NBER WORKING PAPER SERIES

INCORPORATING SALES AND ARRIVALS INFORMATION IN DEMAND ESTIMATION

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

Working Paper 29530
http://www.nber.org/papers/w29530

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021, Revised April 2023

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.

Incorporating Sales and Arrivals Information in Demand Estimation
Ali Hortaçsu, Olivia R. Natan, Hayden Parsley, Timothy Schwieg, and Kevin R. Williams
NBER Working Paper No. 29530
December 2021, Revised April 2023
JEL No. C10,C11,C13,C18,L93

ABSTRACT

We propose a demand estimation method 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 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 as well as a measure of market sizes or consumer arrivals. After presenting simulation studies, we demonstrate the method in an empirical application of air travel demand.

Ali Hortaçsu
Kenneth C. Griffin Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
hortacsu@uchicago.edu

Olivia R. Natan
Haas School of Business
2200 Piedmont Avenue
Berkeley, CA 94720
olivia.natan@berkeley.edu

Hayden Parsley
University of Texas at Austin
parsley@utexas.edu

Timothy Schwieg
University of Chicago
5757 S University Ave
Chicago, IL 60637
timothyschwieg@uchicago.edu

Kevin R. Williams
Yale School of Management
165 Whitney Avenue
New Haven, CT 06520
and NBER
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. Situations with zero transactions are problematic for workhorse demand models because these models require observed market shares to be strictly positive in order to be estimable. 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. Our model combines Poisson arrivals and discrete choice demand models that accommodate 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 sampling error in product shares. This occurs because purchases opportunities tend to grow too slowly relative to the number of products offered. In contrast, our approach explicitly models small market sizes, which allows for low purchase rates and a significant amount of zero-sale observations. The data requirements are conventional market-level data as well as a measure of observed market sizes, such as information on consumer arrival intensity. These data are becoming increasingly available in economics and marketing and not need pertain only to e-commerce. For example, arrival 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 extend our methodology to discrete random coefficients and apply it to the study of airline markets, where the daily demand for flights is low—product-level zeros sales exceed 85%. We compare demand estimates across estimation approaches and use our model to explore the underlying forces that cause cyclical demand for air travel.

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 (and thus, empirical shares) to be strictly positive. In estimation, this inversion is not possible in the presence of zero-sale observations, as it requires taking the log of empirical shares equal to zero. Dropping the zeros is known to create a selection bias in the demand estimates (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. In our model, market sizes are modeled through Poisson distributions. Under our modeling assumptions, 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), Lima (2021), Adam, He, and Zheng (2020), Li (2019), and Gandhi, Lu, and Shi (2023), we explicitly leverage market-size variation as a source of zero sales. An alternative approach would be to fix the market size and model the multinomial distribution of sales, similar to Conlon and Mortimer (2013).

Our approach combines methodologies that allow for sampling error in product shares, price endogeneity, and rich unobserved heterogeneity into a single framework. Market participation is modeled according to flexible Poisson distributions. That is, consumers consider all products or do not participate in the market at all. This contrasts with individual-

level models of limited information, where consumers search across subsets of products (e.g., Amano, Rhodes, and Seiler, 2022; Abaluck, Compiani, and Zhang, 2022). Our approach relates to Burda, Harding, and Hausman (2012), who consider Poisson demand with rich individual-level heterogeneity, and Vulcano, Van Ryzin, and Ratliff (2012), who suggest using choice set variation in Poisson demand estimation. However, both of these works abstract from endogenous variables.[1] Similarly, a growing literature in empirical industrial organization (e.g., Buchholz, 2021; Williams, 2022) and operations (e.g., Newman, Ferguson, Garrow, and Jacobs, 2014; Jain, Rudi, and Wang, 2015; Abdallah and Vulcano, 2021; Wang, 2021) consider Poisson demand. Relative to these works, we allow for prices to have a flexible correlation structure with latent demand characteristics, i.e., prices are endogenous to unobserved product qualities.

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 the data with unobserved shares, we can use the market share inversion of BLP (1995) even though there may exist zero-sale observations and sampling error in product shares. 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 the appendix.

In Section 4, we demonstrate that our estimator can recover accurate and precise parameter values relative to existing methods in several simulation studies. Our estimator provides coverage even if market sizes are very small—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 pref-

[1] See also Chen and Kuo (2001) and Lee, Green, and Ryan (2017) for related work that involves random effects without endogenous characteristics.

erences 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 restricting the flexibility in our treatment of price endogeneity. Lastly, we conduct a set of simulation studies that better represents retail scanner data, where consumers face choice sets with dozens of products.

Finally, we provide an adaptation of our estimator to mass point random coefficients (Kamakura and Russell, 1989; Berry, Carnall, and Spiller, 2006) and consider an empirical application to the airline industry in Section 5. Our empirical setting emphasizes the data requirements for estimation. We use data provided by a large international airline based in the United States. Our sample contains sales quantities, prices, as well as a measure of consumer arrivals. More precisely, we use search query counts at granular levels to inform market sizes in estimation.

The demand for air travel is difficult to estimate because sales are sparse. We find that 85% of observations are zero-sale observations. Moreover, typically just a few consumers arrive per day. Using our method, we estimate mean product elasticities to be -1.2. However, we estimate demand to be far more inelastic by using existing demand approaches. Existing approaches yield price elasticities between -0.13 and 0, with substantial masses of estimates very close to zero. This occurs because imputing (small) shares when sales are zero causes price variation to have no impact on shares due to attenuation bias. Similarly, dropping zeros leads to price elasticity estimates very close to zero, since observations with positive shares feature higher willingness to pay. We find that low arrival rates prevent any product from having consistently well-measured market shares, which causes the estimator of Gandhi, Lu, and Shi (2023) to yield biased demand estimates as well.

With our model estimated, we explore preference heterogeneity across markets and decompose the driving forces of cyclicality in demand for air travel. Our analysis shows that periods of low demand feature both low market participation and consumers with lower willingness to pay. We estimate substantial 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. Preferences across departure times reflect observed price differences. However, this variation is inflated significantly using existing approaches, suggesting consumers are willing to pay thousands of dollars to switch flights. Finally, we discuss why demand cyclicality would be amplified, absent the use of dynamic pricing and frequent capacity adjustments.

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 ($j = 0$). Markets t may be defined across time or other dimensions, such as space. The data may comprise a panel. In the baseline model, consumers are drawn from a continuum of types. We also consider mass-point random coefficients in Section 5.

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 $\beta_i \sim \mathcal{N}(0, I)$ 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 that 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.,

$$A_t \sim \text{Poisson}\left(\lambda_t\right), \tag{1}$$

such that $\lambda_t := \exp(W_t\lambda)$ 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 upon arrival. The matrix W could contain, for example, time, location, or other market-specific covariates, depending on the application.

Our model of market participation is a highly stylized representation of consumer search. Consumers either search among all products or they do not participate in the market.

This is because we do not leverage individual-level data, see e.g., Honka (2014), Armona, Lewis, and Zervas (2021), and Kim, Albuquerque, and Bronnenberg (2010). Our model uses aggregate, market-level data.

Two assumptions allow us to construct analytic expressions for demand: (1) the realizations of arrivals are conditionally independent of preferences ($A \perp \xi, p | X, W$), and (2) consumers solve the above utility maximization problems. With these assumptions, conditional on prices, product characteristics, and observed determinants of market size, demand for each product j is distributed according to a conditionally independent Poisson distribution, i.e.,

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

Importantly, the demand for product j is independent of the demand for product j', conditional on W_t, X_t and p_t[2], which still allows correlation in the observed exogenous determinants of utility and the arrival rates. For example, this may accommodate higher arrival rates on average in markets with higher average demand (e.g., in the case of seasonal demand). X and W could contain the same variables in some applications.

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.

Without small market sizes and sampling error, our method closely follows Jiang, Man-

[2]This result follows from the properties of splitting Poisson processes (Gallager, 1996)

chanda, and Rossi (2009).[3] Unlike in settings with large market sizes, with low arrival rates and sparse sales, we are forced 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 arrivals or arriving 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 arrivals, the probability of $q_{j,t} = 0$ is quite high, but $s_{j,t}$ is never equal to zero. Thus, when we observe few arrivals, we must account for the sampling variation to be expected in sales quantities.

Note that if we did not have endogenous product characteristics, our estimator would closely resemble existing Poisson-logit Maximum Likelihood Estimators (e.g., Burda, Harding, and Hausman, 2012; Vulcano, Van Ryzin, and Ratliff, 2012). These approaches could, in the absence of price endogeneity, accommodate zero sales observations and provide flexible arrival patterns. However, in the presence of endogenous product attributes, such estimators will fail to produce unbiased estimates of key parameters, including the price coefficient(s).

3.1 Accounting for Price Endogeneity

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

$$p_{j,t} = Z'_{j,t}\eta + \upsilon_{j,t}, \tag{3}$$

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

[3] In addition, we extend their approach to mass-point random coefficients and make the pricing equation estimation more flexible using a Dirichlet Process prior and mixture of normal distributions.

We provide additional flexibility by estimating the joint distribution prices and the demand shocks ξ non-parametrically using a Dirichlet process prior, instead of assuming it to be a single joint normal distribution, e.g., in Rossi, Allenby, and McCulloch (2012). The Dirichlet process provides the researcher with more control over the generality of the distribution when data are sparse. We show that 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 price endogeneity, while still remaining non-parametric with respect to the demand unobservables. We comment on this in more detail below.

3.2 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). Appendix B 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. Here we state properties of the Dirichlet process prior 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 together. There are two components to the process: $\tilde{\alpha}$, the tightness parameter, and G_0, a prior distribution for each residual pair. Drawing from the process yields a distribution for the residuals (ξ, υ), and we condition on the drawn distribution for each step of the chain in the rest of estimation. Each residual pair has its own distribution, indexed by parameters θ_n. Many residual pairs will have the same θ_n and be drawn from the same distribution at each step of the chain.

The prior distribution G_0 describes the distribution of a new cluster. It must be diffuse enough to cover the data, but not so diffuse that the likelihood of drawing from it even

at likely parameter values is too low. As with any non-parametric estimator, care must be taken to choose priors such that we can approximate the residuals well. We choose a normal distribution for G_0 so that

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

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. We choose the tightness parameter $\tilde{\alpha}$ to be constant and base it on the data.[4]

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 hyper-parameter $\tilde{\alpha}$. 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 that we

[4]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.

leverage in the rest of estimation to construct several conditional likelihoods. 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 with the same κ_n share the same distribution. Let each cluster be indexed by k, observations in the k^{th} cluster are distributed normally with a shared mean and variance. Observations in the k^{th} cluster are distributed as

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

3.3 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

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

The arrival parameters are informed by both arrival 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 arrival data that informs the distribution. As we parameterize arrivals such that arrival rates are a function of some data, we impose a Log-Gamma prior on λ for a conjugate prior. We sample from λ using a Gibbs step.

Observed market sizes are not large enough to treat the observed market shares, q/A, as equal to the model counterparts. Thus, true product shares are unobserved. Our approach to this measurement error issue is to treat the small market size as the source of the error and observed product shares as samples from their true unobserved distribution.

To sample from this distribution of shares, we augment the data with unobserved values. We condition on the distribution of ξ, based on the current draw of the basis classifier κ and mixture component mean and variance μ_κ and Σ_κ. That is, we leverage the distribution of ξ to construct a distribution of shares. To evaluate whether a drawn value for a share is likely, we invert the distribution of shares to 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 of s.

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.

3.3.1 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. We provide additional details on the computational burden in Section 4.

3.4 Identification

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

3.4.1 Using Arrivals Information to Separate Demand Uncertainty 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 arrival process is straightforward as we have presented it. With arrivals being observed, and under the assumption of conditionally independent consumer arrivals ($A \perp \xi, p | X, W$), the arrival rates are identified from observed variation in arrival data. Our distributional assumption—Poisson—is a robust approach to estimating the con-

ditional mean of consumer arrivals.[5] As noted above, we require some additional restrictions to identify λ since arrivals are measured at the t level. It is impossible to estimate each arrival rate (t-specific) using fixed effects in our setting.[6]

3.4.2 Identification of Share Parameters

With data on consumer arrivals, 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 consider several sets of Monte Carlos where we vary the arrival process, the number of demand covariates, and the number of products in the choice set. We refer to "moderate

[5]The Lévy–Itô decomposition suggests any Lévy process can be decomposed into a Brownian motion and the sum of independent Poisson processes.

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

arrivals" when the Poisson arrival rate is 25. "Small arrivals" occurs when the arrival rate is 5. "Small choice set" means a choice set with 25 products. "Moderate choice set" means a choice set of 45. We map the number of product covariates to the size of the choice set. In addition to these data generating process decisions, we specify parameter values that provide market outcomes that approximate real-world markets. For example, weekly grocery scanner data has many zero-sales observations at the product level, even among top-selling products. Our simulations have a low ratio of arrival 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 of zeros, between 39% and 82%, depending on the simulation.[7,8]

For each Monte Carlo specification, we simulate 100 markets. We define the utility specification to contain random coefficients on three product attributes, given by

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

In our simulations, we draw each vector of exogenous, binary product attributes $X_j^{1:J-1} \sim i.i.d.$ Multinomial$(1, \frac{1}{J})$, that is, a random multinomial vector with one positive entry. These are drawn i.i.d. across products, so multiple products in the same market could have the same $X^{1:J-1}$. The demand shock ξ is drawn from a normal distribution. Marginal cost shifters (c_{jt}), which we will use as instruments, are of dimension two and drawn from $i.i.d.$ Uniform$[0,1]$ distributions. Preference parameters are $\bar{\beta} = [\alpha, \beta^1, \ldots \beta^{J-1}]$, with $\alpha = -2$ and $\bar{\beta}^{1:J-1}$ drawn $i.i.d$ Uniform$[0,1]$. We only simulate random coefficients for

[7]Though arrivals are generated from a single distribution (i.e. $\lambda_t = 5$ or $\lambda_t = 25$, and W_t=1 across markets in the DGP, 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.

[8]For comparison, Dubé, Hortaçsu, and Joo (2021) test their model on data with 42% zero shares, and Gandhi, Lu, and Shi (2023) test their estimator on synthetic data with 52% zero shares. Their bounds estimator is also tested on data with 96% zero observations.

price preferences with a variance equal to 0.2.

Prices are set by maximizing a firm's multi-product profit given the demand function and marginal costs.[9] That is,

$$p_t = \underset{p_{1,t},\dots p_{J,t}}{\operatorname{argmax}} \sum_j (p_{j,t} - c_{j,t}) \cdot q_{j,t}(p), \tag{4}$$

where $q_{j,t}(p) = \lambda_t \cdot s_{j,t}(p_t; X_t, \xi, \alpha, \Gamma, \beta)$. The resulting correlation between the cost shifters and prices ranges from 0.24 to 0.39.

4.2 Monte Carlo Estimates

We contrast our estimation method (Poisson RC) with Berry, Levinsohn, and Pakes (1995) and Gandhi, Lu, and Shi (2023) for the moderate ($\lambda_t = 25$) and small ($\lambda_t = 5$) arrivals data generating process. To implement these alternative approaches, we need to provide the estimators with a measure of empirical shares. We either treat market sizes as observed or we calibrate them to $M = 80$. The latter specification provides insights on what happens without access to arrivals data. Next, we either estimate on disaggregated data or aggregate over every 10 adjacent observations (averaging product attributes, with the interpretation of aggregating over 10 adjacent markets). Given the choice of aggregation and market size, we calculate empirical market shares. In the aggregation case, we average covariates across observations and sum both sales and arrivals. This aggregation is equivalent to typical time-aggregation within a panel or across related units (e.g. retail chain-level market share across a metro area). Finally, for BLP (1995), we must handle zero empirical market shares. We use two ad-hoc solutions to zero empirical shares: either dropping these observations or adjusting them up from the zero bound. In the adjustment case, we replace the empirical shares with their Laplace share equivalent, $s_{jt}^A = \frac{M_t s_{jt} + 1}{M_t + J_t + 1}$, effectively adding one to all quantities and correspondingly increasing the market size. A summary of our estimation approach for an example data generating process, including aggregation

[9] The marginal cost shifters have coefficients $[1.0, 2.0]$ in the marginal cost equation.

method, treatment of market size, zero-share adjustments, and estimation procedures used in our Monte Carlos appears in Table 1.

Table 1: Example Monte Carlo Setup for One Data Generating Process

Data Generating Process	Treatment of Market Size	Aggregation	Treatment of Zeros	Estimation Approach
$\lambda_t = 25$, $J = 25$	Observed	No	Include	Poisson RC
$\lambda_t = 25$, $J = 25$	Observed	No	Drop	BLP (1995)
$\lambda_t = 25$, $J = 25$	Observed	No	Adjust	BLP (1995)
$\lambda_t = 25$, $J = 25$	Observed	Yes	Drop	BLP (1995)
$\lambda_t = 25$, $J = 25$	Observed	Yes	Adjust	BLP (1995)
$\lambda_t = 25$, $J = 25$	Calibrated, M=80	No	Drop	BLP (1995)
$\lambda_t = 25$, $J = 25$	Calibrated, M=80	No	Adjust	BLP (1995)
$\lambda_t = 25$, $J = 25$	Calibrated, M=80	Yes	Drop	BLP (1995)
$\lambda_t = 25$, $J = 25$	Calibrated, M=80	Yes	Adjust	BLP (1995)
$\lambda_t = 25$, $J = 25$	Observed	No	Include	GLS (2023)
$\lambda_t = 25$, $J = 25$	Calibrated, M=80	No	Include	GLS (2023)
$\lambda_t = 25$, $J = 25$	Observed	Yes	Include	GLS (2023)
$\lambda_t = 25$, $J = 25$	Calibrated, M=80	Yes	Include	GLS (2023)

Note: Data Generating Process refers to the size of the arrival process and the number of products. Product shares are generated in an identical manner for different estimators with the same market size. Aggregation and treatment of zeros only occur for BLP (1995) estimation approach. Aggregation treats 10 adjacent observations as the same unit and sums their arrivals and demand and averages covariates and prices.

Monte Carlo results using a moderate arrival rate of 25 ($\lambda_t = 25$) and small choice set ($J = 25$) appear in Table 2. Table 3 shows the results for the low arrival rate of 5 ($\lambda_t = 5$) with small choice sets. Each row in the tables corresponds to an estimator, treatment of market sizes (calibrated or realized arrivals), level of aggregation, and treatment of zero empirical shares (dropping zeros or using Laplace shares). We report the median absolute bias, median relative bias (estimated parameter − true parameter), mean relative bias, and (in parentheses) the 2.5th and 97.5th percentile of the bias across the simulations. We report these measures for mean price preferences (α) and the random coefficient on price (Γ). We report some summary results on exogenous product characteristics parameters β in Figure 1.

Table 2: Monte Carlo Results for $\lambda_t = 25$, $J = 25$, with true values $\alpha = -2$, $\Gamma = 0.2$

				α			Γ		
Estimator	M	Aggregate	Zeros	Median |Bias|	Median Bias	Mean Bias	Median |Bias|	Median Bias	Mean Bias
Poisson-RC	–	–	–	0.09	0.09	0.08 (-0.08, 0.20)	0.16	-0.16	-0.11 (-0.18, 0.08)
BLP (1995)	80	No	Drop	1.27	1.27	1.26 (1.00, 1.35)	0.20	-0.20	-0.16 (-0.20, 0.40)
BLP (1995)	80	No	Adjust	0.43	0.43	0.39 (-0.08, 0.74)	0.99	0.99	1.00 (0.63, 1.56)
BLP (1995)	80	Yes	Drop	0.70	0.69	0.48 (-0.48, 1.60)	0.20	-0.20	0.09 (-0.20, 0.98)
BLP (1995)	80	Yes	Adjust	0.55	0.53	0.27 (-1.89, 1.38)	0.20	-0.20	0.11 (-0.20, 2.03)
BLP (1995)	Realized	No	Drop	1.30	1.30	1.24 (0.50, 1.49)	0.20	-0.20	0.07 (-0.20, 3.77)
BLP (1995)	Realized	No	Adjust	1.09	1.09	1.04 (0.60, 1.31)	0.32	0.32	0.35 (-0.20, 1.34)
BLP (1995)	Realized	Yes	Drop	0.60	0.48	0.51 (-0.91, 1.93)	0.20	-0.20	-0.04 (-0.20, 1.16)
BLP (1995)	Realized	Yes	Adjust	0.58	0.49	0.44 (-0.89, 1.45)	0.20	-0.20	-0.06 (-0.20, 0.93)
GLS (2023)	80	No	–	0.76	0.76	0.76 (0.55, 1.01)	0.82	0.82	0.80 (0.58, 0.94)
GLS (2023)	80	Yes	–	1.15	1.04	0.87 (-2.61, 2.19)	0.20	-0.20	0.09 (-0.20, 3.03)
GLS (2023)	Realized	No	–	1.10	1.10	1.10 (0.73, 1.47)	0.79	0.79	0.78 (0.59, 0.99)
GLS (2023)	Realized	Yes	–	1.06	1.05	0.99 (-0.44, 2.23)	0.20	-0.20	0.12 (-0.20, 0.90)

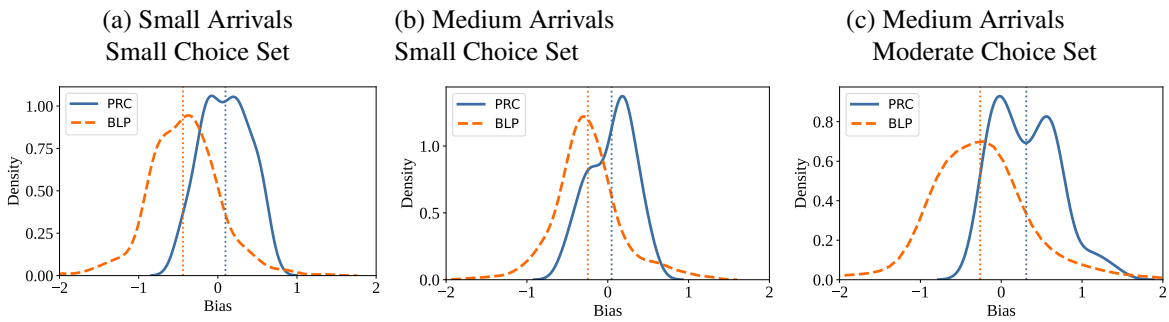
Note: Reported in the table are the median absolute bias, median relative bias (estimated parameter - true parameter), and mean relative bias. In parenthesis, we enclose the 2.5 percentile and the 97.5 percentile of the relative bias for all 100 simulations. Column *M* refers to the market size, which is either handled as part of the model (Poisson-RC), used as data ("Realized"), or calibrated to 80. Column Aggregate refers to whether observations are aggregated prior to estimation. In the case of aggregation, sales, covariates, and arrivals are averaged across 10 adjacent observations. Column Zeros refers to how zero sales are handled - either directly accommodated by the estimator, or observations with zero empirical shares are dropped, or zero empirical shares are replaced with their zero-adjusted (Laplace) shares.

Table 3: Monte Carlo Results for $\lambda_t = 5$, $J = 25$, with true values $\alpha = -2$, $\Gamma = 0.2$

				α			Γ		
Estimator	M	Aggregate	Zeros	Median |Bias|	Median Bias	Mean Bias	Median |Bias|	Median Bias	Mean Bias
Poisson-RC	–	–	–	0.31	0.31	0.30	0.16	-0.16	-0.12
						(0.05, 0.50)			(-0.18, 0.06)
BLP (1995)	80	No	Drop	1.84	1.84	1.22	0.20	-0.20	0.20
						(-0.18, 1.91)			(-0.20, 1.54)
BLP (1995)	80	No	Adjust	1.57	1.57	1.52	0.27	0.27	0.31
						(1.28, 1.69)			(0.08, 0.63)
BLP (1995)	80	Yes	Drop	1.21	1.10	0.38	0.20	-0.20	0.42
						(-5.72, 1.66)			(-0.20, 3.76)
BLP (1995)	80	Yes	Adjust	1.15	1.14	0.90	0.20	-0.20	-0.00
						(0.14, 1.59)			(-0.20, 1.00)
BLP (1995)	Realized	No	Drop	1.86	1.86	1.85	0.20	-0.20	-0.06
						(1.66, 2.02)			(-0.20, 1.06)
BLP (1995)	Realized	No	Adjust	1.75	1.75	1.71	0.20	0.02	0.15
						(1.49, 1.82)			(-0.20, 0.84)
BLP (1995)	Realized	Yes	Drop	1.25	1.24	1.20	0.20	-0.20	0.37
						(-0.68, 2.92)			(-0.20, 5.65)
BLP (1995)	Realized	Yes	Adjust	1.11	1.10	1.08	0.20	-0.20	0.11
						(0.15, 1.93)			(-0.20, 2.20)
GLS (2023)	80	No	–	1.07	1.07	1.05	0.79	0.79	0.77
						(0.71, 1.36)			(0.56, 0.90)
GLS (2023)	80	Yes	–	0.79	0.79	0.73	0.82	0.82	0.82
						(-0.19, 1.41)			(0.60, 1.00)
GLS (2023)	Realized	No	–	1.51	1.51	1.52	0.80	0.80	0.80
						(1.29, 1.75)			(0.53, 1.05)
GLS (2023)	Realized	Yes	–	1.51	1.51	1.47	0.78	0.78	0.76
						(0.18, 3.23)			(0.59, 0.89)

Note: Reported in the table are the median absolute bias, median relative bias (estimated parameter - true parameter), and mean relative bias. In parenthesis, we enclose the 2.5 percentile and the 97.5 percentile of the relative bias for all 100 simulations. Column *M* refers to the market size, which is either handled as part of the model (Poisson-RC), used as data ("Realized"), or calibrated to 80. Column Aggregate refers to whether observations are aggregated prior to estimation. In the case of aggregation, sales, covariates, and arrivals are averaged across 10 adjacent observations. Column Zeros refers to how zero sales are handled - either directly accommodated by the estimator, or observations with zero empirical shares are dropped, or zero empirical shares are replaced with their zero-adjusted (Laplace) shares.

Figure 1: Estimation Bias in Product Characteristics



Note: Reported in each figure is the distribution of parameter bias for the first exogenous characteristic (β_1) across simulations for different estimators. We report only results the first element of $\bar{\beta}$ since each element of both X and $\bar{\beta}$ is generated $i.i.d$—results are interchangeable across the exogenous product attributes. The blue lines report the distribution of mean parameter bias across simulations using Poisson-RC. The orange dashed lines report the distribution of mean parameter bias across simulations *and estimation adjustments* for BLP (1995). That is, it reports the bias distribution across all BLP (1995) adjustments (aggregation and disaggregation, dropping zeros or adjusting zeros) and across simulations.

We find that our Bayesian estimator of the Poisson random coefficient model performs well. For the mean price sensitivity parameter α, under the moderate arrivals data generating process, our estimator's point estimates remain close to the truth. Under the small arrivals data generating process, we see an increase in the bias but our estimator still performs well given the large fraction of observed zero purchases. For the random coefficient on price, our estimator performs better than the alternative estimators but we still observe the estimates attenuating somewhat toward zero. This is driven by in part by the strength of the instruments—the correlation between the instruments and price vary from 0.2-0.4, or in line with moderate strength instrument results in Conlon and Gortmaker (2020).

Additionally, for smaller choice sets (panels (a) and (b) of Figure 1), we find that our estimator produces unbiased estimates of the coefficients on product characteristics. Estimates on these parameters are noisier than estimates of the effect of price, but we outperform alternative methods in terms of average bias. In particular, BLP (1995) ad-hoc estimation approaches produce persistently biased estimates of the exogenous product characteristics.

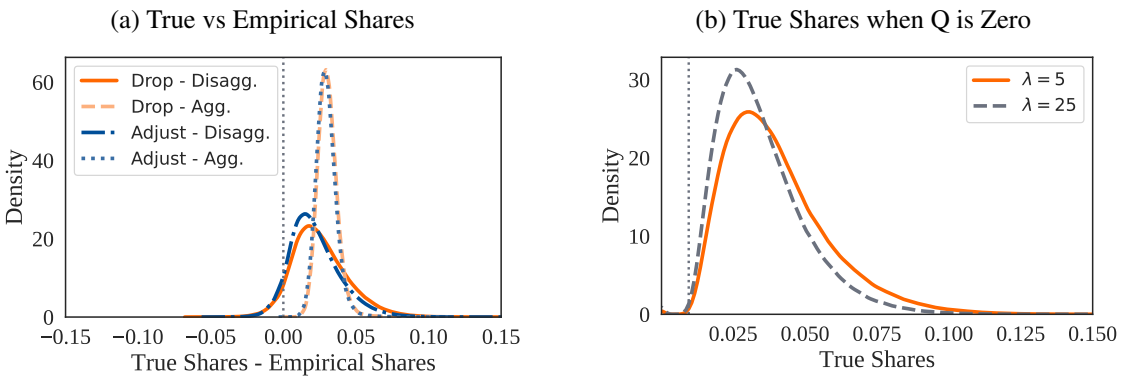
We find that the two common solutions to zero sales observations fail to capture the true parameters. Dropping zero sales observations attenuates the price coefficients (rows with

zeros set to "Drop" in Table 2 and Table 3). This is more severe in the case with the smaller market size. Although the drop-zero estimates sometimes perform better in estimating the price coefficient than the adjusted zeros methods, this approach fails to capture other parameters accurately, including both the mean and variance of the random coefficient. Across various BLP (1995) specifications, we find that the median relative bias for the random coefficient variance is -0.20, which in part is due to constraining the parameter value to be positive in estimation. Removing this constraint leads to a much more dramatic bias due to unconstrained estimation which results in large negative values.

Adjusting product shares when they are equal to zero results in bias (rows with zeros set to "Adjust" in Table 2 and Table 3), but this adjustment biases results on average less than dropping the zero share observations. Adjusting zero shares yields parameter estimates which consistently bias both the mean and variance in the random coefficient on price. 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 2a plots the distribution of the difference between true shares and adjusted shares. Dropping zeros results in a distribution of empirical shares that are lower than the truth. Adjusting zero shares with a small value also understates the true share. Consequently, price sensitivity estimates are attenuated since imputing a tiny share is inaccurate when the zeroes occur due to few consumers arriving. Figure 2b shows the conditional distribution of product shares when sales are zero. Imputing an arbitrary small value performs poorly because imputed shares are on the large end of common imputed definitions 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 understate the price sensitivity of consumers.

An alternative solution to minimize the frequency of zero shares is to aggregate the data. We find that applying share adjustments after aggregation results in fewer adjusted shares, but it does not improve the performance of such estimators (rows with aggregate

Figure 2: True Shares Compared to Share Adjustments



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 ε, which is set by $s_t^A = \frac{M_t s_t + 1}{M_t + J_t + 1}$. Drop 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 the average ε

equal to Yes in Table 2 and Table 3). Instead, aggregating the data removes variation in prices, shares, and the instruments, which results in additional bias in some parameters. This approach results in smaller shares on average compared to the disaggregated results.

We find that, in general, using observed market size realizations improves the results of BLP (1995) where zero shares are dropped, however, our main results hold: the BLP (1995) estimator with zero adjustments performs poorly when market sizes are small.

Our Poisson RC estimation method uses an unknown number of mixture of normal distributions to approximate any joint distribution of (ξ, υ). We find that our approach is able to recover this joint distribution well.[10] In implementing the Dirichlet process (DP) prior on the correlation between prices and ξ, our approach requires minimal ex ante specification—all components have the same prior mean and variance. In addition, the method requires a single prior parameter governing the variability of DP components.

Our Monte Carlo results show that the Poisson RC model can accurately measure consumer preferences and heterogeneity under very small market sizes. Our method accounts for the sampling error to be expected in sales. Alternative solutions to the zeros problem

[10] 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 B.

conduct inference only under large market sizes. For example, we find that the approach of Gandhi, Lu, and Shi (2023) produces significant bias in the price coefficient and noisier estimates than the Poisson RC model in these small market settings. 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 largely driven by a small market size which causes all empirical shares to be noisy measures of true shares. Note that Gandhi, Lu, and Shi (2023) also suggest a partial identification strategy. This approach does not require the presence of safe products. Additionally, the partial identification approach of Dubé, Hortaçsu, and Joo (2021) could be feasible if point estimates are not required.

We also note that our approach is not as general as typical moment-based estimators, e.g., BLP (1995), which only assume $E[\xi Z] = 0$. Our pricing equation does not correspond directly to typical models of differentiated products competition, where prices may depend on demand shocks of all products. This is because the pricing equation we leverage does not match directly the supply model generating the data in our Monte Carlos. Petrin and Train (2010) provide a detailed discussion of the limitations of this approach. At the same time, our approach does satisfy exclusion and relevance for our instruments. Moreover, specifying a full supply model that would be more efficient than our approach (e.g., Yang, Chen, and Allenby, 2003). Nonetheless, as we have shown, our flexible pricing equation is able to accurately recover demand fundamentals.

4.2.1 Estimator Performance in Larger Choice Sets

Many empirical applications feature larger choice sets. We test our estimator in settings where $J = 45$. The rest of the data generating process remains unchanged. Increasing the number of products does slow estimation—in nearly doubling the number of products, estimation takes 3-4 times longer. This is driven primarily because of greater computational burden at each step of the chain (e.g., more share inversions). We expect that using

our approach in settings with hundreds of products may be infeasible unless using highly optimized code.

The results of our estimator and competing methods are presented in Table 4. Qualitatively, estimates using our Poisson RC method are similar to those in smaller choice sets. In this setting, however, our estimator results in some small bias in the estimation of the coefficients on exogenous characteristics (Figure 1, panel c). As in our previous simulations, implementing ad-hoc fixes for BLP (1995) do not allow us to recover sensible estimates and typically return price parameters which understate price sensitivity significantly. Similarly, we find that the estimator of Gandhi, Lu, and Shi (2023) results in significant bias due to the lack of "safe products."

Table 4: Monte Carlo Results for $\lambda_t = 25$, $J = 45$, with true values $\alpha = -2, \Gamma = 0.2$

				α			Γ		
Estimator	M	Aggregate	Zeros	Median |Bias|	Median Bias	Mean Bias	Median |Bias|	Median Bias	Mean Bias
Poisson-RC	–	–	–	0.19	0.19	0.18	0.14	-0.13	-0.09
						(0.00, 0.29)			(-0.18, 0.11)
BLP (1995)	80	No	Drop	1.60	1.60	1.60	0.20	-0.20	-0.20
						(1.55, 1.65)			(-0.20,-0.20)
BLP (1995)	80	No	Adjust	1.32	1.32	1.31	0.30	0.30	0.31
						(1.12, 1.43)			(0.10, 0.63)
BLP (1995)	80	Yes	Drop	0.37	0.35	0.32	0.20	-0.20	-0.10
						(-0.24, 0.82)			(-0.20, 0.79)
BLP (1995)	80	Yes	Adjust	0.48	0.47	0.45	0.20	-0.20	-0.09
						(0.04, 0.96)			(-0.20, 0.94)
BLP (1995)	Realized	No	Drop	1.64	1.64	1.62	0.20	-0.20	0.03
						(1.32, 1.79)			(-0.20, 2.06)
BLP (1995)	Realized	No	Adjust	1.44	1.44	1.41	0.20	0.09	0.15
						(1.17, 1.54)			(-0.20, 0.94)
BLP (1995)	Realized	Yes	Drop	0.47	0.24	-0.15	0.20	-0.20	1.42
						(-2.69, 1.42)			(-0.20, 10.80)
BLP (1995)	Realized	Yes	Adjust	0.52	0.40	0.33	0.20	-0.20	0.18
						(-0.75, 1.32)			(-0.20, 2.46)
GLS (2023)	80	No	–	1.14	1.14	1.12	0.79	0.79	0.78
						(0.81, 1.32)			(0.55, 0.90)
GLS (2023)	80	Yes	–	0.41	0.36	0.39	0.80	0.80	0.80
						(-0.30, 1.11)			(0.57, 1.06)
GLS (2023)	Realized	No	–	1.38	1.38	1.36	0.83	0.83	0.83
						(1.05, 1.56)			(0.70, 0.95)
GLS (2023)	Realized	Yes	–	0.70	0.70	0.78	0.78	0.78	0.73
						(-0.31, 2.14)			(0.04, 0.91)

Note: Reported in the table are the mean, the median, and the median absolute difference between the point estimates and the true parameters. In parenthesis, we enclose the 2.5 percentile and the 97.5 percentile for each statistic across 500 simulations. Column *M* refers to the market size, which is either handled as part of the model (Poisson-RC), used as data ("Realized"), or calibrated to 80. Column Aggregate refers to whether observations are aggregated prior to estimation. In the case of aggregation, sales, covariates, and arrivals are averaged across 10 adjacent observations. Column Zeros refers to how zero sales are handled - either directly accommodated by the estimator, or observations with zero empirical shares are dropped, or zero empirical shares are replaced with a their zero-adjusted shares.

4.2.2 Results with Misspecified Models

Table 5: Monte Carlo Results for Misspecified Distributions

	$A \sim \text{NegBinom}(25, 0.5)$	Misspecified Residual
α	0.11	0.07
	(-0.06, 0.25)	(-0.06, 0.22)
Γ_{11}	-0.16	-0.04
	(-0.18, 0.06)	(-0.18, 0.11)

Note: Reported in the table are the median, the 2.5 percentile, and the 97.5 percentile of the difference between the point estimate and the true parameter across 100 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.

We also test our estimator to two forms of misspecification. Results are shown in Table 5. We simulate a misspecified arrival process (overdispersed with twice the variance of our $\lambda_t = 25$ setup) and a less flexible form of correlation between the pricing error and the demand shock (estimating the correlation structure between all pairs of ξ and υ using a single normal distribution). We find that our estimator performs nearly identically in the case of overdispersion. We also find that using a less flexible form of price endogeneity also provides nearly identical estimates on average, however, the tails of the (bias) distribution are more dispersed. We have found that when the conditional expectation of ξ given υ is approximately linear, or when the correlation summarizes the dependence well, a normal approximation performs 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.[11] Our primary aim is to show how to adapt our method to a relevant empirical setting. We also show how our method can flexibly estimate preferences over time and compare these estimates to existing approaches.

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. 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.

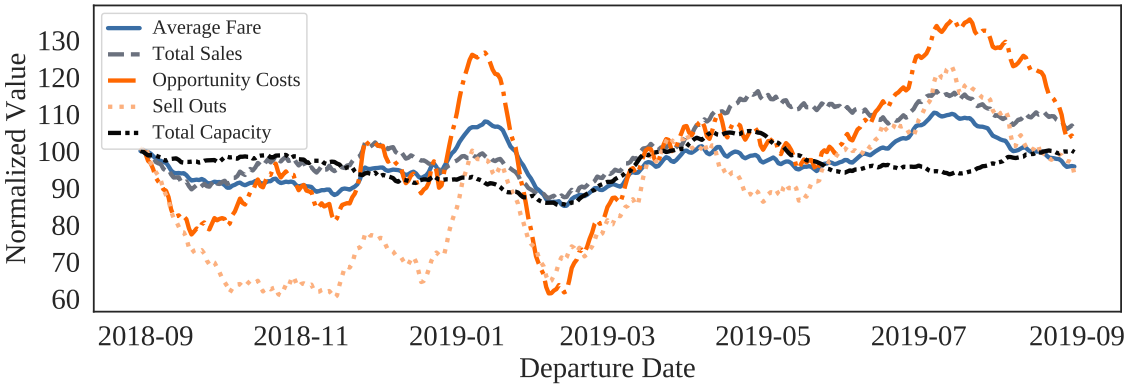
The key additional market-level data for estimation is a measure of market sizes. We calculate consumer arrivals using the number of consumers who initiate search requests on the air carrier's website using consumer clickstream data. In our setting, consumers arrive at the air carrier's website and their 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 just on one website, but consumers may shop via online travel agencies, such as Expedia. We cannot directly measure searches made from other sites. However, because we observe all bookings, we account for searches made via the unobserved sites through scaling factors. Each scaling factor is based on the fraction of sales directly through the airline and the average number of passengers per booking. If we know observed arrivals account for 50% of total bookings, assuming consumers who shop elsewhere have the same distribution of preferences, we can scale up estimated arrival rates by two.[12]

In Figure 3, we plot the 30-day moving average of bookings, fares, fraction of sold out

[11] The airline has elected to remain anonymous.

[12] In-depth summary analysis of the data and how unobserved searches are accounted for can be found in Hortaçsu, Natan, Parsley, Schwieg, and Williams (2023).

Figure 3: 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.

flights, total capacity for all routes in our sample from August 2018 to August 2019. Also included in the graph is a measure of opportunity cost, which has the interpretation of the marginal cost in this setting. We will use this measure as an instrument for demand. The figure shows that 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.

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. That is, we index markets by m, t. We extend the estimator to allow for discrete support random coefficients. Following Berry, Carnall, and Spiller (2006), we assume consumers are one of two discrete 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 and varies by days until the departure date. 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

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

As before, we assume that observed product characteristics $X_{j,t,m}$ are uncorrelated with the unobserved product characteristic $\xi_{j,m,t}$. These exogenous characteristics include departure time, week, and day of week fixed effects. We include week fixed effects in the utilities to flexibly capture seasonal variation in the value of travel. The consumer types differ in their 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

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

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 composition 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 flexible, non-monotonic relationship between

the composition of consumer types and time while producing values bounded between 0 and 1. Depending on the application, this function can be adjusted accordingly.

In addition to allowing for discrete random coefficients, we also 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 $1 - F_q(q_{j,t,m}|\cdot)$, where F_q is the cumulative distribution function of the above Poisson.[13]

We parameterize arrival rates by a set of multiplicative fixed effects across markets m and time t. That is, $\lambda_{m,t} = \exp(W_t\lambda_t + W_m\lambda_m)$, 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 us to distinguish seasonal market participation from seasonal preferences.

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[14]. We lever-

[13] We do not model choice variation within an m, t because arrival/booking rates are low. See Conlon and Mortimer (2013) for a method that accounts for choice set variation within a market.

[14] Opportunity costs during a specific period t depend on past pricing decisions. We include a set of fixed effects in both the exogenous characteristics and in the instrument set which capture persistent differences in demand across departure dates. Conditional on these fixed effects, we assume that demand shocks are independent over time. 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 $\sum_{D'} Q_{D,D'}$, where $Q_{D,D'}$ is the the total number of bookings in period t, across all flights, for all routes where the origin is the original route's

age the expiry of advance purchase discounts since these changes alter prices in a predetermined fashion, regardless of realizations of demand shocks. The opportunity cost of capacity directly 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 $A \rightarrow B$, 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 $B \rightarrow C$ and $B \rightarrow D$ are unrelated to unobserved shocks to demand for the focal route $A \rightarrow B$. 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 $A \rightarrow B \rightarrow C$ or $A \rightarrow B \rightarrow D$ rises. This increase in opportunity cost for connecting tickets also raises the opportunity cost of capacity on the $A \rightarrow B$ leg, which raises the price on $A \rightarrow B$. Our instruments are strong: the correlation between price and the opportunity cost of capacity is 0.72, and the correlation between price and the onward connecting traffic measure is 0.31. Pseudo first stage regressions (presented in Table 6) with our selected specification have a R^2 of 0.83.[15] We report results on the first stage in 6 and compare our results to alternative sets of instruments below.

For this application, we specify a time-varying block structure on the pricing equation, and (ξ, υ) have a block-varying joint normal distribution. That is, within a days-from-departure block, (ξ_t, υ_t) are distributed jointly normal, and this distribution may vary across blocks. 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.[16]

destination. If the route's origin is the hub, we calculate the total number of inward bound bookings, which would be; $\sum_{O'} Q_{O',O}$. Where $Q_{O',O}$ is the total bookings from all routes where the original route's origin is the destination.

[15]We denote these pseudo first stage regressions as we present frequentist OLS estimates of the first stage model we adapt in our estimator.

[16]Our misspecified specifications in Section 4 provide an example where a restrictive distribution of this

Table 6: Pseudo-First Stage Regressions

	(1)	(2)	(3)	(4)	(5)
Onward Connecting Traffic	0.597	0.591		0.377	0.369
	(0.010)	(0.011)		(0.006)	(0.006)
Opportunity Cost			11.069	11.308	11.383
			(0.059)	(0.056)	(0.057)
Opportunity Cost^2			7.288	6.506	6.483
			(0.073)	(0.072)	(0.073)
Opportunity Cost^3			-4.287	-3.747	-3.732
			(0.053)	(0.053)	(0.053)
Opportunity Cost^4			1.545	1.408	1.399
			(0.063)	(0.062)	(0.063)
Opportunity Cost^5			0.162	0.122	0.140
			(0.073)	(0.073)	(0.073)
APD FEs	N	Y	N	N	Y
DoW FEs	Y	Y	Y	Y	Y
Week FEs	Y	Y	Y	Y	Y
Departure Time FEs	Y	Y	Y	Y	Y
Adjusted R^2	0.300	0.302	0.798	0.825	0.829
F-Stat	389.003	369.252	3358.269	3958.341	3851.889

Pseudo-first stage results for our instruments. Columns 1 and 2 exclude polynomial terms of the opportunity cost measured by the airline's algorithm. Column 3 excludes the onward connecting traffic term, and both columns 3 and 4 exclude fixed effects for advance purchase discounts. Column 5 includes the full specification used for estimation. Fixed effect indicators denote inclusion of advance purchase discount, day of week, week of year, and departure time fixed effects, respectively.

5.3 Estimation Procedure

We modify our estimator 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 and the different likelihood function, 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**.

type performs relatively well compared to the fully flexible estimator.

Algorithm 2 Hybrid Gibbs Sampler - Discrete RC

1: **for** $c = 1$ to C **do**
2: Update arrivals λ (Gibbs)
3: Update shares $s(\cdot)$ (Metropolis-Hastings)
4: (**NEW**) Update price coefficients α (Metropolis-Hastings)
5: (**NEW**) Update consumer distribution γ (Metropolis-Hastings)
6: Update linear parameters β (Gibbs)
7: Update pricing equation η (Gibbs)
8: Update basis classifier κ (Gibbs)
9: Update mixture component parameters Σ_k, μ_k (Gibbs)
10: **end for**

5.3.1 Updating price coefficients, α_L, α_B

In the two-type discrete random coefficient model of Berry, Carnall, and Spiller (2006), the price coefficients α_L, α_B only affect utility linearly if we condition on consumer type. We cannot directly use the techniques developed in Jiang, Manchanda, and Rossi (2009). We propose an alternative approach that samples from the posterior distribution with a Metropolis-Hastings step. To sample from this distribution, we need to construct the likelihood.

The conditional likelihood of the price coefficients $\alpha = (\alpha_L, \alpha_B)$ is be constructed in an analogous manner to sampling shares $s(\cdot)$ (see Section A.1.2). For any given potential values of price sensitivities, conditional on other parameters and shares, we can invert the demand system to recover a unique demand shock ξ and its implied likelihood. This likelihood is the basis of the likelihood for a particular α value. Conditional on shares, η, Σ, μ, and κ, we compute the distribution of ξ and determine the likelihood of a particular draw of α. The likelihood is given by

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

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 a label-switching problem, we also impose that $\alpha_L > \alpha_B$. 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.

5.3.2 Updating probabilities on consumer types, γ

In order to update the parameters on the probability distribution of consumer types γ_t, we assume that the consumer distribution shifts over days from departure, t. In other settings, this dimension could be adapted to allow preferences to vary over geographic space or alternative observed covariates.

In the two-type setting, we allow for the probability of business ("high") type to change over t. enforcing a smooth function. This is achieved by using a polynomial basis. We construct a sieve estimator for γ, which allows us to sample over the distribution of sieve coefficients (here, ψ) rather than sampling directly from the distribution of γ.[17] The polynomial approximation maintains a simple candidate distribution when sampling ψ. To ensure we sample values of γ in $(0, 1)$, the polynomial basis is transformed by the logistic function.[18] 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.

The likelihood computation is similar to the price-coefficient likelihood, as α and γ

[17] This is computationally simpler, and it has an identical implied posterior distribution.

[18] The role of the logistic functional form enforces that all γ values lie in the interval $(0, 1)$, but does not restrict the possible shapes of γ over t. This does not impose restrictions beyond smoothness of γ over t—in our application, over time. Alternative link functions are feasible, since the method samples directly ψ rather than the implied γ values.

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

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

We sample particular values of ψ, and their implied γ, using a Metropolis-Hastings Step.[19]

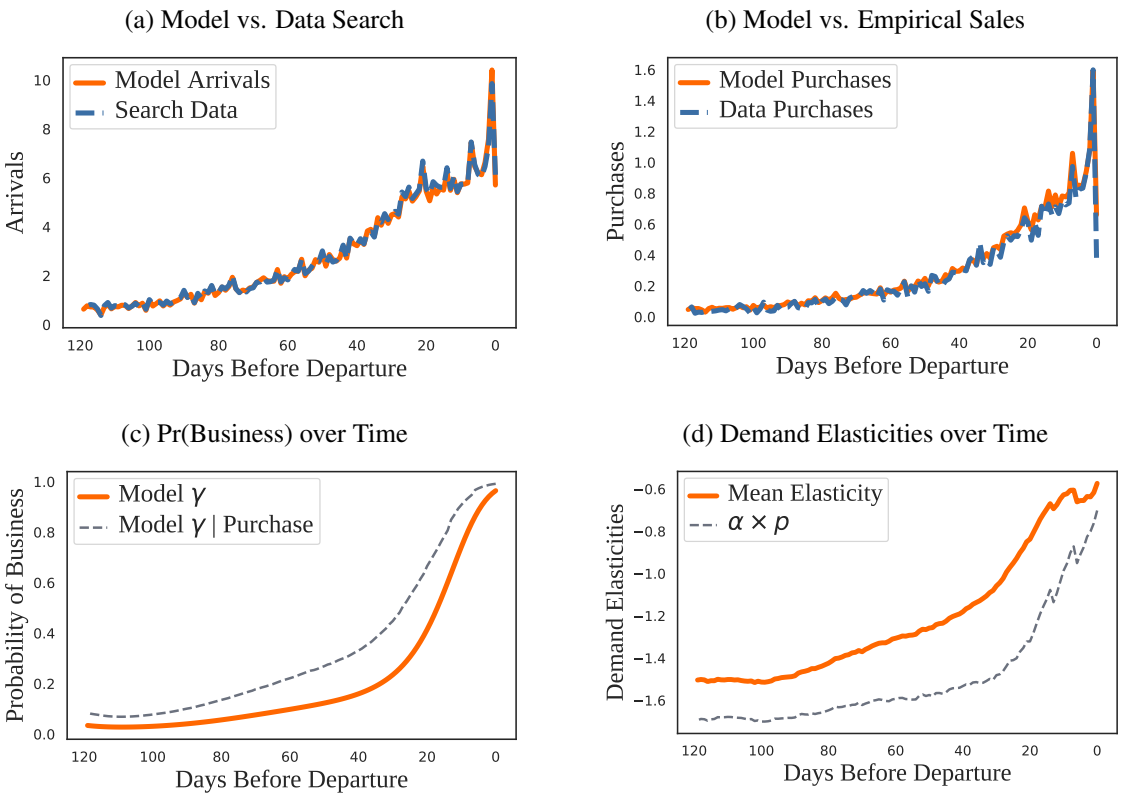
5.4 Empirical Results

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, and this air carrier is the only firm to operate non-stop flights on the route.

Measures of in-sample fit of estimation results are shown graphically in Figure 4. 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 4(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 (closer to zero) price elasticities as the departure date approaches, holding price constant. However, the

[19] We impose a flat prior on ψ, though alternative priors may be imposed since this step does not require conjugacy. The candidate is drawn using a normal distribution centered around the previous accepted value using the above likelihood. We note that since this uses a Metropolis-Hastings step, care must be taken in tuning the candidate distribution for efficient estimation.

Figure 4: Demand Estimates



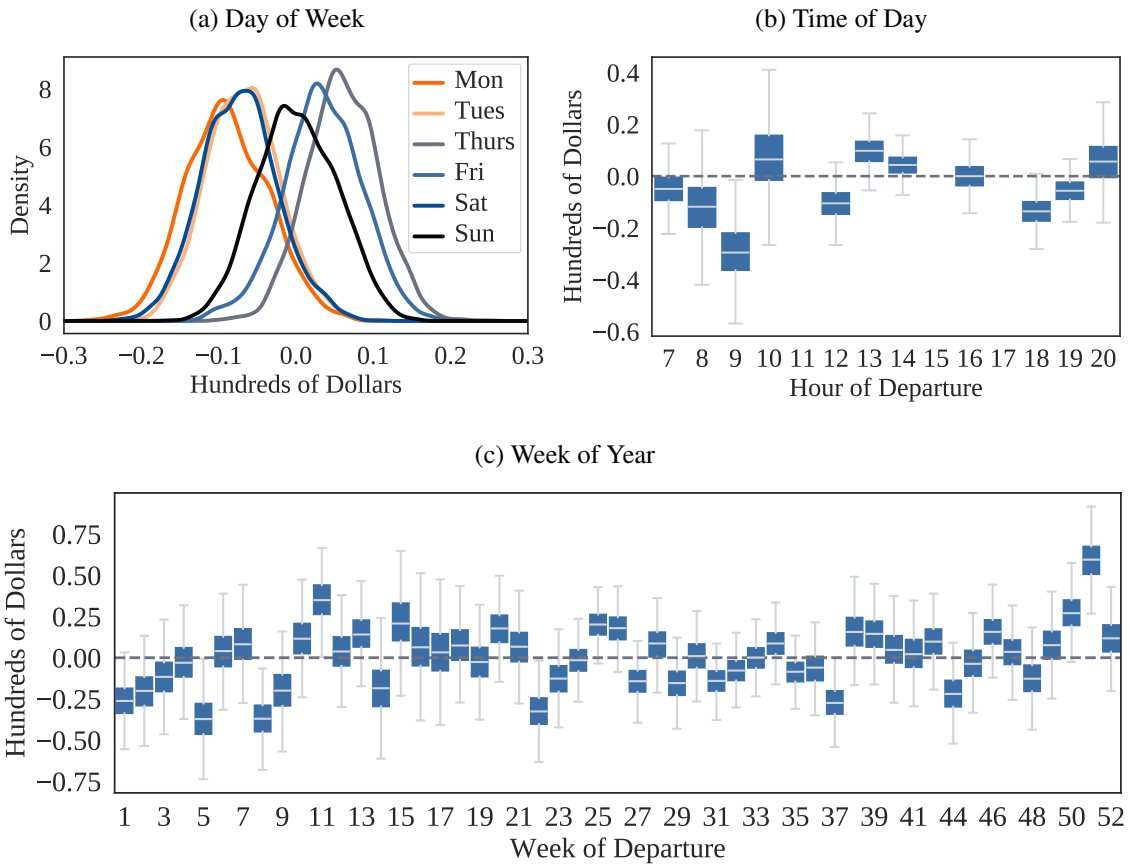
Note: The horizontal axis of all plots denotes the negative time index, e.g. zero corresponds to the last day before departure. (a) Normalized model fit of searches with data searches. (b) Model fit of product sales with empirical sales. (c) Fitted values of γ_t over time, along with the probability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

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 (2023) for evidence on pricing on the inelastic side of demand).

In Figure 5, 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.[20] Leisure types are four times more price sensitive than business type passengers.

Figure 5: 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).

Willingness to pay for travel displays considerable seasonality in this market. Figure 5(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

[20]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.

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.

Our demand estimates are robust to the set of instruments used in estimation. For example, while using only onward connecting traffic to instrument for price reduces the fit of the first-stage, we obtain similar demand results. We estimate average own-price elasticities of -1.11 (s.d. 0.33) without using the opportunity cost of capacity, whereas our baseline specification yields an average elasticity of -1.21 (s.d. 0.30). Adding advance purchase indicators to the set of instruments hardly changes the results. Using solely opportunity costs yields an average elasticity of -1.25 (s.d. 0.09).

We contrast our estimates with typical zero share fixes and assumptions about market size. Table 7 summarizes price elasticity estimates for a model based on Berry, Carnall, and Spiller (2006) and compares these results 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 arrivals 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 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 own-price elasticities do not bunch at zero. The other models yield unrealistic measures of willingness to pay, e.g., consumers are willing to pay up

Table 7: Own-Price Elasticities Across Models with Zero Share Adjustments

Zeros	Market Size	Aggregation	Mean	Std. Dev.	Median	5th	95th
Drop	Calibrated	Disaggregated	-0.002	0.001	-0.001	-0.004	-0.001
Drop	Calibrated	Aggregated	0.000	0.000	0.000	0.000	0.000
Drop	Observed	Disaggregated	-0.083	0.026	-0.083	-0.131	-0.038
Drop	Observed	Aggregated	0.000	0.000	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	-0.035	0.016	-0.031	-0.062	-0.018
Adjusted	Observed	Aggregated	-0.016	0.008	-0.014	-0.028	-0.006
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 "Adjusted" denotes replacing empirical shares with their Laplace share equivalent $s_{jt}^A = \frac{M_t s_{jt}+1}{M_t+J_t+1}$. 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.

to $3000 dollars to move from a Monday flight to a Wednesday flight. Our model suggests the variance in preferences across days of the week is more similar to observed price differences—about $20.

In addition to providing sensible elasticity estimates, our approach allows us to decompose the sources of cyclicality in demand for air travel. Are consumers willing to pay more to travel in popular departure markets? Are these markets larger? We find that market participation (λ_D) and preferences (shares at average prices) are positively correlated (correlation coefficient of 0.16). That is, more consumers arrive and more consumers purchase after arriving for popular departure dates. Given this correlation in preferences and arrivals, we decompose which of these sources of variation drives the variability in demand over the year.

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 or along departure dates, holding prices and arrivals constant. We conduct a similar exercise for arrivals, where

we hold prices, unobserved quality, and preferences fixed. Our decomposition suggests on average sales variability explained by changes in preferences is 25% more than the variation explained by arrivals. Preference variability is highest close to the departure date. Decomposing this variation across departure dates, 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. Therefore, preference changes are more important than market size changes in explaining cyclical demand for air travel.

These results contrast with grocery markets (Chevalier, Kashyap, and Rossi, 2003), where periods of peak demand have lower 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. In our setting, we find that both the intensity of demand and willingness to pay move in the same direction over the calendar year. However, our data also show that on the supply side, both capacities and prices also respond upward in periods of peak demand (see Figure 3). Absent these supply-side responses, this suggests that cyclicality would be even higher, as in the movie industry case.

6 Conclusion

We propose a 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 unbiased estimates, 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 outcomes and one additional data column—measures of consumer arrival intensity. These data are becoming increasingly available to researchers, with relevant applications in e-commerce, retailing, and transportation, among others.

References

Abaluck, J., G. Compiani, and F. Zhang (2022): "A Method to Estimate Discrete Choice Models that is Robust to Consumer Search," Working Paper.

Abdallah, T., and G. Vulcano (2021): "Demand estimation under the multinomial logit model from sales transaction data," *Manufacturing & Service Operations Management*, 23(5), 1196–1216.

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

Amano, T., A. Rhodes, and S. Seiler (2022): "Flexible Demand Estimation with Search Data," Working Paper.

Armona, L., G. Lewis, and G. Zervas (2021): "Learning Product Characteristics and Consumer Preferences from Search Data," 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.

Chen, Z., and L. Kuo (2001): "A note on the estimation of the multinomial logit model with random effects," *The American Statistician*, 55(2), 89–95.

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., and J. Gortmaker (2020): "Best practices for differentiated products demand estimation with pyblp," *The RAND Journal of Economics*, 51(4), 1108–1161.

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.

Gallager, R. G. (1996): *Poisson processes*. Springer.

Gandhi, A., Z. Lu, and X. Shi (2023): "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 (2023): "Organizational structure and pricing: Evidence from a large us airline," Working Paper.

Jain, A., N. Rudi, and T. Wang (2015): "Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need," *Operations Research*, 63(1), 134–150.

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.

Kamakura, W. A., and G. J. Russell (1989): "A probabilistic choice model for market segmentation and elasticity structure," *Journal of marketing research*, 26(4), 379–390.

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

Lee, J. Y., P. J. Green, and L. M. Ryan (2017): "On the" Poisson Trick" and its Extensions for Fitting Multinomial Regression Models," *arXiv preprint arXiv:1707.08538*.

Li, J. (2019): "Compatibility and investment in the us electric vehicle market," Working Paper.

Lima, L. (2021): "Demand Estimation with Zeros: Moving Costs in the US," Working Paper.

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

Newman, J. P., M. E. Ferguson, L. A. Garrow, and T. L. Jacobs (2014): "Estimation of choice-based models using sales data from a single firm," *Manufacturing & Service Operations Management*, 16(2), 184–197.

Petrin, A., and K. Train (2010): "A control function approach to endogeneity in consumer choice models," *Journal of marketing research*, 47(1), 3–13.

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.

Vulcano, G., G. Van Ryzin, and R. Ratliff (2012): "Estimating primary demand for substitutable products from sales transaction data," *Operations Research*, 60(2), 313–334.

Wang, R. (2021): "Consumer choice and market expansion: Modeling, optimization, and estimation," *Operations Research*, 69(4), 1044–1056.

Williams, K. R. (2022): "The welfare effects of dynamic pricing: Evidence from airline markets," *Econometrica*, 90(2), 831–858.

Yang, S., Y. Chen, and G. M. Allenby (2003): "Bayesian analysis of simultaneous demand and supply," *Quantitative marketing and economics*, 1(3), 251–275.

A Estimation Routine using Dirichlet Process Prior

A.1 Markov Chain Monte Carlo Details

A.1.1 Sampling Arrival Parameters

To update the parameters describing the arrival rate of consumers, we use arrival 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 $\lambda_t \cdot s_{j,t}$.

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,

$$\lambda_t := \exp(W_t \lambda), \tag{A1}$$

where W is a full rank matrix (composed of 0 and 1s if using fixed effects). Other specifications are possible.

Arrivals are distributed Poisson,

$$A_t \sim \text{Poisson}(\lambda_t). \tag{A2}$$

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

$$q_{j,t} \sim \text{Poisson}(\lambda_t s_{j,t}). \tag{A3}$$

We note that a conjugate prior choice for λ_t is log-Gamma distribution such that each element $\exp(\lambda_q) \sim \Gamma(k, \zeta)$. Therefore, the posterior distribution of $\exp(\lambda_q)$ is then given by

$$\exp(\lambda_q) \sim \Gamma\left(\sum_{t \in Q}\left(A_t + \sum_j q_{j,t}\right) + k, \frac{\zeta}{1 + \zeta\left(\sum_{t \in Q}(2 - s_{0,t})\right)}\right). \tag{A4}$$

A.1.2 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

$$\left.\begin{array}{ll} \xi_{j,t} & = f^{-1}\left(s_{j,t} \,|\, \beta, \Gamma, X_t\right) \\ \upsilon_{j,t} & = p_{j,t} - Z'_{j,t}\eta \end{array}\right\} |\kappa = k \sim \mathcal{N}_{\text{iid}}(\mu_k, \Sigma_k) \quad \text{such that} \quad \Sigma_k = \begin{pmatrix} \sigma^2_{k,11} & \rho_k \\ \rho_k & \sigma^2_{k,22} \end{pmatrix}. \tag{A5}$$

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

$$f_{s_{j,t}}(x) = f_{\xi_{j,t}}\left(f^{-1}(x)\right) \cdot \left|\mathscr{J}_{\xi_{j,t} \to s_{j,t}}\right|^{-1}. \tag{A6}$$

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

$$\xi|\upsilon,\kappa=k\sim\mathcal{N}\left(\mu_{k,2}+\frac{\rho_k\upsilon}{\sigma^2_{k,11}},\sigma^2_{k,22}-\frac{\rho_k^2}{\sigma^2_{k,11}}\right), \tag{A7}$$

where

$$\begin{pmatrix}\xi\\ \upsilon\end{pmatrix}|\kappa=k\sim\mathcal{N}(\mu_k,\Sigma_k),\quad \Sigma_k=\begin{pmatrix}\sigma^2_{k,11} & \rho_k\\ \rho_k & \sigma^2_{k,22}\end{pmatrix}. \tag{A8}$$

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

$$\prod_{j=1}^{J(t)}\left[\phi\left(\frac{f^{-1}(s_{j,t})-\frac{\rho_k\upsilon}{\sigma^2_{k,11}}}{\sqrt{\sigma^2_{k,22}-\frac{\rho_k^2}{\sigma^2_{k,11}}}}\right)\right]\cdot\left|\mathcal{J}_{\xi_t\to s_t}\right|^{-1}, \tag{A9}$$

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

$$q_{j,t}\sim\text{Poisson}(\lambda_t s_{j,t}). \tag{A10}$$

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

$$\ell(s_{\cdot,t}) = \prod_{j=1}^{J(t)} \left[\phi \left(\frac{f^{-1}(s_{j,t}) - \frac{\rho_k \upsilon}{\sigma^2_{k,11}}}{\sqrt{\sigma^2_{k,22} - \frac{\rho_k^2}{\sigma^2_{k,11}}}} \right) \frac{\lambda_t s_{j,t})^{q_{j,t}} \exp(-\lambda_t s_{j,t})}{q_{j,t}!} \right] \cdot \left| J_{\xi_t \to s_t} \right|^{-1}. \tag{A11}$$

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,

$$\begin{gathered} u_{i,j,t} = x_{j,t}\beta_i + \xi_{j,t} + \epsilon_i \\ \text{and} \\ \beta_i = \begin{pmatrix} \bar{\beta}_{1:L} + \Gamma U_i \\ \bar{\beta}_{L+1:K} \end{pmatrix}, \end{gathered} \tag{A12}$$

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

$$s_{j,t}|\upsilon_{j,t},\kappa_{j,t},\Gamma \sim \mathcal{N}\left(\frac{\mu_{\kappa_{j,t},2}+\rho_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},11}}\upsilon, \sigma^2_{\kappa_{j,t},22}-\frac{\rho^2_{\kappa_{j,t}}}{\sigma_{\kappa_{j,t},11}}\upsilon_{j,t}\right)\mathcal{J}^{-1}_{\xi\to s}. \tag{A13}$$

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

$$\Gamma_{ij} = \begin{cases} \exp(r_{ij}), & \text{for } i = j \\ r_{ij}, & \text{for } i < j \\ 0 & \text{otherwise .} \end{cases} \tag{A14}$$

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

$$\prod_{j=1}^{J(t)}\left[\phi\left(\frac{f^{-1}(s_{j,t})-\mu_{\kappa_{j,t},2}-\frac{\rho}{\sigma_{11}^2}\upsilon_{j,t}}{\sigma_{\kappa_{j,t},2|1}}\right)\right]|J_{\xi_{j,t}\to s_{j,t}}|^{-1}\times\prod_{i\leq j}\frac{1}{\psi_{ij}}\phi(r_{ij}). \tag{A15}$$

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 step searching 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

$$\delta_{j,t}=x_{j,t}\bar{\beta}+\xi_{j,t}. \tag{A16}$$

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 (A16) 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.

$$\frac{\delta_{j,t}-\mu_{\kappa_{j,t},2}-\frac{\rho_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},11}}\upsilon}{\sigma_{\kappa_{j,t},2|1}}=\frac{1}{\sigma_{\kappa_{j,t},2|1}}x_{j,t}\bar{\beta}+U^{\beta}_{j,t}, \tag{A17}$$

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

$$\hat{\delta} = \frac{\delta_{j,t} - \mu_{\kappa_{j,t},2} - \frac{\rho_{\kappa_{j,t}}}{\sigma^2_{\kappa_{j,t},11}} \upsilon}{\sigma_{\kappa_{j,t},2|1}}. \tag{A18}$$

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

$$V_N = (V_0^{-1} + \hat{X}'\hat{X})^{-1}. \tag{A19}$$

A.1.3 Sampling Price-Endogeneity Parameters

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

$$p_{j,t} = Z_{j,t}\eta + \upsilon_{j,t}. \tag{A20}$$

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

$$\frac{p_{j,t} - \mu_{\kappa_{j,t},1} - \frac{\rho_{\kappa_{j,t}}}{\sigma_{\kappa_{j,t},22}}\xi_{j,t}}{\sigma_{\kappa_{j,t},1|2}} = \frac{1}{\sigma_{\kappa_{j,t},1|2}} x_{j,t}\bar{\eta} + U^{\eta}_{j,t}. \tag{A21}$$

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

$$q_i = \frac{1}{\tilde{\alpha}+N-1}\Pr((\upsilon_{j,t}, \xi_{j,t})|\theta_i) \text{ for } i \neq 0, \tag{A22}$$

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.[21] As shown in Murphy (2007), this quantity is distributed multivariate t. Applying our priors, the form is given by:

$$\int \Pr(x|\theta_i) G_0(d\theta_i) \sim t_{\nu-1}(0, \frac{\frac{1}{\nu v} I(a_\mu+1)}{a_\mu(\nu-1)}). \tag{A23}$$

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

[21] We impose standard conjugate priors for computational ease, so

$$\mu|\Sigma \sim \mathcal{N}(0, {a_\mu}^{-1}\Sigma), \quad \text{and} \quad \Sigma \sim IW(\nu, \nu v I),$$

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 v to determine the location via

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

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 fewer 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:

$$\begin{aligned} \Sigma_k &\sim IW(\nu+1, V+S) \\ \mu_k|\Sigma_k &\sim \mathcal{N}(\tilde{\mu}, \frac{1}{1+a_\mu}\Sigma_k), \end{aligned} \tag{A24}$$

with

$$\begin{aligned} S &= \left(Y_k - \iota\bar{\mu}_k'\right)'\left(Y_k - \iota\bar{\mu}_k'\right), + a_\mu\left(\tilde{\mu}_k - \bar{\mu}\right)'\left(\tilde{\mu}_k - \bar{\mu}\right), \\ \tilde{\mu} &= (1+a_\mu)^{-1}(\bar{y}_k + a_\mu\bar{\mu}), \end{aligned} \tag{A25}$$

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 3 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 $\kappa_n \sim$ 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

$$\begin{aligned} Y_k &= \iota\mu' + U, \\ \text{where} \\ U &\sim \mathcal{N}(0, \Sigma_k). \end{aligned} \tag{A26}$$

The posterior covariance and conditional mean of the components are then

$$\begin{aligned} \Sigma_k &\sim IW(\nu + n_k, V + S) \\ \text{and} \\ \mu_k|\Sigma_k &\sim \mathcal{N}(\tilde{\mu}, \frac{1}{n_k + a_\mu}\Sigma_k), \end{aligned} \tag{A27}$$

where we define

$$\begin{aligned} S &= \left(Y_k - \iota\bar{\mu}_k'\right)'\left(Y_k - \iota\bar{\mu}_k'\right) + a_\mu\left(\tilde{\mu}_k - \bar{\mu}\right)'\left(\tilde{\mu}_k - \bar{\mu}\right), \\ \tilde{\mu} &= (n_k + a_\mu)^{-1}(n_k\bar{y}_k + a_\mu\bar{\mu}), \\ \text{and} \\ \bar{y}_k &= \frac{1}{n_k}Y_k'\iota. \end{aligned} \tag{A28}$$

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 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 bivariate 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

1:	**for** $c = 1$ to C **do**	
2:	Update arrivals λ	(Gibbs)
3:	Update shares $s(\cdot)$	(Metropolis-Hastings)
4:	Update linear parameters β	(Gibbs)
5:	Update nonlinear parameters Γ	(Metropolis-Hastings)
6:	Update pricing equation η	(Gibbs)
7:	(**NEW**) Update unconditional mixture weights π	(Gibbs)
8:	(**NEW**) Update component classifier κ	(Gibbs)
9:	Update mixture component parameters Σ_k, μ_k	(Gibbs)
10:	**end for**	

B.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

$$\begin{aligned} \pi &\sim \text{Dirichlet}(\tilde{\alpha}) \\ \tilde{\alpha}_k &= n_k + \bar{\alpha}_k. \end{aligned} \tag{A29}$$

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

$$\bar{\pi}_{j,t,k} = \frac{\pi_k \phi_k((\upsilon_{j,t}, \xi_{j,t})}{\sum_{i=1}^{K} \pi_i \phi_i(\upsilon_{j,t}, \xi_{j,t})}, \tag{A30}$$

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.