TWO-SIDED MARKETS SHAPED BY PLATFORM-GUIDED SEARCH

KWOK HAO LEE
Department of Strategy and Policy, National University of Singapore (NUS) Business School and NUS Artificial Intelligence Institute

LEON MUSOLFF
Business Economics and Public Policy Department, The Wharton School, The University of Pennsylvania

1. DATA DESCRIPTION

Our primary data source is a repricing company that administers offer listings on behalf of third-party merchants on Amazon Marketplace. The company provides us with three data sets of interest spanning one and a half years. Firstly, we have access to the notifications sent to the company programmatically by Amazon. These notifications are sent if there are any changes on any of the offers on the listings that the company administers (e.g., price updates, exits or entries). Secondly, the company is informed of any sales on its customers' listings. Finally, we have regular data on the sales ranks of each product (i.e., market, not offer).

To begin with, the notifications are our primary source of information about the alternatives available on any given market at any given time. For each notification and offer, we observe (i) a prominence dummy, (ii) whether the offer is prominence eligible, (iii) whether Amazon fulfills it, (iv) its list price and (v) shipping price, (vi) the feedback count, (vii) the fraction of positive feedback for the associated seller, and (viii) the time to dispatch.[1] The main challenge of this dataset is its uneven time resolution: a notification is sent whenever any offer characteristic *other than prominence* changes. Thus, while we have infinite temporal resolution on, e.g., prices, we lack this precision in prominence assignments.

The demand data contain, for each merchant registered with the repricing company, the time of every sale the merchant made during its period of registration.[2]

We combine these datasets using the following procedure. First, for each product, we proxy for the period that our seller was registered with the repricing company. To do so, we declare the product *observed* on any dates between the first and last observed sale. Second, we merge the notifications with the sales data by associating each notification with the sales we observe between the time it was sent and 48 hours later or when the next notification arises, whichever comes first. Because the period during which sales are associated with a notification varies across notifications, we record this duration for use in market size calculations.

After merging, we obtain a dataset at the 'offer x notification'-level with 204,396,069 observations. After replacing Amazon's feedback count with the maximum value in the dataset[3], summary statistics for our final data are provided in Table I(a). Further information (e.g., on the number of unique products) is provided in Table I(b). For instance, from 2018-08-26 to 2020-03-23 and across 47,403 products, we observe 803,849 sales which we can attribute to a notification at most 48 hours before the sale.

[1]Whenever we refer to price, we mean the sum of list price and shipping price.

[2]Though sales data linked to sellers on Amazon are rare, a caveat with this dataset is that we cannot independently verify *when* a merchant is registered with the company.

[3]This replacement was made because our data does not contain reliable information about the feedback count for Amazon. This imputation cannot influence our demand or prominence predictions, as we are controlling for an Amazon dummy in the relevant models. However, if we are significantly underestimating the feedback count of Amazon, it could influence the interpretation of the Amazon coefficient in, e.g., the prominence estimation.

TABLE I

SUMMARY STATISTICS AND DESCRIPTIVES.

	N	Avg.	Std. Dev.	Min.	50th	Max.
Price/MSRP	204,396,069	1.18	0.44	0.36	1.07	5.25
FBA	204,396,069	0.55	0.50	0.00	1.00	1.00
Amazon	204,396,069	0.02	0.13	0.00	0.00	1.00
Log (Feedback Count)	204,396,069	7.30	2.82	0.00	7.19	15.69
Time To Ship (in Days)	204,396,069	0.91	1.23	0.00	0.00	4.50
Prominence Dummy	204,396,069	0.09	0.28	0.00	0.00	1.00
Sales	19,409,013	0.04	0.57	0.00	0.00	448.00
Duration (in hours)	204,396,069	3.09	5.45	0.25	1.02	48.00

(a) Offer-Level Summary Statistics.

# Unique Products	47,403
# Merchants	63,620
# Markets (Product x Time)	19,409,013
# Sales	803,849
# Markets with Sales	411,401
Earliest Market	2018-08-26 01:18:08
Latest Market	2020-03-23 23:41:28
Earliest Market w/ Sale	2018-08-26 01:20:25
Latest Market w/ Sale	2020-03-23 22:21:50

(b) Descriptives.

Notes: These tables display summary statistics for the dataset used in prominence and consumer choice estimation. For the left table, the unit of observation is an offer on a given market, so a (merchant, time, product page). The right table displays counts of products, merchants, markets, and sales; as well as the first and last markets we see, with or without a sale.

2. ESTIMATION WITH PRICE ENDOGENEITY

Our mainline estimates make use of identifying Assumptions 6 and 8 in the main text, which imply $\mathbb{E}[\xi_{jt}|p_{jt}] = \mathbb{E}[\xi_{jt}]$ and $\mathbb{E}[\xi^r_{jt}|p_{jt}] = \mathbb{E}[\xi^r_{jt}]$. However, a typical worry in demand estimation is that unobserved quality ξ_{jt} and price p_{jt} may be correlated. Hence, we also develop estimators that rely on quality merely being time-invariant, i.e., Assumptions 7 and 9 in the main text. These estimators allow for arbitrary correlation between an offer's unobserved quality and price, as long as quality and price do not co-move within the same offer over time. This assumption is plausible because *observable* offer quality is mostly time-invariant. Indeed, the R^2 from regressing time until dispatch, log feedback count, being fulfilled by Amazon, and being sold by Amazon on offer fixed-effects are 0.97, 1.00, 0.98, and 1.00 respectively.

2.1. *Offer FE under Full Observability*

We now illustrate how to condition out alternative-specific fixed effects when estimating a multinomial logit discrete choice model. We first consider the case when the chosen alternative is observed (as for the prominence algorithm estimation). Subsequently, we extend our results to the case where there is only one alternative for which we observe whether it is chosen. This section generalizes results in Chamberlain (1980) which themselves are based on Rasch (1960, 1961). See Arellano and Honoré (2001) for a modern treatment.

Recall that the Buybox's mean utility from alternative j on product page w at time τ is given by $\delta^r_{jw\tau} = \mathbf{x}'_{jw\tau}\beta^r - \alpha^r p_{jw\tau} + \xi^r_{jw}$, where in contrast to the main text we have imposed Assumption 7, i.e., offer quality $\xi^r_{jw} \equiv \xi^r_{jw\tau}$ is time invariant.

Mean utilities on each offer are combined with a Type-1 extreme value shock $\epsilon_{jw\tau}$ to yield a utility index $v^r_{jw\tau} = \delta^r_{jw\tau} + \epsilon^r_{jw\tau}$. Prominence is assigned to the offer with the highest utility index. Let $y^r_{jw\tau} = \mathbf{1}\{v^r_{jw\tau} \geq \max_k v^r_{kw\tau}\}$ be the prominence dummy. Using McFadden (1981), this multinomial logit model can be transformed into a standard binary logit model by appropriate conditioning. Indeed, consider two alternatives j and j' that have a non-zero probability of being featured. Then,

$$\mathbb{P}\left(y^r_{jw\tau} = 1 | y^r_{jw\tau} + y^r_{j'w\tau} = 1\right) = \frac{1}{1 + \exp\left(\delta^r_{j'w\tau} - \delta^r_{jw\tau}\right)}. \quad (1)$$

But as argued by Chamberlain (1980), a fixed effect in a binary logit model can be eliminated by conditioning on an appropriate sufficient statistic. Fix two periods τ, τ' and consider the event

$C = \{y^r_{jw\tau} + y^r_{jw\tau'} = 1, y^r_{jw\tau'} + y^r_{j'w\tau'} = 1\}$. Then

$$\begin{aligned}\mathbb{P}(y^r_{jw\tau} = 1 | y^r_{jw\tau} + y^r_{j'w\tau} = 1, C) &= \frac{\mathbb{P}(y^r_{jw\tau} = 1, y^r_{jw\tau} + y^r_{j'w\tau} = 1|C)}{\mathbb{P}(y^r_{jw\tau} + y^r_{j'w\tau} = 1|C)} \\ &= \frac{1}{1 + \exp[(\delta^r_{j'w\tau} - \delta^r_{j'w\tau'}) - (\delta^r_{jw\tau} - \delta^r_{jw\tau'})]}.\end{aligned} \tag{2}$$

But neither $\delta^r_{jw\tau} - \delta^r_{jw\tau'}$ nor $\delta^r_{j'w\tau} - \delta^r_{j'w\tau'}$ contain ξ^r_{jw}. Thus, to consistently estimate β^r and α^r in the presence of ξ^r_{jw} we maximize the log likelihood function

$$\mathcal{L}(\alpha^r, \beta^r) = \sum_{w=1}^{|\mathcal{W}|} \sum_{\tau=1}^{T_w} \sum_{j=1}^{|\mathcal{J}_{w\tau}|} \sum_{j', \tau' \in Z_{jw\tau}} y^r_{jw\tau} \ln \left(\frac{1}{1 + \exp[(\delta^r_{j'w\tau} - \delta^r_{j'w\tau'}) - (\delta^r_{jw\tau} - \delta^r_{jw\tau'})]} \right). \tag{3}$$

Here, $Z_{jw\tau}$ is the set of all potential offers $j' \neq j$ and times $\tau' \neq \tau$ that satisfy $y^r_{jw\tau} + y^r_{j'w\tau} = 1$, $y^r_{jw\tau} + y^r_{jw\tau'} = 1$ and $y^r_{jw\tau'} + y^r_{j'w\tau'} = 1$. In practice, we estimate the model under the restriction $\beta^r = \mathbf{0}$ as non-price offer characteristics vary little.

2.2. *Offer FE under Partial Observability*

We observe sales for exactly one alternative j for each product page. Thus we cannot exactly follow the previous subsection: forming the required conditioning set $Z_{jw\tau}$ requires observations on at least two alternatives. However, at the cost of some power, we can exploit the high-frequency nature of our data to construct an estimator that is consistent in the presence of arbitrary fixed effects. Recall that a consumer's mean utility is given by $\delta_{jw\tau} = \mathbf{x}'_{jw\tau}\beta - \alpha p_{jw\tau} + \xi_{jw}$, where, relative to the main text, we impose Assumption 9, i.e. that offer quality $\xi_{jw} \equiv \xi_{jw\tau}$ is time-invariant. Mean utilities on each offer are combined in the usual fashion with a Type-I Extreme Value shock $\epsilon_{ijw\tau}$ to yield a utility index $v_{ijw\tau} = \delta_{jw\tau} + \epsilon_{ijw\tau}$. For now, assume all consumers are sophisticated, so each consumer simply chooses her preferred option. Let $y_{ijw\tau} = \mathbf{1}\{v_{ijw\tau} \geq \max_k v_{ikw\tau}\}$ indicate whether consumer i chooses alternative j at time τ on product page w. Then

$$\begin{aligned}\mathbb{P}(y_{ijw\tau} = 1 | y_{ijw\tau} + y_{ijw\tau'} = 1) &= \frac{\mathbb{P}(y_{ijw\tau} = 1)\mathbb{P}(y_{ijw\tau'} = 0)}{\mathbb{P}(y_{ijw\tau} = 1)\mathbb{P}(y_{ijw\tau'} = 0) + \mathbb{P}(y_{ijw\tau} = 0)\mathbb{P}(y_{ijw\tau'} = 1)} \\ &= \frac{1}{1 + \dfrac{\sum_{k\neq j} \exp(\delta_{kw\tau})}{\sum_{k \neq j} \exp(\delta_{kw\tau'})} \exp(\delta_{jw\tau'} - \delta_{jw\tau})}.\end{aligned} \tag{4}$$

In general, $\sum_{k\neq j} \exp(\delta_{kw\tau}) \neq \sum_{k\neq j} \exp(\delta_{kw\tau'})$. However, when $x_{kw\tau} = x_{kw\tau'}$ and $p_{kw\tau} = p_{kw\tau'}$ for all $k \neq j$, then equality holds. Thus, we restrict attention to pairs of periods between which all offers for which we do not observe sales remain unchanged. This procedure yields the following likelihood function:

$$\mathcal{L}(\alpha, \beta) = \sum_{w=1}^{|\mathcal{W}|} \sum_{(\tau, \tau') \in X_w} \ln \left(\frac{1}{1 + \exp(\delta_{jw\tau'} - \delta_{jw\tau})} \right), \tag{5}$$

where $X_w = \{(\tau,\tau')|\forall k \neq j : x_{kw\tau} = x_{kw\tau'}, p_{kw\tau} = p_{kw\tau'}\} \cap \{(\tau,\tau')|y_{ijw\tau} = 1, y_{ijw\tau'} = 0\}$. In practice, we estimate the model under the restriction $\beta = \mathbf{0}$ as there is very little variation in non-price offer characteristics.

3. DETAILS OF SIMULATED METHOD OF MOMENTS

In this section, we describe the technical details underlying the SMM procedure. First, the pricing game equilibrium is computed. Next, in the entry game, we smooth over expected gross profit discontinuities in entrants via importance sampling (Ackerberg, 2009). Then, we simulate markets and form moments. Thereafter, we accelerate optimization by concentrating out the wholesale cost parameters and employing DFO-LS, a derivative-free Gauss-Newton optimization method. Finally, we display measures of model fit. Where evident, we suppress market subscripts t.

3.1. *Solving the Pricing Game*

Fix a market p and entrants $\mathcal{J}$. Each entrant's type is (c_j, q_j, q_j^r), where $q_j = \mathbf{x}_j'\beta + \xi_j$ and $q_j^r = \mathbf{x}_j'\beta^r + \xi_j^r$ measure the attractiveness of the entrant to consumers and the prominence algorithm, respectively. Entrants are fully informed about each other's types and treat price as a strategic variable. Thus, the first-order condition ("BLP-markup equation") in seller j's price is

$$\phi s_j(\mathbf{p},\mathbf{q}) + [\phi p_j - c_j]\frac{\partial s_j}{\partial p_j} = 0. \tag{6}$$

An equilibrium of the pricing game comprises a vector of market prices $\mathbf{p}$ satisfying Equation (6) for all entrants $j \in \mathcal{J}$. To ease notation, suppress the dependence of all variables on $\mathbf{q}$, the matrix of buy box and demand qualities across merchants in a market. Following Morrow and Skerlos (2011), decompose the Jacobian matrix of market shares: $\frac{\partial \mathbf{s}}{\partial \mathbf{p}'} = \Lambda(\mathbf{p}) - \Gamma(\mathbf{p})$, where $\Lambda(\cdot)$ is a diagonal matrix and $\Gamma(\cdot)$ contains the factors common to both the diagonal and off-diagonal elements of the Jacobian. Then, the *ζ-markup equation* is written: $\mathbf{p} = \phi^{-1}\mathbf{c} + \zeta(\mathbf{p})$, with $\zeta(\mathbf{p}) = \Lambda(\mathbf{p})^{-1}\operatorname{diag}(\Gamma(\mathbf{p}))(\mathbf{p} - \phi^{-1}\mathbf{c}) - \Lambda(\mathbf{p})^{-1}\mathbf{s}(\mathbf{p})$, for nonsingular $\Lambda(\mathbf{p})$. We chain iterations based on the ζ-markup and BLP-markup equations to find a fixed point, which improves convergence relative to iterating on the BLP-markup equation alone (Morrow and Skerlos, 2011, 329).[4]

3.2. *Solving the Entry Game*

The information set[5] of potential entrants $j \in \mathcal{N}$ is $\mathcal{I}_j = \{c_j, F\}$. As profits are strictly decreasing in own costs, firms play cutoff entry strategies, i.e., $\chi_j = 1\{c_j \le c_j^*\}$. From the perspective of prospective entrant j, entry is profitable if and only if $\mathbb{E}_{\mathbf{q},c_{-j}}[\pi_j(c_j, q_j; \mathbf{c}_{-j}, \mathbf{q}_{-j}, \chi_{-j})] \ge F$. We focus on symmetric equilibria: $c_j^* = c^*$ for all $j \in \mathcal{N}$. This c^* satisfies a zero-profit condition at the cost cutoff:

$$V(c^*) = \mathbb{E}_{\mathbf{q},c_{-j}}[\pi_j(c^*, q_j; \mathbf{c}_{-j}, \mathbf{q}_{-j}, \chi_{-j}^*)] = F; \quad \chi_k^*(c) = 1\{c_k \le c^*\}\ \forall k \neq j. \tag{7}$$

[4] In our model, an equilibrium exists but is unlikely to be unique. With two types of consumers, the pricing game is no longer supermodular, so the equilibrium we find depends on starting values. Our estimation procedure gives the entrant with the highest adjusted prominence algorithm attractiveness, $q_j^r - (\alpha^r/R_t)c_j$, a lower starting price than the others.

[5] Henceforth, we suppress writing fixed costs F in the information set: merchants within a market t face the same fixed costs.

The LHS of this equation is a function $V(c)$ in the candidate cutoff c. We approximate $V(c)$ by an average across simulation draws, i.e., $\hat{V}(c) = \frac{1}{S}\sum_{s=1}^{S} \pi_j(c^*, q_j^s; \mathbf{c}_{-j}^s, \mathbf{q}_{-j}^s, \chi_{-j}^*)$. Applying standard root-finding techniques to Equation (7) yields c^*.

3.3. *Simulating a Market*

We combine Subsections 3.1 and 3.2 to simulate market-level outcomes. We employ the algorithm presented in Figure 1b. We begin by drawing a scalar fixed cost F. Next, we draw S vectors of wholesale costs $\mathbf{c}_s \in \mathbb{R}^{|\mathcal{N}|}$, demand qualities $\mathbf{q}_s \in \mathbb{R}^{|\mathcal{N}|}$, and algorithm qualities $\mathbf{q}_s^r \in \mathbb{R}^{|\mathcal{N}|}$. Using the algorithm in Subsection 3.2, we obtain the entry cutoff c^*. Next, referring to simulation draw $s = 1$, we find the set of successful entrants $\mathcal{J} = \{j \in \mathcal{N} : c_{js} \leq c^*\}$ and calculate their equilibrium prices, profits, and market shares using Subsection 3.1.

3.4. *Aggregating Markets to Moments*

Following Ackerberg (2009), we aggregate market-level outcomes to moments. Fix a market t. Our model postulates a relationship f between (market-level) observables[6] $x_t = (A_t, R_t)$, unobservables u_t, and outcomes y_t. In particular, at the true parameter vector θ_0 we have $y_t = f(x_t, u_t, \theta_0)$. We list the outcomes included in y_t in the Section 5.3 in the main text. If the data $\{x_t, y_t\}_{t=1}^T$ are generated by our model at the true θ_0, then

$$\theta = \theta_0 \implies \mathbb{E}[y_t - \mathbb{E}[f(x_t, u_t, \theta)|x_t]|x_t] = 0. \tag{8}$$

The reverse implication also holds as long as our model parameters are identified. As econometricians, we do not observe the true value of the unobservables u_t. In our model, these shocks include (i) fixed costs F_t for each market and (ii) qualities (q_{jt}, q_{jt}^r) as well as wholesale unit costs c_{jt} for each potential entrant on each market. To proceed, we make parametric assumptions on the distributions of these quantities: we assume that (q_{jt}, q_{jt}^r) are drawn from the empirical distribution implied by our prominence algorithm and consumer choice estimation.

Wholesale and fixed costs are drawn following Assumption 10 in the main text. Given some candidate $\theta^s = (\theta_0^F, \theta_x^F, \theta_\sigma^F, \theta_0^c, \theta_x^c, \theta_\sigma^c)$, the conditional distribution of unobservables $p(u_t|x_t, \theta)$ is thus fully specified. In theory, we could use our model and the moment condition(s) in (8) to estimate θ. However, in practice, these expectations are hard to compute: they involve not one but two layers of games for which equilibria must be numerically computed (the entry game and the pricing game). Instead, we take simulation draws $(u_{t1}, \ldots, u_{tS}) \sim p(u_t|x_t, \theta)$ and approximate the expectation by averaging: $\widehat{\mathbb{E}f_t(\theta)} = \frac{1}{S}\sum_s f(x_t, u_{ts}, \theta)$. Estimation proceeds based on the simulated method of moments estimator that sets the simulated moment vector $\hat{G}(\theta)$ close to zero, as in Equation (15) in the main text (McFadden, 1989, Pakes and Pollard, 1989).

3.5. *Concentrating Out Wholesale Cost Parameters*

We partition $\theta = (\theta^F, \theta^c)$ where $\theta^F = (\theta_0^F, \theta_x^F, \theta_\sigma^F)$ and $\theta^c = (\theta_0^c, \theta_x^c, \theta_\sigma^c)$. Define $\tilde{\theta}^F(\theta^c) = \arg\min_{\theta^F} Q_n((\theta^F, \theta^c))$ and $\tilde{\theta}^c = \arg\min_{\theta^c} Q_n((\theta^F(\theta^c), \theta^c))$. Then, $\hat{\theta} = (\tilde{\theta}^F(\tilde{\theta}^c), \tilde{\theta}^c)$. Hence, we "concentrate out" θ^F when searching over θ^c. We use the algorithm presented in Figure 1a.[7]

[6] Here, A refers to market size and R to the manufacturer's suggested retail price.

[7] The intuition is similar to why the linear parameters are concentrated out when estimating mixed logits: estimation is computationally trivial once this step is performed (Berry et al., 1995). For us, $\tilde{\theta}^F(\theta^c)$ can be found without re-solving the pricing game.

```
for p = 1,...,P do
    for s = 1,...,S_F do
        simulate entry for (p,s)
θ^F = θ^F_0
while dist > tol do
    reweight sim data for θ^F
    dist = Ĝ(θ^c, θ^F)
    update θ^F
return Ĝ(θ^c, θ^F)
```

(a) SMM Moments $\hat{G}(\theta^c)$.

```
Data: S_E sets N_{s_e} of pot entrants
for c̃ ∈ grid do
    fix c_1 = c̃
    for s_e = 1,...,S_E do
        J = {1} ∪ {j ∈ N_{s_e} : c_j < c̃, j > 1}
        find π_1(J) using fix-point iter.
    V(c̃) ≈ (1/S_E) Σ_{s_e} π_1(·)
c* solves V(c*) = F
J = {j ∈ N_1 : c_j < c*}
solve pricing game given J
```

(b) Simulation of Entry.

FIGURE 1.—Algorithms Used in SMM Estimation. We provide the two key algorithms used in the SMM estimation procedure. In the left panel, we detail how to find the value of the *outer* SMM objective. In the right panel, we describe entry simulation.

3.6. *Importance Sampling*

To smooth over discontinuities in the entry process, we employ importance sampling Ackerberg (2009).[8] Because θ^c is fixed, we suppress mention of it in our notation. Then, write outcomes directly as a function of fixed cost draws rather than as a function of shocks: $f(x_t, u_t, \theta) = \tilde{f}(F_t)$.

We pursue the following simulation strategy: draw $F_{ts} \sim g(\cdot)$ where $g(\cdot)$ is a heavy-tailed density. Then use our knowledge of the distribution of fixed costs implied by our current parameter guess to re-weight the outcomes at the simulated fixed cost draws. Formally speaking, we compute $\widetilde{\mathbb{E}f_t}(\theta) = \sum_s \tilde{f}(F_{ts}) \frac{p(F_{ts}|x_t,\theta)}{g(F_{ts}|x_t)}$.[9] To avoid additional simulation error in our estimates, we iterate the importance sampling procedure eight times: at each new iteration, we let $g(\cdot|x_t)$ equal the density $p(\cdot|x_t, \theta^F_{prev})$ at the optimal value θ^F_{prev} of the previous iteration.[10]

4. MONTE CARLO STUDIES

To evaluate the robustness of our estimation procedure to violations of our identifying assumptions, we perform Monte Carlo analyses of demand and supply.

4.1. *Monte Carlo: Prominence*

Regarding prominence, the key identification concern is the possibility of correlation between unobserved prominence advantage ξ^r_{jt} and prices. Our Monte Carlo analysis is designed to evaluate the robustness of our estimator to this concern. In Table II, we take covariates x_{jt} from our data, fix the coefficients at those estimated in our mainline analysis, and simulate the prominence outcome by combining the implied mean utilities with correlated extreme value shocks and unobserved $\xi^r_{jw\tau}$ under three assumptions:

1. Panel A: $\xi^r_{jw\tau} \equiv 0$, i.e., there is no unobserved prominence quality. Then we recover the true coefficients.

[8]Finding fixed cost parameters $\tilde{\theta}^F(\theta^c)$ requires a continuous objective. However, entry is a discrete process: for instance, as we decrease F, there is only one entrant until some point at which this number "jumps" to two.

[9]This simulator has two advantages. Firstly, the density for fixed costs we specified is smooth, ensuring that the simulated outcomes will smoothly depend on θ^F. Thus, for instance, as we increase the mean of the fixed cost distribution, the estimator will smoothly put less and less weight on low fixed cost draws. The second advantage is that we only need to compute the market-level outcomes $\tilde{f}(F_{ts})$ once at the beginning of the procedure. As we vary θ^F, the outcomes employed do not vary; they are simply reweighted.

[10]As in Ackerberg (2009), this iteration converges quickly (typically within three steps).

TABLE II

MONTE CARLO OF PROMINENCE ALGORITHM ESTIMATION.

	Truth	Panel A		Panel B		Panel C	
	(0)	(1)	(2)	(3)	(4)	(5)	(6)
Price / MSRP	-24.22	-24.21 (0.048)	-22.84 (0.731)	-14.25 (0.068)	-23.19 (0.700)	-14.07 (0.029)	-13.34 (0.414)
Time to Ship (in Days)	-1.11	-1.10 (0.006)		-0.96 (0.011)		-1.11 (0.007)	
100 $\times$ Log(Feedback Count)	4.33	4.48 (0.066)		2.72 (0.113)		4.49 (0.065)	
FBA	2.56	2.56 (0.013)		2.28 (0.024)		2.54 (0.014)	
Amazon	1.11	1.10 (0.008)		0.80 (0.017)		1.10 (0.008)	
Outside	4.48	4.48 (0.012)		3.57 (0.011)		3.40 (0.010)	
λ	0.09	0.09 (0.001)		0.10 (0.001)		0.09 (0.001)	
Offer FE?		✗	✓	✗	✓	✗	✓

Notes: This table reports on a Monte Carlo exercise testing the prominence algorithm estimation. We take covariates x_{jt} from the data, fix the coefficients as in (0), and simulate the prominence outcome by combining the implied mean utilities with correlated extreme value shocks and unobserved $\xi^r_{jw\tau}$ under three assumptions. Panel A assumes that there is no demand shock, i.e., $\xi^r_{jwt} \equiv 0$. Panel B assumes that ξ^r_{jwt} is proportional to an offer's average price divided by MSRP. Panel C assumes that ξ^r_{jwt} is proportional to an offer's *current* price divided by MSRP, thus also violating the identifying assumption required for the offer-fixed effects estimator.

2. Panel B: $\xi^r_{jw\tau} \equiv \xi^r_{jw}$, i.e., there is unobserved prominence quality that is time-invariant but endogenous with respect to *average* prices.[11] We recover biased coefficients in the naive estimator but the offer-fixed effects estimator still recovers the true coefficients. Furthermore, unlike in our main analysis, these two estimators recover *different* coefficients, suggesting time-invariant price endogeneity is not a concern in the actual (non-simulated) data.
3. Panel C: $\xi^r_{jw\tau} = 0.4 \times p^z_{jw\tau}$, i.e., prominence quality is endogenous with respect to current prices. Then, both estimators are biased.

We conclude that time-invariant price endogeneity is not a concern (as it could be detected) but our estimators are not robust to time-varying price endogeneity.

4.2. *Monte Carlo: Demand Warmup*

We now explain in a simpler model why the strategy of conditioning on a noisy proxy for the prominence dummy is valid; the notation in this section is separate from the main text. Consider the probit model $y_i = 1\{\alpha + \beta x_i + \epsilon_i > 0\}$, where $x_i \in \{0, 1\}$ is unobserved, $\epsilon_i \sim N(0, 1)$ and (α, β) are scalar parameters. Suppose further we have a noisy proxy p_i such that $x_i | p_i, u_i \sim$ Bernoulli$(p_i + u_i)$ where $u_i \sim F(\cdot)$ is a mean-zero unobservable. As x_i is binary, $F(\cdot)$ turns out to be irrelevant as for any $F(\cdot)$ we have $x_i | p_i \sim$ Bernoulli(p_i). Hence, we

[11]Concretely, let $p^z_{jw\tau}$ be the z-score associated with the offer price divided by MSRP. Then if $\mathcal{T}_w$ is the set of times associated with product web-page w, we set $\xi^r_{jw} = 0.4 \times |\mathcal{T}_w|^{-1} \sum_{\tau \in \mathcal{T}_w} p^z_{jw\tau}$.

TABLE III

MONTE CARLO OF PROBIT WITH UNOBSERVED REGRESSOR (TRUE: 2.0).

	$u_i \equiv 0$	$u_i \sim N(0,1)$	$u_i = \epsilon_i$
Infeasible	2.00 (0.00)	2.00 (0.00)	3.71 (0.00)
Plug-In	1.71 (0.03)	1.71 (0.03)	1.73 (0.03)
Proxy	2.00 (0.06)	1.99 (0.06)	2.03 (0.06)

Notes: These Monte Carlo results compare the proposed proxy-based estimator (which integrates out the uncertainty over the unobserved regressor) with two alternative estimators: the infeasible estimator (which directly uses the unobserved binary x_i) and a plug-in estimator (where p_i is used in place of x_i in the infeasible estimator's likelihood). While the proxy estimator is unbiased, both other estimators are biased under some circumstances.

can estimate this model via maximum likelihood conditional on p_i, i.e.,

$$\begin{aligned} \ell(\alpha,\beta) = \sum_{i=1}^{n} \log \Big[& y_i \big(p_i \ell_1(\alpha,\beta) + (1-p_i)\ell_0(\alpha,\beta)\big) \\ & + (1-y_i)\big(p_i(1-\ell_1(\alpha,\beta)) + (1-p_i)(1-\ell_0(\alpha,\beta))\big)\Big] \\ \text{where} \quad & \ell_1(\alpha,\beta) = \Phi(\alpha+\beta), \quad \text{and} \quad \ell_0(\alpha,\beta) = \Phi(\alpha). \end{aligned} \tag{9}$$

As confirmed by Monte Carlo simulations in Table III ('Proxy' row), this strategy results in a consistent estimator of (α, β) as long as (i) cov$(p_i, \epsilon_i) = 0$, (ii) var$(p_i) > 0$, (iii) p_i is well-calibrated (i.e., $\mathbb{E}[x_i|p_i] = p_i$), and (iv) cov$(\epsilon_i, u_i) = 0$.[12]

In our context, y_i is demand, x_i is a prominence dummy, p_i is the predicted prominence probability, ϵ_i is a demand shock and u_i is a prominence shock. When demand and prominence shocks are uncorrelated, we recover all parameters correctly. When they are correlated, the nonlinearity of the probit means that individual parameters (α, β) are generally biased. However, the response of $\mathbb{E}[y_i|p_i]$ to p_i remains correctly identified. In the demand model, ρ enters linearly as a mixture weight in Equation (3) of the main text and is identified from this response, so $\hat{\rho}$ remains consistent even under correlated shocks—as we confirm in Panel E of Table IV below.

4.3. *Monte Carlo: Demand*

On the demand side, the key identification concerns are the potential endogeneity of price, as well as correlation between unobserved prominence advantage $\xi^r_{jw\tau}$ and unobserved determinants of demand $\xi_{jw\tau}$. To operationalize these concerns in a Monte Carlo analysis, we combine the recovered coefficients from our demand estimates with the covariates $x_{jw\tau}$ in our data and simulated correlated extreme value shocks to generate sales under various assumptions about demand shocks $\xi_{jw\tau}$. The results from six Monte Carlo scenarios are displayed in Table IV:

Regarding endogeneity of price:

- Panel A: No demand shock, i.e., $\xi_{jw\tau} = 0$. Then both our estimator and the one using offer fixed effects recover the true price sensitivity. Note that, as expected, the fixed effect estimator does not yield the same price coefficient as reported in (0) – but this is because the estimator, which cannot distinguish between substitution between inside options and

[12]Just like (i) and (ii) are reminiscent of exclusion and relevance in instrumental variables, (iii) can be thought of as a requirement on the first-stage in an IV strategy. In a linear model, (i)–(iii) alone would suffice, as in standard IV. Condition (iv) is needed because the probit is nonlinear.

substitution to the outside option, is mixing these two sources of substitution. Indeed, as expected, the estimated price sensitivity of -1.31 in (2) lies between the *inside* price sensitivity of -2.95 reported in (1) and the implied inside-vs-outside price sensitivity in that column (which is $-2.95 \times 0.06 \approx -0.19$).

- Panel B: Price endogeneity, constant over time. Here, unobserved determinants of demand $\xi_{jw\tau}$ are proportional to an offer's *average* price divided by MSRP. Then our standard estimator is biased but the method using offer fixed effects recovers the true price sensitivity.
- Panel C: Time-varying price endogeneity. Here, $\xi_{jw\tau}$ is assumed proportional to an offer's *current* price divided by MSRP. Because identifying conditions for both estimators are violated, neither recovers the true price sensitivity.

Note all of these estimators recover the correct fraction of sophisticated consumers.

Regarding prominence endogeneity & proxying:

- Panel D: There are prominence-relevant characteristics uncorrelated with price and demand shocks but which we do not observe, i.e., $\xi^r_{jw\tau} \sim N(0, 0.4^2)$. By analogy to integrating out known measurement error in a continuous variable, one may be concerned this could introduce bias for a binary variable like the prominence dummy, too.[13] However, the Monte Carlo results confirm: with a well-calibrated prominence proxy, the estimates remain unbiased (though their variance increases as the proxy becomes more imprecise). We establish in Section 5, Figure 2(a) that our prominence proxy in the main analysis is well-calibrated.
- Panel E: Unobserved demand quality is correlated with unobserved prominence advantage. Here, we assume $\xi_{jw\tau} \equiv \xi^r_{jw\tau} \sim N(0, 0.4^2)$. The estimate of $\hat{\rho}$ remains unbiased, which seems reasonable because the estimation strategy purges endogenous variation in prominence status by projecting it only on knowns (as with an instrumental variables strategy).
- Panel F: Similar to Panel A, but we replace the eligibility 'instrument' with the short run correlation between prominence and sales. The mismeasurement of prominence leads to an upward bias in the share of sophisticated consumers.

To summarize, we find that our estimate of ρ is robust to prominence endogeneity, and our estimates would detect price endogeneity that is constant over time. However, they are biased if price endogeneity varies over time, and would be biased if we used the *observed*, rather than the *predicted*, prominence dummy (which we do not).

4.4. *Monte Carlo: Supply*

We also verify our supply model by Monte Carlo simulations. Our exercise achieves two goals. First, under our maintained assumptions, we can recover the correct fixed and both third- and first-party wholesale cost parameters. Second, if costs are misspecified because of unmodelled correlation between fixed and wholesale costs, then, as expected, both cost estimates are biased.

Results are displayed in Table V. Columns marked A fix all parameters at their true values, except for the mean fixed cost parameter μ_F and the mean wholesale cost parameters $(\mu_c, \mu_{c,\text{Amz}})$. Columns marked B allow all parameters to vary. The columns (A1) and (B1) validate that we can recover the fixed and wholesale parameters we impose. In columns (A2) and (B2), we estimate the model under misspecified wholesale costs.

[13] As discussed in Section 4.2, this concern is misleading because a binary variable's distribution is fully determined by its mean.

TABLE IV

MONTE CARLO OF DEMAND ESTIMATION.

	Truth (0)	Panel A (1)	Panel A (2)	Panel B (3)	Panel B (4)	Panel C (5)	Panel C (6)	Panel D (7)	Panel E (8)	Panel F (9)
Price / MSRP	-3.00	-2.95 (0.164)	-1.31 (0.534)	8.19 (0.462)	-1.38 (0.543)	10.00 (0.253)	-0.08 (0.398)	-3.62 (0.538)	-0.89 (0.625)	-5.15 (0.116)
Dispatch Time (in Days)	-0.15	-0.13 (0.065)		-0.06 (0.074)		-0.06 (0.061)		0.03 (0.169)	0.22 (0.141)	-0.26 (0.032)
100 $\times$ Log(#Feedback+1)	-4.17	-5.40 (1.027)		-9.01 (1.855)		-5.55 (1.479)		-11.78 (5.903)	-11.42 (7.013)	-0.87 (0.480)
FBA	1.23	1.15 (0.153)		0.58 (0.206)		0.66 (0.149)		2.03 (0.408)	0.73 (0.656)	0.83 (0.074)
Amazon	1.62	1.71 (0.143)		-0.02 (1.071)		1.69 (0.503)		1.89 (0.654)	1.24 (0.910)	0.52 (0.075)
Outside	-3.61	-3.57 (0.018)		-4.34 (0.040)		-4.57 (0.024)		-3.60 (0.046)	-3.19 (0.146)	-3.54 (0.015)
Nesting Coefficient (λ)	0.05	0.06 (0.006)		0.06 (0.006)		0.07 (0.003)		0.04 (0.012)	0.18 (0.152)	0.04 (0.004)
Sophisticates Fraction (ρ)	0.06	0.06 (0.001)		0.06 (0.004)		0.06 (0.002)		0.06 (0.005)	0.07 (0.016)	0.26 (0.003)
Offer FE?		✗	✓	✗	✓	✗	✓	✗	✗	✗

Notes: This table provides results from a Monte Carlo exercise that tests our demand estimation. We sample covariates X from our real data and use the coefficients in (0) to simulate sales under various assumptions about demand shocks ξ_{jwt}. All panels assume a 75% probability that prominence assignments changed since the last observation. Panel A assumes no demand shock, i.e., $\xi_{jwt} \equiv 0$. Under this assumption, both the naive and offer fixed-effects estimators recover the true price sensitivity (i.e., coefficient on Price/MSRP). Panel B assumes that ξ_{jwt} is proportional to an offer's *average* price divided by MSRP. The naive estimator is biased, but we recover the true coefficients once we include offer fixed-effects. Panel C assumes that ξ_{jwt} is proportional to an offer's *current* price divided by MSRP, thus violating the identifying assumption required for the offer-fixed effects estimator; indeed, our estimates are now biased. Panel D assumes var$(\xi^r_{jwt}) > 0$, which does not bias estimates of ρ. Panel E assumes that ξ_{jwt} is exogenous with respect to price, but $\xi_{jwt} \equiv \xi^r_{jwt}$; this prominence endogeneity does not bias $\hat{\rho}$. Panel F is like Panel A but replaces the eligibility instrument with the short-run correlation between prominence and sales; the mismeasurement of prominence leads to an upward bias in the fraction of sophisticates.

For exposition, consider a product t and a seller j. To simulate fixed and wholesale costs, we first draw fixed cost "shocks" $\epsilon_{F,t}$ from a standard normal. Next, we draw wholesale cost "shocks" following $\epsilon_{c,jt}|\epsilon_{F,t} \sim \Sigma\epsilon_{F,t} + \sqrt{1-\Sigma^2}\, N(0,1)$, for some value of Σ in $[0,1]$. The value of Σ is set to zero for columns marked 1, and 0.5 for columns marked 2. Intuitively, a more positive Σ implies wholesale cost shocks are higher on products with higher fixed costs (i.e., fewer entrants).

In columns (A1) and (B1), we find that our estimator delivers excellent point estimates when the model is correctly specified, and the associated confidence intervals generally cover the truth. However, for the misspecified model, if we fix all parameters other than $(\mu_F, \mu_c, \mu_{c,\text{Amz}})$ at their true values, we see in column (A2) that we obtain highly biased estimates for wholesale costs, while mean fixed costs remain well-estimated. If we permit all parameters to vary, as in column (B2), both fixed and wholesale cost parameters are biased away from the truth.

5. OUT-OF-SAMPLE FIT

We display out-of-sample fit for the Buybox, demand, and supply respectively. In each case, we perform an 80-20 train-test split: we train the model on 80% of the data, then test it on the remaining 20%. Our models fit the held-out data well.

5.1. *Out-of-Sample Fit: Prominence and Demand*

We analyze the fit of the demand and prominence models by comparing their out-of-sample predictions to the data. To this end, we first estimate the models on a sample of 80% of products.

TABLE V

MONTE CARLO OF SUPPLY ESTIMATION.

	(A)			(B)		
	Truth	(A1)	(A2)	Truth	(B1)	(B2)
μ_F	4.647	4.654 (0.013)	4.646 (0.012)	4.647	4.660 (0.016)	4.725 (0.023)
$\mu_{F,M}$				1.921	1.919 (0.025)	1.897 (0.037)
σ_F				-0.049	-0.016 (0.030)	0.255 (0.021)
δ_c				0.777	0.772 (0.005)	0.780 (0.003)
σ_c				0.156	0.157 (0.001)	0.146 (0.001)
μ_c	2.258	2.271 (0.012)	2.907 (0.010)	2.258	2.385 (0.045)	2.301 (0.007)
$\delta_{c,AMZ}$				0.474	0.464 (0.008)	0.449 (0.005)
$\sigma_{c,AMZ}$				0.117	0.151 (0.019)	0.108 (0.006)
$\mu_{c,AMZ}$	3.487	3.487 (0.019)	3.715 (0.014)	3.487	3.404 (0.060)	4.028 (0.023)

Notes: This table provides results from a Monte Carlo exercise that tests our code which estimates fixed and wholesale cost parameters. We generate a dataset by simulating markets from 47,249 products in our real data. Using upstream demand parameters, the coefficients reported in (0), as well as simulated correlated shocks for fixed and wholesale cost parameters, we generate the dependent variable under different assumptions about the correlation between fixed and wholesale costs. First, fixed cost "shocks" $\epsilon_{F,t}$ are drawn from a standard normal, then wholesale cost "shocks" are drawn $\epsilon_{c,jt}|\epsilon_{F,t} \sim \Sigma\epsilon_{F,t} + \sqrt{1-\Sigma^2}\,N(0,1)$, for some value of Σ in $[0,1]$. These shocks are then transformed into fixed costs via $F_t = (\mu_F + \mu_{F,M} z_{\text{MSRP},t}) \exp(\sigma_F \epsilon_{F,t})$, where $z_{\text{MSRP},t} = \ln \text{MSRP}_t - T^{-1} \sum_{t=1}^{T} \ln \text{MSRP}_t$. Wholesale costs are obtained via $c_{jt} = (\mu_c + \delta_c \times \text{MSRP}_t) \times \exp(\sigma_c \epsilon_{c,jt} - \sigma_c^2/2)$, and analogously for first-party offers. Panel A fixes all parameters except the displayed parameters at their true values; Panel B allows all parameters to vary. Specifications A1 and B1 use data simulated under the assumption of independent wholesale and fixed costs; the resulting estimates are close to the true parameters. Specifications A2 and B2 simulate correlated wholesale and fixed costs; as expected, the estimates that our now misspecified estimator yields in this case are biased.

Then, for each product in the remaining 20%, we predict (i) each offer's probability of being promoted, and (ii) each observed offer's sales.[14] We then compare these predictions to the data.[15]

Our results are Figures 2 (for prominence) and 3 (for sales). Both figures first show a Hosmer-Lemeshow test that speaks to the calibration of the models, binning observations by their predicted value of the outcome variable and comparing the distribution of the outcome variable to the distribution of the predicted values. We find that these predictions lie close to the 45-degree line, indicating that both models are well-calibrated, though the demand model slightly under-predicts sales of the lowest-selling offers. The excellent calibration of the Buybox model reassures us, given that, as discussed in Appendix 4.3, the identification of the fraction of sophisticates in demand relies on the calibration of the Buybox model.

The second panel shows the results of regressing (without any additional controls) the predicted and actual outcomes on binary or binarized offer characteristics (with the exception of "In Buybox?" which uses predicted prominence given that the observed prominence assignment is mismeasured). We find that the models are able to capture the relationship between the outcome variables and the offer characteristics well. We note in particular that the impact of being in the presence of an Amazon offer is matched well by the models.

[14] Note that predicted sales are a function of prominence assignments at the time of sale, which we do not observe; we use the posterior probability of (not) being promoted given our model as the weight on the predicted sales conditional on (not) being promoted.

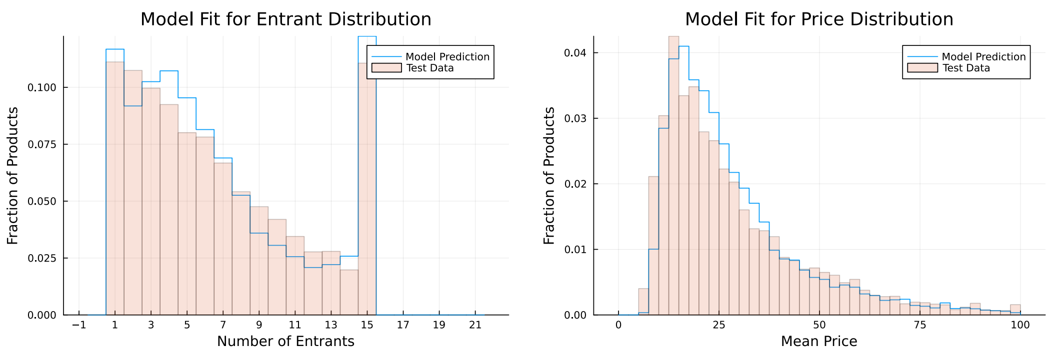
[15] While we could assess fit on the moments that our maximum-likelihood estimation implicitly targets, we could only miss these moments if the model overfits the data. We confirm in unreported tests that this is not the case, i.e., the fit on these moments is good. We focus here instead on more substantive measures of fit.

Finally, the remaining panels assess whether our models correctly capture the causal relationship between offer characteristics and prominence/sales. To this end, we regress both model predictions and data on offer characteristics, controlling for offer fixed effects. If (i) the model is correctly specified and (ii) temporal variation in offer characteristics is uncorrelated with the error term, then the coefficient on the offer characteristics in both regressions should be approximately equal. Furthermore, this exercise verifies that both the data and the model capture intuitive patterns (e.g., that sales rapidly decline as price increases beyond the minimum offer price but fall more slowly as price increases further). The models capture the relationship between the outcome variables and the offer characteristics well. In particular, because of our nested-logit structure, we match well how the number of merchants on a market predicts the probability of being promoted.

5.2. *Out-of-Sample Fit: Supply*

We analyze the fit of the supply model by comparing its out-of-sample predictions of the price and entrant distribution to those in the data.

As for the demand model, we first estimate the supply model on a "training set" comprising 80% of the products. Then, for each product in the remaining 20%, we simulate entry and pricing under the estimated parameters. Finally, we compare the simulated number of entrants and prices to the data. Figure 4 shows that the model is able to replicate the empirical distribution of entrants and prices.



(a) Entrant Distribution

(b) Price Distribution

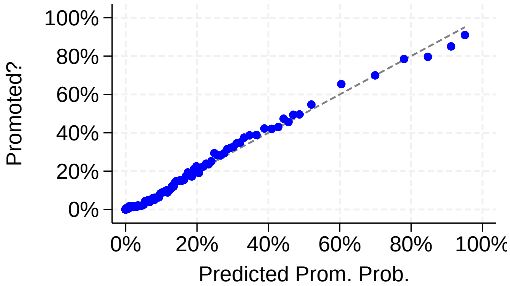
FIGURE 4.—Out of Sample Fit for SMM Model. This figure compares out-of-sample predictions of our entry model to the data. Panel (a) shows the model predicted entrant distribution as an outline overlaid on the entrant distribution in the test data. Panel (b) shows the analogous price distributions for the test data. The entry model is fitted on an 80% random sample of products; this model is then used to simulate market outcomes on the remaining 20% of the data held out from the training set.

6. HETEROGENEITY ANALYSIS

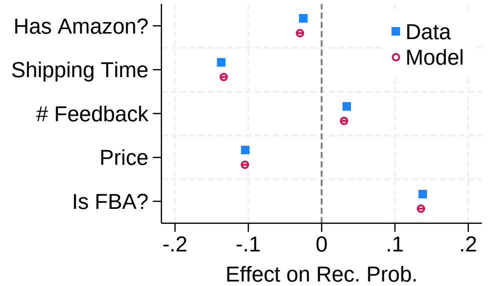
To evaluate the heterogeneity of our results across various partitions of our data, we estimate our model separately for each product category.

6.1. *Heterogeneity: Prominence and Demand*

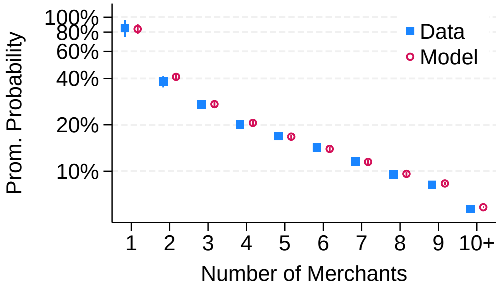
In Table VI, we investigate whether the key parameters of our demand and prominence models vary by (i) product category, or (ii) product characteristics. To this end, we re-estimate



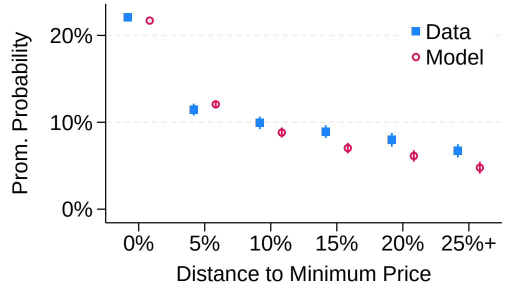
(a) Hosmer-Lemeshow



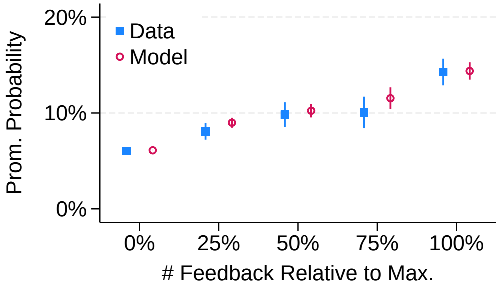
(b) Binarized Offer Characteristics



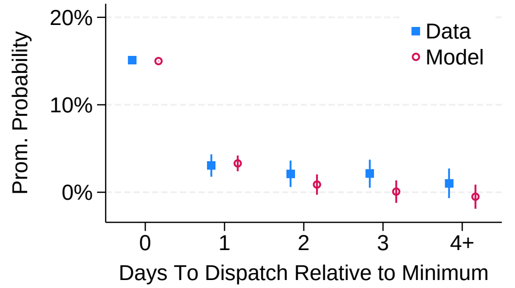
(c) Number of Merchants



(d) Price Relative to Minimum

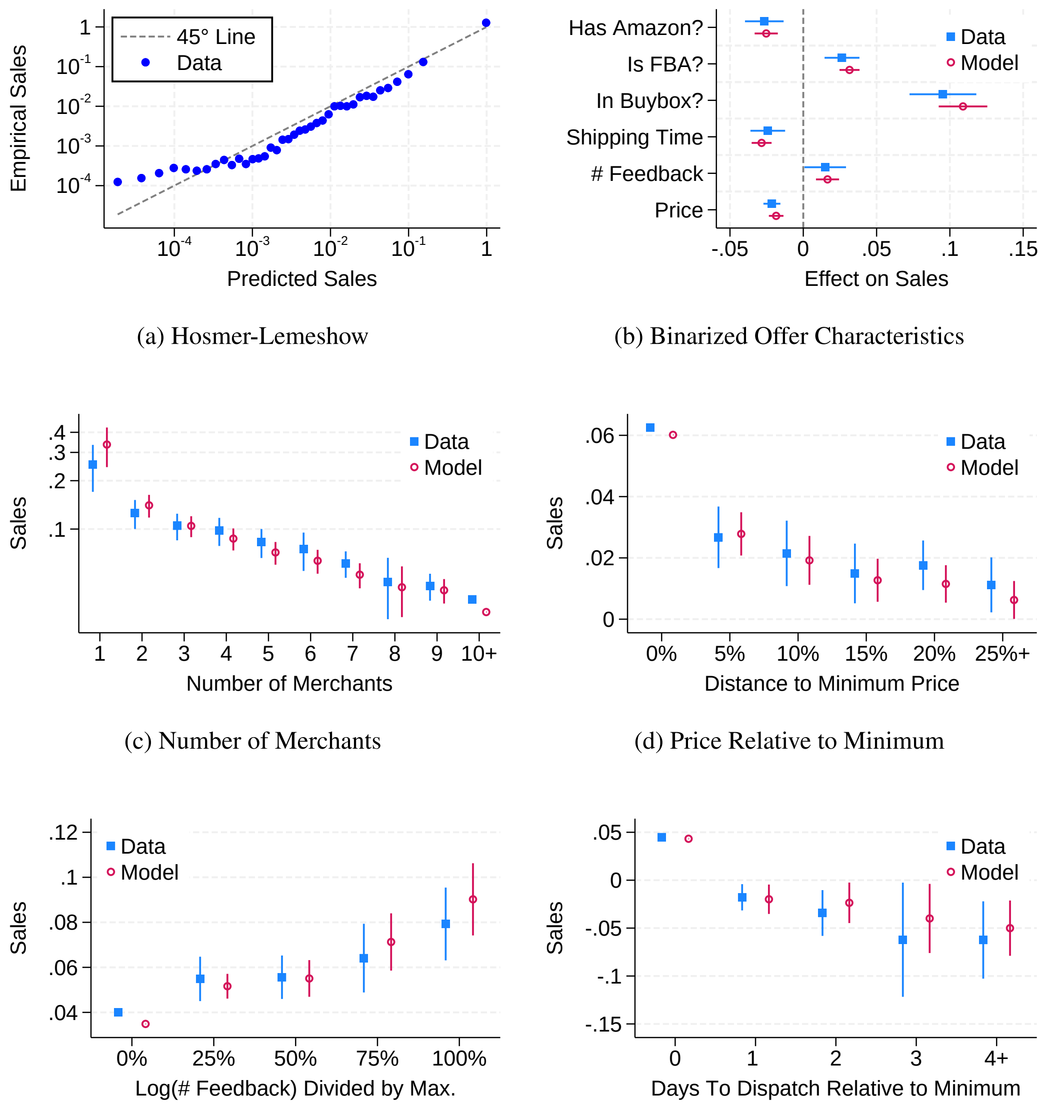


(e) Log(# Feedback) Rel. to Max.



(f) Days To Dispatch Rel. to Min.

FIGURE 2.—Out of Sample Fit for Prominence Model. This figure compares the prominence model's out-of-sample predictions to the data. Panel (a) shows a Hosmer-Lemeshow test, binning observations by predicted probability of being promoted and comparing to actual prominence observations. Panel (b) shows how well the model captures the relationship between prominence assignment (i.e., Buybox wins) and binary offer characteristics like FBA status; continuous characteristics are binarized by comparing to the median value. The underlying regression includes just the dummy and no other controls. Panels (c)-(f) show results from regressing the observed prominence dummy and predicted probabilities on various market- or observed offer characteristics while controlling for offer- fixed effects. Panel (c) examines fit across markets with different numbers of competing merchants. Panel (d) evaluates how well the model captures the relationship between an offer's relative price position and its likelihood of being promoted. Panels (e) and (f) repeat this exercise for feedback count and time to dispatch respectively. All standard errors are clustered at the product level.



(a) Hosmer-Lemeshow

(b) Binarized Offer Characteristics

(c) Number of Merchants

(d) Price Relative to Minimum

(e) Log(# Feedback) Rel. to Max.

(f) Days To Dispatch Rel. to Min.

FIGURE 3.—Out of Sample Fit for Demand Model. This figure compares out-of-sample predictions of our demand model to the data. Panel (a) shows a Hosmer-Lemeshow test, binning observations by predicted sales and comparing to actual sales. Panel (b) shows how well the model captures the relationship between sales and binary offer characteristics like FBA status; continuous characteristics are binarized by comparing to the median value, except for "In Buybox?" which uses the predicted prominence given that the observed prominence is mismeasured. The underlying regression includes just the dummy and no other controls. Panels (c)-(f) show results from regressing actual sales and sales predictions on various market- or observed offer characteristics while controlling for offer fixed effects. Panel (c) examines fit across markets with different numbers of competing merchants. Panel (d) evaluates how well the model captures the relationship between an offer's relative price position and its sales. Panels (e) and (f) repeat this exercise for feedback count and time to dispatch respectively. All standard errors are clustered at the product level.

prominence and demand models separately for each subset of the data defined by a heterogeneity dimension (first column) taking on a particular value (second column).[16] Instead of reporting estimates directly, we report the following key implied parameters: Amazon's prominence advantage (as multiple of the FBA advantage) in Column (1), the price elasticity of prominence in Column (2), Amazon's price advantage in demand (as multiple of the FBA advantage) in Column (3), the price elasticity of demand in Column (4), and the percentage of sophisticated consumers in Column (5).

To stabilize the estimates, we fix the nesting parameter λ and the impact of all characteristics other than price and being sold by Amazon at their full-sample estimates as these parameters are poorly identified in smaller samples. To obtain standard errors, we perform a parametric bootstrap.

We find some evidence of heterogeneity across product categories in both demand and prominence. In particular, Amazon's demand advantage is higher for products in the Pet, Baby and Office categories, and lower for products in the Fashion, Sports and Books categories. Furthermore, there is a tendency for categories with high demand advantage to have higher prominence advantages, though the relationship is noisy. Overall, the pattern (high Amazon advantages for goods needed quickly) suggests that Amazon offers arrive at the customer's doorstep faster in a way that may not be captured by our dispatch time variable, perhaps because Amazon offers qualify for Prime at a higher rate, or because Amazon's offers spend less time in transit even conditional on Prime status.

Moving on to price elasticities, both prominence and demand are more elastic for Food products, and less elastic for Office products, which makes sense if Office products are mostly purchased by people spending their employer's money.

Finally, regarding the fraction of sophisticated consumers, we find that consumers are more likely to explore alternative offers especially when buying books and health or beauty products. They are least likely to explore when buying tools, sporting goods or office products. This could be driven by heterogeneity in willingness to pay for faster shipping for books (and unwillingness to spend one's time to help one's employer save money in the case of office products).

Moving on to binary product characteristics, we find that Amazon's demand advantage is higher for products with a lower MSRP – plausible, as this advantage is implicitly expressed as a percentage of MSRP, and that percentage would be a larger absolute amount for higher-priced products. The biggest difference when it comes to the fraction of sophisticated consumers is that products with many offers attract more sophisticated consumers – and these products also feature a much elevated prominence price elasticity. Again, these findings are reasonable: there is a higher payoff from examining all offers when there are more offers to examine, and from Amazon's perspective, there is no need to worry about incentivizing entry of additional merchants when there are already many offers.

To conclude, we emphasize that this heterogeneity analysis should be interpreted with caution due to the convenience sample of products that underlies our analysis. While we have good coverage of different kinds of sellers across categories, once we zoom in on a particular category or heterogeneity dimension, we are left with only a small number of products, limiting power. In particular, the underlying estimates of the price/MSRP coefficient in demand, α, are not always statistically distinguishable from zero. Furthermore, after conditioning on a heterogeneity dimension, the remaining products are not necessarily representative of the population of products in that category. Therefore, the results in this section and the next should be interpreted with caution.

[16]The predicted prominence probabilities required to estimate our demand model are also re-estimated for each subset.

TABLE VI

HETEROGENEITY OF PROMINENCE AND DEMAND MODELS

Dimension	Value	Prominence (1)	Prominence (2)	Demand (3)	Demand (4)	Demand (5)
Category	Baby	0.32 (0.21)	-22.90 (3.86)	2.84 (0.45)	-13.43 (2.21)	7.89 (3.01)
Category	Beauty	0.15 (0.04)	-18.61 (0.68)	1.91 (0.17)	-14.06 (0.68)	23.88 (1.67)
Category	Books	0.42 (0.05)	-21.92 (1.04)	0.56 (0.72)	-10.18 (0.71)	26.57 (6.71)
Category	Fashion	0.59 (0.11)	-24.44 (1.32)	0.99 (0.70)	-17.69 (0.72)	4.88 (0.74)
Category	Food	0.29 (0.09)	-31.18 (1.36)	2.19 (0.23)	-16.56 (0.77)	14.99 (2.98)
Category	Health	0.35 (0.09)	-21.00 (1.16)	1.49 (0.28)	-15.27 (0.93)	22.74 (3.39)
Category	Home	0.17 (0.04)	-19.26 (1.17)	1.92 (0.22)	-13.33 (0.62)	6.82 (0.95)
Category	Office	0.84 (0.15)	-12.51 (1.41)	2.45 (0.20)	-9.80 (0.99)	8.87 (1.66)
Category	Pet	1.07 (0.41)	-14.11 (2.19)	2.95 (0.45)	-10.49 (1.22)	10.13 (1.56)
Category	Sports	0.27 (0.05)	-25.31 (1.54)	0.81 (0.45)	-18.08 (0.98)	3.28 (0.54)
Category	Tools	0.48 (0.14)	-18.23 (1.30)	1.35 (0.36)	-13.25 (0.84)	4.89 (0.98)
Category	Toys	0.95 (0.13)	-24.29 (0.89)	2.51 (0.16)	-12.64 (0.49)	11.08 (0.92)
Category	Video Games	0.45 (0.48)	-23.88 (2.62)	2.32 (0.47)	-12.03 (0.75)	10.27 (2.19)

Panel A: Heterogeneity by Category.

Dimension	Value	Prominence (1)	Prominence (2)	Demand (3)	Demand (4)	Demand (5)
Durable	No	0.33 (0.03)	-21.15 (0.72)	1.08 (0.16)	-14.32 (0.37)	5.79 (0.37)
Durable	Yes	0.42 (0.04)	-21.53 (0.67)	1.15 (0.17)	-13.63 (0.37)	3.48 (0.37)
High item volume	No	0.50 (0.05)	-19.29 (0.71)	1.39 (0.14)	-12.52 (0.33)	6.88 (0.46)
High item volume	Yes	0.48 (0.04)	-22.82 (0.94)	0.96 (0.24)	-14.10 (0.40)	3.79 (0.43)
High msrp	No	0.43 (0.04)	-20.19 (0.64)	1.51 (0.14)	-12.49 (0.29)	8.75 (0.59)
High msrp	Yes	0.45 (0.04)	-23.80 (0.81)	0.75 (0.21)	-15.93 (0.42)	3.49 (0.26)
High num offers	No	0.40 (0.04)	-15.45 (0.63)	1.07 (0.14)	-11.67 (0.33)	3.39 (0.21)
High num offers	Yes	0.44 (0.04)	-27.64 (0.77)	1.84 (0.14)	-14.70 (0.30)	12.01 (0.80)
High package weight	No	0.42 (0.04)	-20.94 (0.65)	0.68 (0.19)	-13.80 (0.33)	4.30 (0.30)
High package weight	Yes	0.48 (0.04)	-20.96 (0.80)	1.35 (0.16)	-13.59 (0.35)	5.66 (0.43)
Working hours	No	0.37 (0.03)	-21.42 (0.70)	1.17 (0.12)	-13.89 (0.31)	4.88 (0.27)
Working hours	Yes	0.46 (0.04)	-20.82 (0.62)	1.11 (0.11)	-13.57 (0.30)	5.25 (0.28)

Panel B: Heterogeneity by Product Characteristics.

Notes: We re-estimate prominence and demand models separately for each subset of the data defined by a heterogeneity dimension (first column) taking on a particular value (second column). Column (1) reports the Amazon advantage in the prominence algorithm (expressed as a multiple of the FBA advantage). Column (2) reports the average price elasticity of prominence. Column (3) reports the Amazon advantage in demand (again expressed as a multiple of the FBA advantage). Column (4) reports the price elasticity of demand (taking into account the impact of prominence responding to price). Column (5) reports the percentage of sophisticated consumers. To dichotomize products into durable and non-durable, we asked Claude Sonnet 3.5 to estimate the fraction of products sold in each category on Amazon that are durable; we consider a category durable if the fraction is above 50%. The item volume refers to the product of the width, height and length of the product. Working hours refer to 9AM to 5PM in Chicago, excluding weekends. All binarized variables are dichotomized at the median value of the full sample unless otherwise indicated. Standard errors are clustered at the product level.

6.2. *Heterogeneity: Implications for Counterfactual Analyses*

What do our findings suggest for the heterogeneous impacts of search guidance, as well as potential antitrust remedies, across product categories? First, we re-estimate the pricing and entry model in the body of the paper assuming the Buybox and demand processes in each category are governed by the parameters recovered in Panel A of Table VI.[17] Next, we repeat the main counterfactual analyses from Section 7 of the main text. Finally, we display the percentage effects of these interventions on consumer surplus by category.

The results are summarized in Table VII. The pooled consumer surplus results are qualitatively similar to those of the main counterfactual: search guidance helps consumers; softening price competition harms them; and handicapping Amazon fails to deliver substantial consumer surplus gains. Henceforth, we focus on differential impacts by category. First, search guidance is most valuable to Baby (34.33%), Fashion (24.37%), Video Games (14.06%), and Books (10.92%), though not for Office products (-2.86%). Second, making comparisons on a per-product basis, handicapping Amazon may improve consumer surplus for Office products (1.98%), but is unlikely to deliver gains elsewhere. Finally, if Amazon softens price competition to extract more seller fees, consumer surplus declines the most for products in the Office (-2.44%), Beauty (-1.78%) and Health (-1.54%) categories, but may benefit consumers of Books somewhat (2.2%). As before, we caution against taking these results as precise estimates of welfare impacts

[17]To stabilize SMM convergence, our indirect inference for first- and third-party wholesale prices aims to match only the constant and standardized MSRP coefficients in their respective regressions of offer prices on market characteristics.

across categories, given that our limited sales data implies noisy category-level consumer choice estimates.

TABLE VII

HETEROGENEOUS EFFECTS ON CONSUMER SURPLUS (% RELATIVE TO BASELINE)

Category	N	Value of Guidance	Handicap Amazon	Soften Pricing
Baby	529	34.33%	0.12%	-0.43%
Beauty	4220	1.85%	0.33%	-1.78%
Books	5676	10.92%	-0.06%	2.2%
Fashion	12766	24.37%	0.03%	0.1%
Food	4259	1.74%	0.19%	-1.11%
Health	2502	0.29%	0.61%	-1.54%
Home	3393	2.75%	0.08%	-0.82%
Office	1186	-2.86%	1.98%	-2.44%
Pet Supplies	667	1.58%	0.42%	-1.32%
Sports	3653	5.46%	0.06%	-0.38%
Tools	1075	3.03%	0.44%	-1.44%
Toys	6210	2.24%	0.43%	-0.98%
Video Games	571	14.06%	0.2%	-0.26%

Notes: We simulate counterfactuals separately for each subset of the data defined by a product category taking on a particular value (first column). The number of products in that category, N, is also displayed (second column). After using the prominence and demand estimates in Panel A of Table VI, we estimate an entry and pricing process as in the main text, then simulate three main counterfactuals: the value of search guidance relative to if consumers sorted strongly on price, i.e., if the prominence algorithm had three times the price sensitivity and did not select on non-price characteristics (Column 3); forbidding Amazon from promoting its own offer beyond what is predicted by our observables (Column 4); and if Amazon halved its price coefficient to soften price competition (Column 5). All consumer surplus figures reported are from the "Long-Run" scenario in the main text, where merchants adjust both pricing and entry. Pooled consumer surplus results are suppressed because they are qualitatively similar to those of the main counterfactual.

7. COUNTERFACTUALS

7.1. *Computing Counterfactual Outcomes*

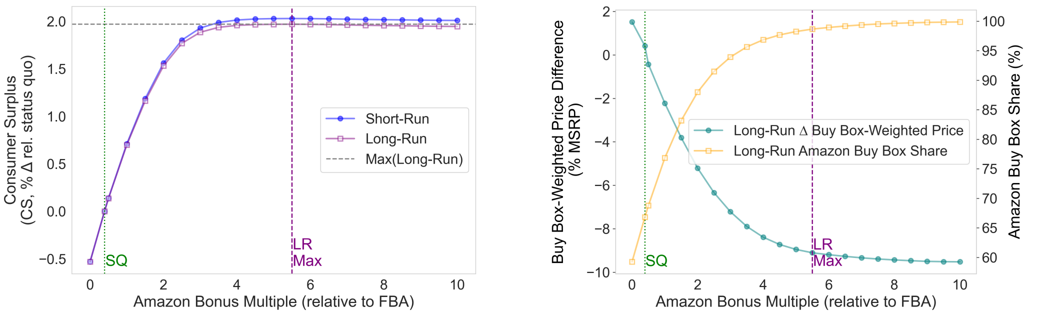
We compute a set of welfare and platform-wide quantities. For the components of welfare on market t, consumer surplus to unsophisticated agents is computed as $CS_{t,U} = [(1-\rho) \times A_t]/\alpha_t \times \sum_{j \in \mathcal{J}_t} r_{jt} \ln[\exp(\delta_{jt}) + \exp(\delta_{0t})]$, while those to sophisticated agents is computed as $CS_{t,S} = (\rho \times A_t)/\alpha_t \times \ln\left[\exp(\delta_{0t}) + [\sum_{j \in \mathcal{J}_t} \exp(\delta_{jt}/\lambda)]^\lambda\right]$. Producer surplus is written $PS_t = A_t \sum_{j \in \mathcal{J}_t} s_{jt} \times (\phi p_{jt} - c_{jt}) - N_t F_t$, and platform intermediation fees are evaluated as $\text{Fees}_t = A_t(1-\phi) \sum_{j \in \mathcal{J}_t} s_{jt} p_{jt}$. Welfare is thus $CS_{t,U} + CS_{t,S} + PS_t + \text{Fees}_t$.

As for other mean market outcomes, the number of entrants is $\bar{N}_t$; sales per month are written $\overline{\text{Sales}} = \frac{1}{T} \sum_{t=1}^{T} A_t \sum_{j \in \mathcal{J}_t} s_{jt}$. Dollar quantities as % of MSRP are price: $\bar{p} = \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{N_t \times \text{msrp}_t} \sum_{j \in \mathcal{J}_t} p_{jt}\right]$; and mean minimum price: $\bar{p}_{\min} = \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{msrp_t} \min_{j \in \mathcal{J}_t} p_{jt}\right]$.

7.2. *Is Amazon Self-Preferencing if Amazon Prices are Fixed?*

In our paper, our leading counterfactual analyses assume that the price of the Amazon offer, where it exists, is set to maximize short-run profits. However, one could also take the view that the platform sets the price of its offers with some long-run objective in mind: for instance, as a benchmark to encourage comparable third-party offers to price more competitively or in an attempt to attract more consumers to the platform. As discussed in the main text, such a long-run objective is consistent with both criticisms of the company (Khan, 2016) and statements by the company's leadership (Rose, 2013).

Therefore, we now assess the robustness of our result that Amazon is not self-preferencing to the assumption that Amazon's retail division sets prices to maximize short-run profits. To this end, in Figure 5, we run the same counterfactual analysis as in Figure 5 of the main text, but



(a) Consumer Surplus

(b) Price & Amazon Prom. Share

FIGURE 5.—If Amazon does not adjust its prices, how much should it guide consumers to its own offers? The left panel shows the impact on consumer surplus as we vary the weight that Amazon's prominence algorithm places on Amazon's own offers, holding fixed the price of the Amazon offer, where it exists. The right panel shows the corresponding effects on prices and the average share of the Amazon offer in prominence, conditional on the Amazon offer being present. The vertical dashed line represents the value of $\hat{\beta}^{r}_{Amazon}/\hat{\beta}^{r}_{FBA}$ that maximizes the long-run consumer surplus, while the vertical dotted line represents the Status Quo.

now we assume that the price of the Amazon offer is fixed at the Status Quo level. Under this alternative assumption, we find that consumer surplus would be maximized by a much higher prominence weight on Amazon's own offers than in Section 7.3 of the main text. Indeed, we find that Amazon is insufficiently steering consumers towards its own offers in the Status Quo, and hence that the platform is not self-preferencing. This result is driven by strong consumer preferences for the Amazon offers, but also by their comparatively low prices: as we see in Figure 5b, the Buybox-weighted price falls as we increase the prominence weight on Amazon's offers.

BIBLIOGRAPHY

ACKERBERG, DANIEL A (2009): "A new use of importance sampling to reduce computational burden in simulation estimation," *Quantitative Marketing and Economics*, 7 (4), 343–376. [4, 5, 6]

ARELLANO, MANUEL AND BO HONORÉ (2001): "Panel data models: Some recent developments," in *Handbook of Econometrics*, ed. by J.J. Heckman and E.E. Leamer, Elsevier, vol. 5, chap. 53, 3229–3296, 1 ed. [2]

BERRY, STEVEN, JAMES LEVINSOHN, AND ARIEL PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, 63 (4), 841–890. [5]

CHAMBERLAIN, GARY (1980): "Analysis of covariance with qualitative data," *The Review of Economic Studies*, 47 (1), 225–238. [2]

KHAN, LINA M (2016): "Amazon's antitrust paradox," *Yale Law Journal*, 126, 710–805. [17]

MCFADDEN, DANIEL (1981): "Econometric models of probabilistic choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by Charles F. Manski and Daniel McFadden, MIT Press, 198–272. [2]

——— (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 57 (5), 995–1026. [5]

MORROW, W ROSS AND STEVEN J SKERLOS (2011): "Fixed-point approaches to computing Bertrand-Nash equilibrium prices under mixed-logit demand," *Operations Research*, 59 (2), 328–345. [4]

PAKES, ARIEL AND DAVID POLLARD (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica*, 57 (5), 1027–1057. [5]

RASCH, GEORG (1960): *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*, Nielsen & Lydiche. [2]

——— (1961): "On general laws and the meaning of measurement in psychology," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 4, 321–333. [2]

ROSE, CHARLIE (2013): "Amazon," Television broadcast, 60 Minutes. Aired December 1 on CBS. [17]