log likelihood falls below 10^{-6} or until 50 iterations are reached.⁷

We initialise using principal components or by hot-starting from a previous run. Full details of the log-likelihood, Kalman filter and smoother, and M-step updates appear in Appendix E.

5.2 Stage 1: Extensive-margin switching decision

Stage 1 decisions are made before the period’s factor realisation, using yesterday’s information F_{t-1} . This timing assumption ensures Stage 1 does not induce selection on period- t price shocks in Stage 2.

Each day, j chooses one of three market configurations: RTB-only ($k = 1$), PMP-only ($k = 2$), or both markets ($k = 3$). To enter PMP a publisher must have active deal IDs. The costs and

⁷The algorithm centres and scales the data before iterating. After convergence, we recover intercepts as $\mu_{1j} = \tilde{\mu}_{1j} + \sigma_{13,j} \cdot \bar{M}_{1j}$ and $\mu_{2j} = \tilde{\mu}_{2j} - \sigma_{23,j} \cdot \bar{M}_{2j}$, where \bar{M}_{mj} is the time-series mean of the Mills ratio. The means μ are computed once at the end from the centred data and estimated selection parameters.

inertia of maintaining or changing these relationships are the focus of Stage 1.

The Stage 1 decision has a nested two-tier structure. In the upper tier, j observes last period's choice k and decides whether to stay or switch, comparing the payoff from remaining with yesterday's choice against the expected payoff from switching (the inclusive value of the alternatives) net of a switching cost. In the lower tier, conditional on switching, j chooses between the two remaining options.

The deterministic payoffs for each market configuration, formed as conditional expectations of latent price indices given F_{t-1} , are

$$\begin{aligned} V_{1tj} &= \beta_j \bar{y}_{1tj} \\ V_{2tj} &= \kappa_{2tj} + \beta_j (\bar{y}_{1tj} + \bar{y}_{3tj}) \\ V_{3tj} &= \kappa_{3tj} + \beta_j (\bar{y}_{1tj} + y_{3tj}^+) \end{aligned} \quad (6)$$

where $\bar{y}_{1tj} = \mathbb{E}[y_{1tj} \mid F_{t-1}]$ is expected RTB revenue, $\bar{y}_{3tj} = \mathbb{E}[y_{3tj} \mid F_{t-1}]$ is expected PMP attractiveness for a product not yet routing impressions optimally, and $y_{3tj}^+ = \mathbb{E}[y_{3tj} \Phi(y_{3tj}) + \phi(y_{3tj}) \mid F_{t-1}]$ is the selection-corrected expected payoff when both markets are active and the publisher routes each impression optimally. The fixed costs $(\kappa_{2tj}, \kappa_{3tj})$ capture the net attractiveness of PMP-only and both-markets relative to RTB-only. They include the costs of identifying and negotiating PMP deals as well as additional investment in first-party data and verification. $\beta_j > 0$ scales the revenue terms.

Conditional on switching, j faces a binary logit over the two available destinations. The expected payoff from switching away from k , the inclusive value, is

$$\begin{aligned} IV_{1tj} &= V_{2tj} + \log(1 + e^{V_{3tj} - V_{2tj}}) \\ IV_{2tj} &= \beta_j \bar{y}_{1tj} + \log(1 + e^{\kappa_{3tj} + \beta_j y_{3tj}^+}) \\ IV_{3tj} &= \beta_j \bar{y}_{1tj} + \log(1 + e^{\kappa_{2tj} + \beta_j \bar{y}_{3tj}}) \end{aligned}$$

Staying with option k yields payoff $V_{NS,ktj}$ scaled by α_j . Switching yields the inclusive value scaled by λ_j plus a switching cost c_{stj} :

$$\begin{aligned} V_{NS,1tj} &= \alpha_j \bar{y}_{1tj}, \quad V_{NS,2tj} = \alpha_j (\bar{y}_{1tj} + \bar{y}_{3tj}), \quad V_{NS,3tj} = \alpha_j (\bar{y}_{1tj} + y_{3tj}^+) \\ V_{S,ktj} &= c_{stj} + \lambda_j IV_{ktj} \end{aligned}$$

Under i.i.d. Type I extreme value shocks across the stay/switch decision, the probability of switching away from current option k is

$$\Pr(S \mid k)_{tj} = \frac{e^{c_{stj} + \lambda_j IV_{ktj}}}{e^{V_{NS,ktj}} + e^{c_{stj} + \lambda_j IV_{ktj}}} \quad (7)$$

The parameters $\alpha_j > 0$ and $\lambda_j \in (0, 1)$ allow stickiness (large α_j) and switching responsiveness (large λ_j) to vary independently, which is necessary to fit observed persistence.

Each parameter decomposes additively into site and format fixed effects. Entry and switching costs also carry a time-varying B-spline profile

$$\begin{aligned}\log \alpha_j &= \alpha_{s_j} + \alpha_{z_j} \\ \log \beta_j &= \beta_{s_j} + \beta_{z_j} \\ \text{logit}(\lambda_j) &= \lambda_{s_j} + \lambda_{z_j}\end{aligned}$$

$$\kappa_{2tj} = \kappa_{2,s_j} + \kappa_{2,z_j} + B(t)^\top \gamma_{\kappa_2}$$

with analogous decompositions for κ_{3tj} and c_{stj} , where $B(t) = (B_1(t), \dots, B_L(t))^\top$ is a vector of L piecewise-constant B-spline basis functions on equispaced knots over $[1, T]$ ($L = 30$ in our specification). The log and logit transforms ensure $\alpha_j > 0$, $\beta_j > 0$, and $\lambda_j \in (0, 1)$.

The Stage 2 DFM delivers, for each product-day (j, t)

$$\bar{y}_{3tj} = \mu_{3j} + \lambda_{3j}^\top \sum_{s=1}^p A_s f_{t-s}, \quad \sigma_{y_{3tj}}^2 = \lambda_{3j}^\top \Sigma_\eta \lambda_{3j} + \sigma_{\xi_{3,j}}^2.$$

The incremental term y_{3tj}^+ is evaluated by Gauss-Hermite quadrature over the conditional normal.

For estimation, let $d_{kl,tj} = 1$ if product j is in market channel k on day $t - 1$ and channel l on day t . The log-likelihood contribution for (j, t) is

$$\log L_{tj} = \sum_{k=1}^3 d_{kk,tj} \log \Pr(k | k)_{tj} + \sum_{k=1}^3 \sum_{l \neq k} d_{kl,tj} \left[\log \Pr(S | k)_{tj} + \log \Pr(l | k)_{tj} \right], \quad (8)$$

where the first term covers staying with option k and the second covers switching from k to l . Under the logistic structure, the log stay probability is $\log \Pr(k | k) = -\log(1 + e^{V_{S,k} - V_{NS,k}})$ and the log switch probability is $\log \Pr(S | k) = (V_{S,k} - V_{NS,k}) - \log(1 + e^{V_{S,k} - V_{NS,k}})$. Conditional destination probabilities $\Pr(l | k)$ follow a binary logit over the two available destination payoffs. The total log-likelihood sums over all product-days (j, t) . Parameters $\theta = (\alpha., \beta., \lambda., \kappa_{2.,}, \kappa_{3.,}, c_{s.,}, \gamma_{\kappa_2}, \gamma_{\kappa_3}, \gamma_{c_s})$ — comprising site and format fixed effects for each scale and cost parameter, plus B-spline coefficients — are estimated by maximum likelihood using Ipopt. Switching costs are identified from transition frequencies between market configurations, conditional on the PMP attractiveness inputs $(\bar{y}_{3tj}, y_{3tj}^+)$.

5.3 Inference and limitations

The model is supply-side and partial-equilibrium. We take demand and the auction environment as given: publisher revenue in each market is summarised by the latent price indices (y_{1tj}, y_{2tj}) , with demand-side forces absorbed into the common factors. We abstract from bid-level auction outcomes and model prices at the daily product level; the specific auction mechanism — first-price, second-price, or reserve-price rules — affects prices only through the observed price indices and is not separately identified. Counterfactual exercises hold demand-side behaviour fixed and trace publisher supply responses to changes in the price environment.

We use the estimated model to compute supply elasticities and conduct counterfactual analyses. The results are presented in Chapter 6, which documents substantial supply responsiveness (median elasticities 4–6), declining elasticity over time, and heterogeneity across formats and site categories. The model is estimated with maximum likelihood, but we do not report standard errors for brevity.

6 Results

We now report the structural results. We begin with the extensive margin, then turn to common factors, selection, and supply elasticities. We close with heterogeneity across site categories.

Three findings stand out. First, the effective cost of entering PMP swung by roughly \$0.26 per CPM across the sample, from $-\$0.12$ in 2018 Q2 to $+\$0.14$ in 2020 Q3. This shift is equivalent to about 4% of the mean PMP CPM.

Second, selection accounts for a material share of observed PMP prices in the early period. About 30% of video PMP price variance is attributable to selection, versus 14% for display. That gap closes towards zero by late 2020 as adoption broadens and marginal entrants resemble the average product more closely.

Third, selection-corrected supply elasticities are large, median 3.4 for display and 7.3 for video, and decline over the period from roughly 8 in mid-2018 to below 4 by late 2020. Ignoring selection overstates these elasticities by roughly a quarter for display and by roughly a third for video.

6.1 Model fit

Fitted-versus-observed correlations exceed 90% for both RTB and PMP across display and video formats (further diagnostics are shown in Appendix B). We select $r = 60$ latent factors using a scree plot of eigenvalues and then focus on the most important factors for economic interpretation in Section 6.3. Stage 1 fit, assessed by comparing model-implied long-run regime shares with end-of-sample observed shares, is documented in Appendix B.2.

6.2 Entry and participation

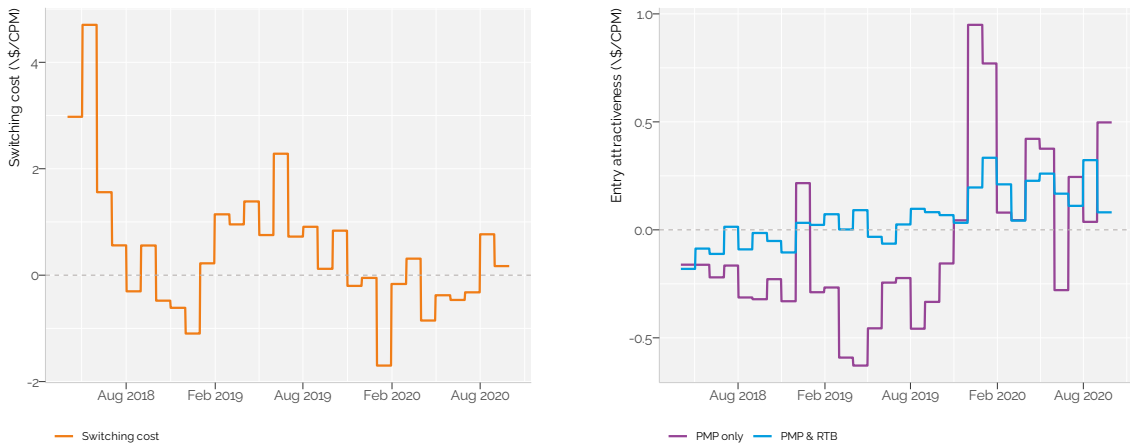
Each day, publishers in our model choose whether to operate in RTB only, PMP only, or both markets. In Stage 1, we estimate entry barriers and switching costs that govern movement across these configurations. The payoffs use Stage 2 expectations about PMP attractiveness, where $\bar{y}_{3,jt}$ is the expected PMP uplift for a product that cannot condition on impression-level realisations and $y_{3,jt}^+$ is the selection-corrected uplift when both markets are active and the publisher routes each impression to the better option. We begin by discussing the extensive margin and the role that unobserved costs and frictions play in PMP participation. Stage 1 model fit is documented in Appendix B.2.

6.2.1 Declining entry barriers

Figure 8 plots the time-varying components of entry attractiveness and switching cost, both converted to dollar-CPM equivalents.⁸

Effective entry costs fell over the sample. The combined entry-attractiveness measure,⁹ moved from a median of $-\$0.12$ per CPM in 2018 Q2 to $+\$0.14$ per CPM in 2020 Q3, a decline in effective barriers of roughly $\$0.26$, equivalent to about 4% of the mean PMP CPM. Because the entry threshold depends on whether PMP revenue can cover these effective costs, a rising attractiveness series means publishers with progressively lower CPMs find entry worthwhile, consistent with lower deal-setup costs, improved SSP tooling, and growing buy-side familiarity with private marketplace formats.¹⁰

Switching inertia has fallen towards zero. Adtech has substantially reduced the friction of switching between different marketplaces. We return to this in our counterfactual analysis in Section 7.1.



(a) Switching cost

(b) Entry attractiveness

Figure 8: Time-varying switching cost (left) and entry attractiveness (right), expressed as \$/CPM equivalents. Positive switching cost indicates friction. Rising entry attractiveness indicates declining effective barriers to PMP participation.

Entry and switching costs vary substantially across site categories. We document this in

⁸Entry costs are converted from utility units to \$/CPM by dividing by the median revenue scale β_j , which is denominated in utility units. The switching cost uses $\lambda_j \beta_j$ as the denominator, since the switching payoff is scaled by the inclusive-value weight λ_j before entering the stage-1 choice probability.

⁹Computed as a probability-weighted average $\bar{c}_t = \bar{w}_t \kappa_{2t} + (1 - \bar{w}_t) \kappa_{3t}$, where \bar{w}_t is the panel-averaged probability that a switching publisher chooses PMP-only rather than both-markets operation, and both cost parameters are divided by the median revenue-scale $\tilde{\beta}$ to convert from utility units to \$/CPM. This is a derived summary of two structural parameters, not a primitive of the model. It closely tracks the cost contribution to the Stage 1 inclusive value IV_1 (the log-sum-exp of destination payoffs that enters switching probabilities), confirming that the probability-weighted average is a faithful approximation of the structural object that governs entry decisions.

¹⁰For context, observable data and technology fees in our sample are of similar magnitude — roughly 3–5% of CPM — so the structural entry cost decline is comparable in scale to measured adtech fees.

Section 6.6 alongside the elasticity heterogeneity results. The entry decline documented here is the structural mechanism behind the extensive-margin contribution to PMP share growth shown in Figure 6 (see Section 4.2.5).

6.3 Common factors

The model estimates 60 latent factors that capture aggregate demand conditions affecting all products simultaneously. These factors absorb the correlated unobservables that would otherwise confound the selection and elasticity estimates.

We label the 10 leading factors by deseasonalising their time series, regressing them on known market events, and checking whether loadings are asymmetric across channels or formats. Table 2 summarises the result for the top ten factors. The leading factors load differently on RTB and PMP, supporting our interpretation of the two channels as distinct markets responding to the same external shocks. The long tail is unsurprising given the complexity of the market, particularly during the COVID pandemic, and likely captures category-specific advertiser demand and short-term audience cycles. Appendix C presents factor loadings, variance decomposition, and the event-based naming exercise in detail.

Table 2: Factor labels, top ten factors by variance explained.

Factor	Var%	Event Response	Deal Type	Format	Name
f01	0.9%	Trend	Balanced	Video-heavy	Trend-Video
f02	0.6%	ITP	PMP-dominant	Balanced	ITP-PMP
f03	0.6%	GAM Rollout	RTB-dominant	Display-heavy	GAM-RTB-Display
f04	0.5%	COVID	Balanced	Balanced	COVID
f05	0.5%	Trend	Balanced	Balanced	Trend
f06	0.4%	ITP	RTB-dominant	Video-heavy	ITP-RTB-Video
f07	0.4%	Seasonal	PMP-dominant	Video-heavy	Seasonal-PMP-Video
f08	0.4%	GDPR	RTB-dominant	Balanced	GDPR-RTB
f09	0.4%	Mixed	PMP-dominant	Display-heavy	Mixed-PMP-Display
f10	0.3%	CCPA	Balanced	Video-heavy	CCPA-Video

Note:

Event response from de-seasonalised regressions; channel and format bias from loading asymmetry. Major marketwide trends are themed - ITP (Apple’s Intelligent Tracking Prevention in Safari — updates roughly every 6 months from 2017); GDPR (EU’s General Data Protection Regulation - April 2018); CCPA (California’s Consumer Privacy Act - Jan 2020); GAM (Google switch to first price auction - Sep 2019); COVID (COVID-19 lockdowns - March 2020).

As an illustration of how we use the event-labelled factors, Figure 9 compares mean signed exposure (loading \times event coefficient) for factors that respond significantly to ITP version count, by deal type.¹¹ For each of these ITP-responsive factors, we multiply the product-level loadings by the estimated ITP effect and average across products. Positive values indicate that prices rise

¹¹As a reminder, Apple’s Intelligent Tracking Prevention (ITP) restricted cross-site tracking in Safari browsers. It was introduced in 2017 and progressively tightened over seven updates, finally eliminating cross-site tracking in 2020.

with ITP; negative values indicate that prices fall. The figure shows that ITP-responsive factors affect RTB and PMP differently. Some factors push RTB prices up while PMP prices fall, or vice versa. The analysis supports the hypothesis in Duckworth et al. (2023) that RTB is more sensitive to ITP than PMP because it relies more heavily on third-party tracking.

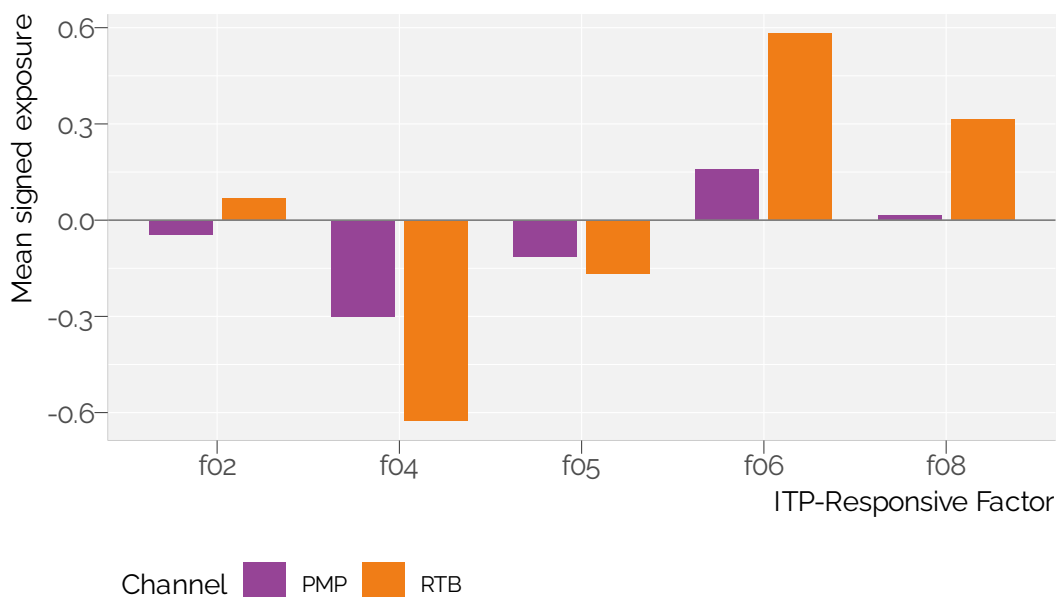


Figure 9: ITP factor exposure, mean signed exposure (loading $\times \beta$) by channel.

6.4 Selection and sorting

Having established that the factor structure captures aggregate dynamics, we turn to selection. Our model shows that selection effects are material for PMP in particular, but have diminished over time. Controlling for selection materially affects our price and elasticity estimates.

Selection operates through the allocation index y_{3jt} , which summarises non-price reasons to route impressions to PMP rather than RTB. Because y_{3jt} maps directly into PMP share, higher values imply a higher probability that an impression is allocated to PMP even when prices are held fixed.

We decompose this index as $y_{3jt} = \mu_{3j} + \lambda'_{3j} f_t + \xi_{3jt}$, where λ_{3j} is the $r \times 1$ vector of factor loadings for product j 's selection equation (the stacked matrix Λ is defined in Section 5.1.1.2). The product-specific mean μ_{3j} captures persistent differences in PMP attractiveness such as audience quality, contractual commitments, and publisher-specific sales strategy. The inverse Mills ratio, scaled by the selection-variance parameter σ_{23j} , then corrects the PMP price equation for the fact that the set of impressions routed to PMP is endogenously selected.

6.4.1 Determinants of the selection index

We use the same observable characteristics introduced in Chapter 4 — site category, format, scale, viewability, and data fees — to interpret all structural parameters in this section. We first ask which products are systematically more likely to use PMP, holding prices fixed. The time-invariant component of the selection index, μ_{3j} , captures these non-price determinants of PMP participation on the intensive margin. Figure 10 regresses μ_{3j} on observable product characteristics. The selection index varies with video versus display and with site category: video products and more aspirational content (e.g. Beauty & Fitness, Home & Garden, Food & Drink) are more likely to use PMP conditional on price.¹² Conversely, more ‘informational’ content (e.g. Real Estate, Business, News, Finance) are less likely to use PMP.

The negative correlation with impression-level information might appear surprising at first. However, data fees are paid by the advertiser and viewability data is provided to assist comparisons across sites in automated auctions. PMP deals are typically bundled with more information from the publisher (e.g. from logged in subscribers) and with quality dimensions (e.g. premium pages), so they are substitutes for this measure.

6.4.2 Evolution of selection: Mills ratios over time

Figure 11 plots the median inverse Mills ratio for PMP over time, faceted by display and video. The figure shows a clear decline in both formats, with the Mills ratio approaching zero by late 2020. This pattern indicates that selection effects diminish as PMP scales. Early adopters possess unobserved quality advantages that justify PMP participation. As adoption broadens, marginal entrants resemble the average product more closely. As we see below, elasticity estimates that ignore selection are biased upward in the early period but approximately unbiased by 2020.

Selection effects are more substantial for PMP and for video. Table 3 summarises the variance share of the selection correction — $\sigma_{m3j}^2 \times \text{Var}(M_{mtj})$ — by market, format, and year. Selection accounts for a larger fraction of PMP price variance than RTB, and video products show consistently higher shares than display products. Impression-weighted means (which up-weight the largest publishers) generally exceed the unweighted medians, indicating that high-volume products tend to exhibit stronger selection pressure.

¹²Advertising is often closely matched to the content on these sites (sometimes referred to as ‘native’). PMP deals can be used by publishers to sell more advertising around these content types.

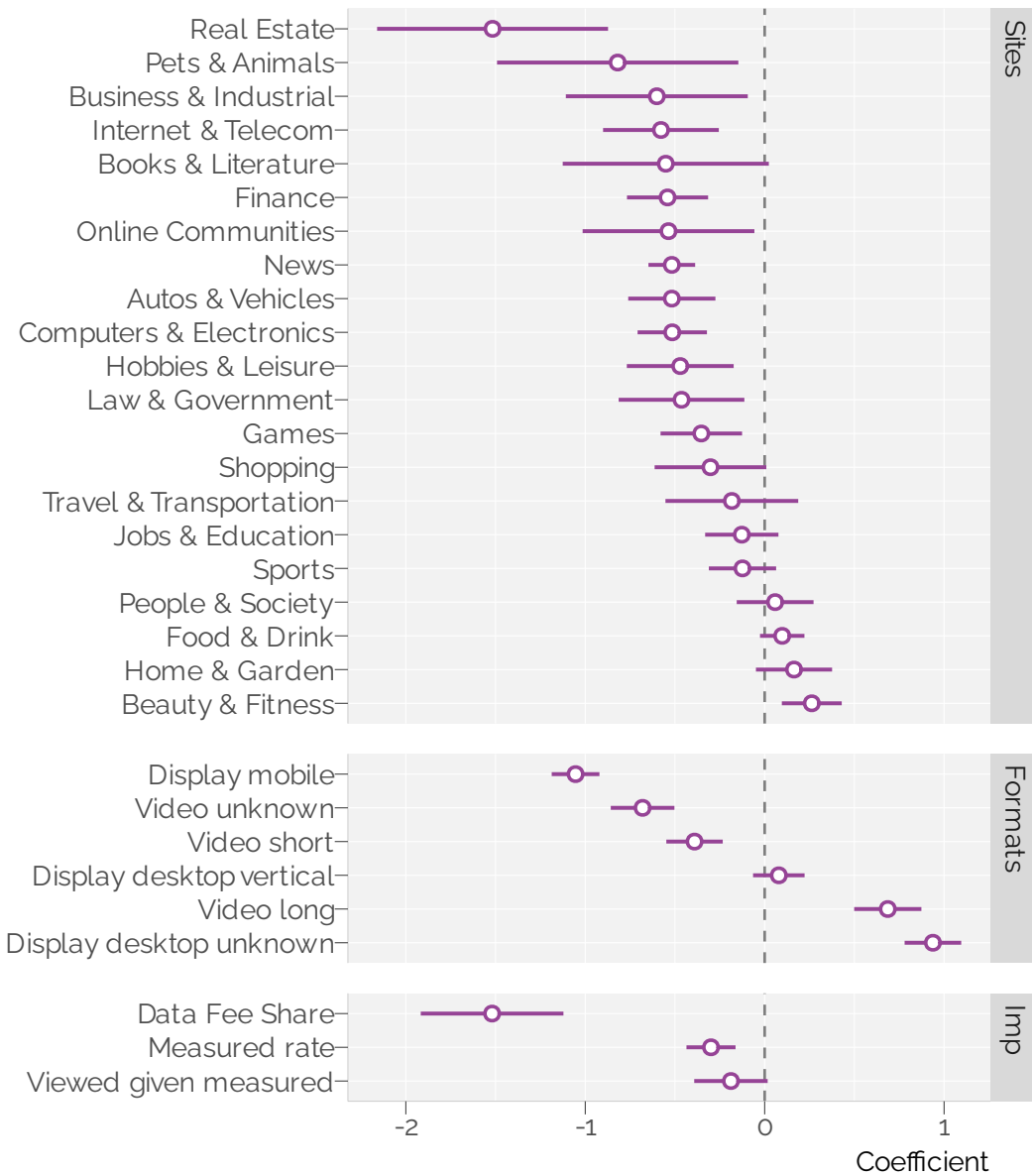


Figure 10: Variates associated with selection index μ_3 , coefficient estimates from regression on observable product characteristics.

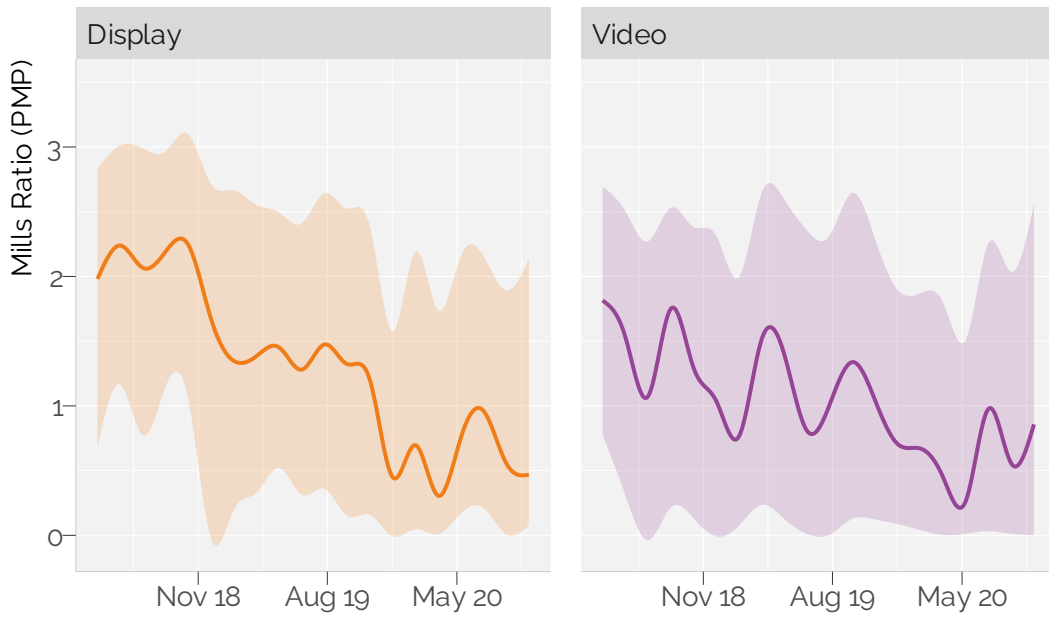


Figure 11: Mills ratio (PMP selection) over time, by creative format, median and 20th–80th percentile range over product space. Smoothed with cubic spline with $k=20$.

Table 3: Variance share of the selection correction ($\sigma_{m3j}^2 \times \text{Var}(M_{mtj})$), by market, format, and year.

Format	2018*		2019		2020*	
	Median	Mean	Median	Mean	Median	Mean
PMP						
Display	1.1%	12.2%	1.9%	9.8%	1.5%	8.5%
Video	3.4%	29.4%	6.7%	30.7%	6.6%	28.8%
RTB						
Display	0.0%	0.9%	0.3%	1.8%	0.4%	3.3%
Video	0.0%	12.1%	0.4%	14.5%	1.7%	14.8%

Note:

Variance share = $\sigma^2 \times \text{Var}(\lambda)$ within calendar year, capped at 1. Mean is impression-weighted. * Partial year.

Viewable impressions are more likely to be routed to PMP (see Appendix C.1.3), suggesting that publishers direct higher-quality supply towards private deals on the intensive margin.

6.4.3 Information asymmetry

The parameter σ_{23j} captures information asymmetry in impression allocation. It is the covariance between the unobserved shock that shifts an impression towards PMP and the unobserved shock to the PMP price. When $\sigma_{23j} > 0$, higher-value impressions are disproportionately routed through PMP, inflating observed PMP prices relative to the latent index. When $\sigma_{23j} < 0$, the pattern runs in the opposite direction, consistent with PMP being used as a residual channel in weaker demand states. For example, publishers might route cookie-less impressions to PMP. We estimate both signs across products, with substantial cross-sectional dispersion.

The covariance σ_{23j} conflates two conceptually distinct objects. Since $\text{Var}(\varepsilon_3) = 1$, we have $\sigma_{23j} = \rho_{23j} \sqrt{\sigma_{22j}}$, where ρ_{23j} is the correlation between the allocation and PMP price shocks — the pure information-advantage parameter — and $\sqrt{\sigma_{22j}}$ captures within-day PMP price dispersion, which amplifies the scope for quality-based routing regardless of whether the publisher possesses any informational advantage. To separate these two objects we proxy σ_{23j} using impression-level CPM variance from an independent data source.¹³ The construction, match rates, and diagnostics are in Appendix C.2. Regressions of both σ_{23j} and ρ_{23j} on observable product characteristics — site category, format, scale, viewability, and data fees — show no systematic patterns. If adverse selection is present, it does not operate along observable dimensions: publishers in the same content category or format do not systematically route better impressions to PMP. Asymmetries are idiosyncratic — present in individual products but not predictable from the characteristics available to advertisers.

6.4.4 Bounds on σ_{12}

Our focus is supply elasticity, but potential adverse selection between PMP and RTB is a theme in the literature. Balocco et al. (2025) find that PMP presence depresses RTB quality and prices. In principle, our model can bound σ_{12j} , the covariance between unobserved shocks to RTB and PMP prices. Appendix C.3 shows the bounds are wide and uninformative. This may partly reflect our data being more aggregated than the impression-level data in Balocco et al. (2025). It is also consistent with the covariance being genuinely heterogeneous — some publishers route higher-value impressions to PMP (positive σ_{12j}), others use PMP as a residual channel for lower-value inventory (negative σ_{12j}) — and the aggregate signal averages near zero.

This lack of identification does not affect our elasticity estimates because the formulae depend on σ_{23j} (identified via the selection correction) and on the marginal variances, but not on σ_{12j} directly.

6.5 Supply elasticities

With selection and factors in hand, we quantify how publishers adjust PMP supply in response to price changes. The supply elasticity $\eta_{s_2, PMP}$ quantifies the percentage change in PMP impressions supplied in response to a 1% increase in the PMP price, holding the RTB price fixed. Cross-price elasticities $\eta_{s_2, PRTB}$ capture substitution between channels. To our knowledge, no prior study has estimated supply-side elasticities in programmatic advertising over time while accounting for endogenous selection.

¹³We use intraday impression-level data from a Demand-Side Platform for the same US advertisers in 2019 Q2 over a 90-day window.

6.5.1 Elasticity definition and computation

Under the structural model, PMP allocation share follows $s_{2tj} = \Phi(y_{3jt})$, where Φ is the standard normal CDF and y_{3jt} is the selection index. The PMP own-price elasticity is

$$\eta_{s_2,p_2} = \frac{\partial s_{2tj}}{\partial \bar{p}_{2tj}} \cdot \frac{\bar{p}_{2tj}}{s_{2tj}}$$

Computing the elasticity requires a selection-corrected derivative because a price change shifts the allocation index, which alters the Mills ratio, which feeds back into the observed price.¹⁴ Omitting this correction biases the elasticity upward when $\sigma_{23j} < 0$ and downward when $\sigma_{23j} > 0$. In the early period, where Mills ratios are large, $\sigma_{23j} < 0$ products dominate among high-Mills observations, producing a net 30-40% overstatement in uncorrected estimates. The gap closes by 2020 as Mills ratios approach zero (see Appendix D.3).

6.5.2 Main elasticity estimates

Table 4 presents selection-corrected elasticity estimates.¹⁵ The median own-price elasticity for display inventory is 3.4, indicating that a 10% increase in PMP prices elicits a 34% increase in PMP impressions supplied. Video elasticities are higher (median 7.3), consistent with greater supply flexibility in video inventory. Cross-price elasticities are positive and substantial, confirming that PMP and RTB are substitutes from the publisher’s perspective. When RTB prices rise, publishers shift impressions towards RTB, reducing PMP supply. Dispersion is considerable within both formats. The interquartile range for display spans [1.3, 7.1] and for video spans [2.4, 17.1], with some publishers adjusting supply aggressively and others responding more sluggishly.

Table 4: Selection-corrected supply elasticities.

Format	N	Own-Price			Cross-Price	
		Mean	Median	SD	Mean	Median
Display	170730	6.57	3.43	9.17	-3.68	-2.52
Video	47849	11.32	7.27	15.99	-17.88	-7.04

Note:

Mean is impression-weighted. Sample restricted to days when the product was active in both the PMP and RTB markets. A small number of extreme observations ($|\eta| > 100$) are excluded.

¹⁴Formally, price changes induce selection changes via $\bar{p}_{2tj} \rightarrow y_{2tj} \rightarrow s_{2tj} \rightarrow M_{2tj} \rightarrow y_{2tj}$. The selection-corrected derivative is $\partial s_{2tj} / \partial \bar{p}_{2tj} = \phi(y_{3tj}) / [1 - \sigma_{23j} \cdot (y_{3tj} + M_{2tj}) / s_{2tj} \cdot \phi(y_{3tj})]$, where ϕ is the standard normal PDF. See Appendix D for derivation.

¹⁵Elasticity estimates focus on product-days where both PMP and RTB are observed on the same day, i.e. where the publisher actively allocates between the two markets. We exclude days with missing prices (single-market participation) or corner shares (s_{2tj} near 0 or 1). As a robustness check, we impute some excluded observations using the factor model; see Appendix D. Imputation typically flattens estimated elasticity because it adds days when publishers were committed to one market and less responsive to price.

Elasticities of this magnitude indicate that publishers possess substantial discretion in allocating inventory across channels. This discretion likely stems from the ability to adjust PMP floor prices, change participation rates, or shift impressions to RTB when PMP demand weakens. The high elasticities also imply that small changes in relative prices can induce large shifts in supply, which is consistent with the volatile PMP share dynamics documented in Section 4.2. For comparison, baseline (uncorrected) elasticities are higher: the median display baseline is 4.3 versus the corrected 3.4 (a 26% reduction), and for video 11.7 versus 7.3 (a 37% reduction). The selection correction is therefore economically significant, particularly for video.

These elasticity estimates are identified within the structural model rather than by an external price shifter. Intuitively, we use within-product price variation after absorbing common demand shocks with the factor structure and correcting for selection into PMP participation. The identifying assumption is therefore that the remaining price variation is informative about supply rather than driven by omitted product-level demand shocks.

6.5.3 Declining elasticity over time

Figure 12 plots the median own-price elasticity over time. Elasticity declined from approximately 8 in mid-2018 to below 4 by late 2020 for display inventory. Video elasticities follow a similar downward trajectory, though with greater period-to-period variation.

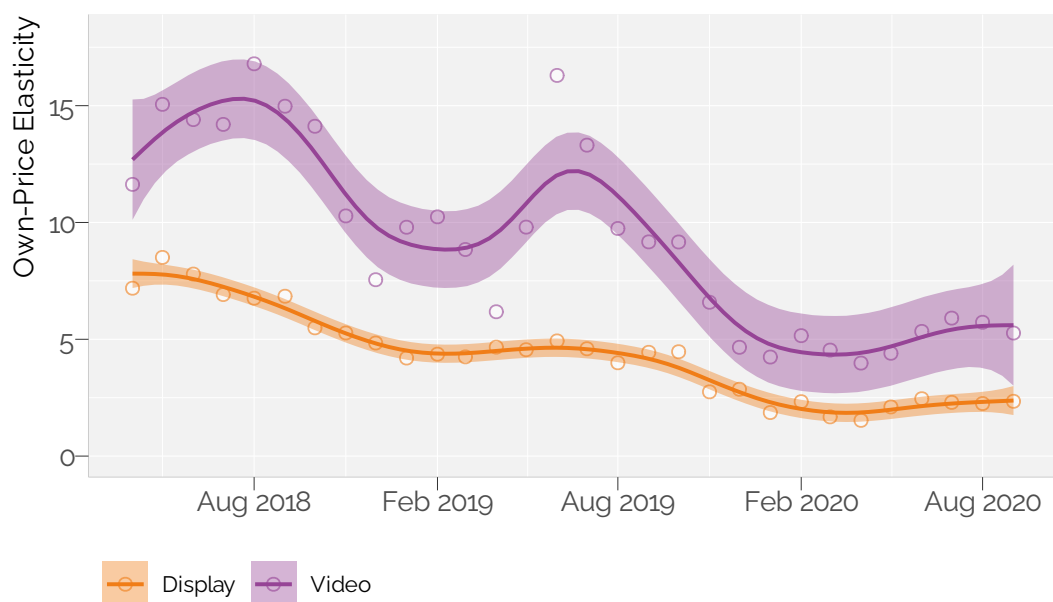


Figure 12: Evolution of median PMP own-price elasticity over time, smoothed with cubic spline with $k = 30$.

Supply elasticities are not constant. They decline as markets mature and as the compositional mix of participants shifts. We now examine where that variation sits across publisher types before turning to counterfactual scenarios.

6.6 Heterogeneity across site categories

The aggregate results mask systematic variation across the publisher population. The selection index μ_{3j} , supply elasticity, and entry barriers co-vary across site categories. This co-movement is central to interpretation because it identifies which categories are inframarginal and which are close to participation thresholds.

6.6.1 Selection, elasticity, and publisher type

Figure 13 plots category-level median μ_{3j} against impression-weighted mean supply elasticity. The negative relationship is partly mechanical. When μ_{3j} is high, a product sits near the top of the probit CDF, where the density $\phi(y_{3j})$ is small, so marginal price changes move PMP share less. But the pattern also has clear economic content. Categories with high unconditional PMP affinity, such as Beauty & Fitness, Home & Garden, and Food & Drink, are largely inframarginal. Categories with low or negative μ_{3j} , such as News, Finance, and Technology, sit closer to the participation margin and therefore reallocate more when relative prices move.

The value of this relationship is not just the negative slope. It shows which categories sit near the participation margin and which are structurally attached to PMP. That distinction matters more for policy than the average elasticity because the marginal categories are the ones that reallocate when relative conditions change.¹⁶ The selection correction is also larger for high- μ_{3j} categories. Their early Mills ratios are higher, so ignoring selection overstates their elasticities by more than it does for categories that are already closer to the margin.

6.6.2 Entry and switching costs by site category

Entry barriers also vary systematically across site categories. Appendix Section C.4 shows that News, People & Society, and Home & Garden sit on the lower-friction side of the Stage 1 distribution, whilst Travel & Transportation, Law & Government, and Computers & Electronics sit on the high-friction side. These differences identify which publishers are most exposed to falling adtech frictions and therefore matter most for the CF1 counterfactual in Section 7.1.

6.6.3 Implications for news publishers

News sites combine low μ_{3j} and high elasticity, placing them closer to the participation margin than categories that are structurally attached to PMP. The Stage 1 estimates point in the same direction: news sits on the lower-friction side of the distribution. That positioning means news publishers are more likely to respond when relative conditions move, which is consistent with selective PMP use for premium segments or high-demand periods rather than default allocation. It also foreshadows why category position in the model matters for the counterfactual exercises.

6.7 Results summary

Taken together, the structural results show that PMP growth reflects changing participation frictions, changing selection, and substantial but declining supply responsiveness.

PMP share growth reflects both margins. In display, the extensive margin contributes more because falling participation frictions bring new publishers into PMP. In video, where PMP is

¹⁶Category-level elasticities by format are reported in Appendix Appendix D.

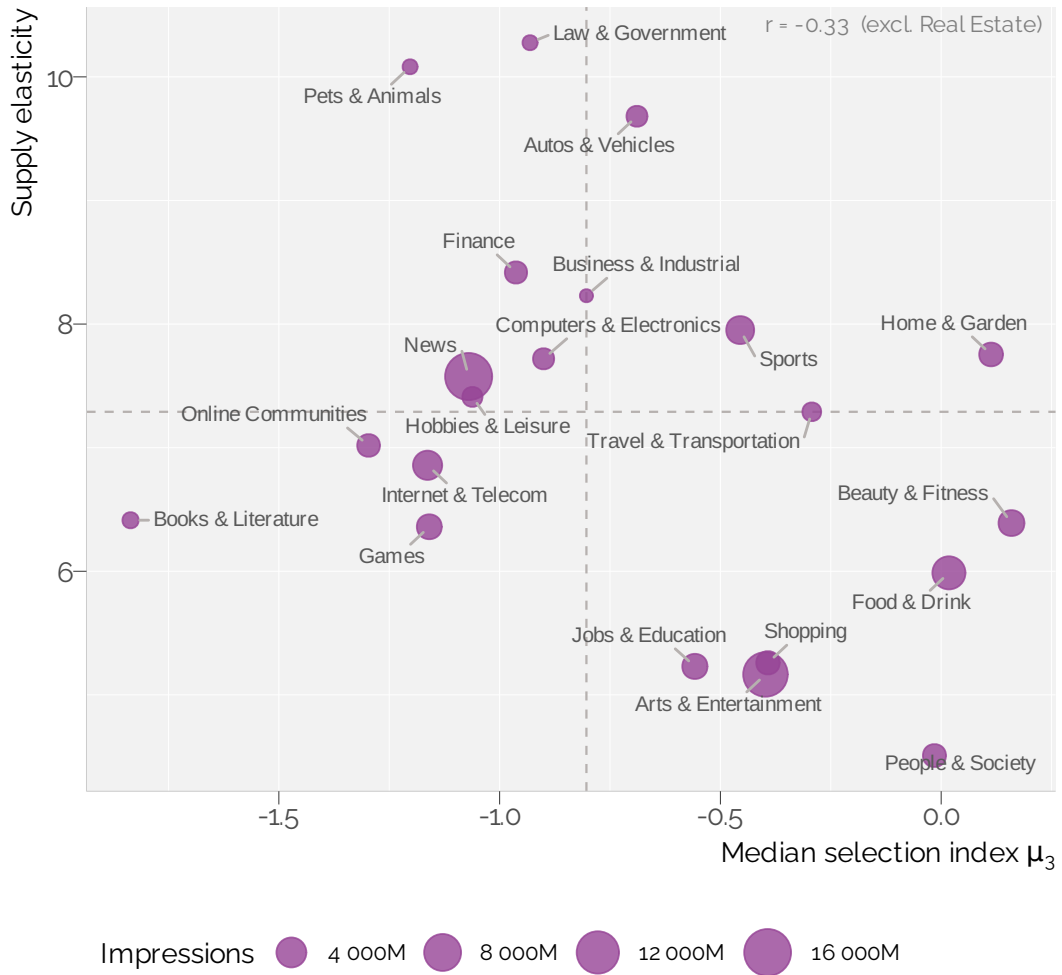


Figure 13: Selection index μ_3 vs supply elasticity by site category, higher μ_3 indicates stronger unconditional PMP affinity. Point size proportional to total impressions, dashed lines at category medians, Real Estate excluded (small sample).

already more mature, reallocation within the set of participants plays a relatively larger role. Selection matters most early in the sample and becomes less important as PMP matures. Early PMP adopters have stronger unobserved quality advantages, so ignoring selection would overstate elasticity. As PMP broadens, those advantages become less distinctive and the correction shrinks.

Supply remains elastic throughout the sample, but less so as the market matures. Publishers retain substantial scope to reallocate impressions across channels, yet that responsiveness declines as PMP broadens and fewer publishers sit at the most price-sensitive margin.

The factor structure identifies privacy regulation as an asymmetric demand shock. ITP-responsive factors load differently on RTB and PMP products. This pattern is consistent with Duckworth et al. (2023) and supports treating privacy rollout as a structural shift in the relative value of the two channels rather than a uniform demand shock.

The counterfactuals that follow use these structural elasticity and entry estimates directly. Appendix Appendix D reports robustness to excluded-observation imputation, and Appendix Section D.4 shows how the elasticity magnitudes move when we vary the selection correction. We now use these estimates in partial-equilibrium counterfactuals that vary the supply-side primitives governing participation and allocation.

7 Counterfactuals

The previous section shows that PMP growth reflects several distinct supply-side forces. Entry barriers fall over time, selection on unobserved quality attenuates, and supply elasticities remain economically large even as they decline with market maturity. We now ask what these estimates imply under alternative supply-side conditions.

We treat the counterfactuals as partial-equilibrium scenario exercises. We hold the common factors, factor loadings, and product set fixed throughout, and vary only a small set of supply-side primitives. The results therefore describe the mechanical implications of alternative parameter values for participation, allocation, and revenue. They are not full-equilibrium exercises.

We study three counterfactuals. First, we ask how much of observed PMP growth is explained by declining adtech setup frictions on the extensive margin (CF1). Second, we consider an as-if deterioration in the non-price advantages that support PMP allocation, motivated by the possibility that AI makes high-quality content or quality signals easier to produce (CF2). Third, we consider the mirror as-if case in which AI makes commoditised RTB supply easier to produce and so weakens the RTB price block (CF3). These are motivating interpretations of parameter shifts, not literal forecasts.

Table 5: Headline counterfactual findings, PMP share, long-run participation, and net revenue under benchmark shocks.

Counterfactual	Primitive shift	PMP share and participation			Net revenue (%)	
		Mean PMP share	Steady-state PMP active		Display	Video
		Δ mean PMP share (pp)	Display (pp)	Video (pp)		
CF1-entry	κ_2, κ_3 frozen post- t_0	—	-17.7	-25.1	—	—
CF1-full	κ_2, κ_3, c_s frozen post- t_0	—	-17.9	-25.2	—	—
CF2: AI lowers PMP non-price advantage	$\Delta\mu_3 < 0$	-7.4	-17.3	-15.8	-34.3	-10.4
CF3: AI weakens RTB quality	$\Delta\mu_1 < 0,$ $\Delta\mu_3 = -\Delta\mu_1$	7.6	28.7	29.2	15.0	-4.3

CF2 and CF3 use the benchmark $0.5 \times \text{sd}(\mu_{3,j})$ shock.

Participation columns report differences against baseline P(PMP active).

CF1-entry and CF1-full affect Stage 1 participation frictions only, so revenue effects are not reported in this headline table.

Under CF2, prices are fixed and revenue changes come mainly from reallocation.

Under CF3, revenue changes reflect both reallocation and RTB price-block movement.

7.1 CF1, adtech cost decline

PMP adoption requires publishers to incur real setup costs. They must establish trading relationships with intermediaries, configure ad-server integrations, negotiate deal structures, and maintain the data and measurement needed to monitor contracts. Our Stage 1 model captures these forces through the time-varying entry-cost splines κ_{2t} and κ_{3t} , and through the switching cost c_{st} . We estimate that these entry frictions fall materially over 2018–2020, which is consistent with the steady standardisation of ad-tech tooling over the period. We therefore ask what steady-state publisher participation would have looked like had this decline not occurred.

CF1 isolates this extensive-margin channel in two versions. We freeze κ_{2t} and κ_{3t} at their January 2019 values after t_0 , and then evaluate steady-state PMP activity under the estimated transition system. In CF1-entry, only the entry-cost splines are frozen, whilst c_{st} follows its estimated path. In CF1-full, we also freeze c_{st} , which gives a broader barrier-persistence scenario. The benchmark effect is a 2.6 to 3.1 percentage point reduction in long-run PMP participation. This implies that declining ad-tech setup frictions explain a meaningful, but not dominant, share of observed PMP growth.

CF1-entry and CF1-full are close. This suggests that entry barriers account for most of the long-run level effect, whilst switching frictions matter more for transition speed than for the steady-state outcome.

7.2 CF2 and CF3, opposing intensive-margin shocks

CF2 and CF3 are paired intensive-margin experiments. In economic terms, they represent two different ways in which AI could shift relative PMP and RTB attractiveness. In CF2, we consider an as-if deterioration in the non-price advantages that support PMP allocation, motivated by the possibility that AI makes high-quality content or quality signals easier to produce. In CF3, we

consider the mirror as-if case in which AI makes commoditised RTB supply easier to produce and so weakens the RTB price block.

The two experiments differ in the structural mapping. CF2 applies a negative shift to μ_3 only and keeps prices fixed in the main mapping. CF3 applies a negative shift to μ_1 with compensating movement in μ_3 , so that the implied switching-cost term is held fixed.¹⁷ We calibrate both experiments on the cross-product dispersion in $\mu_{3,j}$, using $0.5 \times sd(\mu_{3,j})$ as the benchmark and $0.25 \times sd(\mu_{3,j})$ as sensitivity.

At the benchmark magnitude, the two intensive-margin experiments move PMP share in opposite directions. CF2 lowers mean PMP share by 7.4 percentage points, whilst CF3 raises it by 7.6 percentage points. The same sign pattern carries through to long-run participation. Under CF2, steady-state PMP participation falls by 17.3 percentage points in display and 15.8 percentage points in video. Under CF3, it rises by 28.7 percentage points in display and 29.2 percentage points in video.

For revenue, the comparison differs in an economically important way. Under CF2, prices are fixed by construction, whilst c_s is still allowed to evolve along its estimated path. In practice that extra channel does little, so the revenue effect is best read as reallocation between PMP and RTB at the existing price schedule. Under the structural main mapping, net revenue falls by 34.3% in display and 10.4% in video. Under CF3, the RTB price block moves with the μ_1 shock, so revenue changes reflect both reallocation and a change in relative prices. Net revenue then rises by 15.0% in display and falls by 4.3% in video. We treat these as the headline structural mappings, and view alternative mappings only as sensitivity.

Category heterogeneity helps interpret the aggregate results. Under CF2, categories whose PMP advantage depends more on scarce quality signals lose more PMP share when those signals become easier to reproduce. Under CF3, the broad ranking shifts, but not in a mechanical one-for-one way. News is exposed in CF2 and suffers when content or quality signals can be copied more cheaply, but it also performs strongly in CF3, when weaker RTB quality puts more weight on its relative PMP advantage. Categories that are less exposed to commoditised RTB supply pressure gain more when the RTB price block weakens. This is consistent with the earlier evidence that news publishers are relatively elastic and rely less on persistent PMP-specific non-price advantages than more inframarginal categories.

8 Conclusions

We study why PMPs expand and how the PMP premium should be interpreted when publishers allocate unobserved impression quality across channels. We estimate a supply-side model in which publishers choose market participation and, conditional on operating in both channels, route impressions between RTB and PMP. Our central conclusion is that PMP growth is driven less by a persistent intrinsic premium than by falling entry frictions and changing selection on unobserved quality. Correcting prices for that selection is therefore first order. Once we do

¹⁷The compensation matters because, in the model, a change in μ_1 alone does not correspond to a pure RTB price-block shock. It also changes the implied non-price term that governs relative PMP attractiveness. We therefore offset the μ_1 shift with a compensating change in μ_3 so that CF3 isolates RTB price pressure rather than mechanically changing switching frictions at the same time.

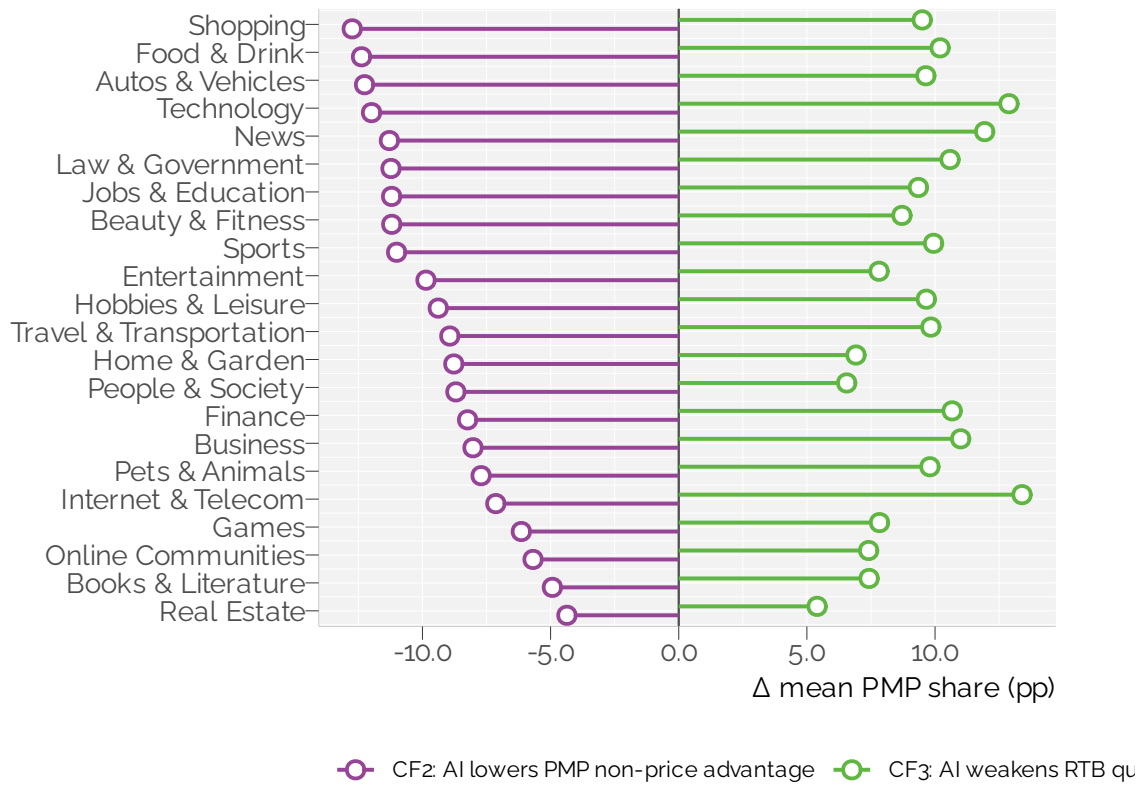


Figure 14: CF2 and CF3, change in mean PMP share by site category under the two AI-motivated scenarios. Lollipops report impression-weighted category-level changes, averaged over post- t_0 days and pooled across display and video.

so, the narrowing PMP premium no longer looks like simple erosion of marketplace value. It reflects changes in who participates and what inventory is allocated.

The estimates point to three mechanisms. First, effective entry barriers into PMP fall materially over time. Entry costs decline by about \$0.26 per CPM between 2018 Q2 and 2020 Q3, around 4% of the mean PMP price, and counterfactual steady-state participation is meaningfully lower when those costs are held fixed. Second, selection on unobserved quality is economically large early in the sample but attenuates as PMP participation expands. It accounts for roughly 30% of video PMP price variance early on, but approaches zero by late 2020. Third, selection-corrected supply elasticities are large, with medians of 3.4 for display and 7.3 for video, but decline over time. Taken together, these estimates show that falling adtech costs are a first-order driver of PMP growth.

These results speak to a broader cycle in open web markets. Open auctions struggle when quality is hard to verify and information is asymmetric. RTB lowered access barriers, but adverse selection followed. PMPs arose as an institutional response, bundling access and disclosure into curated trading relationships and temporarily restoring informational rents to publishers with superior inventory. We find that this response is effective but not permanent. As adoption widens, entry barriers fall, the selection advantage of early participants dissipates, and the premium compresses. By the end of our sample, PMP no longer commands a systematic price premium once we correct for selection. The open web is approximately back where it started. It is broader and cheaper to access, but its quality guarantees are weaker.

One implication is that market design and regulation can shift market structure when they change verifiability. Privacy and platform shocks do not operate as uniform demand shifts because they affect RTB and PMP differently, and publishers respond on both the extensive and intensive margins. More generally, interventions that improve measurement and monitoring in open auctions can substitute for curation by making quality easier to contract on, whilst changes that worsen verifiability can tilt trade towards private mechanisms. These are not welfare conclusions. They are implications of a supply-side model estimated in partial equilibrium.

The analysis has four main limitations. We estimate the supply side in partial equilibrium, so we absorb advertiser demand into common factors and interpret the elasticities accordingly. We also treat first-party data, viewability measurement, and monitoring capacity as fixed product characteristics rather than choice variables. Stages 1 and 2 are estimated sequentially rather than jointly, and we do not report formal standard errors for the structural parameters. The probit allocation rule and linear factor structure are maintained assumptions. These choices are standard given the scale of the problem, but they bound what we can learn from the counterfactual exercises. Within those limits, we show that PMP growth is best understood as a supply-side response to changing frictions and changing selection.

Appendix A — Diode Dataset

This appendix documents the construction of our estimation sample from the Diode programmatic advertising panel. It covers the raw data extract, the modelling subsample, and external validation against comScore and Association of National Advertisers (ANA) benchmarks. The main text (Chapter 4) summarises the sample dimensions and stylised facts; readers interested in data construction and representativeness will find the details here.

A.1 Diode programmatic advertising dataset

Our programmatic advertising data is drawn from Diode, a commercial database owned and operated by a large advertising agency. Diode has been collecting data from advertiser campaigns since 2016. It has data on nearly 2bn auctions per day, measuring impressions, prices, clicks, and conversions by advertiser, campaign, site, creative format, and user device, browser and operating system.

The underlying data is advertiser focused. For each advertiser campaign, Diode collects impression-level data on ads purchased and sold by advertising intermediaries on behalf of advertisers and sites. The data is then cleaned, aggregated to daily level, and stored in tables for easy retrieval and reporting. To manage data storage costs and complexity, these aggregated tables focus on specific margins, such as “users and devices” and “sites”. Consequently, aggregation loses part of the joint distribution between sites, advertisers, and users. In this paper, we focus on the “Site Table”.

The “Site Table” contains data on daily costs, impressions, clicks and impression viewability broken down by country, site, deal type, ad format as well as advertiser campaign, Demand Side Platform (DSP) and Supply Side Platform (SSP). Data is available for over 45 countries since mid 2017.

Our longer-term goal is to model both demand and supply, so we preserve these dimensions within practical constraints. Our model sample is constructed in two steps: a broad “raw” data extract from which we create an aggregated subset for modelling.

A.1.1 Raw data extract

The raw data extract is a random sample of 1,131 advertisers running 14,390 campaigns in 42 markets. It captures 164bn advertising impressions purchased on more than 40 thousand sites.

To make the extract more manageable (i.e. fewer than 1bn rows), we aggregate the tail of sites into three groups.

- **Head:** Individual named sites that account for around 70% of impressions in each market. In the US, our focal market, 639 sites account for 67% of impressions.
- **Categorised:** Sites that could be aggregated into one of 24 site categories. These are sites where it was possible to identify a site category using the HTTP Archive categorisation.
- **Uncategorised tail:** Some sites could not be categorised easily because, for example, URLs were disguised or anonymised. In the US, about 18% of impressions ran on sites that we categorise as “unknown”.

The resulting dataset has about 720M rows.

A.1.2 Site model

The focus of this paper is to model the adoption of Private Marketplaces (PMP) by sites. We have focused on the US, the largest market in the Diode dataset.

The modelling data is a subset of the raw extract. We apply three restrictions.

- Focus on the US only from April 2018 to September 2020
- Marginalise out the demand side (advertisers and campaigns) and intermediaries (both Supply Side Platforms (SSPs) and Demand Side Platforms (DSPs))
- Simplify creative to 7 formats across video and display.

Without country, campaign, DSP, and SSP interactions, the modelling dataset has 2.5M rows. We describe it fully in the main text.

A.2 External validity

There is no widely accepted, publicly available data source on programmatic advertising that allows us to benchmark CPM distributions by PMP status and site. We therefore validate our data in two steps. First, we compare programmatic CPMs with The Association of National Advertisers Programmatic Media Supply side study (ANA, 2023a). Second, we examine the correlation between top sites in our data (measured by total impressions) and top sites in comScore data (measured by reach). We do not expect a one-to-one match because consumer reach and ad impressions are not the same object, but we do expect a positive correlation.

Our data is broadly consistent with these benchmarks, closely matching the distribution of CPMs and showing positive correlation with top comScore sites. There is evidence that PMP deals are slightly more prevalent in our data.

A.2.1 ANA programmatic supply chain study

The Association of National Advertisers Programmatic Media Supply side report is an in-depth study of the programmatic supply chain in the US (ANA, 2023a). It collected log-level impression data for 21 US advertisers buying more than 35 billion impressions. The study was conducted in 2022 H2.

We compare these data with the latest year in our dataset, 2019-2020. The overall distribution of CPMs in Diode closely matches the ANA study. The median in the ANA dataset is just under \$3 CPM. In our raw extract, the median CPM is \$3.11 and 14% of impressions are below \$1. The median is slightly higher in our modelling dataset (see Figure 15).

A.2.2 comScore

The ANA study does not provide site-level information. We use comScore’s consumer panel data to check whether our sites are representative of the US market. This is an imperfect approach because consumer panel data is only a proxy for availability of advertising inventory at site level. Some popular sites and apps may carry very little advertising, e.g. Netflix.

We match about half of the comScore top 200 websites. Figure 16 shows a reasonable correlation between top sites in comScore and Diode.

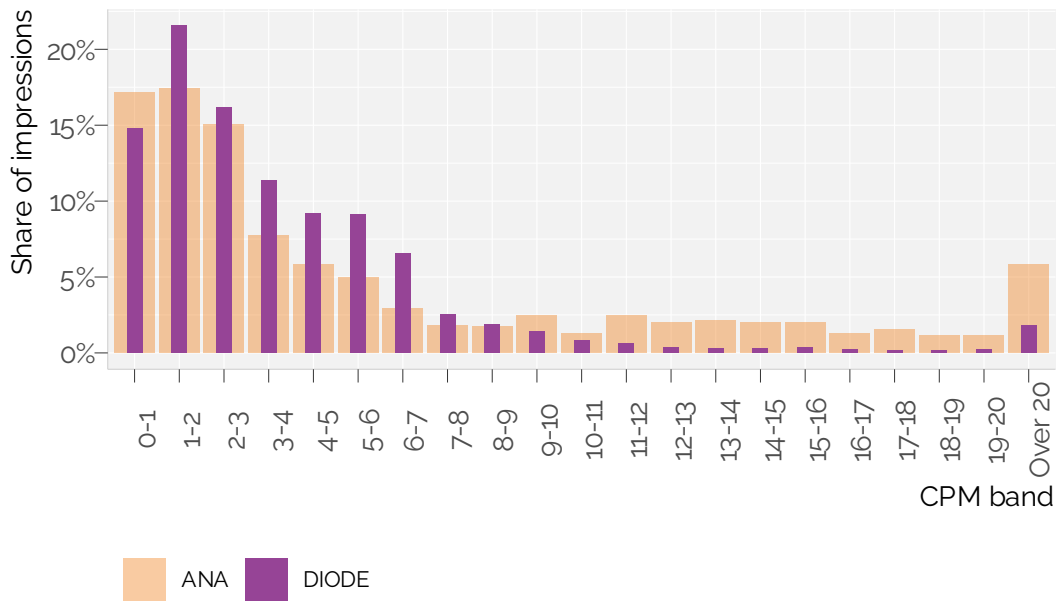


Figure 15: CPM distribution, DIODE raw vs ANA study, US 2019–2020.

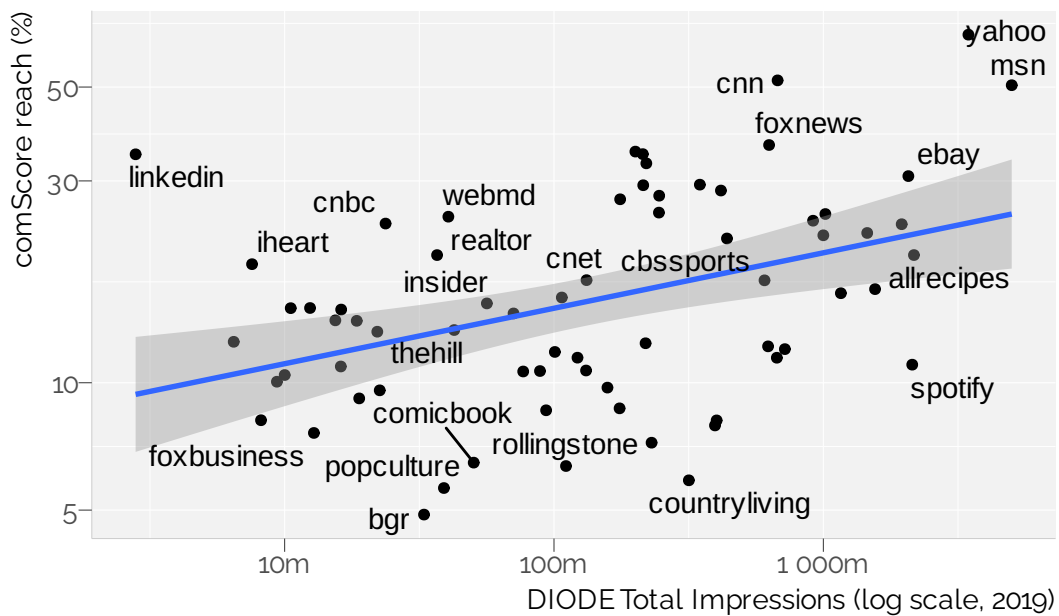


Figure 16: Representativeness of DIODE site data, comparison with comScore reach, 2019.

Appendix B — Model Fit and Validation

This appendix evaluates the overall fit of the intensive (Stage Two) and extensive (Stage One) models. The main text (Section 6.1) summarises the headline fit statistics; here we provide supporting diagnostics.

B.1 Stage Two: Intensive margin model

The Dynamic Factor Model (DFM) produces fitted values for three endogenous variables: RTB price (y_1), PMP price (y_2), and the selection index (y_3). Fitted values are constructed from the estimated parameters μ_j , factor loadings Λ_j , factors f_t , and inverse Mills ratios (for the PMP equation only).

Table 6 summarises the correlation between fitted and observed values, and Figure 17 shows fitted versus observed CPMs with normalised 2D density contours, faceted by equation (RTB, PMP) and creative format (display, video).

The model fit is acceptable. Outliers in the upper tail (observed CPM > \$50) are systematically under-predicted. This is expected: the factor model regularises extreme values towards the conditional mean, and products with very high CPMs likely possess idiosyncratic characteristics not fully captured by the 60 factors. Because elasticity estimates are computed at the product level and averaged, these tail outliers do not disproportionately influence the main results.

Table 6: Model fit, correlation between fitted and observed values.

Model	N	Observed, %	Pearson	Spearman
RTB Price (y)	1,848,861	59.1%	0.96	0.95
PMP Price (y)	1,038,581	33.2%	0.97	0.95
Selection Index (y)	899,803	28.8%	0.97	0.96

Note:

Observed, %: share of product-day observations with fitted values available. Fitted PMP prices require observed PMP participation to compute the Mills correction. Coverage is therefore lower for PMP (58%) than RTB (82%). The selection index has near-complete coverage (>99%) because it is observed whenever either channel is active.

B.2 Stage One. Extensive margin model

The Stage 1 model assigns each product daily switching probabilities across regimes. Applying these probabilities forward indefinitely yields a long-run distribution, the share the market converges to regardless of where it starts. Figure 18 compares this implied long-run distribution with the actual regime shares over the last 30 days of the sample. The two should be close if the market was near its resting point by late 2020. We find broad consistency for the both-markets and RTB-only regimes, the two largest groups. The PMP-only share is small, consistent with publishers preferring dual-market presence over exclusive PMP commitment.

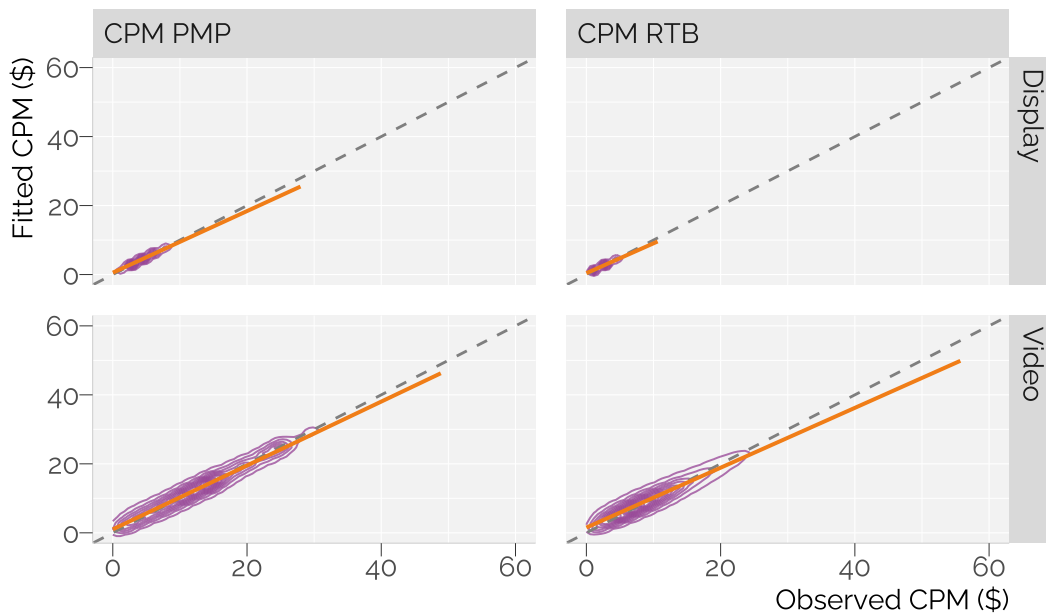


Figure 17: Fitted vs observed CPM by equation and creative format.



Figure 18: Stage 1 model validation, long-run regime shares (model-implied long-run equilibrium) vs end-of-sample observed shares (last 30 days).

Appendix C — Structural Parameters

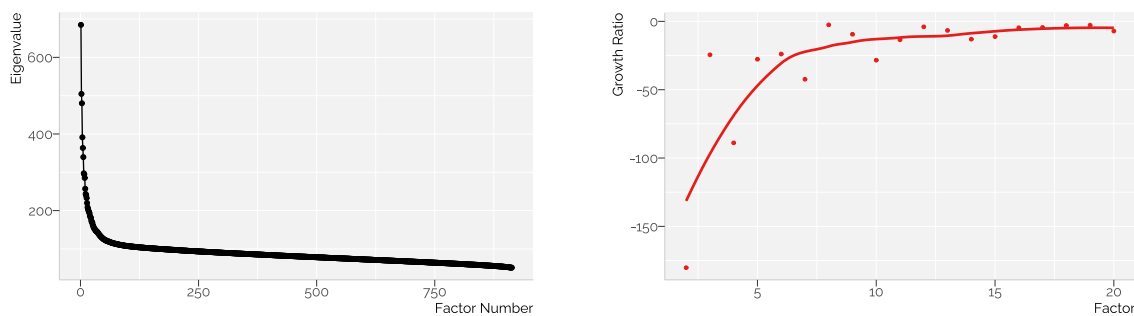
This appendix analyses the structural parameters that supplement Chapter 6. It covers factor analysis and naming, viewability and the inverse Mills ratio, joint determinants of σ_{23} and ρ_{23} , bounds on σ_{12} , and Stage-1 category heterogeneity in entry barriers and switching inertia.

C.1 Factor analysis

The dynamic factor model estimated $K = 60$ latent factors, where the number of factors K is chosen by the marginal contribution to explained variance. These factors capture common trends and shocks affecting all products, allowing us to control for correlated unobservables without imposing parametric restrictions on the form of these shocks.

C.1.1 Eigenvalue scree plot

From these $K = 60$ factors, we focus our naming exercise on the top 10. Figure 19 shows that there is a distinct elbow around factor 10. We focus the naming exercise on the first 10 factors, but retain all 60 so that idiosyncratic variation remains limited and elasticity estimates are less exposed to omitted common shocks.



(a) Eigenvalues

(b) Growth ratio (factors 2–20)

Figure 19: Scree plot, eigenvalues (left) and growth ratio (right).

C.1.2 Factor naming and event correlation

We label the leading factors by first deseasonalising and then correlating them with known market events and observable aggregate time series (e.g., Google Trends for “GDPR”, stock market indices, COVID-19 case counts). Figure 20 shows event-study coefficients for level shifts after key events. For example, Apple’s Intelligent Tracking Prevention (ITP) is associated with factors 4 and 5, and Google’s roll-out of first-price bidding (GAM) is associated with factor 2.

Whilst many factors map to specific events, some combine several influences. This is expected because factors are linear combinations of all prices and participation decisions, and several confounded shocks (technology changes, advertiser demand shifts, platform policy updates)

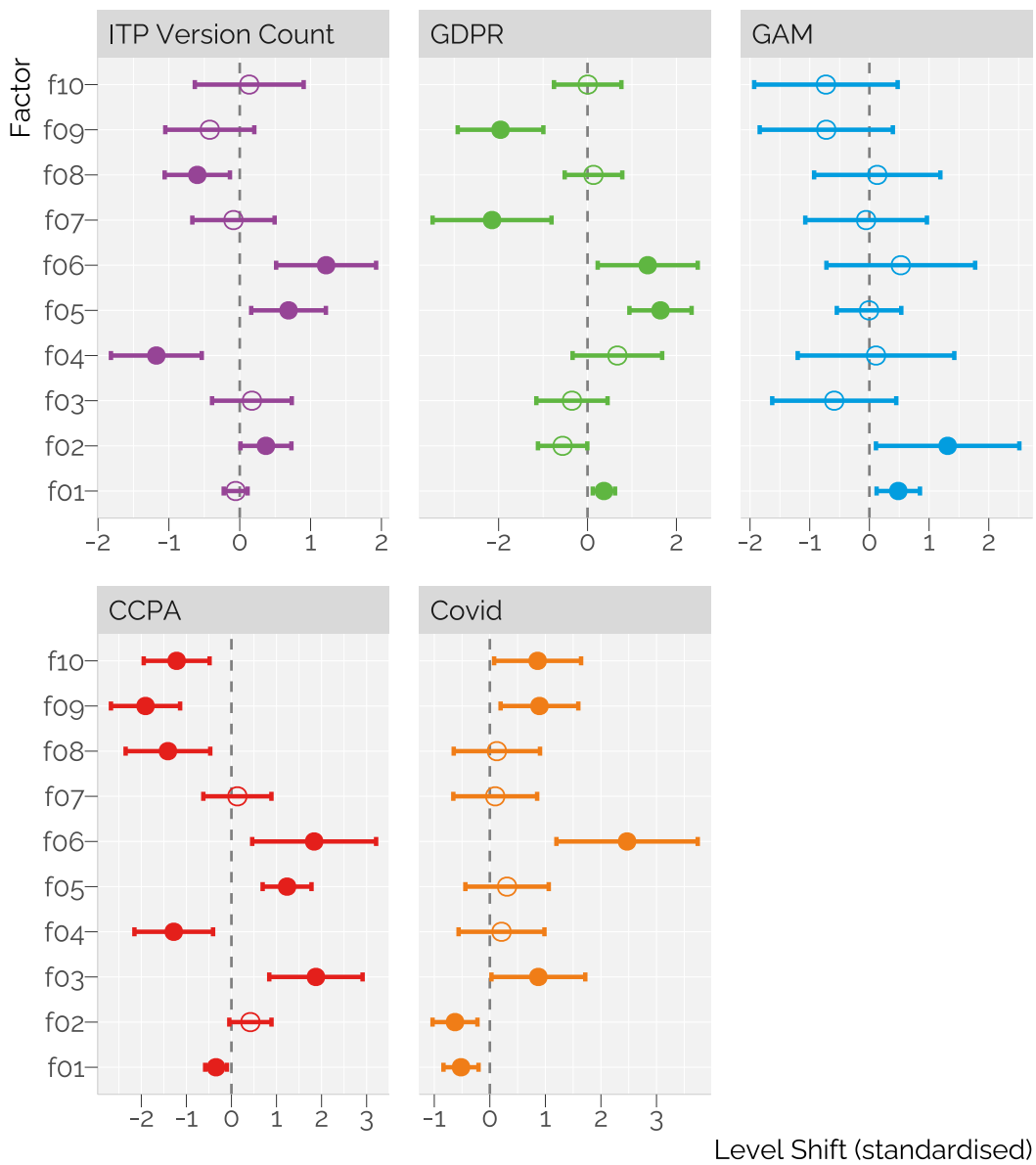


Figure 20: Event effects on factor levels.

occur at the same time. The methodological point is that factors control for these shocks without requiring us to name or model each one explicitly.

C.1.3 Viewability and selection intensity

Higher daily viewability is associated with a higher inverse Mills ratio. On days when a product delivers more viewable impressions, it is more likely to route those impressions to PMP (Table 7). This pattern is consistent with publishers using real-time viewability signals to direct better inventory to private deals, the information-asymmetry channel captured by a positive σ_{23} .

Table 7: Impact of viewability on Mills ratio, different fixed effects specifications.

	Video PMP		Display PMP	
	(1)	(2)	(1)	(2)
Measurable rate	0.193*** (0.002)	0.141*** (0.002)	0.243** (0.082)	0.128* (0.052)
Viewable given measured	0.674*** (0.008)	0.123*** (0.009)	0.334*** (0.012)	-0.097*** (0.008)
Num.Obs.	199703	199703	616085	616085
R2	0.198	0.509	0.211	0.600
FE: Period	✓	✓	✓	✓
FE: Site	Site Category	Product-level	Site Category	Product-level

Note:

Dependent variable: Inverse Mills Ratio. (1) date FE with site category as regressor (category coefficients omitted). (2) product and date FE (within-product, within-day variation). Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.2 Information asymmetry: σ_{23} and ρ_{23}

The parameter σ_{23} is the covariance between the impression-level shock that pushes an impression towards PMP and the impression-level shock to the PMP price. It captures the degree of information asymmetry in the allocation decision. When $\sigma_{23} > 0$, impressions with above-average unobserved value are disproportionately routed through PMP — consistent with publishers using private information to direct premium inventory to private deals. When $\sigma_{23} < 0$, the pattern runs in the opposite direction. PMP is used as a residual channel in weaker demand states, with higher-value impressions remaining in RTB. For example, publishers might route cookie-less Safari impressions to PMP. Both patterns are plausible in this market and we find evidence of both signs in the data, with the cross-sectional distribution approximately centred at zero.

C.2.1 Decomposition: σ_{23} versus ρ_{23}

Since $\text{Var}(\varepsilon_3) = 1$ (the allocation shock variance is normalised), we have $\sigma_{23} = \rho_{23}\sqrt{\sigma_{22}}$, where $\rho_{23} \in [-1, 1]$ is the correlation between the allocation shock and the PMP price shock, and σ_{22} is the within-day variance of the PMP price shock. This decomposition reveals that σ_{23} conflates two economically distinct objects.

- ρ_{23} captures the publisher’s pure information advantage: how well their routing decision predicts unobserved impression value, irrespective of how much within-day price heterogeneity exists.
- $\sqrt{\sigma_{22}}$ captures the opportunity set: a product with highly dispersed intraday CPMs exhibits a large σ_{23} even if the correlation ρ_{23} is only moderate.

Interpreting σ_{23} as a measure of information asymmetry therefore requires controlling for σ_{22} . Regressing σ_{23} directly on site characteristics will conflate informational ability with price dispersion if, for example, large or niche sites tend to have more volatile intraday CPMs. Working with ρ_{23} strips out the scale effect and isolates the publisher’s ability to exploit private information.

C.2.2 Recovering σ_{22} and ρ_{23}

The intraday variance σ_{22} is not identified from daily-average data alone. We proxy it using independent impression-level data from a Demand-Side Platform (Xandr, US, 2019 Q2), which contains intraday CPMs at the domain-format-deal-type level. For each product we compute the within-cell variance of CPMs and aggregate across the 90-day window. Where a direct site-by-format match exists we use it; otherwise we fall back to site-level, tag-level, or overall averages. This creates two datasets: an exact-match sample and an imputed sample. We obtain direct matches for approximately 60% of sites and 70% of formats; the remainder rely on aggregated proxies. We refer to products with direct matches as the “exact” sample throughout this appendix. Because σ_{22} for unmatched products is based on averages, estimates of ρ_{23} for those products are attenuated. The exact-match sample therefore provides the more reliable basis for inference.

C.2.3 Determinants of σ_{23} and ρ_{23}

Table 8: Joint F-test, product characteristics do not predict σ_{23} or ρ_{23} .

	σ_{23}		ρ_{23}	
	All	Exact	All	Exact
N	2,913	801	2,913	801
k	27	25	27	25
R^2	0.009	0.04	0.005	0.033
F	0.95	1.3	0.56	1.05
$p(F)$	0.545	0.146	0.967	0.393

OLS with video format, site category, log impressions, viewability rates, and data fee. H_0 : all slope coefficients jointly zero. $F = (R^2/k) / ((1 - R^2)/(n - k - 1))$. All sample uses imputed σ_{22} ; Exact sample uses direct site \times format Xandr matches.

Table 8 presents joint regressions of σ_{23} and ρ_{23} on observable product characteristics across the full sample and the exact-match sample. Both σ_{23} and ρ_{23} are regressed on a common baseline specification (video, site category, log impressions). The results are qualitatively similar across the two objects, indicating that the absence of systematic patterns is not driven solely by scale variation in σ_{22} . Site categories do not predict either parameter, and neither do impression volumes, viewability rates, or data fees. R^2 values are negligible in all specifications, confirming that observable characteristics explain virtually none of the cross-sectional variation.

The absence of systematic patterns suggests that these effects vary idiosyncratically across products rather than along observable group dimensions. For example, first-party data on logged-in users may increase $|\rho_{23}|$, but we do not observe this capability directly and it is more likely to vary at site level than at site-category level. The sign of ρ_{23} may also vary within a site or product over time as inventory mix and deal structure change. In some states, high-quality impressions are selected for PMP, while in others lower-quality impressions are selected (for example, impressions in Safari browsers without cross-site tracking).

C.3 Bounds on σ_{12}

The parameter σ_{12} represents the covariance between unobserved shocks to RTB and PMP prices. The impact of PMP availability on RTB prices is a recurring theme in the literature. (Balocco et al., 2025) use data from a single Demand-Side Platform (DSP) for one week and show that PMP depresses RTB prices through an adverse-selection effect. Because we never observe the same impression sold in both channels, σ_{12} is not point-identified. We can bound it from the positive semi-definiteness of Σ_j only if we have external estimates of the intraday variances ($\sigma_{11j}, \sigma_{22j}$).

C.3.1 Bounding methodology

For each product j , the covariance matrix of unobserved shocks is

$$\Sigma_j = \begin{pmatrix} \sigma_{11j} & \sigma_{12j} & \sigma_{13j} \\ \sigma_{12j} & \sigma_{22j} & \sigma_{23j} \\ \sigma_{13j} & \sigma_{23j} & 1 \end{pmatrix}$$

Only $\text{Var}(\varepsilon_{3tji}) = 1$ is normalised (the allocation shock). As discussed above in the context of ρ_{23} , the intraday variances ($\sigma_{11j}, \sigma_{22j}$) are not identified from daily average data. We therefore use the impression-level data from Xandr to estimate these variances. We compute bounds numerically for each product where these external variance estimates are available (the “exact” sample). Despite this additional data, the bounds are wide and uninformative, and we can sign fewer than 10% of products.

This lack of identification for σ_{12} does not affect our elasticity estimates because the elasticity formulae depend on σ_{23} (which is identified via the selection correction) and on $\text{Var}(\xi_{3,jt})$ (the day-to-day idiosyncratic variance of the allocation index), but not on σ_{12} or the intraday variances (σ_{11}, σ_{22}) directly.

C.4 Heterogeneity in category-level entry barriers and switching inertia (Stage 1)

The extensive-margin model maps site- and format-level intercepts into PMP-only entry attractiveness (κ_2 , in \$/CPM equivalents) and switching inertia ($-c_s/(\lambda\beta)$ in \$/CPM, with λ trimmed near zero before conversion), using the same probability-weighted destination weights as in the structural analysis chapter. Figure 21 plots site-category medians. Categories towards the right face lower implicit barriers to PMP-only listing at a given PMP price; higher vertical positions indicate more friction when re-configuring market participation. The cloud slopes downward. Categories where PMP-only entry is less attractive also tend to exhibit higher switching inertia, which suggests that the two estimated objects co-vary systematically across site categories. We interpret that pattern as evidence that common category characteristics, such as deal-making capacity or buy-side demand, may shape both margins.

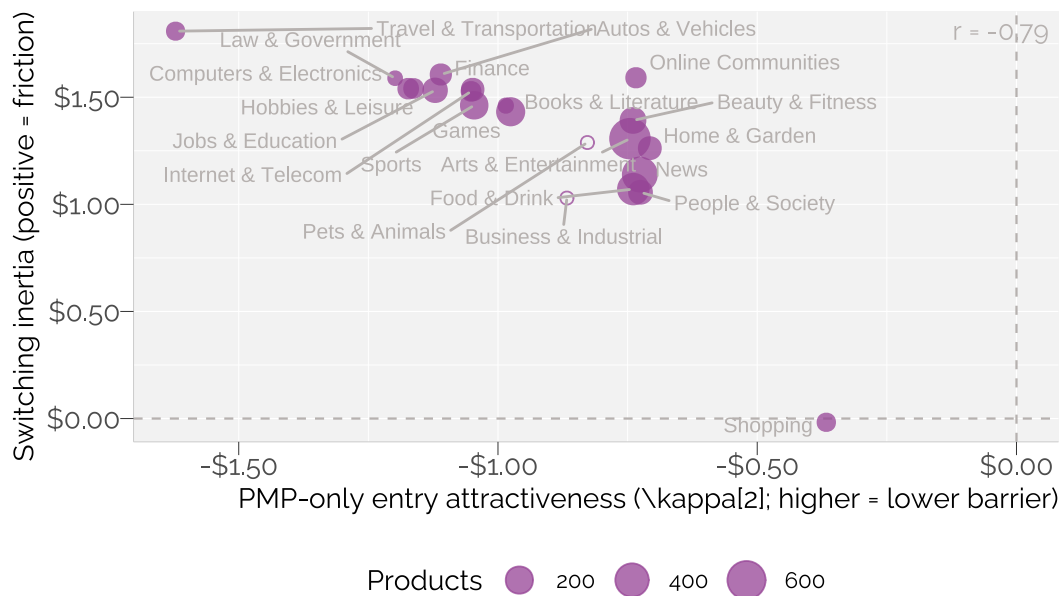


Figure 21: Site category medians: PMP-only entry attractiveness (κ_2) vs switching inertia ($-c_s/(\lambda\beta)$), \$/CPM equivalents. Point size proportional to number of products; open circles flag $n < 20$. Real Estate is excluded because of small sample size.

Appendix D — Elasticity Estimates

This appendix documents the robustness of the elasticity estimates in Section 6.5. It derives the selection-corrected elasticity formula, compares baseline versus corrected estimates, discusses coverage and sample construction, and presents cross-price elasticity and distribution details.

D.1 Elasticity definition and identification

The PMP own-price supply elasticity measures how the PMP allocation share s_{2tj} (the fraction of impressions sent to PMP) responds to changes in observed PMP price \bar{p}_{2tj}

$$\eta_{s_2, p_2} = \frac{\partial s_{2tj}}{\partial \bar{p}_{2tj}} \cdot \frac{\bar{p}_{2tj}}{s_{2tj}}$$

where $s_{2tj} = \Phi(y_{3tj})$ is the probit allocation share. To compute $\partial s_{2tj} / \partial \bar{p}_{2tj}$, we first obtain the derivative of s_{2tj} with respect to the latent PMP index y_{2tj} , then correct for the fact that observed price \bar{p}_{2tj} is not equal to y_{2tj} because of selection.

Selection feedback. Under the Heckman-style observation equation, observed PMP price is $\bar{p}_{2tj} = y_{2tj} + \sigma_{23,j} M_{2tj}$, where $M_{2tj} \equiv \phi(y_{3tj}) / \Phi(y_{3tj})$. The complication is that M_{2tj} depends on s_{2tj} , and s_{2tj} depends on y_{2tj} through the allocation index $y_{3tj} = y_{2tj} - y_{1tj} - \text{cost}_{tj}$. So when y_{2tj} rises, both the direct effect (higher price) and the indirect effect (higher $s_{2tj} \rightarrow$ lower Mills ratio) operate. Ignoring this feedback overstates the responsiveness of allocation to observed price and thus overstates the elasticity. We derive the correct derivative by inverting the chain rule.

Derivation. The allocation index satisfies $\partial y_{3tj} / \partial y_{2tj} = 1$. Under baseline normalisation ($\sigma_\eta = 1$), the probit derivative is $\partial s_{2tj} / \partial y_{2tj} = \phi(y_{3tj})$.¹⁸ Differentiating $M_{2tj} = \phi(y_{3tj}) / \Phi(y_{3tj})$ yields $\partial M_{2tj} / \partial y_{3tj} = -M_{2tj} (y_{3tj} + M_{2tj})$. The total derivative of observed price with respect to y_{2tj} is therefore

$$\frac{d\bar{p}_{2tj}}{dy_{2tj}} = 1 + \sigma_{23,j} \frac{\partial M_{2tj}}{\partial y_{3tj}} \cdot \frac{\partial y_{3tj}}{\partial y_{2tj}} = 1 - \sigma_{23,j} M_{2tj} (y_{3tj} + M_{2tj})$$

because rising y_{2tj} increases s_{2tj} , lowers the Mills ratio, and partially offsets the rise in \bar{p}_{2tj} . Inverting via the chain rule:

$$\frac{\partial s_{2tj}}{\partial \bar{p}_{2tj}} = \frac{ds_{2tj}/dy_{2tj}}{d\bar{p}_{2tj}/dy_{2tj}} = \frac{\phi(y_{3tj})}{1 - \sigma_{23,j} M_{2tj} (y_{3tj} + M_{2tj})}$$

Substituting $M_{2tj} = \phi(y_{3tj}) / s_{2tj}$ yields the equivalent form

$$\frac{\partial s_{2tj}}{\partial \bar{p}_{2tj}} = \frac{\phi(y_{3tj})}{1 - \sigma_{23,j} \cdot \frac{y_{3tj} + M_{2tj}}{s_{2tj}} \cdot \phi(y_{3tj})}$$

¹⁸If within-day quality heterogeneity were non-trivial ($\Delta\gamma \neq 0$), the aggregate share would require integrating $\Phi((y_{3tj} + \Delta\gamma\sigma_j\epsilon)/\sigma_\eta)$ over $\epsilon \sim N(0, 1)$, approximated with Gauss-Hermite quadrature at $Q = 20$ nodes. All reported elasticity estimates use the analytic formula with $\Delta\gamma = 0$.

The selection parameters ($\sigma_{13,j}, \sigma_{23,j}$) are identified from three sources: the non-linearity of Mills ratios (which prevents absorption into the factor structure), corner-solution variation (days with $s_{2tj} \approx 0$ or $s_{2tj} \approx 1$ pin down the parameters via the non-linear squeeze on prices), and the Heckman-style exclusion restriction that y_{3tj} affects observed prices only through the Mills correction. The cross-price covariance $\sigma_{12,j}$ is not point-identified because we never observe the same impression in both markets.

D.2 Trend in elasticity over time

Table 9: Median own-price elasticity by year.

Format	2018	2019	2020
Display	6.24	4.05	2.12
Video	12.97	8.45	5.09

Note:

Median own-price elasticity computed for days where products are active in both PMP and RTB markets.

D.3 Baseline versus selection-corrected elasticities

The main text presents selection-corrected elasticities, which account for unobserved quality differences via the inverse Mills ratio. For robustness, Table 10 presents baseline elasticities that omit the Mills correction. These baseline estimates are computed by differentiating the PMP supply function with respect to PMP prices while setting the Mills ratio contribution to zero.

Table 10: Baseline (uncorrected) supply elasticities.

Format	N	Own-Price		Cross-Price	
		Mean	Median	Mean	Median
Display	527820	6.73	4.29	-3.50	-2.52
Video	160011	18.46	11.74	-11.95	-9.53

Note:

Baseline estimates set the Mills ratio contribution to zero (no selection correction). Mean is impression-weighted. Sample restricted to days when the product was active in both the PMP and RTB markets. A small number of extreme observations ($|\eta| > 100$) are excluded.

Baseline elasticities are systematically higher than selection-corrected elasticities. The median own-price elasticity for display is 4.3 (baseline) versus 3.4 (corrected), and for video it is 11.7 (baseline) versus 7.3 (corrected), a 30-40% difference. The bias direction depends on $\sigma_{23,j}$: for products with $\sigma_{23,j} > 0$ (unobserved quality raises both price and allocation), the corrected

elasticity exceeds the baseline; for products with $\sigma_{23,j} < 0$ (supply dilution, where greater allocation accompanies lower per-impression prices), the corrected elasticity is lower. When we ignore selection, we conflate genuine price responsiveness with this composition effect.

D.4 Sensitivity to σ_{23} strength

The selection-corrected elasticity depends on the estimated $\sigma_{23,j}$ through the denominator of the derivative $\partial s_{2tj} / \partial \bar{p}_{2tj}$. To assess how sensitive our conclusions are to the magnitude of the selection correction, we scale $\sigma_{23,j}$ by $\omega \in \{0, 0.5, 1.0, 1.5, 2.0\}$ and recompute the elasticity at each value. $\omega = 0$ removes selection entirely (equivalent to the baseline in the preceding section), $\omega = 1$ is our central estimate, and $\omega = 2$ doubles the correction. If the headline elasticity estimates are robust, the $\omega \in \{0.5, 1.5\}$ bounds should bracket the central estimate by less than 30%.

Figure 22 shows that as we increase ω , the supply elasticity for both video and display formats falls, and the gap between the two narrows as the selection effect dominates.

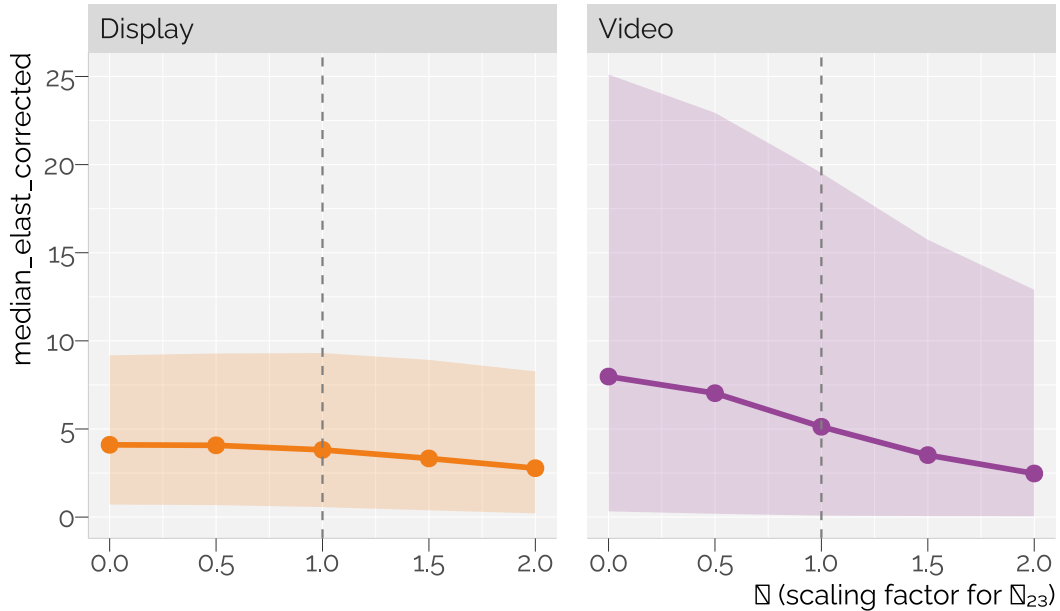


Figure 22: Sensitivity of selection-corrected supply elasticity to σ_{23} scaling (ω), ribbon shows interquartile range across products, line shows median. $\omega = 1$ is the central estimate.

D.5 Coverage and sample

Elasticities are computed at the product-day level and then aggregated. This restricts the sample to around 30% of the original sample. While it logically makes sense to only compute elasticities for product-days where both PMP and RTB are observed, we can use the Dynamic Factor Model to impute the elasticity for product-days where only one market is observed. The main results are robust to alternative handling of these observations.

We can partially fill excluded observations using the factor model: for corner shares or selection-correction failures, use factor-implied y_3 or baseline (uncorrected) elasticity as fallback. Table 11 compares the main sample with the imputed sample. Median and IW mean elasticities are similar across samples, indicating that the exclusion of problematic observations does not materially distort the headline estimates.

Table 11: Robustness check, main vs imputed sample (baseline fallback when selection correction fails).

Format	sample	N	Median	IW Mean
Display	Main	170,730	3.4	6.6
Display	Imputed	527,820	3.7	6.9
Video	Main	47,849	7.3	11.3
Video	Imputed	160,011	9.0	14.5

Appendix E — Model Derivations and Technical Details

This appendix collects the technical derivations referenced in Chapter 5. It covers expected profits for the extensive-margin choice, the Jacobian of the data-to-latent mapping, the log-likelihood and EM algorithm, treatment of missing values, Stage 1 gradients, and extensions (dynamic switching, binary logit).

E.1 Expected profits for the “both markets” option

Publishers choosing to operate in both markets optimally allocate each impression based on its realised quality ε_{3tji} . We derive expected profits *before* observing individual ε 's but *after* observing the allocation index y_{3tj} .

Conditional on y_{tj} and impression quality ε_{tji} , profits from RTB and PMP are

$$\begin{aligned}\pi_{1tji} &= y_{1tj} + \varepsilon_{1tji} \\ \pi_{2tji} &= y_{2tj} + \varepsilon_{2tji} - (y_{2tj} - y_{1tj} - y_{3tj} + \varepsilon_{2tji} - \varepsilon_{1tji} - \varepsilon_{3tji}) = y_{1tj} + y_{3tj} + \varepsilon_{1tji} + \varepsilon_{3tji}\end{aligned}$$

The optimal rule allocates to PMP if $\varepsilon_{3tji} > -y_{3tj}$.

Expected profits by market:

RTB (impressions with $\varepsilon_{3tji} < -y_{3tj}$):

$$\mathbb{E}[\pi_1 \mid y_{3tj}, \varepsilon_{3tji} < -y_{3tj}] = y_{1tj} - \sigma_{13,j} \frac{\phi(-y_{3tj})}{\Phi(-y_{3tj})}$$

PMP (impressions with $\varepsilon_{3tji} \geq -y_{3tj}$):

$$\mathbb{E}[\pi_2 \mid y_{3tj}, \varepsilon_{3tji} \geq -y_{3tj}] = y_{1tj} + y_{3tj} + (1 + \sigma_{13,j}) \frac{\phi(-y_{3tj})}{\Phi(y_{3tj})}$$

Weighting by market shares $\Phi(-y_3)$ and $\Phi(y_3)$:

$$\mathbb{E}[\pi \mid y_{3tj}] = \Phi(-y_{3tj})\bar{\pi}_{1tj} + \Phi(y_{3tj})\bar{\pi}_{2tj} = y_{1tj} + y_{3tj}(1 - \Phi(-y_{3tj})) + \phi(-y_{3tj})$$

Define $y_{3tj}^\dagger = y_{3tj}(1 - \Phi(-y_{3tj})) + \phi(-y_{3tj})$. Taking expectations conditional on F_{t-1} :

$$\mathbb{E}[\pi_{3tj} \mid F_{t-1}] = \mathbb{E}[y_{1tj} \mid F_{t-1}] + y_{3tj}^\dagger$$

where $y_{3tj}^\dagger = \mathbb{E}[y_{3tj}^\dagger \mid F_{t-1}]$. This quantity requires Gauss-Hermite quadrature because y_{3tj}^\dagger is non-linear in y_{3tj} .

E.2 Jacobian of the transformation

The mapping $H_j(\bar{p}_j)$ from observed data to latent indices has Jacobian:

$$\frac{\partial H_j}{\partial \bar{p}_j} = \begin{bmatrix} 1 & 0 & \sigma_{13,j} \frac{\partial M_{1tj}}{\partial y_3} \\ 0 & 1 & -\sigma_{23,j} \frac{\partial M_{2tj}}{\partial y_3} \\ 0 & 0 & 1 \end{bmatrix}$$

where

$$\frac{\partial M_{1tj}}{\partial y_3} = \frac{\phi(y_3)}{\Phi(-y_3)} \left(-y_3 + \frac{\phi(y_3)}{\Phi(-y_3)} \right), \quad \frac{\partial M_{2tj}}{\partial y_3} = -\frac{\phi(y_3)}{\Phi(y_3)} \left(-y_3 - \frac{\phi(y_3)}{\Phi(y_3)} \right)$$

The determinant is $\det \left(\frac{\partial H_j}{\partial \bar{p}_j} \right) = 1$, so the Jacobian term vanishes in the log-likelihood.

E.3 Log-likelihood and EM algorithm

Let $H_{t*}(x_t)$ denote the stacked latent indices for non-missing observations on date t , and J_{t*} the number of non-missing elements. The log-likelihood of X_T is

$$L_X(X_T) = \int e^{\sum_{t=1}^T g_{x_{t*}}(x_{t*}|f_t) + g_f(f_t|F_{t-1}) + g_0(F_0)} dF_T$$

where

$$\begin{aligned} g_{x_{t*}}(x_{t*}|f_t) &= -0.5 (H_{t*}(x_t) - \mu_{t*} - \Lambda_{t*} f_t)^T \Sigma_{\xi,t*}^{-1} (H_{t*}(x_t) - \mu_{t*} - \Lambda_{t*} f_t) \\ &\quad - 0.5 \log \det(\Sigma_{\xi,t*}) + \log \det \left| \frac{\partial H(x_t)}{\partial x_t} \right| - \frac{J_{t*}}{2} \log(2\pi) \\ g_f(f_t|F_{t-1}) &= -0.5 \left(f_t - \sum_{s=1}^p A_s f_{t-s} \right)^T \Sigma_{\eta}^{-1} \left(f_t - \sum_{s=1}^p A_s f_{t-s} \right) - \log \det(C_{\eta}) - \frac{r}{2} \log(2\pi) \\ g_0(F_0) &= -0.5 (F_0 - \mu_0)^T \Sigma_0^{-1} (F_0 - \mu_0) - \log \det(C_0) - \frac{rp}{2} \log(2\pi) \end{aligned}$$

The EM algorithm alternates in two steps. First, in the *E-step*, we use the Kalman filter and smoother to compute $\mathbb{E}[f_t | X_T, \theta]$ and $\mathbb{E}[f_t f_t^T | X_T, \theta]$. Second, in the *M-step*, we maximise the expected complete-data log-likelihood. Factor loadings and selection parameters are updated via equation-by-equation OLS, and VAR parameters via moments of the smoothed factors. We iterate until $\max |\theta_i - \theta_{i-1}| < \tau$ or the relative log-likelihood change falls below tolerance.

E.3.1 Factor normalisation

We normalise $\mathbb{E}[f_t] = 0$ and $\Sigma_{f_0} = I_r$. All observed variables are centred to mean 0 and scaled to variance 1, providing a data-driven rotation that resolves factor scale indeterminacy.

E.4 Treatment of missing values

Several cases yield missing values.

RTB only: If product j is observed only in RTB, then μ_{2j} , μ_{3j} , Λ_{2j} , Λ_{3j} , $\sigma_{13,j}$, $\sigma_{23,j}$ and corresponding covariance blocks are undefined for that product.

PMP only: Analogously, if product j is observed only in PMP, then μ_{1j} , μ_{3j} , Λ_{1j} , Λ_{3j} , $\sigma_{13,j}$, $\sigma_{23,j}$ are undefined.

Zero variance: If \bar{p}_{1j} or \bar{p}_{2j} has zero variance, the corresponding loadings and selection parameters cannot be identified. The Kalman filter omits the affected rows from the measurement equations on each date.

E.5 Stage 1 gradient and Hessian

Let $(s_{1tj}, s_{2tj}, s_{3tj})$ be the multinomial logit choice probabilities and $(d_{1tj}, d_{2tj}, d_{3tj})$ the observed choice indicators. The Stage 1 log-likelihood is

$$\log L(\theta) = T^{-1} \sum_{t,j} (d_{1tj} \log s_{1tj} + d_{2tj} \log s_{2tj} + d_{3tj} \log s_{3tj})$$

The gradients are

$$\frac{\partial \log L}{\partial \beta_j} = T^{-1} \sum_t [\bar{y}_{3tj}(d_{2tj} - s_{2tj}) + y_{3tj}^+(d_{3tj} - s_{3tj})]$$

$$\frac{\partial \log L}{\partial \kappa_{2j}} = T^{-1} \sum_t (d_{2tj} - s_{2tj})$$

$$\frac{\partial \log L}{\partial \kappa_{3j}} = T^{-1} \sum_t (d_{3tj} - s_{3tj})$$

References

- Abraham, I., Athey, S., Babaioff, M., & Grubb, M. D. (2020). Peaches, lemons, and cookies: Designing auction markets with dispersed information. *Games and Economic Behavior*, *124*, 454–477. <https://doi.org/10/gmwwq6>
- Adshead, S., Forsyth, G., Wood, S., & Wilkinson, L. (2019). *Online Advertising in the UK* (p. 110) [Tech. {{Rep}}]. Plum.
- Alcobendas, M., & Zeithammer, R. (2023). *Slim Shading in Ad Auctions: Adjustment of Bidding Strategies to First-Price Rules* [Working {{Paper}}].
- ANA. (2023a). *Programmatic Media Supply Chain Transparency Study*. Association of National Advertisers.
- ANA. (2023b). *Programmatic Media: Supply Side Transparency Study - First Look*. Association of National Advertisers.
- Araman, V. F., & Popescu, I. (2010). Media Revenue Management with Audience Uncertainty: Balancing Upfront and Spot Market Sales. *M&SOM*, *12*(2), 190–212. <https://doi.org/10.1287/msom.1090.0262>
- Arnosti, N., Beck, M., & Milgrom, P. (2016). Adverse Selection and Auction Design for Internet Display Advertising. *American Economic Review*, *106*(10), 2852–2866. <https://doi.org/10/f9bnkd>
- Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, *70*(1), 191–221. <https://www.jstor.org/stable/2692167>
- Balocco, F., Lu, Y., Li, T., & Gupta, A. (2025). Lemon Ads: Adverse Selection in Multichannel Display Advertising Markets. *Management Science*. <https://doi.org/10.1287/mnsc.2022.03407>
- Balseiro, S. R., Besbes, O., & Weintraub, G. Y. (2015). Repeated Auctions with Budgets in Ad Exchanges: Approximations and Design. *Management Science*, *61*(4), 864–884. <https://www.jstor.org/stable/24550370>
- Bergemann, D., Heumann, T., Morris, S., Sorokin, C., & Winter, E. (2022). *Optimal Information Disclosure in Auctions* (Working {{Paper}} DP 16858).
- Bhattacharya, V., Roberts, J. W., & Sweeting, A. (2014). Regulating bidder participation in auctions. *The RAND Journal of Economics*, *45*(4), 675–704. <https://doi.org/10.1111/1756-2171.12067>
- Breitmar, N. A., Harding, M., & Lamarche, C. (2023). Using Grouped Data to Estimate Revenue Heterogeneity in Online Advertising Auctions. *AEA Papers and Proceedings*, *113*, 161–165. <https://doi.org/10.1257/pandp.20231095>
- CHEQ. (2020). *Ad Fraud 2020 - Economic cost of bad actors on the internet*.
- Choi, H., & Mela, Carl F. (2019a). *Display Advertising Pricing in Exchange Markets* (p. 53) [Working {{Paper}}].
- Choi, H., & Mela, Carl F. (2019b). Monetizing Online Marketplaces. *Marketing Science*. <https://doi.org/10/gg6rnr>
- Choi, W. J., & Sayedi, A. (2022). Open and Private Exchanges in Display Advertising. *Marketing Science*. <https://doi.org/10.1287/mksc.2022.1399>
- Cohen, M. C., Désir, A., Korula, N., & Sivan, B. (2022). Best of Both Worlds Ad Contracts: Guaranteed Allocation and Price with Programmatic Efficiency. *Management Science*. <https://doi.org/10.1287/mnsc.2022.4542>
- D’Annunzio, A., & Russo, A. (2023). Intermediaries in the Online Advertising Market. *Market-*

- ing Science. <https://doi.org/10.1287/mksc.2023.1435>
- Despotakis, S., Ravi, R., & Sayedi, A. (2021). First-Price Auctions in Online Display Advertising. *Journal of Marketing Research (JMR)*, 58(5), 888–907. <https://doi.org/10.1177/00222437211030201>
- Devaux, R. (2023). *Display Advertising: How Context Matters?*
- Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics, Annals Issue on Forecasting*, 164(1), 188–205. <https://doi.org/10.1016/j.jeconom.2011.02.012>
- Duckworth, S., Mysliwski, M., & Nesheim, L. (2023). *Taking the Biscuit: How Safari Privacy Policies Affect Online Advertising* (Version 1). Cemmap. <https://doi.org/10.47004/wp.cem.2023.0423>
- Gentzkow, M., Shapiro, J. M., Yang, F., & Yurukoglu, A. (2024). Pricing Power in Advertising Markets: Theory and Evidence. *American Economic Review*, 114(2), 500–533. <https://doi.org/10.1257/aer.20220943>
- Goke, S., Weintraub, G. Y., Mastromonaco, R., & Seljan, S. (2022). *Bidders' Responses to Auction Format Change in Internet Display Advertising Auctions* (arXiv:2110.13814). arXiv. <https://arxiv.org/abs/2110.13814>
- Google. (2019). Rolling Out First Price Auctions to Google Ad Manager Partners. In Google. GroupM. (2022). *This Year Next Year 2022* [Tech. {{Rep}}]. GroupM.
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153–161. <https://doi.org/10.2307/1912352>
- Hristakeva, S., & Mortimer, J. H. (2023). Price Dispersion and Legacy Discounts in the National Television Advertising Market. *Marketing Science*. <https://doi.org/10.1287/mksc.2023.1442>
- IAB. (2014). *The Programmatic Handbook* [Tech. {{Rep}}]. Internet Advertising Bureau.
- Kim, A., Mirrokni, V., & Nazerzadeh, H. (2021). Deals or No Deals: Contract Design for Online Advertising. *Operations Research*, 69(5), 1450–1467. <https://doi.org/10.1287/opre.2020.2087>
- Lauermaun, S., & Wolinsky, A. (2025). Auctions with Frictions: Recruitment, Entry, and Limited Commitment. *Rev Econ Stud*, rdaf052. <https://doi.org/10.1093/restud/rdaf052>
- Levin, J., & Milgrom, P. (2010). Online Advertising: Heterogeneity and Conflation in Market Design. *American Economic Review*, 100, 603–607. <https://doi.org/10.1257/aer.100.2.603>
- Li, J., Yuan, Y., & Qin, R. (2014). Information disclosure in real-time bidding advertising markets. *Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics*, 139–143. <https://doi.org/10.1109/SOLI.2014.6960708>
- Media, J. (2020). *The Little Black Book of Private Marketplaces*.
- Sayedi, A. (2018). Real-Time Bidding in Online Display Advertising. *Marketing Science*, 37(4), 553–568. <https://doi.org/10.1287/mksc.2017.1083>
- Sayedi, A., Jerath, K., & Baghaie, M. (2018). Exclusive Placement in Online Advertising. *Marketing Science*, 37(6), 970–986. <https://doi.org/10.1287/mksc.2018.1098>
- Stock, J. H., & Watson, M. W. (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460), 1167–1179. <https://www.jstor.org/stable/3085839>
- Wang, J., Zhang, W., & Yuan, S. (2017). Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting. *arXiv:1610.03013 [Cs]*. <https://arxiv.org/abs/1610.03013>
- Zeithammer, R., & Choi, W. J. (2025). Auctions of Auctions. *Management Science (INFORMS)*, 71(9), 7347–7365. <https://doi.org/10.1287/mnsc.2024.05233>