11-18-2021

Incorporating Search and Sales Information in Demand Estimation

Ali Hortaçsu

Olivia R. Natan

Hayden Parsley

Timothy Schwieg

Kevin R. Williams
Yale University, kevin.williams@yale.edu

INCORPORATING SEARCH AND SALES INFORMATION IN DEMAND ESTIMATION

By

Ali Hortaçsu, Olivia R. Natan, Hayden Parsley, Timothy Schwieg, Kevin R. Williams

November 2021

COWLES FOUNDATION DISCUSSION PAPER NO. 2313

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Incorporating Search and Sales Information in Demand Estimation

Ali Hortaçsu, *University of Chicago and NBER*
Olivia R. Natan, *University of California, Berkeley*
Hayden Parsley, *University of Texas, Austin*
Timothy Schwieg, *University of Chicago, Booth*
Kevin R. Williams, *Yale School of Management and NBER**

November 2021

Abstract

We propose an approach to modeling and estimating discrete choice demand that allows for a large number of zero sale observations, rich unobserved heterogeneity, and endogenous prices. We do so by modeling small market sizes through Poisson arrivals. Each of these arriving consumers then solves a standard discrete choice problem. We present a Bayesian IV estimation approach that addresses sampling error in product shares and scales well to rich data environments. The data requirements are traditional market-level data and measures of consumer search intensity. After presenting simulation studies, we consider an empirical application of air travel demand where product-level sales are sparse. We find considerable variation in demand over time. Periods of peak demand feature both larger market sizes and consumers with higher willingness to pay. This amplifies cyclicality. However, observed frequent price and capacity adjustments offset some of this compounding effect.

JEL Classification: C10, C11, C13, C18, L93
Keywords: Discrete Choice Modeling, Demand Estimation, Zeros, Bayesian Methods, Cyclical Demand, Airline Markets

---

*The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. We thank the anonymous airline for giving us access to the data used in this study. Under the agreement with the authors, the airline had "the right to delete any trade secret, proprietary, or Confidential Information" supplied by the airline. We agreed to take comments in good faith regarding statements that would lead a reader to identify the airline and damage the airline's reputation. All authors have no material financial relationships with entities related to this research. Emails: hortacsu@gmail.com, olivia.natan@berkeley.edu, timothyschwieg@uchicago.edu, parsley@utexas.edu, kevin.williams@yale.edu

1 Introduction

Digitization has brought about unparalleled opportunities to gather, store, and process micro data. While these data allow study of research questions that were not answerable with more aggregated data, these micro data sets create new challenges in analysis. For example, when studying infrequently measured aggregate sales data, researchers would seldom encounter data periods devoid of transactions. However, with frequently measured disaggregated data, many or most observations may contain no purchases. A particularly relevant example is e-commerce, where consumers commonly choose among a relatively large set of products, or where few consumers consider purchasing any product for a given market definition. For example, in air travel demand, only a few individuals search for a given flight itinerary each day. Situations with zero transactions are problematic for workhorse demand models because these models require observed market shares to be strictly positive. This practical estimation challenge raises the concern that these standard demand models are conceptually inappropriate for high-frequency, detailed micro data sets.

In this paper, we propose an approach to modeling and estimating discrete choice demand suitable for data environments with sparse sales. We show how commonly used random coefficients logit demand models can be recast as Poisson demand systems. The recast model retains random coefficients, flexible latent product characteristics, and endogenous prices. Although alternative recent methodological advances propose solutions to the zeros problem by considering inference under large market sizes, in many cases, the number of consumers will never grow in a way that reduces this sampling error. This occurs because purchases opportunities tend to rise too slowly relative to observed purchases. In contrast, our approach explicitly models the arrival of consumers to the market thereby allowing for low purchase rates and a significant amount of zero-sale observations. Estimation requires conventional market-level data as well as a measure of market sizes, such as information on consumer search intensity. These data are becoming increasingly available in economics and marketing and not need pertain only to e-commerce. For example, search intensity may involve foot-traffic statistics or a proxy for market sizes, such as the number of consumers who purchased a particular good, e.g., milk in the grocery context. We present simulation studies to compare to alternative approaches to handling sparse sales. We show that our approach performs well in situations where alternative methods produce biased demand estimates. Finally, we apply our methodology to an empirical setting of consumers shopping for airline tickets. In this setting, product-level zeros exceed 85%. We flexibly estimate demand over time and use the model estimates to decompose the underlying forces that cause cyclical demand with a large seasonal component. We show that periods of peak demand feature both larger market sizes and consumers with higher willingness to pay. This correlation amplifies demand variability, which, absent frequent price and capacity adjustments, would be even higher.

In Section 2, we consider the workhorse demand model of Berry, Levinsohn, and Pakes (1995), henceforth BLP (1995), which is often used to flexibly estimate substitution patterns across differentiated products. In the BLP (1995) approach, empirical product shares are matched to their model counterparts via a market share inversion that requires all sales quantities to be strictly positive. If zero shares are observed and dropped from the analysis, estimates will suffer from a selection bias (Berry, Linton, and Pakes, 2004). Although aggregation of the data may be possible, this may smooth over the heterogeneity of interest—in our empirical application, we find that aggregation also yields implausible estimates of demand. Our approach to modeling demand is to take random coefficients logit model and explicitly model the market size through Poisson distributions. With Poisson arrivals, our modeling assumptions imply that demand is also distributed Poisson. This allows us to rationalize zero sale observations and account for the sampling distribution of sales driven by small market sizes. Relative to other proposed solutions to the zeros problem, including Quan and Williams (2018), Dubé, Hortaçsu, and Joo (2021), Adam, He, and Zheng (2020), and Gandhi, Lu, and Shi (2019), we explicitly leverage market-size variation as a source of zero sales.[1]

We model aggregate product-level demand with complete information. In related work,

[1]An alternative approach would be to fix the market size and model the multinomial distribution of sales, similar to Conlon and Mortimer (2013).

Burda, Harding, and Hausman (2012) propose an individual-level Poisson demand model that allows for rich individual-level heterogeneity. Recent empirical work have also considered Poisson models to characterize aggregate demand—for example, Williams (2021) in airlines and Buchholz (2021) for taxis. Relative to these works, we allow for prices to have a flexible correlation structure with latent demand characteristics, i.e., price correlated with unobserved product qualities. Our proposed arrival process can be considered a stylized model of search in that consumers consider all products or do not participate in the market at all. Because we do not use individual-level search data, we abstract from modeling search with information acquisition, e.g., in the style of Weitzman (1979).

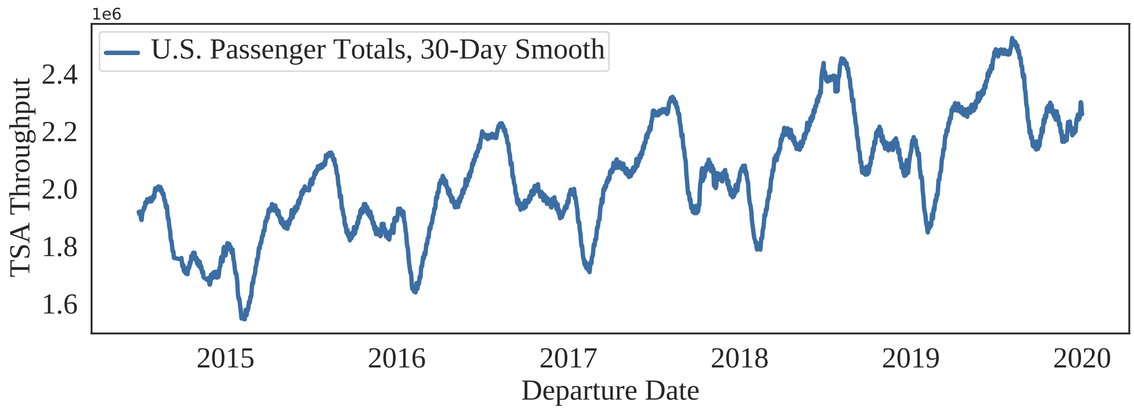
We develop a Bayesian instrumental variables estimator for the model in Section 3. We build on the methods proposed by Jiang, Manchanda, and Rossi (2009) by adding an explicit model of market size that accommodates unobserved product shares (Poisson demand), discrete or continuous random coefficients, and a flexible treatment of addressing price endogeneity. By augmenting our data with unobserved shares, we can use market share inversion of BLP (1995) even though there may exist zero-sale observations. We present two approaches to handling price endogeneity that use limited information pricing equations. The first treatment considers a non-parametric relationship between price and the demand unobservables by applying a Dirichlet process prior. We also present a semi-nonparametric treatment using a mixture normal model.

In Section 4, we demonstrate that our estimator can recover accurate and precise parameter values relative to typical methods in several simulation studies. Our estimator provides coverage in very small market sizes—for example, when arrival rates are five individuals per market. In this setting, we show that common zero-share solutions introduce considerable bias to the demand estimates and overstate the dispersion of preferences within and across markets. In addition, we show that our estimator can be robust to some forms of misspecification, including misspecifying the dispersion of consumer arrivals and misspecifying the correlation structure between the demand unobservables and price.

Finally, we consider the empirical application of the demand for air travel in Section 5.

Data from the Transportation Security Administration (TSA) show that there is high variability in the demand for air travel over time (Figure 1). Demand follows a clear cyclical pattern—more individuals travel in the summer and fewer travel in the winter. There is also consistent within-year variation in spring and fall.

Figure 1: Demand Variability for Air Travel over Time using TSA Throughput



Note: 30-day moving average of the total volume of passengers that pass through a TSA checkpoint in any airport. Total calculated from all U.S. airports daily between 2014 and 2020. The TSA Throughput data are weekly public reports that contain hourly checkpoint totals for all security checkpoints in the United States. We extract passenger counts from these reports to derive a measure of daily travel.

To understand the driving forces behind this cyclical pattern, we apply our methodology to booking, pricing, and search data provided by a large international airline based in the United States. Unlike Chevalier, Kashyap, and Rossi (2003), who document countercyclical pricing in grocery retail, we observe that peak demand also corresponds to periods with high prices and high "opportunity costs" of remaining capacity, as measured by the airline's pricing system. We estimate a mass-point random coefficient version of the model in order to decompose the drivers of demand variability. In our empirical setting, 85% of observations have no sales, and search activity is typically fewer than a few searches per day. Using our approach, we estimate mean product elasticities to be -1.3, with variation in preferences across travel itineraries: passengers are willing to pay $96 more for the most popular week for travel over the least popular week on an identical route. On the other hand, existing demand approaches yield price elasticities between -0.13 and 0, with substantial masses very close to zero. This occurs because imputing (small) sales when shares are zero causes price variation to have no impact on shares. As a result, standard estimators return highly inelastic demand. Similarly, dropping zeros leads to price elasticity estimates very close to zero, since observations with positive shares have higher average preferences and thus lower price sensitivity.

Our demand estimates allow us to decompose the drivers of cyclical demand noted above. We find that both peaks and troughs in demand reflect changes to both the intensity of demand (consumer arrivals) and preference for travel (willingness to pay). These changes are positively correlated over time, though changes to preferences account for about two-thirds of the variation in aggregate demand. That is, periods of peak (off-peak) demand feature both more (less) interest in travel as well as consumers with higher (lower) willingness to pay. We show that prices and capacities adjust over time which act to dampen these compounding forces. In closely related work, Einav (2007) shows that seasonality in the movie industry is the result of two, different compounding forces: higher quality movies are released, and overall interest in movies is higher, during peak periods. Price ridigities amplify these effects, whereas in our application, dynamic pricing and increased capacities during peak periods counteract the forces that drive cyclicality in the airline industry.

2 Model of Consumer Demand

We model aggregate demand using the widely applied random coefficients logit demand model. Consumers, indexed by i, arrive in market t and make a discrete choice, choosing among market-specific products ($j \in J_t$) and an outside option ($j = 0$). In the baseline model, consumers are drawn from a continuum of types. We also consider mass-point random coefficients in an extension (see Section 5 and Appendix C).

2.1 Utility Specification

We assume that indirect utilities are linear in product characteristics and are given by

$$u_{i,j,t} = \begin{cases} X_{j,t}\beta_i + \xi_{j,t} + \varepsilon_{i,j,t}, & j \in J_t \\ \varepsilon_{i,0,t}, & j = 0 \end{cases},$$

where X are product characteristics, including price, ξ are unobserved (to the econometrician) product characteristics that are potentially correlated with price, and ε are independent and identically distributed error terms. We assume these errors are distributed type-1 extreme value. The random coefficients are assumed to be distributed jointly normal across consumers and are independent of characteristics. That is,

$$\beta_i = \bar{\beta} + \Gamma b_i,$$

where $b_i \sim \mathcal{N}(0, I)$ is a multivariate, standard normal distribution, and Γ is the Cholesky decomposition of a positive definite matrix. This allows for a general variance pattern between demand parameters and the random coefficients. Some of the parameters may be linear, meaning there are no associated random coefficients with these characteristics.

All consumers solve a straightforward utility maximization problem; consumer i chooses product j if, and only if,

$$u_{i,j,t} \geq u_{i,j',t}, \forall j' \in J \cup \{0\}.$$

The distributional assumption on the idiosyncratic error term leads to analytical expressions for the individual choice probabilities of consumers. In particular, the probability that consumer i purchases product j is equal to

$$s_{i,j,t} = \frac{\exp\left(X_{j,t}\beta_i + \xi_{j,t}\right)}{1 + \sum_{k \in J_t} \exp\left(X_{k,t}\beta_i + \xi_{k,t}\right)}.$$

Integrating over all consumers, we obtain product market shares, which are equal to

$$s_{j,t} = \int_i s_{i,j,t}\, dF(b_i).$$

2.2 Distribution on Market Sizes

We model the distribution on market sizes using Poisson distributions, i.e.,

$$A_t \sim \text{Poisson}\Big(\lambda_t\Big),$$

such that $\lambda_t = \exp(W_t \tau)$, and W_t is a full-rank matrix. Note that we assume arrivals are measured specific to t rather than specific to j, t. That is, all arriving consumers have full information about all products. Because arrivals are t-specific, this necessarily means that it is impossible to estimate λ_t using fixed effects. For this reason, we highlight the formulation $\lambda_t = \exp(W_t \tau)$. For example, W_t may contain day-of-week or time-of-day indicators depending on the application.

Our model of market participation is a highly stylized representation of consumer search. Consumers either search among all products (with perfect information) or they do not participate in the market. This contrasts with some of the empirical literature that use individual-level data, e.g., Honka (2014) and Kim, Albuquerque, and Bronnenberg (2010). However, it is common in the search literature to condition on market participation, whereas we model the distribution of market sizes.

Two assumptions allow us to construct analytic expressions for demand: (1) arrivals are conditionally independent of preferences, (2) and consumers solve the above utility maximization problems. With these assumptions, conditional on prices and product characteristics, demand for each product j is distributed according to a conditionally independent Poisson distribution, i.e.,

$$q_{j,t} \sim \text{Poisson}\Big(\exp(W_t \tau) \cdot s_{j,t}\Big).$$

Importantly, the demand for product j is independent of the demand for product j', conditional on X_t and p_t.

3 Demand Estimation

We propose a Bayesian estimator (Poisson RC) that can be used to recover a rich set of demand parameters. Our approach differs from frequentist estimators that use market-level data in two key ways. First, we directly accommodate small market sizes by providing an alternative to the empirical share inversion step of Berry (1994) and BLP (1995). Our estimator uses data augmentation to directly sample from the distribution of unobserved demand shocks, thus only requiring an inversion of model shares. Second, our estimator scales well to many markets and rich demand covariates, including many fixed effects.

In the absence of stochastic arrivals and in large markets, our method would be identical to the estimator used in Jiang, Manchanda, and Rossi (2009). Unlike markets with many consumers, where market shares are observed, our setting with low arrival rates and sparse sales force us to treat market shares as unobserved. That is, we cannot simply average observed sales and equate them to market shares because zero market shares can be due to zero consumer searches or searching consumers choosing not to purchase. Observed sales in the data are $q_{j,t}$, which are not only a function of the product shares, but also the number of people that arrive in each time period. This is important because in periods with low search, the probability of $q_{j,t} = 0$ is quite high, but $s_{j,t}$ is never equal to zero. Thus, when we observe few searches, we must account for the sampling variation to be expected in sales quantities.

3.1 Accounting for Price Endogeneity

To account for price endogeneity, we model pricing through a limited information pricing equation with observed components and an unobserved component. Using a set of instruments $Z_{j,t}$, we specify

$$p_{j,t} = Z'_{j,t}\eta + \upsilon_{j,t},$$

where υ is unobservable. We allow the aggregate demand shock ξ to be correlated with price through υ and follow the standard Bayesian framework for simultaneity with discrete choice models (Rossi and Allenby, 1993; Jiang, Manchanda, and Rossi, 2009; Rossi, Allenby, and McCulloch, 2012).

We provide additional flexibility by estimating the joint distribution between the unobservables flexibly, instead of assuming it to be a single joint normal distribution, e.g., in Rossi, Allenby, and McCulloch (2012). We provide two methods for estimating more general distributions. First, we suggest a non-parametric estimator for the joint distribution of price and unobservable demand using a Dirichlet process. Second, we provide a mixture normal model for semi-nonparametric approximation of the joint distribution. The mixture normal provides the researcher with more control over the generality of the distribution when data are sparse. These estimators allow for a sufficiently general approximation of equilibrium behavior between unobserved components of demand and price while still allowing for likelihood-based estimation. Alternatively, the econometrician can fully specify a supply-side model and explicitly model how prices are endogenous to the system.

The mixture normal model approximates the joint distribution with a finite number of basis functions. The Dirichlet process extends this in two directions. First, it allows for alternative base distributions. Second, it allows for an arbitrary number of clusters. The Dirichlet process has support for as many clusters of distributions as data, but the number is disciplined by a tightness parameter. We choose a normal distribution for the base distribution for computational ease. For both specifications, observations in the k^{th} cluster are distributed normally with a shared mean and variance. We discuss the implementation details in Appendix A.[2] For observations in the k^{th} cluster, their distribution is given by

$$\begin{pmatrix} \upsilon_{j,t} \\ \xi_{j,t} \end{pmatrix} \Big| \kappa = k \sim \mathcal{N}(\mu_k, \Sigma_k)$$
$$\text{s.t. } \Sigma_k = \begin{pmatrix} \sigma^2_{k,11} & \rho_k \\ \rho_k & \sigma^2_{k,22} \end{pmatrix}.$$

3.2 Estimation Procedure

We sample the implied posterior distributions using the hybrid-Gibbs sampler outlined below. We split estimation into several distinct parts: consumer arrivals, product shares, preference parameters, and the price endogeneity parameters. Arrival parameters allow us to rationalize zero sale observations; share draws allow us to recover the preference parameters and unobserved product qualities; the price endogeneity parameters allow for price to be correlated with the unobserved components of demand. A more detailed treatment of the estimation procedure can be found in Appendix A.

Algorithm 1 Hybrid Gibbs Sampler: Non-parametric estimation of pricing errors

for $c = 1$ to C **do**	
Update arrivals λ	(Gibbs)
Update shares $s(\cdot)$	(Metropolis-Hastings)
Update linear parameters β	(Gibbs)
Update nonlinear parameters Γ	(Metropolis-Hastings)
Update pricing equation η	(Gibbs)
Update basis classifier κ	(Gibbs)
Update mixture component parameters Σ_k, μ_k	(Gibbs)
end for	

The arrival parameters are informed by both search data and purchase data. Purchase data is distributed Poisson with rate $\lambda \cdot s$. Purchases provide a noisy signal of arrivals, along with the exogenous search data that informs the distribution. As we parameterize arrivals

[2] Appendix D provides details for the more restrictive (though computationally cheaper) finite mixture of normal distributions that can be used to semi-nonparametrically model the joint distribution of ξ and υ.

such that $\lambda = \exp(W_t \tau)$, we impose a Log-Gamma prior on τ for a conjugate prior. We sample from τ using a Gibbs step.

We assume that the observed market sizes are not large enough to treat the observed market shares, $\frac{q}{A}$, to be equal to the model equivalents. Thus, shares are unobserved, and we must treat observed product shares as samples from a distribution. While search data identify the size of the market, the shares are identified by variation of purchases within each market and the characteristics of each good given demand parameters. The distribution of ξ implies a distribution of shares. We invert the distribution and compute the likelihood of the value of ξ implied by the draw of s, conditional on demand parameters. We use a Metropolis-Hasting step to sample from the posterior distribution.

Given the draws of shares, the utility parameters that are common across all consumers are sampled using a straightforward Bayesian-IV regression. We recover the mean utility using the contraction mapping used in BLP (1995) and sample β using a Gibbs step. For the nonlinear parameters, the sampling step is not as straightforward. We follow Rossi, Allenby, and McCulloch (2012) and draw a candidate value of Γ from a set of Cholesky decompositions of positive-definite variance matrices. We then sample from the posterior using a Metropolis-Hastings step in a similar manner to shares. Given the joint distribution between υ and ξ, we can perform another Bayesian IV regression to update the pricing equation coefficients, η, in a similar manner as sampling β. We draw from η using a standard Gibbs draw from an IV regression.

3.2.1 Scaling

One of the benefits of utilizing Bayesian methods is that our model scales to a large parameter spaces. We have found that our estimation methodology scales well to rich arrival process specifications, many product characteristics, and rich levels of unobserved heterogeneity. We demonstrate this in our empirical application, which involves hundreds of parameters. In general, increasing the dimensionality of the parameter space along these lines is computationally cheaper than adding heterogeneity-specific parameters such as a richer specification of random coefficients.

Scaling to environments with large J is computationally more difficult because share inversions and chain sampling take longer as the choice set grows. Additionally, as the choice set grows, the average product share declines mechanically. With lower average shares, our estimator requires more data as each observation becomes less informative when simulating shares.

3.3 Identification

We discuss two main identification challenges and their solutions: separating preferences from arrivals and identifying product shares.

3.3.1 Using Search Information to Separate Consumer Arrivals from Preferences

The difficulty in estimating a model with small market sizes is separably identifying shocks to arrivals from shocks to preferences. For example, if a smaller number of consumers than is typical arrive to the market in a period with high prices, a researcher investigating demand based on sales quantity alone may incorrectly infer that demand is quite elastic (few people bought). However, conditioning on the fact that few consumers considered purchasing may lead to the opposite conclusion—the fact that few consumers arrived suggests observing few sales may happen with high probability, even if consumers are not price sensitive.

Identification of the search process is straightforward as we have presented it. Under the assumption of (1) independent consumer search and (2) observed arrivals, the Poisson distributional assumption is a robust approach to estimating the conditional mean of consumer search. As noted above, we require some additional restrictions to identify λ since if arrivals are measured for each t, it is impossible to estimate each arrival rate (t-specific) using fixed effects.[3]

[3]It may be the case that a researcher cannot perfectly measure search intensity. This is also true in our empirical application where we do not observe all searches. However, as we show, we can account for unobserved searches relatively easily using properties of the Poisson distribution. For example, if a researcher is confident that she observes 20% of searches, then estimated arrival rates can be scaled up using a scaling

3.3.2 Identification of Share Parameters

With data on consumer search, we directly observe variation in the market size and are able to condition on it to pin down the preference parameters. The variation that is used to identify the preference parameters is the same variation commonly cited in the literature on estimating demand for differentiated products using market-level data (Berry, Levinsohn, and Pakes, 1995; Berry and Haile, 2014, 2016). The parameters governing preferences for the exogenous characteristics are identified from the variation of products offered across markets, and the price coefficients are identified from exogenous variation introduced by instruments.

4 Simulation Studies

We study the performance of our estimator on sparse sales data through Monte Carlo experiments. We compare our estimation approach to the performance of available zero-share fixes.

4.1 Data Generating Process

We conduct two sets of Monte Carlos. The first considers moderate arrivals (Poisson arrival rate of 25), and the second considers small arrivals (Poisson arrival rate of 5). These specifications are used to approximate real-world markets—our empirical application is closer to the latter—where relatively few people consider purchasing any product. We specify a low ratio of search to assortment size so that empirical market shares (q/A) have a large sampling error relative to the demand function. Our specifications create a substantial number

factor equal to five.

of zeros: 59% in the first specification and 78% in the second specification.[4,5]

For each Monte Carlo specification, we simulate $1,000$ markets. Markets are structured as a panel, with 100 groups of 10 related markets. The number of products in each group of markets is drawn at random from $[1,2,3]$. We define the utility specification to contain random coefficients on three product attributes, given by

$$\begin{aligned} u_{i,j,t} &= X_{j,t}\beta_i + \xi_{j,t} + \varepsilon_{i,j}, \\ \beta_i &= \bar{\beta} + \Gamma b_i, \\ X_{j,t} &= [p_{j,t}, X_j^1, X_j^2]. \end{aligned}$$

In our simulations, we draw $X_j^1 \sim i.i.d.$ Uniform$[0,1]$, $X_j^2 \sim i.i.d.$ Uniform$[0,1]$, and these non-price attributes are fixed within each group of 10 markets for a product. The demand shock ξ is drawn from a "banana distribution," meaning that the residual in the pricing equation and unobserved product quality follow a non-linear relationship (see Figure 4 for a contour plot of this distribution). Finally, prices are drawn from a distribution correlated with ξ and as a function of exogenous instruments Z.[6] Preference parameters are $\bar{\beta} = [-2.0, .5, -.25]$ with variance $0.2I$, where I is a 3x3 identity matrix.

We contrast our estimation method with BLP (1995) and Gandhi, Lu, and Shi (2019). For BLP (1995), we treat zeros either by dropping them or adding an arbitrary amount of sales (1e-6) to each observed zero. We either aggregate over intervals of length 10 or estimate on the disaggregated data. In the aggregation case, we average covariates across observations and sum both sales and arrivals. Finally, we treat market sizes as observed

[4]Though arrivals are generated from a single distribution (i.e. $\lambda = 5$ or $\lambda = 25$), we treat markets as belonging to panel groups and estimate group-specific pooled λ parameters. This is analogous to how one might estimate market size at a weekly level in many applications. In our setting, it allows us to show the robustness of preference estimates to limited observations of the arrival process.

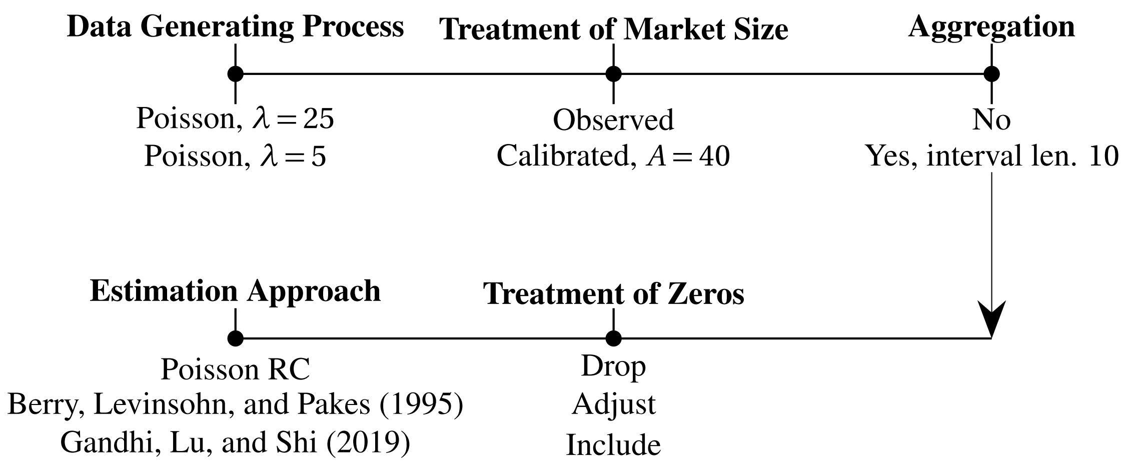
[5]For comparison, Dubé, Hortaçsu, and Joo (2021) test their model on data with 42% zero shares, and Gandhi, Lu, and Shi (2019) test their estimator on synthetic data with 52% zero shares. Their bounds estimator is also tested on data with 96% zero observations, however, in this specification, only one parameter is estimated and the rest are calibrated at the truth.

[6]Instruments (dimension 2) are drawn from $i.i.d.$ Uniform$[0,1]$ distributions, and they have coefficients $\eta = [1.0, 2.0]$ in the pricing equation.

or we calibrate them to $A = 40$. The latter specification approximates what would happen without access to search data.

A summary of our data generating processes, treatment of market size, aggregation method, zero-share adjustments, and estimation procedure used in our Monte Carlos appears in the figure below.

Figure 2: Monte Carlo Setup



Note: Details of Monte Carlo setup. Aggregation and treatment of zeros is considered only for the BLP (1995) estimation approach.

4.2 Monte Carlo Estimates

We begin by comparing our approach (Poisson RC) to the estimation approach of BLP (1995) using common zero-share adjustments. We focus on results using calibrated market sizes ($A = 40$). Results using observed arrivals appear in Appendix B. Monte Carlo results using a moderate arrival rate of $\lambda = 25$ appear in Table 1. Table 2 shows the results for the low arrival rate of $\lambda = 5$. All results are presented relative to the truth, i.e., results are reported in terms of parameter bias. In parentheses, we report the 2.5th and 97.5th percentile of the bias across the 1,000 simulations.

In both tables, the first column reports results for our estimator (Poisson RC). The next two columns reports results of using the disaggregated/aggregated data when we drop zeros from the analysis. The final two columns again use the disaggregated/aggregated data, where sales quantities equal to zero are replaced by 1e-6.

Table 1: Monte Carlo Results for $\lambda = 25$, Poisson RC and BLP with Calibrated Arrivals

	Poisson RC	BLP, Drop 0's		BLP, Adjust 0's	
		Disagg.	Agg.	Disagg.	Agg.
α	-0.10	0.26	-0.54	-5.32	-1.13
	(-0.33, 0.11)	(-3.43, 1.15)	(-6.25, 1.16)	(-23.53, -0.64)	(-10.51, 1.15)
β_1	-0.02	-0.77	-1.31	-3.32	-1.68
	(-0.29, 0.26)	(-4.92, 0.01)	(-8.89, 0.35)	(-25.38, 0.98)	(-13.47, 0.66)
β_2	-0.02	-0.50	-1.23	-4.12	-1.94
	(-0.31, 0.25)	(-4.69, 0.35)	(-8.28, 0.38)	(-32.11, 0.39)	(-14.76, 0.41)
Γ_{11}	-0.03	0.63	-0.01	1.38	0.03
	(-0.12, 0.21)	(-1.24, 4.72)	(-2.40, 2.98)	(-2.61, 8.38)	(-3.01, 4.28)
Γ_{22}	0.04	0.16	0.41	1.04	0.61
	(-0.08, 0.30)	(-2.92, 5.20)	(-5.33, 7.53)	(-10.20, 18.94)	(-6.19, 10.89)
Γ_{33}	0.04	0.19	0.38	0.84	0.51
	(-0.07, 0.41)	(-2.43, 4.81)	(-4.82, 7.00)	(-11.47, 16.57)	(-7.15, 11.78)

Note: Reported in the table are the mean, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 1000 simulations. Column "BLP, Drop 0s, Disagg" drops observations with zero empirical shares $s = q/A$. Column "BLP, Drop 0s, Agg" aggregates sales and arrivals across 10 adjacent observations, then drops observations with zero aggregate shares. Covariates and instruments are averaged across observations. Column "BLP, Adjust 0s, Disagg" replaces zero empirical shares with a small number $\varepsilon = 1e^{-6}$. Column "BLP, Adjust 0s, Agg" aggregates sales and arrivals across 10 adjacent observations, then replaces zero aggregate shares with $1e^{-6}$. Covariates and instruments are averaged across observations.

We find that our Bayesian estimator of the Poisson RC model performs well. For moderate market size ($\lambda = 25$), we have good coverage of 95% credibility intervals for most parameters. For very small markets ($\lambda = 5$), the coverage for our price coefficient decays somewhat, but point estimates remain close to the truth. Coverage is slightly worse for some elements of the random coefficient variance Γ: on average, the credibility interval contains the truth only 80-88% of the time. This is not unique to our estimator—Gandhi and Houde (2015) show that identification is challenging in standard settings even with sufficient instruments. We find that estimates of Γ, the nonlinear parameters, are more sensitive to the choice of prior than other parameters. Overall, the Poisson RC produces little bias in all parameters for both specifications.

In contrast, we find that the two common solutions to zero sales observations fail to

Table 2: Monte Carlo Results for $\lambda = 5$, Poisson RC and BLP with Calibrated Arrivals

	Poisson RC	BLP, Drop 0's		BLP, Adjust 0's	
		Disagg.	Agg.	Disagg.	Agg.
α	-0.17	1.20	-1.11	-5.52	-6.00
	(-0.54, 0.17)	(-1.11, 1.71)	(-10.89, 1.49)	(-23.13, 0.11)	(-32.87, 1.24)
β_1	-0.01	-1.43	-2.41	-5.14	-6.41
	(-0.50, 0.48)	(-7.54, -0.22)	(-14.12, 0.17)	(-37.99, 0.54)	(-44.76, 1.61)
β_2	-0.02	-1.06	-2.43	-5.40	-6.86
	(-0.51, 0.44)	(-9.23, 0.38)	(-16.18, 0.37)	(-36.57, 0.40)	(-52.48, 0.98)
Γ_{11}	-0.02	0.59	0.03	1.80	0.35
	(-0.12, 0.27)	(-1.32, 4.57)	(-4.14, 3.50)	(-3.95, 9.18)	(-8.43, 10.35)
Γ_{22}	0.05	0.05	0.66	0.81	2.28
	(-0.07, 0.36)	(-4.31, 4.77)	(-5.93, 9.30)	(-10.23, 14.66)	(-13.44, 26.85)
Γ_{33}	0.04	0.25	0.37	0.61	1.29
	(-0.07, 0.33)	(-4.09, 6.77)	(-6.36, 8.21)	(-10.99, 16.85)	(-16.31, 24.95)

Note: Reported in the table are the mean, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 1000 simulations. Column "BLP, Drop 0s, Disagg" drops observations with zero empirical shares $s = q/A$. Column "BLP, Drop 0s, Agg" aggregates sales and arrivals across 10 adjacent observations, then drops observations with zero aggregate shares. Covariates and instruments are averaged across observations. Column "BLP, Adjust 0s, Disagg" replaces zero empirical shares with a small number $\varepsilon = 1e^{-6}$. Column "BLP, Adjust 0s, Agg" aggregates sales and arrivals across 10 adjacent observations, then replaces zero aggregate shares with $1e^{-6}$. Covariates and instruments are averaged across observations.

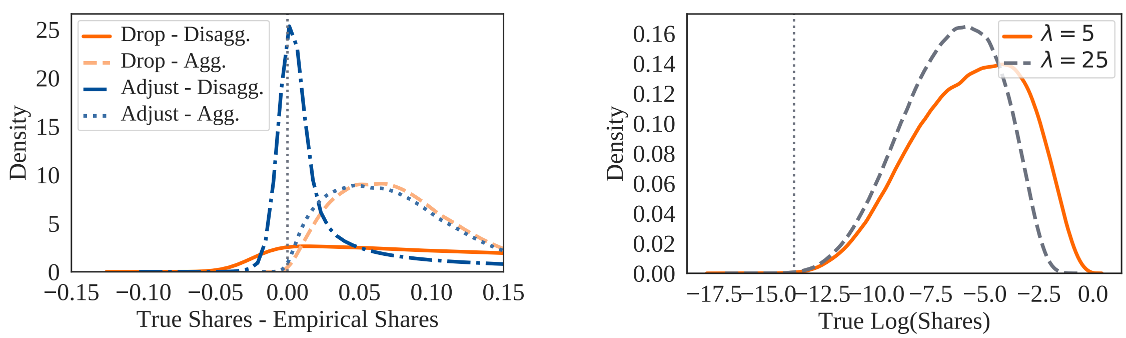
capture the true parameters. Dropping zero sales observations attenuates the price coefficients with disaggregate data (columns two and three in Table 1 and Table 2) . This is more severe in the case with the smaller market size. Although the drop-zero estimates perform better in estimating the price coefficient, this approach fails to capture other parameters accurately, including both the mean and variance of random coefficients. Dropping zero sales observations after aggregation has the opposite effect on price coefficients; estimates of the price coefficient are biased downward.

Adjusting product shares when they are equal to zero results in more extreme bias (columns four and five in Table 1 and Table 2). The parameter estimates consistently misestimate the mean and variance in the random coefficients on price and exogenous characteristics. The distribution of product shares when sales are zero is centered below the distribution of shares when sales are positive. Dropping the zero shares creates selection since zero shares reflect higher prices or lower demand shocks (Berry, Linton, and Pakes, 2004). Figure 3a plots the distribution of the true shares minus the adjusted shares. Dropping zeros results in adjusted shares that are higher the truth. Adjusting zero shares with a small value understates the true share. Consequently, price effects are overstated since imputing a tiny share is inaccurate when the zeroes arise due to few consumers. Figure 3b shows the conditional distribution of log product shares when sales are zero. Imputing an arbitrary small value ($\varepsilon = 1e^{-6}$) performs poorly because imputed shares are on the large end of common imputed definition of zero shares, but they are lower than true shares when quantity sold is zero. As a result, observations with small true shares (e.g. when price is higher than usual) will be imputed to have *even smaller* shares. This leads estimates to overstate the price sensitivity of consumers.

Figure 3: True Shares Compared to Share Adjustments

(a) True vs Empirical Shares

(b) True Log(Shares) when Q is Zero

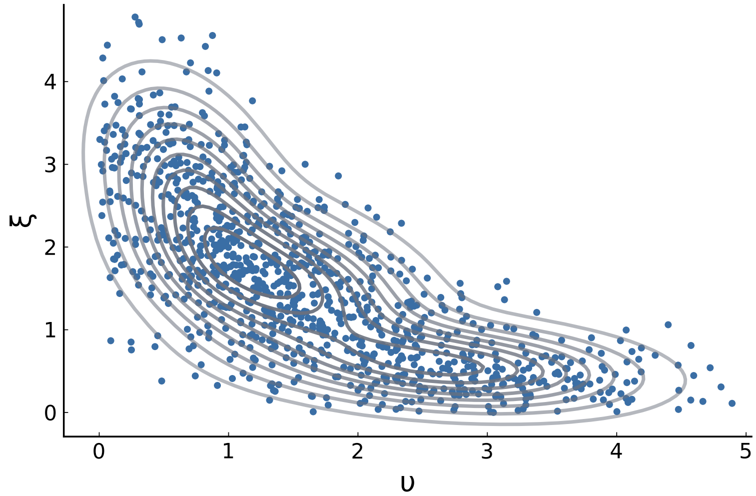


Note: (a) Distributions of the difference between the "true" model shares that generated the data from the various zero-share adjustments. Adjusted shares refers to taking any observations where the empirical shares would equal zero and replace it with an arbitrary small number $\varepsilon = 1e^{-6}$. Drop 0's refers to dropping all observations where the empirical shares are equal to zero. (b) The density of the log of the model shares when quantity sold is zero. Plotted in the dashed grey line is $\log(\varepsilon)$.

An alternative solution to minimize the frequency of zero shares is to aggregate across time periods. Applying share adjustments after aggregation results in fewer adjusted shares, but it does not improve the performance of such estimators (the third and fifth columns in Table 1 and Table 2). Instead, aggregating across sales periods removes price variation and results in considerable bias in some parameters.

Our Poisson RC estimation method uses an unknown number of mixture of normal distributions to approximate any joint distribution of (ξ, υ). Figure 4 compares the generating distribution with the sampled distribution. Our non-parametric using a Dirichlet process prior is able to recover this joint distribution well.[7]

Figure 4: Recovering Joint Distribution of (ξ, υ)



Note: Contours show the generating joint distribution of (ξ, υ) and scatter plot shows sampled (ξ, υ).

Table 5 and Table 5 in Appendix B repeat these Monte Carlos, but use observed arrivals instead of calibrated arrivals. We find that using observed market size realizations improves the results of BLP where zero shares are dropped, however, our main results hold: the BLP estimator with zero adjustments performs poorly when market sizes are small.[8]

Our Monte Carlo results show that the Poisson RC model can accurately measure consumer preferences and heterogeneity under very small market sizes. This is because the method accounts for the sampling error to be expected in sales. Alternative solutions to the zeros problem conduct inference under large market sizes. For example, in Appendix B, we also report results using observed market sizes and the estimator from Gandhi, Lu, and Shi (2019). These appears appear in Table 6 and Table 7 We find that the approach of

[7]In the case that the researcher is certain about the number of components to use to approximate this distribution, we provide an extension to that allows for semi-nonparametric estimation of the distribution of residuals in Appendix D.

[8]In one case ($\lambda = 25$, disaggregated data, observed arrivals, BLP), the BLP estimator outperforms our Poisson RC estimator on average for the price coefficient. However, noise across simulations is very high compared to the Poisson RC model.

Gandhi, Lu, and Shi (2019) still produces significant bias in the price coefficient and noisier estimates than the Poisson RC model. We hypothesize that this is due to the lack of “safe products” used in estimation. Safe products are products in which empirical shares are observed with minimal measurement error in sample (the identities of these products need not be known). In our simulations, zero empirical shares are not being driven by a large tail of products, but instead are driven by a small market size, which causes all empirical shares to be noisy measures of true shares.[9]

In addition to these simulations, we test our estimator to two forms of misspecification in Appendix B. Results are shown in Table 8. We simulate a misspecified search process (overdispersed with twice the variance of our $\lambda = 25$ setup) and a less flexible form of correlation between the pricing error and the demand shock (estimating the correlation structure between ξ and υ using a single normal distribution instead of a mixture of normal distributions). We find that our estimator performs nearly identically in the case of overdispersion: the use of a Poisson distribution is robust to overdispersed arrivals. We also find that using a less flexible form of price endogeneity also provides nearly identical estimates. When the conditional expectation of ξ given υ is approximately linear, or when the correlation summarizes the dependence well, a normal approximation will perform well. In cases where there is strong dependence but low correlation—such as ξ being a symmetric function of υ—this simpler specification may be restrictive and lead to biased demand estimates. Symmetry may be unrealistic because it implies that demand shocks are associated with both low and high prices.

5 Empirical Application to the Airline Industry

We use our approach to estimate the demand for air travel with data from a large international air carrier based in the United States.[10] Our objectives are twofold. First, we aim to

[9]A partial identification strategy is an alternative solution, which Gandhi, Lu, and Shi (2019) suggests pursuing. This approach does not require the presence of safe products.

[10]The airline has elected to remain anonymous.

flexibly estimate preferences over time and compare these estimates to existing approaches. Second, we decompose the underlying forces that cause cyclical demand in this industry.

5.1 Data

We use data from the air carrier's booking system to construct the quantity of tickets sold and the price paid for every flight, each day before departure. The sample contains hundreds of origin-destination pairs over two years of data.

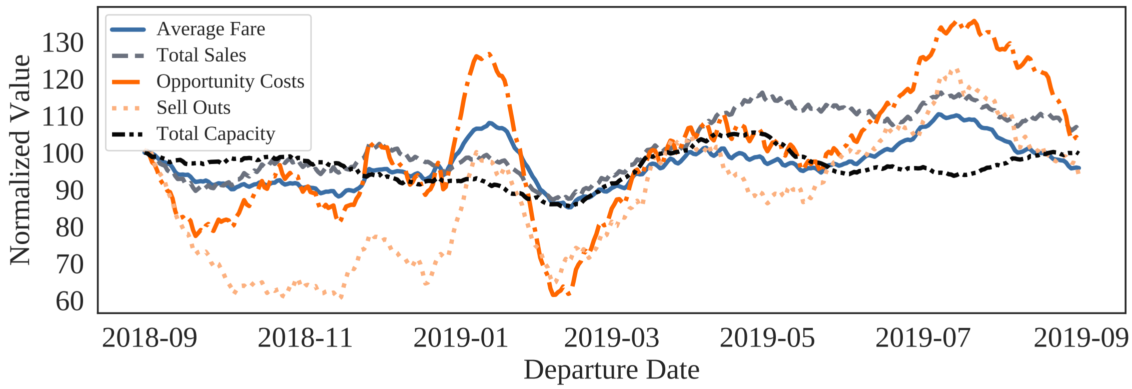
Our market definition is an origin-destination, day before departure. For this analysis, we concentrate on nonstop bookings. In addition to prices and quantities, we also extract basic product characteristics, such as the departure time for each flight and the date of departure.

We calculate consumer searches using the number of consumers who initiate travel requests on the air carrier's website using consumer clickstream data. Consumers arrive at the air carrier's website and their activity within a browsing session is tracked. We then aggregate search activity to the level of origin-destination-search date-departure date. Note that we measure search for the direct channel, which is one of three booking channels. Consumers can also purchase via an Online Travel Agency, such as Expedia, or via a travel agent. We cannot directly measure searches made from other channels. However, because we observe all bookings, we account for searches made via the unobserved channels through a market-specific scaling factor. Each scaling factor is based on the fraction of sales through the direct channel and the average number of passengers per booking. We use this aggregate search activity as a measure for our market size.

In-depth summary analysis of the data and how unobserved searches are accounted for can be found in Hortaçsu, Natan, Parsley, Schwieg, and Williams (2021).

In Figure 5 we plot the 30-day moving average of bookings, fares, opportunity cost of capacity, fraction of sold out flights, and total capacity for all routes in our sample from August 2018 to August 2019. Opportunity costs measure the shadow value of a seat as recorded from output of the airline's pricing optimization program. The figure shows that

Figure 5: Demand, Prices, Opportunity Costs, and Capacity for Markets in Data



Note: 30-day moving average of sales, fares, opportunity costs, probability of flights selling out, and total capacity by departure date. Values are normalized by the respective value for the first departure date in our data, 08-01-2018.

all these variables are positively correlated. For example, all curves peak around the winter holiday season as well as during summer. During peak periods, demand is high. At the same time, on the supply-side, prices, opportunity costs, flight capacity, and the percentage of flights that sell-out are also high. There is some within season variability that matches the patterns observed in the longer horizon TSA traffic displayed in Figure 1. Using the Poisson RC model, we will to decompose changes in the intensity of demand from changes in willingness to pay over time. To do so, we adapt the baseline Poisson RC model to the airline setting.

5.2 Empirical Specification

We define a market (m) as an origin, destination, and departure date tuple and let the time index (t) denote days until the departure date. We extend the estimator to allow for discrete support random coefficients (see Appendix C for the estimation details). Following Berry, Carnall, and Spiller (2006), we assume consumers are one of two types, corresponding to leisure (L) travelers and business (B) travelers. An individual consumer is denoted as i and her consumer type is denoted by $\ell \in \{B, L\}$. The probability that an arriving consumer is a business traveler is equal to γ_t. These types need not correspond to the consumer's purpose for travel; they merely are commonly used names for discrete consumer types. The less-price-sensitive type is typically referred to as business. We assume the indirect utilities are linear in product characteristics and given by

$$u_{i,j,t,m} = \begin{cases} X_{j,t,m}\beta - p_{j,t,m}\alpha_{\ell(i)} + \xi_{j,t,m} + \varepsilon_{i,j,t,m}, & j \in J(t,m) \\ \varepsilon_{i,0,t,m}, & j = 0 \end{cases}.$$

The product characteristics that are assumed to be uncorrelated with the unobserved product characteristic $\xi_{j,m,t}$ are contained in $X_{j,t,m}$. These exogenous characteristics include departure time, week, and day of week fixed effects. We include week fixed effects in the utilities to capture seasonal variation in the value of travel. The consumer types differ in their in preferences on price, $\alpha_{\ell(i)}$, and we assume that $\xi_{j,m,t}$ is correlated with price. Given our assumption on $\varepsilon_{i,j,t,m}$, the probability that consumer i wants to purchase product j is equal to

$$s^{i}_{j,t,m} = \frac{\exp\left(X_{j,t,m}\beta - p_{j,t,m}\alpha_{\ell(i)} + \xi_{j,t,m}\right)}{1 + \sum_{k \in J(t,d)} \exp\left(X_{k,t,m}\beta - p_{k,t,m}\alpha_{\ell(i)} + \xi_{k,t,m}\right)}.$$

Since consumers are one of two discrete types, we define $s^{L}_{j,t,m}$ as the conditional choice probability for leisure type consumers (and $s^{B}_{j,t,m}$ for business types). Integrating over consumer types, we have

$$s_{j,t,m} = \gamma_t s^{B}_{j,t,m} + (1-\gamma_t) s^{L}_{j,t,m}.$$

In this mass-point random coefficient model, we parameterize the change in the compisition of consumers as follows. We assume γ_t is equal to

$$\gamma_t = \frac{\exp\left(f(t)\right)}{1 + \exp\left(f(t)\right)}.$$

where $f(t)$ is an orthogonal polynomial basis of degree 5 with respect to days from departure. This parametric assumption allows for a non-monotonic relationship between the composition of consumer types and time while producing values bounded between 0 and 1.

We adjust the likelihood to account for the possibility of binding capacity constraints (sell-out events). In particular, when capacity is binding, we observe a right-censored estimate of the true number of individuals that wished to purchase. That is, for a given capacity $C_{j,t,m}$,

$$q_{j,t,m} = \min\left\{\tilde{q}_{j,t,m}\,,\, C_{j,t,m}\right\},$$
$$\tilde{q}_{j,t,m} \sim \text{Poisson}(\lambda_{t,m} \cdot s_{j,t,m}).$$

Note that when the capacity constraint is observed to bind, the likelihood contribution is instead $1 - F_q(q_{j,t,m}|...)$, where F_q is the cumulative distribution function of the above Poisson.

We parameterize arrivals as a set of multiplicative fixed effects across markets m and time t. That is, $\lambda_{t,m} = \exp(W_t \tau_t + W_m \tau_m)$, where W_t is a dummy matrix with a column for each day from departure and W_m is a dummy matrix with a column for each departure market m (an origin, destination, departure date tuple). This parameterization approach allows us to capture general increases in market size towards departure across all seasonal markets. In addition, we then have two sources of seasonal variation in participation and preferences built directly into the model, which will enable the decomposition of seasonality that we perform below.

Finally, we instrument for price to address endogeneity. We use the opportunity cost of capacity for a given flight, advance purchase discount indicators, and the number of inbound or outbound bookings from a route's hub airport as our instruments[11]. We leverage the expiry of advance purchase discounts since these changes alter prices in a pre-set fashion, regardless of realizations of demand shocks. The shadow value of capacity directly

[11]For a route with origin O and destination D, where D is a hub, the total number of outbound bookings from the route's hub airport is defined as the following; $\sum_{i=1}^{K} Q_{D,D'}$. Where $Q_{D,D'}$ is the the total number of bookings in period t, across all flights, for all K routes where the origin is the original route's destination. If the route's origin is the hub, we calculate the total number of inward bound bookings, which would be; $\sum_{i=1}^{K} Q_{O',O}$. Where $Q_{O',O}$ is the total bookings from all K routes where the original routes origin is the destination.

influences price-setting, as residual variation (after our use of fixed effects) is driven by bookings on onward itineraries. The total number of inbound or outbound bookings to a route's hub airport captures the change in opportunity cost for flights that are driven by demand shocks in other markets. For example, consider a flight from $A \to B$, where B is a hub which serves many markets. We construct all onward traffic from B onward to other destinations C or D. We assume that unobserved, systematic demand shocks are independent across routes, so shocks to demand for travel from $B \to C$ and $B \to D$ are unrelated to unobserved shocks to demand for the focal route $A \to B$. Pricing decisions across routes are related via capacity: if a positive shock to demand out of hub B is realized, the opportunity cost to provide service from $A \to B \to C$ or $A \to B \to D$ rises. This increase in opportunity cost for connecting tickets also raises the opportunity cost of capacity on the $A \to B$ leg, which raises the price on $A \to B$. Finally, we use estimates of the shadow value of remaining capacity that is observed as output of the airline's pricing algorithm.

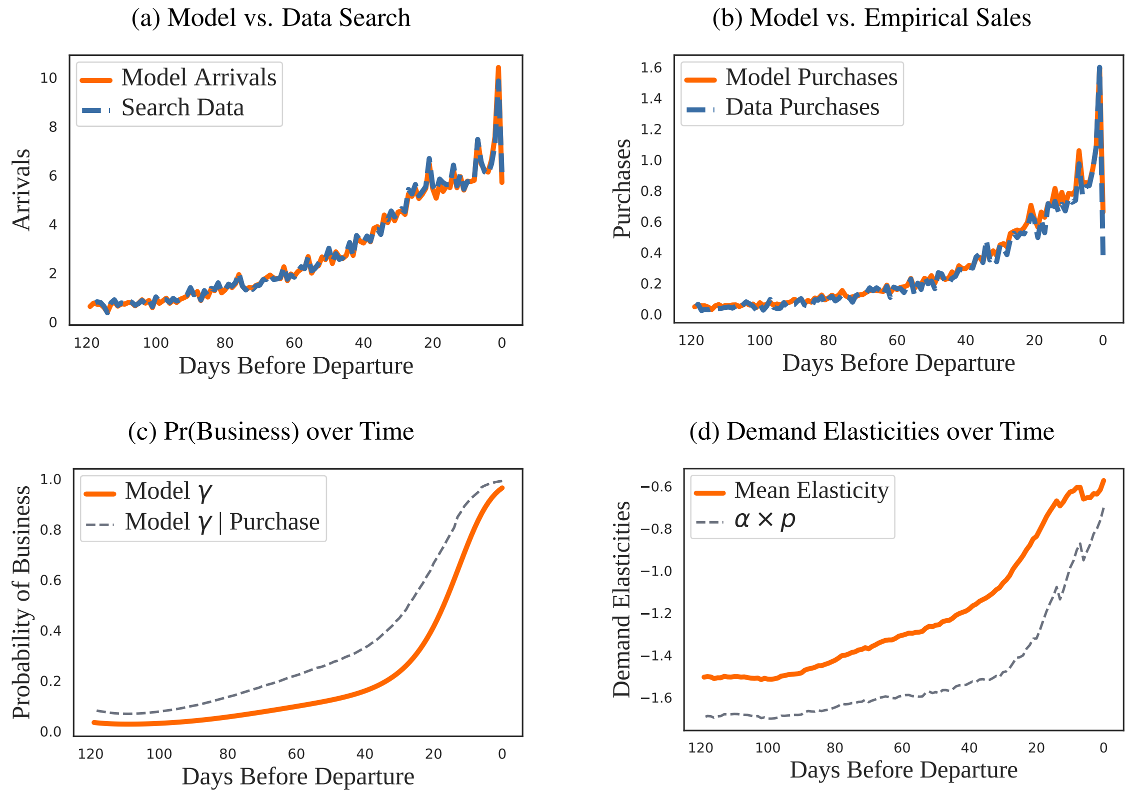
We specify a time-varying block structure on the pricing equation, and (ξ, υ) have a block-varying joint normal distribution. Though such a specification may appear more restrictive than the Dirichlet process prior, this specification allows us to tailor our specifications to our empirical context where the pricing equation clearly changes over time due to advance purchase discounts.[12]

5.3 Empirical Results

We provide detailed demand results for an origin-destination pair in the sample and compare our results to existing methods of estimating demand. We select an average market in terms of zero sale observations: 85% of observations involve zero sales versus the sample average of 88%. Most departure markets for this route have 1 or 2 daily flights. Interested readers can find a summary of demand results for many markets in Hortaçsu, Natan, Parsley, Schwieg, and Williams (2021).

[12] Our misspecified specifications in Section 4 provide an example where a restrictive distribution of this type performs relatively well compared to the fully flexible estimator.

Figure 6: Demand Estimates



Note: The horizontal axis of all plots denotes the negative time index, e.g. zero corresponds to the last day before departure. (a) Normalized model fit of searches with data searches. (b) Model fit of product shares with empirical shares. (c) Fitted values of γ_t over time, along with the probability a consumer is business conditional on purchase. (d) Mean product elasticities over time, along with the least and most elastic flights.

Measures of fit of estimation results are shown graphically in Figure 6. Panel (a) shows the average market size (Poisson distribution means) across the booking horizon. Most consumers arrive very close to departure. Our estimates fit searches (scaled for unobserved searches) and sales quantities well. Note that the average market size is about 5 searches per day, which is in line with our Monte Carlo exercises. The composition of consumers changes considerably over the booking horizon, as shown in Figure 6(c). Well in advance of departure, passengers are entirely composed of price-sensitive, leisure passengers. Close to departures, arriving passengers are almost entirely less price-sensitive, business travelers. The changing composition of customers yields smaller price elasticities as the departure date approaches. However, the average price rises precipitously close to departure, which yields marginally higher elasticities close to departure (panel d). Note that our estimates suggest demand elasticities may be less than one close to the departure date (see Hortaçsu, Natan, Parsley, Schwieg, and Williams (2021) for evidence on pricing on the inelastic side of demand).

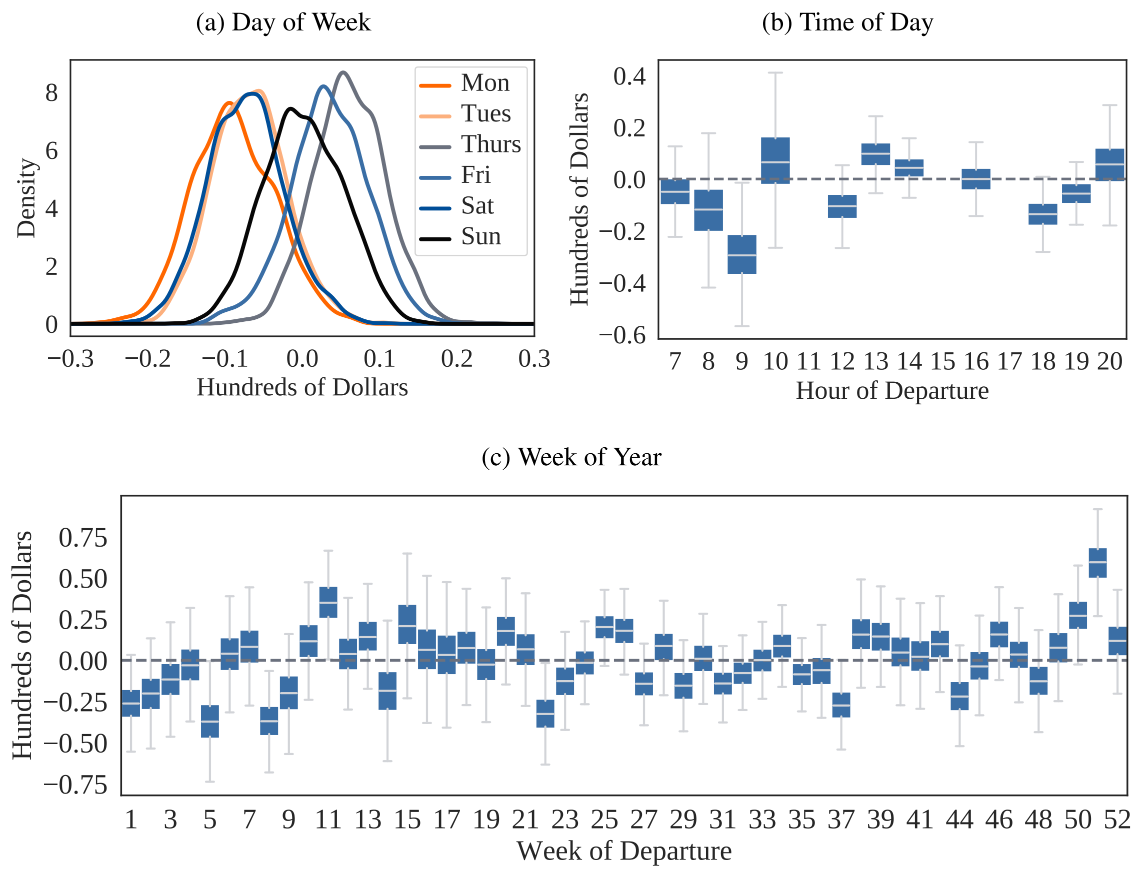
In Figure 7, we graphically show our estimates of preferences over product characteristics, scaled by the price coefficient for leisure travelers. This origin-destination pair features lower demands for earlier days of the week. Preferences for flight time are less differentiated, with 9am and 6pm being the least preferred travel times. Noisy and small estimates suggest that relative time of day preference is not a large source of variation in demand within market.[13] Leisure types are four times more price sensitive than business type passengers.

Willingness to pay for travel displays considerable seasonality in this market. Figure 7(c) shows variation in the valuation of travel by week of year. The most popular week is valued \$96 more than the least popular week for travel, for the same route. However, estimates of these preferences are relatively noisy. Only 18% of the week of year preferences have 95% credibility intervals that exclude 0. Arrival rates also vary seasonally and towards the departure date. 69% of departure-date arrival fixed effects are significant, and 85% of day-from-departure arrival fixed effects are significantly different than the day of departure.

Variation in willingness to pay may be amplified or compressed when combined with supply responses. For example, if the firm prices countercyclically (setting low prices in periods of low demand), we might expect prices to be highly correlated with willingness to pay across departure markets. We find some evidence of such responses in seasonal pricing. Figure 8 plots smoothed average fares, sales, and consumer surplus among leisure travelers. Prices are countercyclical, which may reflect the role of capacity-limited pricing with fixed capacity across the year. We further explore this variation below in Section 5.4.

[13] These estimates report the ratio of preference β to price sensitivity α, so a confidence interval overlapping with zero is not a direct measure of significance. However, for time of day and day of week, many preference parameters have credibility intervals containing zero.

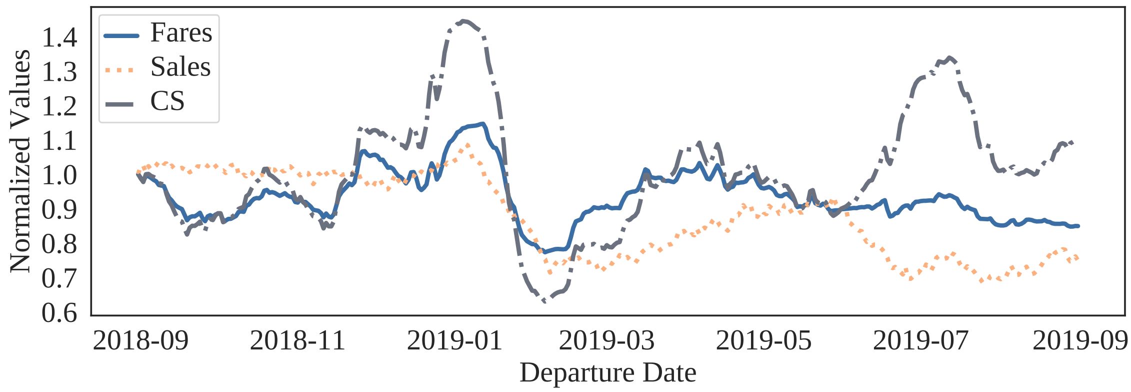
Figure 7: Relative Willingness to Pay for Flight Attributes



(a) Kernel Density Estimates of a leisure consumer's willingness to pay to change flight day of week from Wednesday. (b) Leisure consumer's willingness to pay to change time of flight from 3 pm. (c) Leisure consumer's willingness to pay relative to a flight departing in the 43rd week of year (end of October).

We contrast our estimates with typical zero share fixes and assumptions about market size. Table 3 summarizes price elasticity estimates for a model based on Berry, Carnall, and Spiller (2006) and compares to our estimator. We vary how zero sales are handled (either dropped or imputed with a small market share) and how the market size is constructed (using observed search or fixing the total market size), similar to our Monte Carlos. We find that dropping zero shares yields extremely inelastic estimates and this is worse when we use calibrated arrivals. Since 85% of the flight observations consist of zero sales, imputing these values with a small value replaces most of the observed shares with identical values. The lack of variation in empirical shares drives the estimator to boundary solutions where

Figure 8: Seasonal Variation in Fares, Demand, and Consumer Surplus



Fares and sales series are averaged within a departure market and normalized to the first departure date in our data. Consumer Surplus (CS) is computed as the inclusive value of inside goods for Leisure (L) type passengers, averaged across the selling horizon, and is based on realized price and attribute levels. This series is also normalized to the first departure date in the data.

the entire market is composed of perfectly price inelastic consumers. While the strategy of dropping observations where sales are equal to zero provides non-zero estimates of own-price elasticities, they remain unreasonably inelastic.

One alternative to our daily measurement of demand would aggregate sales over time to reduce the zero-sales frequency. In this empirical context, such aggregation would smooth over price changes. We find that estimating weekly demand also produces extremely inelastic demand. Across all nine models estimated, only the Poisson RC model produces demand estimates where product elasticities do not bunch at zero.

5.4 Decomposing the Drivers of Demand Variability

In Figure 9(a), we plot the demeaned distributions of arrival rates and mean utilities under average prices. There is considerable variation in both series reflecting the heterogeneity in demand across departure dates in the sample. In panel (b), we plot the normalized mean of the departure date component to arrivals and product shares at average prices across departure dates. There are two noticeable patterns. First, there is significant variability in the series across departures. Second, both series co-move. Departure dates with a larger volume of arriving consumers have, on average, a higher willingness to pay through the

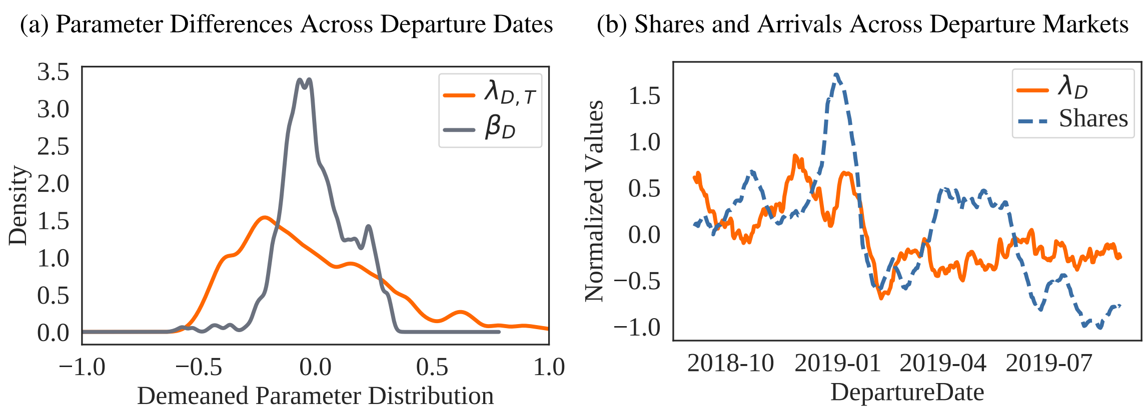
Table 3: Own-Price Elasticities Across Models with Zero Share Adjustments

Zeros	Market Size	Aggregation	Mean	Std. Dev.	Median	5th	95th
Drop	Calibrated	Disaggregated	0.000	0.000	0.000	0.000	0.000
Drop	Calibrated	Aggregated	0.000	0.000	0.000	0.000	0.000
Drop	Observed	Disaggregated	-0.083	0.030	-0.077	-0.136	-0.037
Drop	Observed	Aggregated	-0.001	0.035	-0.000	-0.000	-0.000
Adjusted	Calibrated	Disaggregated	0.000	0.000	0.000	0.000	0.000
Adjusted	Calibrated	Aggregated	0.000	0.000	0.000	0.000	0.000
Adjusted	Observed	Disaggregated	0.000	0.000	0.000	0.000	0.000
Adjusted	Observed	Aggregated	-0.000	0.000	-0.000	-0.000	-0.000
Poisson Random Coefficients			-1.215	0.433	-1.162	-2.007	-0.555

The table presents summary statistics for the realized own-price elasticities from estimating the model as in Berry, Carnall, and Spiller (2006), ignoring the arrival process and employing zero share adjustments to the empirical shares. The first set of adjustments involve the handling of empirical share observations; "Drop" indicates an adjustment where the zero observations are dropped and $\varepsilon = 1e^{-6}$ denotes replacing the zero observations with a small number. The second set of adjustments involve selecting the market size in order to calculate the empirical market shares. The first method is to use the search data series as the values for the arrival process, the second is to calibrate the market size to be 40 for all observations, and the last method involves uses the total number of observed arrivals aggregated across different days from departure. The final row summarizes posterior mean of price elasticities for our estimator.

week fixed effects.

Figure 9: Seasonal Parameter Variation



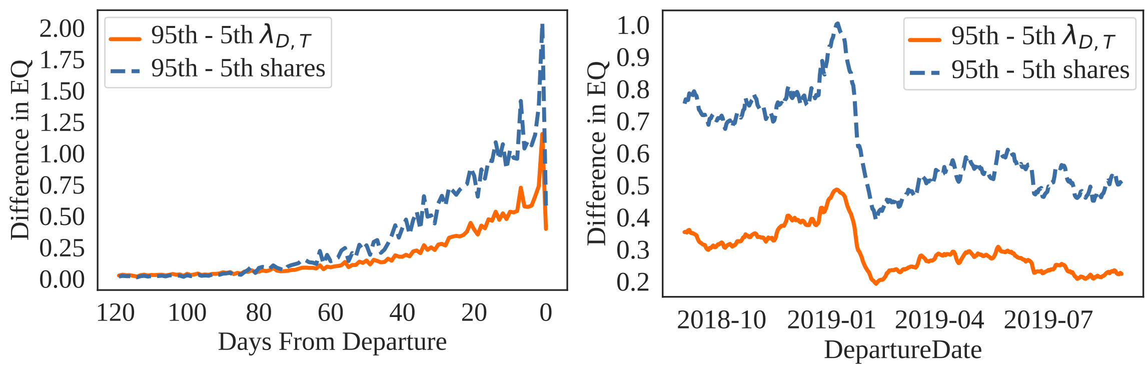
Note: Panel (a) plots the distribution of normalized parameter values across weeks of the year for arrival and preference parameters. For arrival parameters, this distribution abstracts from the variation in arrivals towards the departure date, since our parameterization restricts this shape to be identical across departure dates. Panel(b) plots a 30 day moving average of average normalized product shares (computed at average prices) and normalized arrival rates across markets.

Using our demand estimates, we decompose the variability of demand over time due to the changing intensity of demand (arrivals) versus willingness to pay (preferences). In

Figure 10(a), we compute the change in expected demand driven by preferences by moving from the 5th to the 95th percentile in shares along days from departure, holding prices and arrivals constant. We conduct a similar exercise for arrivals, where we hold prices, unobserved quality, and preferences fixed. In Figure 10(b), we perform the same calculations across departure dates.

Figure 10: Inter-Quantile Range for Variation in Arrivals vs Shares

(a) IQR Across Days from Departure (b) IQR Across Departure Dates



Note: Panel (a) plots the 30-day moving average difference in expected sales from 5th percentile to the 95th percentile of one parameter, holding all others constant. Letting $Q_{(E[q]\|\cdot)}(0.05)$ $Q_{(E[q]\|\cdot)}(0.95)$ represent the expected percentile functions. The solid orange line, plots $Q_{(E[q_{j,t,m}]\|s_{j,t,m},p)}(0.95) - Q_{(E[q_{j,t,m}]\|s_{j,t,m},p)}(0.05)$ averaged across days from departure. The dashed blue line plots $Q_{(E[q_{j,t,m}]\|\lambda_{t,m},p)}(0.95) - Q_{(E[q_{j,t,m}]\|\lambda_{t,m},p)}(0.05)$ averaged across days from departure. Panel(b) repeats the exercise across travel departure markets. All points are evaluated at average prices.

If changes in willingness to pay and the intensity of demand were of equal importance in explaining the variation in demand across days from departure and departure dates, the arrival (blue) and share (lines) would coincide. We note two regularities. First, arrival and preference variation are correlated - as we might expect based on the correlation in parameters noted above. Second, in this market, the sales variation induced by product shares is always greater than the variation from arrivals. Our decomposition suggests on average sales variability explained by changes in preferences is 25% more than the variation explained by arrivals. Preferences variability is highest close to the departure date. Decomposing this variation across departure dates (panel b), we find that changes to preferences account for two times more variability in expected demand across the selling period within departure dates than changes in arrivals. That is, two-thirds of the variation in demand over time is due to changes in willingness to pay. Preference changes thus are more important than market size changes in explaining cyclical demand for air travel.

Unlike in grocery markets (Chevalier, Kashyap, and Rossi, 2003), we find that periods of peak demand also have higher prices. Our findings complement the work of Einav (2007) in the movie industry where two compounding forces drive cyclicality. In movies, peak demand corresponds to periods where consumers have higher willingness to pay and firms release movies with higher quality. For the airline industry, we find that both the intensity of demand and willingness to pay move in the same directions over the calendar year. However, as previously noted, our airline data show that on the supply side, both capacities and prices also respond upward in periods of peak demand (see Figure 5). Absent these supply-side responds, cyclicality would be even higher as in the movie industry case.

6 Conclusion

We propose a new method to estimate product-level demand with small market sizes. Our approach allows for many zero sale observations, endogenous prices, and rich unobserved consumer heterogeneity. We derive a Bayesian IV estimator to recover random coefficients logit demand parameters with Poisson arrivals. We show through simulation studies that this method can outperform typical zero-sales adjustments and provide good coverage, even in very small markets or under a misspecified pricing function.

Our approach can be applied to many settings where granular demand estimates are necessary in order to evaluate counterfactuals or address firms' decision making. The key data requirements in our approach are traditional market-level data as well as measures of consumer search intensity. These data are becoming increasingly available to researchers, with relevant applications not only including e-commerce—as we demonstrate with our empirical application—but also offline markets where a notion of arrivals are obtainable.

References

Adam, H., P. He, and F. Zheng (2020): "Machine Learning for Demand Estimation in Long Tail Markets," Working Paper.

Berry, S., M. Carnall, and P. Spiller (2006): "Airline hubs: costs, markups and the implications of customer heterogeneity," *Advances in airline economics*, 1, 183–214.

Berry, S., and P. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82(5), 1749–1798.

——— (2016): "Identification in Differentiated Products Markets," *Annual Review of Economics*, 8, 27–52.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, pp. 242–262.

Berry, S. T., O. B. Linton, and A. Pakes (2004): "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *The Review of Economic Studies*, 71(3), 613–654.

Blackwell, D., J. B. MacQueen, et al. (1973): "Ferguson distributions via Pólya urn schemes," *The annals of statistics*, 1(2), 353–355.

Buchholz, N. (2021): "Spatial Equilibrium, Search Frictions, and Dynamic Efficiency in the Taxi Industry," *The Review of Economic Studies*, rdab050.

Burda, M., M. Harding, and J. Hausman (2012): "A Poisson mixture model of discrete choice," *Journal of econometrics*, 166(2), 184–203.

Chevalier, J. A., A. K. Kashyap, and P. E. Rossi (2003): "Why don't prices rise during periods of peak demand? Evidence from scanner data," *American Economic Review*, 93(1), 15–37.

Conlon, C. T., and J. H. Mortimer (2013): "Demand Estimation under Incomplete Product Availability," *American Economic Journal: Microeconomics*, 5(4), 1–30.

Dubé, J.-P., A. Hortaçsu, and J. Joo (2021): "Random-coefficients logit demand estimation with zero-valued market shares," *Marketing Science*.

Einav, L. (2007): "Seasonality in the US motion picture industry," *The Rand journal of economics*, 38(1), 127–145.

Gandhi, A., and J.-F. Houde (2015): "Measuring Substitution Patterns in Differentiated Products Industries - The Missing Instruments," Working paper.

Gandhi, A., Z. Lu, and X. Shi (2019): "Estimating Demand for Differentiated Products with Zeroes in Market Share Data," Working Paper.

Honka, E. (2014): "Quantifying search and switching costs in the US auto insurance industry," *The RAND Journal of Economics*, 45(4), 847–884.

Hortaçsu, A., O. R. Natan, H. Parsley, T. Schwieg, and K. R. Williams (2021): "Organizational structure and pricing: Evidence from a large us airline," Discussion paper.

Jiang, R., P. Manchanda, and P. E. Rossi (2009): "Bayesian Analysis of Random Coefficient Logit Models using Aggregate Data," *Journal of Econometrics*, 149(2), 136–148.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010): "Online demand under limited consumer search," *Marketing science*, 29(6), 1001–1023.

Murphy, K. P. (2007): "Conjugate Bayesian analysis of the Gaussian distribution," Working Paper.

Quan, T. W., and K. R. Williams (2018): "Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail," *The RAND Journal of Economics*, 49(4), 877–913.

Rossi, P. E. (2014): *Bayesian Non- and Semi-parametric Methods and Applications*. Princeton University Press.

Rossi, P. E., and G. M. Allenby (1993): "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30(2), 171–182.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2012): *Bayesian Statistics and Marketing*. John Wiley & Sons.

Weitzman, M. L. (1979): "Optimal search for the best alternative," *Econometrica:*, pp. 641–654.

Williams, K. R. (2021): "The Welfare Effects of Dynamic Pricing: Evidence from Airline Markets," *Econometrica*.

A Estimation Routine

A.1 The Dirichlet process Prior for (ξ, υ)

We use a Dirichlet process prior to allow for an arbitrary number of distributions to be mixed together to approximate the joint distribution of (ξ, υ). Rather than provide a rigorous treatment of the Dirichlet process prior, we state its properties and detail the algorithm used to implement the method within our MCMC algorithm.

The Dirichlet process can be viewed as a distribution over distributions. Each residual pair (ξ, υ) is drawn from some distribution, and the Dirichlet process clusters these distributions appropriately. There are two components to the process: $\tilde{\alpha}$, the tightness parameter, and G_0, a prior distribution for each residual pair. Draws from this process are then distributions of the residuals, which we condition for the rest of estimation. As such, each residual pair has its own distribution, indexed by parameters θ_n; many of which will be equal. The details of the distribution θ_n are given in the following section.

The prior distribution G_0 gives the distribution of a new cluster, unconditional on any data. It must be diffuse enough to cover the data, but not so diffuse that the likelihood of drawing it is too low. As with any non-parametric estimator, care must be taken to choose priors such that we approximate the errors well. We choose a normal distribution for G_0 so that

$$G_0 \sim \mathcal{N}(\mu, \Sigma).$$

We impose standard conjugate priors for computational ease, so

$$\mu | \Sigma \sim \mathcal{N}(0, {a_\mu}^{-1} \Sigma),$$

and

$$\Sigma \sim IW(\nu, \nu \upsilon I),$$

where prior parameters ν determines the tightness of the Inverse-Wishart distribution, and a_μ determines the scale of variance of means. We allow prior parameter υ to determine the location via

$$\text{mode}(\Sigma) = \frac{\nu}{\nu+2} \upsilon I.$$

We choose the tightness parameter $\tilde{\alpha}$ to be constant, based on the data. Rossi (2014) provides a more in-depth look at the choice of hyper-parameters, including treatments where $\tilde{\alpha}$ is random, determined by data, and a further parameterization of a, ν, υ. We omit these for simplicity.

We are interested in the conditional distribution of θ. We follow the Blackwell-MacQueen Pólya Urn representation (Blackwell, MacQueen, et al., 1973) where

$$\theta_n | \theta_{-n} \sim \frac{\tilde{\alpha} G_0 + \sum_{j \neq n} 1_{\theta_j}}{\tilde{\alpha} + N - 1},$$

which is a mixture distribution between G_0 and 1_{θ_j}, a point-mass located at θ_j. Given other draws from the Dirichlet process, a new draw has a positive probability of drawing the same parameters as previous draws, and $\frac{\tilde{\alpha}}{\tilde{\alpha}+N-1}$ probability of a new draw from the distribution G_0. By drawing the classifiers from a Dirichlet process, there is always a possibility of introducing a new normal distribution for each data point, but this is disciplined by the choice of the prior distribution, which has a tendency to cluster observations together.

The Dirichlet process produces a prior that is very similar to what one might select for a finite mixture of normal model, but with two substantial differences. The number of mixtures is changing with every step of the chain, and there is a positive probability of adding another mixture component. The prior probability of adding another mixture to the model is governed by the hyperparameter $\tilde{\alpha}$. In practice, this means that for K existing clusters, the prior probability of the n^{th} data point being drawn from the existing clusters is $\frac{N_k}{\tilde{\alpha}+N-1}$, where N_k is the number of data points currently in cluster K. The prior probability of a new cluster is $\frac{\tilde{\alpha}}{\tilde{\alpha}+N-1}$.

Conditional on θ_n, the residuals are distributed bivariate normal, a key fact used in estimation. Because of the tendency of the Dirichlet process to cluster distributions together, we index these clusters by $\kappa_{j,t}$, which is sufficient for θ_n. All residual pairs points where κ_n

are equal share the same distribution, and conditional on κ are distributed bivariate normal.

A.2 Markov Chain Monte Carlo Details

A.2.1 Sampling Arrival Parameters

To update the parameters describing the arrival rate of searching consumers, we use search and quantity data. We define the likelihood to be the joint probability of observing A_t and $q_{j,t}$, conditional on $s_{j,t}$. Arrivals are distributed Poisson. Conditional on shares, we split the arrival process (with rate λ_t) by the shares to obtain the distribution of quantities sold. Each purchase is drawn from a Poisson distribution with rate $\lambda \cdot s_{j,t}$.

Because data on arrivals may be sparse—perhaps only a single data point per market—we suggest parameterizing the arrival rate with a series of fixed effects whenever possible,

$$\lambda_t = \exp(W_t \tau),$$

where W is a full rank matrix composed of 0 and 1s. Other specifications are possible.

Arrivals are distributed Poisson,

$$A_t \sim \text{Poisson}(\lambda_t). \tag{1}$$

Note that purchase quantities also depend on arrivals. Using the properties of the Poisson distribution, we have

$$q_{j,t} \sim \text{Poisson}(\lambda_t s_{j,t}). \tag{2}$$

We note that a conjugate prior choice for λ_t would be the Gamma distribution, but as our parameter of choice is τ_q, we impose a log-Gamma prior such that $\exp(\tau_q) \sim \Gamma(k, \zeta)$. Therefore, the posterior distribution of $\exp(\tau_q)$ is then given by

$$\exp(\tau_q) \sim \Gamma\left(\sum_{t\in Q}\left(A_t + \sum_j q_{j,t}\right) + k, \frac{\zeta}{1+\zeta\left(\sum_{t\in Q}(2-s_{0,t})\right)}\right).$$

A.2.2 Sampling Shares and Utility Parameters

Updating shares. The Dirichlet process allows for complex distributions of (ξ, υ) to be approximated by a series of normal distributions through a component classifier κ. Conditional on this classifier, each pair of residuals (ξ, υ) are distributed bivariate normal. We apply the standard treatment of simultaneity by conditioning on the variance structure of the normal and the respective residuals. The following sections condition upon κ and derive the sampler for a multivariate normal joint distribution of the demand shock and pricing residual. In the final sections we discuss sampling the classifier and the component means and variances.

Conditional on $\beta, \Gamma, \kappa, \mu, \Sigma$, and υ, the shares are an invertible function of ξ. The conditional distribution of ξ is also normal, which implies a distribution of shares. We compute the likelihood of any particular set of share draws by inverting the demand system for these shares. We derive a distribution of shares via a standard change of variables theorem.

Since ξ is assumed to be correlated with price, we follow the Bayesian framework for simultaneity with discrete choice models (Rossi and Allenby, 1993; Jiang, Manchanda, and Rossi, 2009; Rossi, Allenby, and McCulloch, 2012). Using a set of exogenous and relevant instruments $Z_{t,d}$, we assign

$$\left.\begin{aligned} \xi_{j,t} &= f^{-1}\left(s_{j,t} \,|\, \beta, \Gamma, X_t\right) \\ \upsilon_{j,t} &= p_{j,t} - Z'_{j,t}\eta \end{aligned}\right\} |\kappa = k \sim \mathcal{N}_{\text{iid}}(\mu_k, \Sigma_k) \quad \text{such that} \quad \Sigma_k = \begin{pmatrix} \sigma^2_{k,11} & \rho_k \\ \rho_k & \sigma^2_{k,22} \end{pmatrix}.$$

For notational parsimony, we omit the conditioning statement, but note that each function is implicitly conditioned on the other demand parameters. We refer to the share equation as $s_{j,t,d} = f(\xi_{j,t,d})$. Since f is invertible, the density of $s_{j,t,d}$ is given by

$$f_{s_{j,t}}(x) = f_{\xi_{j,t}}\left(f^{-1}(x)\right) \cdot \left|\mathscr{J}_{\xi_{j,t} \to s_{j,t}}\right|^{-1}.$$

With this notation, $\mathscr{J}_{\xi_{j,t} \to s_{j,t}}$ represents the Jacobian matrix of model shares with respect to ξ and $\left|\mathscr{J}_{\xi_{j,t} \to s_{j,t}}\right|^{-1}$ denotes the inverse of the determinant of the Jacobian.

Since υ and ξ are assumed to be jointly normal, knowing υ provides information about the magnitude of the demand shock. This joint normality does not factor into the Jacobian of the shares distribution, because neither $s_{j,t}$ or $\xi_{j,t}$ are in the pricing equation and it is assumed it to be a linear system. However, we must use the correct conditional distribution for ξ. Conditioning on both η and Σ is enough to pin down the the correlation structure between ξ and υ, and to "observe" υ as well. Drawing on the structure of the bivariate normal distribution, we have

$$\xi|\upsilon,\kappa=k\sim\mathcal{N}\left(\mu_{k,2}+\frac{\rho_k\upsilon}{\sigma^2_{k,11}},\sigma^2_{k,22}-\frac{\rho^2_k}{\sigma^2_{k,11}}\right),$$

where

$$\begin{pmatrix}\xi\\ \upsilon\end{pmatrix}|\kappa=k\sim\mathcal{N}(\mu_k,\Sigma_k),\quad \Sigma_k=\begin{pmatrix}\sigma^2_{k,11} & \rho_k\\ \rho_k & \sigma^2_{k,22}\end{pmatrix}.$$

One interpretation of this treatment of simultaneity is that price gives information about the realized demand shock ξ, so the conditional distribution of ξ is higher or lower depending on the unobserved component υ that influences price.

The conditional distribution of shares is then given by

$$\prod_{j=1}^{J(t)}\left[\phi\left(\frac{f^{-1}(s_{j,t})-\frac{\rho_k\upsilon}{\sigma^2_{k,11}}}{\sqrt{\sigma^2_{k,22}-\frac{\rho^2_k}{\sigma^2_{k,11}}}}\right)\right]\cdot\left|\mathscr{J}_{\xi_t\to s_t}\right|^{-1},$$

where $\phi(\cdot)$ is the standard normal density function.

Shares directly shape the distribution of sales. The distribution of purchases is a split Poisson process given by

$$q_{j,t}\sim\text{Poisson}(\lambda_t s_{j,t}).$$

Since the Poisson draw is only dependent on the demand parameters through the shares, $q_{j,t}$ is conditionally independent of ξ. Thus the likelihood of a particular market's shares is given by the product of the density of ξ and the mass function of $q_{j,t}$, given by

$$\ell(s_{\cdot,t}) = \prod_{j=1}^{J(t)} \left[\phi \left(\frac{f^{-1}(s_{j,t}) - \frac{\rho_k \upsilon}{\sigma_{k,11}^2}}{\sqrt{\sigma_{k,22}^2 - \frac{\rho_k^2}{\sigma_{k,11}^2}}} \right) \frac{(\lambda_t s_{j,t})^{q_{j,t}} \exp(-\lambda_t s_{j,t})}{q_{j,t}!} \right] \cdot \left| J_{\xi_t \to s_t} \right|^{-1}. \qquad (3)$$

The posterior likelihood is constructed by taking the product of the each market's inversion multiplied by the likelihood contribution of each product's quantity sold. There is no conjugate prior distribution, so we sample from the posterior using a Metropolis-Hastings step.

Our candidate distribution for share draws is a transformation of a normal distribution added to ξ. This allows for easy tuning of the candidate distribution via the variance of the normal. However, as a complication, the candidate distribution is not reversible. That is $q(a|b) \neq q(b|a)$. As a result, we reweight the Metropolis-Hastings step according to the implied p.d.f. to make the chain reversible.

Updating distribution of consumer types, Γ We use a random coefficients demand specification, where demand parameters can be grouped into nonlinear and linear parameters. We order these demand parameters such that the first L parameters are distributed normally, and the remaining $K-L$ are constant across consumers. That is,

$$u_{i,j,t} = x_{j,t}\beta_i + \xi_{j,t} + \epsilon_i$$

and

$$\beta_i = \begin{pmatrix} \bar{\beta}_{1:L} + & \Gamma U_i \\ \bar{\beta}_{L+1:K} & \end{pmatrix},$$

where $U_i \sim \mathcal{N}(0, I_L)$, and Γ is the Cholesky decomposition of a variance matrix. This allows for a flexible covariance between the demand parameters with random coefficients, while maintaining linearity in K parameters. We use the Cholesky decomposition for computational simplicity.

We sample from the posterior distribution of nonlinear demand parameters Γ with a

Metropolis-Hastings step. The distribution of ξ remains unchanged, and we evaluate a candidate Γ in a similar manner as to drawing shares, but without incorporating the likelihood of purchases.

The likelihood of a particular Γ is constructed from the implied distribution of the demand shock ξ from inverting the demand system. The likelihood of the shares, given Γ, is given by

$$s_{j,t}|\upsilon_{j,t},\kappa_{j,t},\Gamma \sim \mathcal{N}\left(\frac{\mu_{\kappa_{j,t},2}+\rho_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},11}}\upsilon,\sigma^2_{\kappa_{j,t},22}-\frac{\rho^2_{\kappa_{j,t}}}{\sigma_{\kappa_{j,t},11}}\upsilon_{j,t}\right)\mathscr{J}^{-1}_{\xi\to s}. \tag{4}$$

We use a short-hand distribution here of a distribution times the Jacobian to mean that the p.d.f. of $\xi_{\cdot,t}$ (evaluated at a set of shares $s_{\cdot,t}$) is the p.d.f. of a normal distribution with those parameters multiplied by the determinant of the Jacobian.

However, a candidate Γ cannot be drawn in a trivial manner, as we must sample from the set of Cholesky decompositions of positive-definite (variance) matrices. We employ the parameterization suggested by Jiang, Manchanda, and Rossi (2009), which lets

$$\Gamma_{ij} = \begin{cases} \exp(r_{ij}), & \text{for } i = j \\ r_{ij}, & \text{for } i < j \\ 0 & \text{otherwise .} \end{cases}$$

This enforces a strictly positive diagonal upper-triangular matrix for any candidate draw r.

We have given the likelihood of the demand residual, to complete the posterior likelihood of Γ, we must also define a prior distribution over Γ. Following Jiang, Manchanda, and Rossi (2009), we impose normal priors over each r. Jiang, Manchanda, and Rossi (2009) explore the implications of this prior specification: $r_{ij} \sim \mathcal{N}(0,\psi^2_{ij})$.

The posterior distribution of Γ is given by

$$\prod_{j=1}^{J(t)}\left[\phi\left(\frac{f^{-1}(s_{j,t})-\mu_{\kappa_{j,t},2}-\frac{\rho}{\sigma^2_{11}}\upsilon_{j,t}}{\sigma_{\kappa_{j,t},2|1}}\right)\right]|J_{\xi_{j,t}\to s_{j,t}}|^{-1}\times\prod_{i\leq j}\frac{1}{\psi_{ij}}\phi(r_{ij}). \tag{5}$$

For alternative utility specifications, the same procedure can be used. Note that only a few non-linear parameters may be estimated in a single step, as a Metropolis-Hastings search in a high dimension traverses the stationary distribution slowly.

Updating type-invariant parameters, $\bar{\beta}$. To sample from the linear demand parameters, we define $\delta_{j,t}$ such that

$$\delta_{j,t} = x_{j,t}\bar{\beta} + \xi_{j,t}. \tag{6}$$

Since $\xi_{j,t}$ has a normal distribution and we impose a normal prior on $\bar{\beta}$, we have a standard Bayesian linear regression after we account for the influence of the pricing residual, and the different variances in each element of $\xi_{j,t}$. We accomplish this by normalizing each component of equation (6) by subtracting the expected value of $\xi_{j,t}$ and diving both sides by the standard deviation. We then perform a Bayesian linear regression on this collection of normalized equations, as these rescaled errors have unit variance. Let $\sigma_{k,2|1} = \sqrt{\sigma^2_{k,22} - \frac{\rho_k^2}{\sigma^2_{k,11}}}$ be the variance of ξ conditional on υ and Σ.

$$\frac{\delta_{j,t} - \mu_{\kappa_{j,t},2} - \frac{\rho_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},11}}\upsilon}{\sigma_{\kappa_{j,t},2|1}} = \frac{1}{\sigma_{\kappa_{j,t},2|1}} x_{j,t}\bar{\beta} + U^{\beta}_{j,t}, \tag{7}$$

where $U^{\beta} \sim \mathcal{N}(0,1)$.

We follow the typical conjugate prior distribution for a linear regression—$\bar{\beta} \sim \mathcal{N}(\bar{\beta}_0, V_0)$. The posterior distribution is then a shrinkage estimator of OLS.

Let

$$\hat{X}_{j,t} = \frac{x_{j,t}}{\sigma_{\kappa_{j,t},2|1}},$$

and

$$\hat{\delta} = \frac{\delta_{j,t} - \mu_{\kappa_{j,t},2} - \frac{\rho_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},11}}\upsilon}{\sigma_{\kappa_{j,t},2|1}}.$$

Then the posterior distribution of $\bar{\beta}$ is $\bar{\beta} \sim \mathcal{N}(\beta_N, V_N)$, where

$$\beta_N = (\hat{X}'\hat{X} + V_0^{-1})^{-1}\left(V_0^{-1}\beta_0 + \hat{X}'\hat{\delta}\right),$$

and

$$V_N = (V_0^{-1} + \hat{X}'\hat{X})^{-1}.$$

A.2.3 Sampling Price-Endogeneity Parameters

Updating pricing equation, η. The pricing equation is given by

$$p_{j,t} = Z_{j,t}\eta + \upsilon_{j,t}. \tag{8}$$

Conditional on shares, Γ, and $\bar{\beta}$, ξ is known, so we use the conditional distribution of υ given ξ to perform another Bayesian linear regression in the same manner as $\bar{\beta}$. We impose a Normal prior, subtract the expected value and divide by the conditional variance.

Define $\sigma_{\kappa_{j,t},1|2} = \sqrt{\sigma^2_{\kappa_{j,t},11} - \frac{\rho^2_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},22}}}$. Then

$$\frac{p_{j,t} - \mu_{\kappa_{j,t},1} - \frac{\rho_{\kappa_{j,t}}}{\sigma_{\kappa_{j,t},22}}\xi_{j,t}}{\sigma_{\kappa_{j,t},1|2}} = \frac{1}{\sigma_{\kappa_{j,t},1|2}} x_{j,t}\bar{\eta} + U^{\eta}_{j,t}. \tag{9}$$

After this normalization, $U^{\eta}_{j,t}$ is a standard normal error term. We draw from η using a standard Gibbs-Sampler draw from a linear regression with unit variance, which is the same process as used for $\bar{\beta}$.

Updating Component Classifier Using the properties of the Dirichlet process, the prior probability of each cluster is weighted by the likelihood of each data point being sampled from the cluster. The posterior distribution of θ is

$$\theta_n|\theta_{-n}, \upsilon_{j,t}, \xi_{j,t}\tilde{\alpha} \sim \frac{q_0 G_0 + \sum_{i\neq n} q_i 1_{\theta_i}}{q_0 + \sum_{i\neq n} q_i},$$

where

$$q_i = \frac{1}{\tilde{\alpha}+N-1} \Pr((\upsilon_{j,t}, \xi_{j,t})|\theta_i) \text{ for } i \neq 0,$$

and

$$q_0 = \frac{\tilde{\alpha}}{\tilde{\alpha}+N-1} \int \Pr((\upsilon_{j,t}, \xi_{j,t})|\theta_i) G_0(d\theta_i).$$

This is a mixture distribution with weights q_0 for a new cluster, and $\pi_i q_i$ for existing clusters, where π_i is the sum of data points in cluster i divided by the total data points. While q_i presents a similar form as a finite mixture model, q_0 is difficult to calculate. Because we assume $G_0 \sim \mathcal{N}$, q_0 is the prior predictive distribution, i.e. the likelihood of a data point over the distribution of possible normal distributions θ_i might take on. As shown in Murphy (2007), this quantity is known to be distributed multivariate t. Applying our priors, the form is given by:

$$\int \Pr(x|\theta_i) G_0(d\theta_i) \sim t_{\nu-1}(0, \frac{\frac{1}{\nu v} I(a_\mu + 1)}{a_\mu(\nu - 1)}).$$

We can evaluate the p.d.f. of θ at each of the residuals to determine the posterior probability of adding a new cluster.

It is important to draw a connection between θ_n and κ, the component classifier. There are at most n unique values of θ_n, and usually far less due to the clustering nature of the Dirichlet process. κ is then drawn from a categorical distribution, with weights q_0 for a new cluster, and $N_k q_k$ for each cluster k. The number of unique values of θ_n is constantly changing, so the size of κ must be adjusted whenever θ changes in every estimation step.

If a new cluster is drawn from the categorical distribution, we must know what distribution to sample. The prior distribution of a new cluster is G_0, but since the residual pair belongs to the cluster, we sample from its posterior distribution. This is the same process as sampling means and variance for a finite mixture basis that contains only a single point—a multivariate Bayesian linear regression. We draw from its posterior distribution in the standard way. Let $Y_k = (\upsilon_k, \xi_k)$, which is the residual pair for the new cluster, the posterior distribution of component variance and mean are:

$$\Sigma_k \sim IW(\nu+1, V+S)$$
$$\mu_k|\Sigma_k \sim \mathcal{N}(\tilde{\mu}, \frac{1}{1+a_\mu}\Sigma_k),$$

with

$$S = \left(Y_k - \iota\bar{\mu}_k'\right)'\left(Y_k - \iota\bar{\mu}_k'\right), + a_\mu\left(\tilde{\mu}_k - \bar{\mu}\right)'\left(\tilde{\mu}_k - \bar{\mu}\right),$$
$$\tilde{\mu} = (1+a_\mu)^{-1}(\bar{y}_k + a_\mu\bar{\mu}),$$

and

$$\bar{y}_k = Y_k'\iota,$$

where ι is a corresponding length vector of all ones.

To combine all of the above steps, we present the following algorithm for updating the component classifier κ.

Algorithm 2 Drawing Component clusters under a DP prior

```
for n = 1 to N do
    Compute probability of new cluster, q_0, for residual pair n
    for k = 1 to K do
        Compute Bayes Factor q_k.
    end for
    Draw classifier κ_n ~ Multinomial(q)
    if q == 0 then
        Draw cluster mean, μ_{K+1} and variance, Σ_{K+1}
        Update K = K + 1
    end if
    Check if a cluster has been orphaned. Adjust K
end for
```

Updating the Component Distributions, Σ_K and μ_k. Conditional on $\kappa_{j,t}$, each pair of residuals is known to come from a particular component of the mixture normal. If we only consider the residual pairs drawn from a particular component k, then it is as if all of the residuals are drawn from the same distribution, and the standard Inverse-Wishart parameterization can be used to draw the variance parameters. We follow that procedure here for each component, with an extra step to allow for each component to have a different mean parameter as well.

Since there is no intercept in the demand parameters, there is an extra degree of freedom in this problem that we use to sample as a mean for each component bivariate normal distribution. We sample from this mean using a multivariate regression with only a constant, since each component distribution is normal. Some care must be made since the residuals are not independent, so we use a Bayesian multivariate regression to correctly sample from their joint distribution. To utilize the standard Bayesian machinery for such a regression, we impose standard (normal) priors to exploit conjugate priors. For any component k, the variance Σ_k has an Inverse Wishart prior $IW(\nu, V)$ and the mean $\mu_k|\Sigma_k$ has a normal prior distribution $\mathcal{N}(\bar{\mu}, a_{\mu}^{-1}\Sigma_k)$. Define the vector $Y_k = (\upsilon, \xi_k)$, which is only the collection of residual pairs such that $\kappa_{j,t} = k$. We can write

$$Y_k = \iota\mu' + U,$$

where

$$U \sim \mathcal{N}(0, \Sigma_k).$$

The posterior covariance and conditional mean of the components are then

$$\Sigma_k \sim IW(\nu + n_k, V + S)$$

and

$$\mu_k | \Sigma_k \sim \mathcal{N}(\tilde{\mu}, \frac{1}{n_k + a_\mu}\Sigma_k),$$

where we define

$$S = \left(Y_k - \iota\bar{\mu}_k'\right)'\left(Y_k - \iota\bar{\mu}_k'\right) + a_\mu\left(\tilde{\mu}_k - \bar{\mu}\right)'\left(\tilde{\mu}_k - \bar{\mu}\right),$$

$$\tilde{\mu} = (n_k + a_\mu)^{-1}(n_k\bar{y}_k + a_\mu\bar{\mu}),$$

and

$$\bar{y}_k = \frac{1}{n_k}Y_k'\iota.$$

The vector ι is a corresponding length vector of all ones, and n_k is the number of observations in cluster k. This is repeated for each component k.

B Additional Monte Carlo Results

Table 4: Monte Carlo Results for $\lambda = 25$, Poisson RC and BLP with Observed Arrivals

	Poisson RC	BLP, Drop 0's		BLP, Adjust 0's	
		Disagg.	Agg.	Disagg.	Agg.
α	-0.10	-0.01	-0.32	-4.66	-0.92
	(-0.33, 0.11)	(-3.48, 0.96)	(-5.66, 1.18)	(-18.27, -0.84)	(-8.77, 1.19)
β_1	-0.02	-0.36	-0.89	-1.54	-1.12
	(-0.29, 0.26)	(-2.14, 0.16)	(-6.81, 0.44)	(-14.87, 1.23)	(-8.54, 0.75)
β_2	-0.02	-0.16	-0.76	-2.33	-1.20
	(-0.31, 0.25)	(-2.44, 0.38)	(-5.92, 0.40)	(-16.15, 0.44)	(-9.50, 0.45)
Γ_{11}	-0.03	0.76	-0.04	1.19	0.19
	(-0.12, 0.21)	(-1.26, 4.80)	(-1.94, 2.55)	(-2.06, 6.85)	(-1.80, 4.19)
Γ_{22}	0.04	0.05	0.46	0.69	0.19
	(-0.08, 0.30)	(-2.16, 3.89)	(-3.25, 7.24)	(-7.81, 10.74)	(-7.42, 7.18)
Γ_{33}	0.04	0.12	0.20	0.68	0.23
	(-0.07, 0.41)	(-1.77, 4.45)	(-3.99, 5.47)	(-7.79, 10.41)	(-6.84, 7.26)

Note: Reported in the table are the mean, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 1000 simulations. Column "BLP, Drop 0s, Disagg" drops observations with zero empirical shares $s = q/A$. Column "BLP, Drop 0s, Agg" aggregates sales and arrivals across 10 adjacent observations, then drops observations with zero aggregate shares. Covariates and instruments are averaged across observations. Column "BLP, Adjust 0s, Disagg" replaces zero empirical shares with a small number $\varepsilon = 1e^{-6}$. Column "BLP, Adjust 0s, Agg" aggregates sales and arrivals across 10 adjacent observations, then replaces zero aggregate shares with $1e^{-6}$. Covariates and instruments are averaged across observations.

Table 5: Monte Carlo Results for $\lambda = 5$, Poisson RC and BLP with Observed Arrivals

	Poisson RC	BLP, Drop 0's		BLP, Adjust 0's	
		Disagg.	Agg.	Disagg.	Agg.
α	-0.17	1.18	0.26	-5.17	-3.78
	(-0.54, 0.17)	(-0.38, 1.67)	(-3.47, 1.67)	(-23.34, -0.36)	(-19.34, 1.41)
β_1	-0.01	-0.36	-0.84	-1.60	-1.77
	(-0.50, 0.48)	(-0.90, 0.10)	(-6.42, 0.26)	(-13.76, 0.99)	(-16.12, 2.50)
β_2	-0.02	0.12	-0.56	-2.33	-2.43
	(-0.51, 0.44)	(-0.50, 0.53)	(-5.17, 0.49)	(-16.27, 0.58)	(-17.50, 1.36)
Γ_{11}	-0.02	0.41	-0.03	1.53	0.28
	(-0.12, 0.27)	(-1.30, 4.63)	(-1.09, 2.06)	(-2.41, 8.77)	(-3.53, 6.73)
Γ_{22}	0.05	-0.06	0.25	0.43	1.36
	(-0.07, 0.36)	(-0.92, 2.58)	(-2.67, 6.51)	(-6.02, 9.65)	(-10.00, 18.09)
Γ_{33}	0.04	-0.10	0.13	0.54	1.04
	(-0.07, 0.33)	(-1.88, 2.36)	(-4.15, 5.73)	(-6.04, 11.23)	(-6.75, 16.05)

Note: Reported in the table are the mean, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 1000 simulations. Column "BLP, Drop 0's, Disagg" drops observations with zero empirical shares $s = q/A$. Column "BLP, Drop 0's, Agg" aggregates sales and arrivals across 10 adjacent observations, then drops observations with zero aggregate shares. Covariates and instruments are averaged across observations. Column "BLP, Adjust 0's, Disagg" replaces zero empirical shares with a small number $\varepsilon = 1e^{-6}$. Column "BLP, Adjust 0's, Agg" aggregates sales and arrivals across 10 adjacent observations, then replaces zero aggregate shares with $1e^{-6}$. Covariates and instruments are averaged across observations.

Table 6: Monte Carlo Results for $\lambda = 25$, Poisson RC and Gandhi, Lu, and Shi (2019)

	Poisson RC	Calibrated		Observed	
		Disagg.	Agg.	Disagg.	Agg.
α	-0.17	1.66	2.15	1.89	2.29
	(-0.54, 0.17)	(1.43, 1.87)	(1.43, 2.56)	(1.48, 2.34)	(1.90, 2.70)
β_1	-0.01	-0.91	-0.74	-0.73	-0.63
	(-0.50, 0.48)	(-1.13, -0.69)	(-1.54, -0.31)	(-1.19, -0.30)	(-1.01, -0.21)
β_2	-0.02	-2.17	-2.98	0.10	-0.63
	(-0.51, 0.44)	(-2.34, -2.01)	(-3.49, -2.46)	(-0.37, 0.51)	(-1.23, -0.03)
Γ_{11}	-0.02	0.72	-0.18	0.74	-0.20
	(-0.12, 0.27)	(0.33, 0.85)	(-0.44, 0.21)	(0.31, 0.89)	(-0.20, -0.20)
Γ_{22}	0.05	0.83	0.02	0.79	-0.20
	(-0.07, 0.36)	(0.70, 1.01)	(-0.52, 1.06)	(0.49, 0.98)	(-0.20, -0.20)
Γ_{33}	0.04	0.83	0.07	0.79	-0.20
	(-0.07, 0.33)	(0.71, 1.00)	(-0.51, 1.17)	(0.54, 1.01)	(-0.20, -0.20)

Note: Reported in the table are the mean, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 1000 simulations. The first column reports our estimator. The column "GLS, Fixed, Disagg" reports results using disaggregate data, a fixed market size ($A = 40$), and the estimator in Gandhi, Lu, and Shi (2019) (GLS). The column "GLS, Fixed, Agg" shows results using averages covariates and total sales across 10 adjacent observations, and sets market size $A = 400$ using the GLS estimator. The column "GLS, Observed, Disagg" reports results using disaggregate data, observed market size, and the GLS estimator. The column "GLS, Observed, Disagg" shows results of the GLS estimator using aggregate observed market size and sales and average covariates across 10 adjacent observations.

Table 7: Monte Carlo Results for $\lambda = 5$, Poisson RC and Gandhi, Lu, and Shi (2019)

	Poisson RC	Calibrated		Observed	
		Disagg.	Agg.	Disagg.	Agg.
α	-0.10	1.81	2.30	1.88	2.35
	(-0.33, 0.11)	(1.54, 2.09)	(1.81, 2.66)	(1.48, 2.23)	(1.97, 2.70)
β_1	-0.02	-0.88	-0.72	-0.85	-0.69
	(-0.29, 0.26)	(-1.12, -0.64)	(-1.22, -0.37)	(-1.22, -0.48)	(-1.09, -0.35)
β_2	-0.02	-1.33	-1.00	-0.71	-0.38
	(-0.31, 0.25)	(-1.59, -1.10)	(-1.61, -0.37)	(-1.10, -0.34)	(-1.04, 0.29)
Γ_{11}	-0.03	0.79	-0.20	0.78	-0.20
	(-0.12, 0.21)	(0.56, 0.90)	(-0.21, -0.18)	(0.54, 0.93)	(-0.21, -0.19)
Γ_{22}	0.04	0.80	-0.07	0.79	-0.16
	(-0.08, 0.30)	(0.61, 0.92)	(-0.22, 0.90)	(0.54, 0.95)	(-0.21, 0.53)
Γ_{33}	0.04	0.81	-0.06	0.78	-0.15
	(-0.07, 0.41)	(0.65, 0.95)	(-0.29, 0.90)	(0.48, 0.95)	(-0.21, 0.61)

Note: Reported in the table are the mean, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 1000 simulations. The first column reports our estimator. The column “GLS, Fixed, Disagg” reports results using disaggregate data, a fixed market size ($A = 40$), and the estimator in Gandhi, Lu, and Shi (2019) (GLS). The column “GLS, Fixed, Agg” shows results using averages covariates and total sales across 10 adjacent observations, and sets market size $A = 400$ using the GLS estimator. The column “GLS, Observed, Disagg” reports results using disaggregate data, observed market size, and the GLS estimator. The column “GLS, Observed, Disagg" shows results of the GLS estimator using aggregate observed market size and sales and average covariates across 10 adjacent observations.

Table 8: Monte Carlo Results for Misspecified Distributions

	$A \sim \text{NegBinom}(25, 0.5)$	Misspecified Residual
α	-0.11	-0.10
	(-0.33, 0.11)	(-0.31, 0.13)
β_1	-0.01	-0.01
	(-0.27, 0.24)	(-0.27, 0.24)
β_2	-0.02	-0.02
	(-0.33, 0.24)	(-0.29, 0.27)
Γ_{11}	-0.03	-0.02
	(-0.12, 0.20)	(-0.12, 0.28)
Γ_{22}	0.02	0.03
	(-0.08, 0.25)	(-0.08, 0.26)
Γ_{33}	0.03	0.03
	(-0.07, 0.31)	(-0.08, 0.31)

Note: Reported in the table are the mean, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 1000 simulations. The first column simulates data in the same manner as our main Monte Carlo experiments, but using only a different arrival process. This process has an identical mean but has twice the variance of the baseline Poisson. The second column presents results when the model is estimated assuming that the joint distribution between ξ and ν is normal and not a mixture, despite being generated from a mixture of normal distributions.

C Discrete Random Coefficients Estimation Details

We modify our estimation approach in order to accommodate discrete unobserved consumer heterogeneity. We consider a two-type model, though this can be extended to more than two points of discrete support. Conditional on sampled parameters, many of our estimation steps remain unchanged. The modified algorithm used for estimation is given below. Adjusted steps relative to the continuous random coefficients case are highlighted with **NEW**.

Algorithm 3 Hybrid Gibbs Sampler

for $c = 1$ to C **do**
- Update arrivals λ (Gibbs)
- Update shares $s(\cdot)$ (Metropolis-Hastings)
- (**NEW**) Update price coefficients α (Metropolis-Hastings)
- (**NEW**) Update consumer distribution γ (Metropolis-Hastings)
- (**NEW**) Update linear parameters β (Gibbs)
- Update pricing equation η (Gibbs)
- Update basis classifier κ (Gibbs)
- Update mixture component parameters Σ_k, μ_k (Gibbs)

end for

C.1 Updating price coefficients, α_1, α_2

In the two-type discrete random coefficient model of Berry, Carnall, and Spiller (2006), the technique in Jiang, Manchanda, and Rossi (2009) cannot be used to sample the distribution of price coefficients. We propose an alternative approach that allows for sampling from the posterior distribution of the pricing coefficients.

The likelihood of the price coefficients $\alpha = (\alpha_1, \alpha_2)$ can be constructed using the same logic as in determining the prior for the share draws. Instead of sampling different shares, α is sampled. Different price sensitivities change the residual ξ, for which we can invert the demand system and evaluate the likelihood. Conditional on η, Σ, μ, and κ we can compute the distribution of ξ and determine the likelihood of a particular draw of α. The likelihood is given by

$$\prod_t \prod_{j=1}^{J(t)} \left[\phi \left(\frac{f^{-1}(s_{j,t}) - \frac{\rho_k v}{\sigma^2_{k,11}}}{\sqrt{\sigma^2_{k,22} - \frac{\rho^2_k}{\sigma^2_{k,11}}}} \right) \right] \cdot \left| J_{\xi_t \to s_t} \right|^{-1}$$

where $\phi(\cdot)$ is the standard Normal density function.

Due to the lack of availability of conjugate priors, we can use any prior distribution on the price coefficients. We impose a log-normal prior on α such that

$$\log(\alpha) \sim \mathcal{N}(\alpha_0, \Sigma_\alpha).$$

To avoid the well-known label-switching problem, we also impose that $\alpha_1 > \alpha_2$. This ensures that there is a single stationary distribution being sampled. This constraint can be viewed as an additional prior placed upon the distribution of α.

C.2 Updating probabilities on consumer types, γ

Next, we update the parameters γ_t, which are the probabilities on consumer types. We assume that there exists some dimension o that the consumer mix is shifting over. In practice this dimension is often time, as consumers become more or less price sensitive. We allow for the probability of business type to change over this dimension, enforcing a smooth function. This is achieved by using a polynomial basis over o. For computational simplicity, we do not directly sample from the distribution of γ. Rather γ is constructed from a sieve estimator, and we sample over the distribution of the coefficients of the sieve estimator—ψ. This ensures that γ is a smooth function over o, while maintaining a simple candidate distribution when sampling ψ. That is,

$$\gamma_t = \text{Logit}(G_o(t)'\psi),$$

where $G_o(t)$ is a vector of orthogonal polynomials evaluated for each market t along dimension m.

The likelihood computation is similar to the price-coefficient likelihood, as α and γ both are inputs into the inversion that we use to compute the likelihood of ξ. We omit a detailed discussion of the likelihood of γ for this reason. The likelihood of ψ given the shares drawn is equal to

$$\prod_t \prod_{j=1}^{J(t)} \left[\phi \left(\frac{f^{-1}(s_{j,t}) - \frac{\rho_k \upsilon}{\sigma_{k,11}^2}}{\sqrt{\sigma_{k,22}^2 - \frac{\rho_k^2}{\sigma_{k,11}^2}}} \right) \right] \cdot \left| J_{\xi_t \to s_t} \right|^{-1}. \tag{10}$$

We impose a Uniform prior on ψ. The logistic functional form enforces that all γ values lie in the interval $(0,1)$. This does not impose strong restrictions beyond smoothness of γ over time. We sample particular values of ψ, and their implied γ using a Metropolis-Hastings Step.[14] We note however, that since this uses a Metropolis-Hastings step, care must be taken in tuning the candidate distribution.

C.3 Updating remaining preferences, β

We define $\delta_{j,t} = X_{j,t}\beta + \xi_{j,t}$ where $\xi_{j,t}$ is taken as the random effects draw from the previous step. As before, we still need to deal with the correlation structure between ξ and price. We impose a normal prior on β so that this is a standard Bayesian linear regression. We must only account for the influence of the pricing residual on the distribution of ξ. We accomplish this by normalizing each component of 6 by subtracting the expected value of $\xi_{j,t}$ and diving both sides by the standard deviation. We then perform a Bayesian linear regression on this collection of normalized equations, as these rescaled errors have unit variance.

Let $\sigma_{k,2|1} = \sqrt{\sigma_{k,22}^2 - \frac{\rho_k^2}{\sigma_{k,11}^2}}$ be the variance of ξ conditional on υ and Σ. It follows that

$$\frac{\delta_{j,t} - \mu_{\kappa_{j,t},2} - \frac{\rho_{\kappa_{j,t}}}{\sigma_{\kappa_{j,t},11}^2}\upsilon}{\sigma_{\kappa_{j,t},2|1}} = \frac{1}{\sigma_{\kappa_{j,t},2|1}} x_{j,t}\bar{\beta} + U_{j,t}, \tag{11}$$

[14] The candidate is drawn using a normal distribution centered around the previous accepted value using the above likelihood.

where $U \sim \mathcal{N}(0,1)$.

We follow the typical conjugate prior distribution for a linear regression—$\beta \sim \mathcal{N}(\bar{\beta}_0, V_0)$. The posterior distribution is then a shrinkage estimator of OLS. Let

$$\hat{X}_{j,t} = \frac{x_{j,t}}{\sigma_{\kappa_{j,t},2|1}},$$

and

$$\hat{\delta} = \frac{\delta_{j,t} - \mu_{\kappa_{j,t},2} - \frac{\rho_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},11}} \upsilon}{\sigma_{\kappa_{j,t},2|1}}.$$

Then the posterior distribution of $\bar{\beta}$ is $\bar{\beta} \sim \mathcal{N}(\beta_N, V_N)$, where

$$\beta_N = (\hat{X}'\hat{X} + V_0^{-1})^{-1} \left(V_0^{-1}\beta_0 + \hat{X}'\hat{\delta}\right)$$
$$V_N = (V_0^{-1} + \hat{X}'\hat{X})^{-1}.$$

This estimation step is the same as the mean utility $(\bar{\beta})$ step for the random-coefficient demand scheme. For parameters with no random coefficients, the posterior draws will be drawn from the same distribution (conditional upon δ being equal).

The draw for product preferences (β) is substantially simpler than the update for α because these preferences are assumed to be constant across consumer types. If these preferences were also type specific, a similar method to the draw of α would be required. Our distributional assumptions that allow us to use a Bayesian regression technique to draw β. This method is efficient, scales to high dimensions easily, and does not require estimation tuning.

D Extension: Finite Mixture Components

For computational speed or researcher preference, one may wish to put some restrictions on the joint distribution of the demand shock and the pricing error. We provide an extension from our more flexible model presented in the body of the paper to allow for a finite number of mixture components. That is, we treat the number of component distributions as fixed, and thus don't need to evaluate whether to add or remove components at each step of the sampler. After updating the unconditional mixture weights for each component, we only need to update the component distribution probabilities for each observation $(\xi_{j,t}, \upsilon_{j,t})$ for the fixed set of components. To do so, we take the current candidate draw of π as a prior and evaluate the likelihood that this observation of $(\xi_{j,t}, \upsilon_{j,t})$ is drawn from that component distribution given the current candidate mean and variance. Rather than clustering means and variance, we augment the data with a classifier for each observation using $\hat{\pi}$, the posterior probabilities. Conditional upon the classifier, the residuals are distributed normal. We then update posterior cluster mean and variances based on the classified observations with a standard Gibbs step.

Algorithm 4 Hybrid Gibbs Sampler: Finite Mixture

```
for c = 1 to C do
    Update arrivals λ                                         (Gibbs)
    Update shares s(·)                                        (Metropolis-Hastings)
    Update linear parameters β                                (Gibbs)
    Update nonlinear parameters Γ                             (Metropolis-Hastings)
    Update pricing equation η                                 (Gibbs)
    (NEW) Update unconditional mixture weights π              (Gibbs)
    (NEW) Update component classifier κ                       (Gibbs)
    Update mixture component parameters Σ_k, μ_k              (Gibbs)
end for
```

D.0.1 Details of Sampling Price-Endogeneity Parameters

Updating Mixing Probabilities We assume a Dirichlet prior on the mixture probabilities, $\pi \sim \text{Dirichlet}(\bar{\alpha})$. Conditional on the classifier κ, we have information about which data points fall into which classifier, and the posterior distribution of π is given by

$$\pi \sim \text{Dirichlet}(\tilde{\alpha})$$
$$\tilde{\alpha}_k = n_k + \bar{\alpha}_k.$$

This gives the unconditional probability that a data point is drawn from classifier k.

Updating Component Classifier This step now skips the need to (a) evaluate new component probabilities and (b) check for orphaned components. Rather than using a classifier κ that is sufficient for all unique values of θ_n, we augment the data with the classifier at each step of the chain. Each residual can be treated as drawn from a single, unobserved normal distribution, simplifying the computation required when evaluating its distribution.

The classification of each data point can be thought of as a multinomial draw with π as the prior probability of each classification. The remaining information can be gathered from the likelihood of each component. We exploit the conjugacy nature of the multinomial distribution and the Dirichlet distribution, so that $\kappa_{j,t}|\pi \sim \text{Mulitnomial}(\bar{\pi}_{j,t})$ and

$$\bar{\pi}_{j,t,k} = \frac{\pi_k \phi_k((\upsilon_{j,t}, \xi_{j,t})}{\sum_{i=1}^{K} \pi_i \phi_i(\upsilon_{j,t}, \xi_{j,t})},$$

where $\phi_k(x)$ is the likelihood of the k^{th} component evaluated at x.

This step is computationally expensive, as the number of computations is $\mathcal{O}(N \times K)$. It requires evaluating the likelihood of each residual at every distribution, this must be evaluated with every draw, as ξ, υ and the mean and variance of each component change each draw. Through careful application of the prior ω, and priors on the mean and variance of the components, complex distributions can be approximated with relatively few mixtures, which can reduce the computational burden of this procedure.

Updating the Component Distributions, Σ_K and μ_k. This step proceeds identically to the Dirichlet Process case.