Demand Estimation With Finitely Many Consumers

Jonas Lieber* Thomas Wiemann†

November 21, 2022

Abstract

Although market shares are frequently estimated via averages of finitely many consumer choices, commonly applied methods for demand estimation are not robust to estimation error in these shares. While non-negligible estimation error in market shares always introduces bias in the demand parameter estimators, the issue becomes most salient when estimated market shares are zero. In the presence of zero shares, widely applied estimators of the random coefficient logit model cannot be computed without ad-hoc data manipulations. This paper proposes a new estimator of demand parameters for settings with endogenous prices and estimated market shares that is robust to zero-valued market shares. The estimator generalizes the constrained optimization program of Dubé et al. (2012) with probabilistic bounds on the estimation error in market shares. We show consistency as the number of markets T grows sufficiently slowly relative to the number of consumers n such that $\log(T)/n \to 0$, and provide confidence intervals under the same regime. Simulations suggest improved finite sample properties of the proposed estimator to conventional alternatives.

*University of Chicago, lieber@uchicago.edu.
†University of Chicago, wiemann@uchicago.edu.

1 Introduction

The problem of estimating demand parameters in settings with endogenous prices arises frequently in empirical economics. A key assumption in commonly applied demand models is that population-level market shares are observed without error. In practice, however, market shares are often estimated as averages of consumer choices. When the number of consumers over which these averages are computed is finite, these shares will be estimated with error. Estimation errors in the market shares propagate to estimation errors in the demand parameters which do not "average out" due to the nonlinearity of the discrete choice model.

One symptom of this is the so-called "zero-market-share problem" which arises when products are not purchased in every market (e.g., Dubé et al., 2021). Zero-valued market shares imply that the conventional demand estimators are not defined and cannot be computed, making the zero-market-share problem particularly salient. In response to zero-valued market shares, researchers often use ad-hoc solutions such as removing market-product combinations with no purchases. As noted in Dubé et al. (2021), researchers may not always be explicitly aware of ad-hoc manipulation of the data. Common datasets such as retail scanner data from IRI and Nielsen, for example, only report products with positive purchases. These data manipulations may however introduce additional biases and render existing theoretical guarantees on conventional estimators inapplicable.

In this paper, we propose a new estimator of demand parameters suitable for settings with estimated and possibly zero-valued market shares. The small but important departure from the demand models of Berry (1994) and Berry et al. (1995) is that we consider the observed market shares to be generated by a finite number of consumers. The estimator is based on a constrained optimization problem constructed by generalizing the mathematical program with equilibrium constraints (MPEC) formulation of Dubé et al. (2012) using known bounds on the estimation error in the observed market shares. We dub the estimator Estimated/Zero-share MPEC (EZ-MPEC) to highlight the applicability of the estimator in settings with estimated and zero-valued market shares. Our theoretical results show consistency of the estimator as the number of markets (T) and the number of consumers in each market (n) grow such that $\log(T)/n \to 0$. We further provide confidence intervals via test inversion under the same regime.

Although we focus on demand estimation based on random coefficient logit models, our results generalize to a larger class of demand estimators, including nonparametric demand estimation as in Tebaldi et al. (2019). To the best of our knowledge, this is the first application of finite-sample concentration bounds to the construction of a demand estimator for settings with estimated market shares and endogenous prices. In simulations, we highlight prevalence of biases arising through estimated market shares and ad-hoc data manipulations as well as illustrate the good performance of the EZ-MPEC estimator.

This paper contributes to the growing literature aiming to resolve the zero-market-share problem. One strand of this literature views the occurrence of zeros in the observed market shares as a rejection of the conventional random coefficient demand model and proposes extensions that motivate population-level market shares of zero. Gandhi et al. (2020) propose an asymptotic regime in which products are either "safe" with a population-level market share bounded away from zero or "risky" with a population-level market share of zero, and Dubé et al. (2021) incorporate consideration sets of consumers. The other strand of literature, to which we contribute primarily, focuses on sampling error of the market shares – and associated positive probability of zero-valued market shares – when these are constructed based on finitely many consumer purchases. To characterize the finite sample uncertainty in the observed market shares, Hortaçsu et al. (2021) consider a Bayesian IV approach and assume consumer arrivals follow a Poisson process. Regardless of the cause of the zero-shares considered by these approaches, however, the estimators proposed in existing literature rely on additional data not conventionally needed for demand estimation. In particular, instruments informative about the identity of safe products, the consumers' consideration sets, or the consumers' search behavior is needed, respectively. In contrast, the estimator proposed in this paper does not make a substantial structural deviation of the popular demand model and requires no additional data beyond the number of consumers in every market.

We also contribute to the literature on demand estimation with estimated market shares.

Berry et al. (2004) and Freyberger (2015) develop asymptotic distributions of the random coefficient logit demand estimator with estimated market shares when the number of products J or the number of markets T grow, respectively, and show that the number of customers n must grow sufficiently quickly to achieve asymptotic normality. Freyberger (2015) further shows that the conventional estimator is not centered at the true value unless $\sqrt{T}/n \to 0$. The bias correction and covariance adjustments suggested by this literature are not applicable, however, when an estimated market share is zero-valued. In contrast, the confidence intervals we provide hold for $\log(T)/n \to 0$ and can be computed regardless of whether zero-valued market shares occur in the data. This appears particularly important as the settings with small numbers of consumers in which the finite-sample corrections of Freyberger (2015) are most likely to be relevant are also the settings in which they are least likely to be applicable due to increased probability of zero-valued market shares in the data.

The rest of the paper proceeds as follows: Section 2 reviews the random coefficient logit model and discusses consequences of estimated market shares. Section 3 presents and discusses the EZ-MPEC estimator. Section 5 provides Monte Carlo simulations to illustrate finite sample performance of the proposed estimator to conventional alternatives. Section 6 concludes.

2 The random coefficient logit model

This section briefly reviews the random coefficient demand model of Berry (1994) and Berry et al. (1995), discusses the finite number of consumers adaptation, and illustrates its consequences for estimation.

We begin by defining the demand model in a single market. Consider a consumer i who chooses an alternative from $\mathcal{J} \equiv \{0, \ldots, J\}$ that maximizes their utility

$$Y_i = \underset{j \in \mathcal{J}}{\operatorname{argmax}}\ X_j^\top \beta_i + \xi_j + \varepsilon_{i,j}, \tag{1}$$

where X_j are observed product characteristics of the jth alternative including prices, β_i is a consumer-specific parameter vector β_i, ξ_j is the corresponding unobserved demand shock, and ε_{ij} is a consumer and product-specific latent utility shock. Throughout, $j = 0$ is the outside option with utility normalized to zero. Let X be the $J \times d_X$ vector with X_j as elements and define ξ analogously. To obtain the random coefficient logit model, we place distributional assumptions on the demand coefficients and latent utility shocks.

Assumption 1. *Consumers choose an alternative from* $\mathcal{J} \equiv \{0, \ldots, J\}$ *via* (1) $\forall i = 1, \ldots, n$. *The latent utility shocks* $\varepsilon_{i,j}$ *are i.i.d. T1EV. Customer preference parameters* β_i *are i.i.d. multivariate normal with parameter* $\theta \equiv (\mu, \Sigma)$.

Integrating over ε_i and β_i, results in the common expression for conditional choice probabilities (CCPs):

$$\pi_j(X, \xi; \theta) \equiv \Pr(Y_i = j | X, \xi; \theta) = \int \frac{\exp(X_j \beta + \xi_j)}{1 + \sum_{k=1}^{J} \exp(X_k \beta + \xi_k)} dF(\beta; \theta). \tag{2}$$

Since the latent demand shocks ξ are unobserved by the econometrician but potentially correlated with observed product characteristics such as prices, researchers frequently leverage instrumental variables Z for estimation of the demand parameters. Assumption 2 states a moment condition frequently employed in practice.

Assumption 2. *There exists an instrument* Z *such that* $E\left[Z^\top \xi\right] = 0$.

To allow for the application of common linear IV methodology despite the non-linear dependence of the CCPs in (2), Berry (1994) shows that for any (X, θ), there exists a bijective map between the value of the CCPs and the latent demand shocks. In particular, for all $s \in (0,1)^J : \|s\|_1 < 1$, there exists a unique $\xi \in \mathbb{R}^J$ such that $\pi(X, \xi; \theta) = s$, where $\pi(X, \xi; \theta)$ denotes the $J \times 1$ vector of CCPs. Replacing the unobserved demand shocks in the moment condition of Assumption 2 with these solutions $\xi(X, s; \theta)$, then motivates GMM estimators that replaces s with the strictly positive CCPs.

In practice the econometrician may not observe market shares sampled directly from the CCPs as postulated in Berry et al. (1995). Instead, market shares are frequently estimated as sample averages over consumer choices. Assumption 3 highlights this small but critical deviation from the conventional demand models with observed CCPs.

Assumption 3. *Observed market shares are sample averages of consumer choices*

$$\hat{S}_j^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{Y_i = j\},\ \forall j \in \mathcal{J}.$$

Settings with observed market shares constructed from a finite sample of consumer purchase decisions as considered here introduce several challenges to demand estimation. In particular, due to the nonlinear nature of the conventional demand model in equation (2), the sampling error in the market shares does not straightforwardly average out and instead introduces an incidental parameter problem. For example, consider the conventional GMM estimator in a setting with many markets given by

$$\hat{\theta}_{blp} = \arg\min_{\theta\in\Theta} \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi(X_t, S_t; \theta)\right)^\top W_T \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi(X_t, S_t; \theta)\right), \tag{3}$$

where is $(S_t, X_t, Z_t) \overset{iid}{\sim} (\pi(X, \xi; \theta_0)), X, Z)$ is a sample of $t = 1, \ldots, T$ markets, and W_T is a positive-definite weighting matrix, often chosen to be $W = (\sum_{t=1}^{T} Z_t Z_t^\top)^{-1}$. Freyberger (2015) characterizes the asymptotic distribution of $\hat{\theta}_{blp}$ when S_t are replaced with estimated market shares $\hat{S}_t^n$ as $T \to \infty$. Similarly, Berry et al. (2004) characterizes the analogue estimator with estimated market shares in a setting with many products.[1] The authors show that unless the number of consumers grow at sufficiently fast rate, the conventional estimator is not $\sqrt{T}$ or $\sqrt{J}$ Gaussian. Freyberger (2015) further shows that $\hat{\theta}_{blp}$ with estimated market shares is not bounded in probability at the rate $\sqrt{T}$ unless $\sqrt{T}/n \to c$ for some finite constant c. The author proposes a bias correction to improve finite sample performance, which is shown to improve performance in Monte Carlo simulations.

In addition to the incidental parameter problem, estimated market shares can be cause for the zero-market-share problem. This is because for any product-market combination, $n\hat{S}_{jt}^{(n)} \sim \text{Binomial}(\pi_j(X_t, \xi_t; \theta_0), n)$, implying a strictly positive probability for the event that some products do not have purchases in every market (i.e., $\Pr(\exists j, t : \hat{S}_{jt} = 0) > 0$). The

[1] Berry et al. (2004) and Freyberger (2015) also accommodate for sampling uncertainty from the Monte Carlo integration of the integral in equation (2). We focus on sampling uncertainty from the estimation of market shares only here, but note that our application of concentration inequalities straightforwardly extends to the Monte Carlo integration error.

occurrence of zero-valued market shares in particular has received increased attention in recent literature due to both the abundance of economic settings with no observed purchases for some products and its severe consequences for conventional estimation approaches. An important aspect of estimators such as $\hat{\theta}_{blp}$ leveraging the demand inversion directly is that they require the set of market shares $(S_t)_{t=1}^T$ to be positive since otherwise $\xi(\cdot)$ is not defined. The infeasiblity of $\hat{\theta}_{blp}$ with estimated zero-valued market shares is particularly evident using the MPEC formulation proposed by Dubé et al. (2012):

$$\begin{aligned} \min_{(\theta,\xi)} \quad & \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right)^\top W_T \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right) \\ \text{s.t.} \quad & S_{jt} = \pi_j(X_t, \xi_t; \theta),\ \forall j, t. \end{aligned} \tag{4}$$

Dubé et al. (2012) show that MPEC is equivalent to the previously defined GMM estimator $\hat{\theta}_{blp}$. Since the domain of the CCPs $\pi_j(\cdot)$ is strictly positive, presence of any vector of market shares S_t that is not strictly positive implies that no feasible solution to the mathematical program exists. Replacing $(S_t)_{t=1}^T$ with their estimated counterparts can thus result in infeasibility of $\hat{\theta}_{blp}$. Further, the bias corrections suggested by Freyberger (2015) cannot be computed for the same reasons.

In practice, researchers often apply ad-hoc manipulations to their data when zero-valued market shares arise to be able to apply conventional random coefficient logit estimators. Popular approaches appear to be: 1) Replacing the zero-valued market shares with an arbitrary small number (e.g., $0.5/n$), or 2) removing the product-market combinations from the sample that are associated with zero-valued market shares (Quan and Williams, 2018; Gandhi et al., 2020; Dubé et al., 2021).[2] The bias of the first approach is likely sensitive to the specific small number used (see, e.g., Dubé et al., 2012). We are more concerned with the second approach of truncating the data, as this can lead to substantial selection bias. In

[2] Alternatively, researchers may choose to aggregate purchases across products and markets until estimated shares are strictly positive. This introduces measurement error and limits the type of question that can be answered as products and markets are combination are often artificial, smoothing across relevant heterogeneity and making the results challenging to interpret. Quan and Williams (2018) give an augmented nested logit model under which local demand heterogeneity is identified using aggregated rather than local market shares.

particular, while by assumption $E[Z_t^\top \xi_t] = 0$ for all markets for identification, the moment conditions $E[Z_t^\top \xi_t | \hat{S}_t^{(n)} > 0]$ do not hold in general because conditional on $\hat{S}_t^{(n)} > 0$ the latent demand shocks are less likely to be negative. The next section develops an alternative estimator that explicitly takes sampling error in the marketing shares into account and can be applied to settings with zero-valued market shares.

3 Estimation

To address the issues arising from estimation errors in market-shares, we combine the MPEC estimator (4) developed by Dubé et al. (2012) with finite-sample confidence intervals on the observed market shares. Throughout, we consider an i.i.d. sample across markets.

Assumption 4. *The data is an i.i.d. sample* $\{(\hat{S}_t^{(n)}, X_t, Z_t)\}_{t=1}^T$ *from* $(\hat{S}^{(n)}, X, Z)$.

We propose to consider any solution to

$$\begin{aligned} \min_{(\theta,\xi)} \quad & \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right)^\top W_T \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right) \\ \text{s.t.} \quad & \hat{S}_{jt}^{(n)} \in C_{n,T}^j(X_t, \xi_t; \theta, \alpha), \ \forall j, t. \end{aligned} \tag{5}$$

Compared to (4), we replace the constraint that observed market shares equal model-implied expected CCPs with the constraint that observed market shares must be contained in a set $C_{n,T}$. As in the equality constraints of (4), $C_{n,T}$ depends on product characteristics, the latent demand shocks, and the demand parameter. In addition, the set depends on the number of consumers n, the number of markets T, and a hyper-parameter $\alpha \in (0, 1)$.[3]

The purpose of the sets $C_{n,T}$ is to bound the sampling errors in the estimation of the market shares. We choose this set to be a joint confidence set that covers the true population-level market shares with probability at least $1 - \alpha$ for any finite number of consumers. These finite-sample confidence intervals provide a probabilistic bound on the deviations of the estimated market shares from their population-values. These joint confidence intervals are constructed in two steps: First, we consider a particular product-market combination and

[3] $C_{n,T}$ further depends on the number of products J. Since J is fixed throughout, we omit this dependence.

derive a probabilistic bound on the estimation error in the observed market share. Second, we adjust the marginal confidence level to achieve uniform coverage at a desired rate. We now discuss both steps in turn.

Multiple methods to derive finite-sample confidence sets exist. In choosing a method choice, the researcher faces a trade-off between tractability and tightness. On one end of the spectrum are confidence-intervals based on Hoeffding's inequality. These confidence intervals lend themselves to straightforward computation of the EZ-MPEC estimator as they correspond to constraints which are linear in the model-implied choice probabilities. EZ-MPEC based on these bounds thus exhibits the same Jacobian and Hessian used for solving infeasible MPEC.[4] However, Hoeffding's inequality is conservative, allowing for "too much" sampling error in the market shares. On the other end of the spectrum, binomial quantiles have exact coverage but imply an increased computational burden on the constrained optimization problem due to the non-linear dependence of the confidence sets on the model-implied CCPs.[5]

Given a choice of bound, the marginal confidence intervals need to be adjusted to achieve joint coverage for all estimated market shares.

Proposition 1. *Let assumptions 1-4 hold. Fix $\alpha \in (0,1)$. Then with Hoeffding's inequality, it holds that*

$$Pr\left(\exists (j,t) \in \{1,\ldots,J\} \times \{1,\ldots,T\} :\ |\hat{S}_{jt} - \pi_j(X_t,\xi_t;\theta_0)| \geq \sqrt{\frac{\log\left(\frac{2J}{1-\sqrt[T]{1-\alpha}}\right)}{2n}}\right) \leq \alpha,$$

$\forall n \in \mathbb{N}_{++}$. *With binomial quantiles, it holds that*

$$Pr\Bigg(\exists (j,t) \in \{1,\ldots,J\} \times \{1,\ldots,T\} :\ \hat{S}_{jt} \notin \Bigg[\frac{1}{n}F^{-1}_{Bin}\left(\frac{1-\sqrt[T]{1-\alpha}}{J}, \pi_j(X_t,\xi_t;\theta_0), n\right),$$
$$\frac{1}{n}F^{-1}_{Bin}\left(1 - \frac{1-\sqrt[T]{1-\alpha}}{J}, \pi_j(X_t,\xi_t;\theta_0), n\right)\Bigg]\Bigg) \leq \alpha,$$

[4]We note that $C_{n,T}$ based on Hoeffding's inequality also allows using finite-sample confidence intervals for the nonparametric demand model proposed in Tebaldi et al. (2019) which requires linearity for tractability. See Appendix A.4 for further illustration.

[5]In Appendix A.5, we give implementation details for feasible MPEC using the binomial quantiles.

$\forall n \in \mathbb{N}_{++}$, where $F_{Bin}^{-1}(\cdot, p, n)$ denotes the quantile function of the binomial distribution with n trials and success probability $p \in (0,1)$. For $J = 1$, the second inequality holds with equality.

Remark 1. *Although we focus here on the random coefficient logit model due to its popularity in demand estimation, we highlight that these bounds do not depend on the specific demand model but apply to any discrete choice setting. In particular, researchers may substitute CCPs implied by any discrete choice model for $\pi(\cdot)$.*

Proposition 1 provides explicit bounds on the estimation error of all market shares simultaneously that hold with probability at least $1 - \alpha$. Since these bounds depend on the true CCPs, $\pi(X_t, \xi_t; \theta_0)$, they can be leveraged in estimation. For example, when using bounds based on Hoeffding's inequality, the EZ-MPEC estimator is given by

$$\begin{aligned}
\min_{(\theta,\xi)} \quad & \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right)^\top W_T \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right) \\
\text{s.t.} \quad & \hat{S}_{jt}^{(n)} - \pi_j(X_t, \xi_t; \theta) \leq \sqrt{\frac{\log\left(\frac{2J}{1-\sqrt[T]{1-\alpha}}\right)}{2n}}, \quad \forall j, t, \\
& \pi_j(X_t, \xi_t; \theta) - \hat{S}_{jt}^{(n)} \leq \sqrt{\frac{\log\left(\frac{2J}{1-\sqrt[T]{1-\alpha}}\right)}{2n}}, \quad \forall j, t.
\end{aligned}$$

4 Asymptotic Properties

This section provides formal consistency and inference results based on large n and T asymptotics. We begin with listing additional assumptions and emphasize that all below are also assumed in Freyberger (2015).[6]

Assumption 5. $\exists \gamma \in (0,1)$ *such that* $P(\pi(X, \xi; \theta_0) \in [\gamma, 1-\gamma]^J) = 1$.

Assumption 6. Θ, $\operatorname{supp} X$, *and* $\operatorname{supp} Z$ *are compact.*

[6]Notation: When A is a matrix, $\|A\| = (A^\top A)^{1/2}$. Else, $\|\cdot\|$ denotes the Euclidean norm. Further, we denote a neighborhood by $\mathcal{N}_{x_0}(\delta) \equiv \{x \in \mathcal{X}_0 : \|x - x_0\| \leq \delta\}$.

Assumption 7. *The matrix* $\frac{1}{T}\sum_{t=1}^{T} Z_t^\top Z_t$ *has full rank and is stochastically bounded, i.e.,* $\forall \varepsilon > 0$ *there exists an* $M(\varepsilon)$ *such that* $\Pr\left(\|\frac{1}{T}\sum_{t=1}^{T} Z_t^\top Z_t\| > M(\varepsilon)\right) < \varepsilon$.

Assumption 8. $\forall \delta > 0$, $\exists M(\delta) > 0$, *such that*

$$\lim_{T\to\infty} \Pr\left(\inf_{\theta\notin\mathcal{N}_{\theta_0}(\delta)} \|G_T(\theta) - G_T(\theta_0)\| \geq M(\delta)\right) = 1,$$

where G_T *is the objective function in* (3).

Assumptions 5 and 6 restrict the support of the population-level market shares, the product characteristics, and the parameter space. Assumption 7 places moment restrictions on the instrument vectors. Finally, Assumption 8 assumes identification of the demand parameter from the moment conditions in Assumption 2.

These assumptions are leveraged to show convergence of the EZ-MPEC estimator to the demand parameters θ_0 for many markets T and many consumers per market n at the rate $\log(T)/n \to 0$. For asymptotic analysis, we let the confidence parameter α depend on (n, T).

Theorem 1. *Let assumptions 1 to 8 hold. If in addition* $\alpha_{n,T} \in (0,1) : \alpha_{n,T} = o_p(1)$ *and* $\log(T) = o_p(n)$, *then* $\forall \epsilon > 0$,

$$\lim_{n,T\to\infty} \Pr\left(\sup_{\tilde{\theta}\in\Theta^*_{n,T}} \left\|\tilde{\theta} - \theta_0\right\| > \epsilon\right) = 0,$$

where $\Theta^*_{n,T}$ *denotes the* $\arg\min$ *of the EZ-MPEC estimator in* (5) *with bounds based on Proposition 1 and hyperparameter* $\alpha_{n,T}$, *and* θ_0 *are the true demand parameters.*

Remark 2. *While Theorem 1 is formulated for EZ-MPEC based on the inequalities of Proposition 1, it also applies to many other methods to derive finite-sample confidence intervals.* (5) *combined with any method to construct finite-sample confidence intervals enjoys the properties of Theorem 1 as long as the confidence intervals are not larger than the confidence intervals based on Hoeffding's inequality. This includes in particular multinomial quantiles.*

To obtain confidence intervals with sufficient coverage of θ_0, we rely on inversion of tests of null hypotheses of the form

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta \neq \theta_0.$$

For this purpose, we propose the test statistic $\hat{G}_T$ given by

$$\begin{aligned} \hat{G}_T(\alpha) \equiv \min_{\{\tilde{\xi}_t\}_{t=1}^T} \quad & T\left(\frac{1}{T}\sum_{t=1}^T Z_t^\top \tilde{\xi}_t\right)^\top \left(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t^\top Z_t Z_t^\top \hat{\xi}_t\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^T Z_t^\top \tilde{\xi}_t\right) \\ \text{s.t.} \quad & \hat{S}_{jt}^{(n)} \in C_{n,T}^j(X_t, \tilde{\xi}_t; \theta, \alpha),\ \forall j, t, \end{aligned} \tag{6}$$

where $(\hat{\xi}_t)_{t=1}^T$ are consistent estimates of the latent demand shocks obtained from computing EZ-MPEC in a first step. The test statistic is akin to the χ^2 test statistic conventionally leveraged in GMM inference of the random coefficient logit models of Berry et al. (1995), but accounts for the sampling error in the market shares. Theorem 2 gives a test based on (6) with controlled size.

Theorem 2. *If the assumptions of Theorem 1 hold, in particular,* $\log(T) = o_p(n)$ *and* $\alpha_{n,T} \in (0,1) : \alpha_{n,T} = o_p(1)$*, then under* H_0

$$\limsup_{n,T\to\infty} E\left[\mathbb{1}\{\hat{G}_T(\alpha_{n,T}) > c_K^{1-\tau}\}\right] \leq \tau,$$

for any $\tau \in (0,1)$*, where* $c_K^{\alpha/2}$ *is the* $1 - \frac{\alpha}{2}$ *quantile of a* χ^2 *distribution with* K *degrees of freedom.*

Importantly, Theorem 1 and Theorem 2 hold as $\log(T) = o_p(n)$. In contrast, Freyberger (2015) shows that the naive estimator that replaces the population-level market shares in infeasible MPEC with their estimates is asymptotically normal only $\sqrt{T} = o_p(n)$. As $\log(T) << \sqrt{T}$ for large T, our estimator is robust to situations when there are much fewer consumers per market. Even stronger, Freyberger (2015) shows that in the asymptotic regime we consider, the naive estimator that replaces the population-level market shares in infeasible MPEC with their estimates incurs a bias that is not bounded in probability when rescaled with $\sqrt{T}$.

5 Monte Carlo Simulation

We conduct Monte Carlo simulations to illustrate the implications of estimated market shares for conventional random coefficient logit estimators that remove zero-valued market shares from their data and highlight improvements of the proposed EZ-MPEC estimator based on binomial quantiles. The setting is similar to the simulations reported in Dubé et al. (2012).

Each of the simulations considers a setting with $J = 5$ products (and an outside option) and $T = 50$ markets. We let the number of consumers per market vary across simulations to analyze settings with varying sampling uncertainty in the estimated market shares. Each consumer makes their purchasing decision with CCP associated with product j given in Assumption 1, where

$$X_j^\top \beta_i = \beta_i^0 + W_j^\top \beta_i^w - p_j \beta_i^p,$$

and $\beta_i = (\beta_i^0, \beta_i^{w\top}, \beta_i^p)$. Throughout, we take

$$W_j = \begin{bmatrix} W_j^1 \\ W_j^2 \\ W_j^3 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.55 & -.25 & 0.2 \\ -.25 & 1.55 & 0.2 \\ 0.2 & 0.2 & 1.55 \end{bmatrix} \right), \quad \text{and} \quad \xi_j \sim N(0,1),$$

and let the endogenous price be given by $p_j = \max(0.1\xi_j + e_j, , 0.01)$, with $e_j \sim N(0,1)$. Note that endogeneity is introduced through dependence of price on the latent demand shock ξ_j. For each product, there is an additional underlying instrument vector Z_j of dimension 6×1 with entries generated $Z_j^m = U(0,1) + 0.25e_j$, where $U(0,1)$ is the realization of a standard uniformly distributed random variable. In estimation, we construct a higher order standard polynomial expansion of Z_j with the product characteristics W_j.[7]

Our analysis applies the proposed EZ-MPEC estimator in (5) where the bounds C_n^α are based on the Binomial quantiles. We choose $\alpha = 0.1$ so that the set of all $J \times T$ sampling error conditions hold jointly with probability at least 0.9.

[7]Following a similar approach as in Dubé et al. (2012), we consider $Z_j, Z_j^2, Z_j^3, W_j, W_j^2, W_j^3, \prod_{m=1}^6 Z_j^m, \prod_{k=1}^3 W_j^k, Z_j \cdot W_j^1, Z_j \cdot W_j^2$. This results in a total of 42 moment conditions.

We begin by considering a data generating process where the demand parameters β_i are fixed across consumers at $\beta^0 = 0$, $\beta^w = (1\ 1\ \text{-}1)^\top$, and $\beta^p = -3$. In this logit setting without random coefficients, demand inversion with population-level market shares motivates a simple two-stage least squares (TSLS) estimator with $\log(S_j/S_0)$ as the second stage left-hand side variable. We compare EZ-MPEC to both the infeasible TSLS estimator using the population-level market shares as well as to the ad-hoc TSLS estimator that removes product-market combinations with zero-valued market shares.

Table 1 presents the results for 1,000 simulations. We focus on the median absolute error (MAE) of the demand parameter corresponding to the endogenous price variable ($\beta^p = -3$) as the evaluation criterion in columns (1)-(3) to assess both centrality and dispersion. Column (4) gives the average share of zero-valued market shares in the sample, which corresponds to the share of the $J \times T$ observations removed from the data in column (2).

TSLS with population-level market shares in column (1) of Table 1 has an MAE of approximately 0.09.[8] In contrast, the TSLS estimator with estimated sample shares (TSLS: Ad-Hoc) has substantially higher MAE as large as 0.44 at 250 consumers per market. This difference reduces as the number of consumers per market grow and corresponding share of zero-valued market shares decreases, yet, even with 5000 consumers per market do not suffice to estimate the market shares at sufficient accuracy to ignore their sampling error.

The EZ-MPEC estimator in column (3) of Table 1 improves over the TSLS estimator with estimated market shares. In particular for markets with a small numbers of consumers between 500-1000, the differences are substantial. For example, at 750 consumers per market, the EZ-MPEC estimator as an MAE that is 0.103 smaller than TSLS: Ad-Hoc. Given that the sampling uncertainty of the infeasible TSLS estimator with population-level market shares corresponds to an MAE of approximately 0.09, this highlights that ignoring sampling error in the market shares can lead to qualitatively very different results. As the number of consumers increase and the share of zero-valued market shares decreases, the incidental parameter and selection bias of the TSLS estimator with estimated market shares decrease,

[8]Note that because consumer markets have no implications for the population-level market shares, any differences in the corresponding TSLS estimator are due to sampling uncertainty of the T markets.

reducing the magnitude of the MAE to that of the EZ-MPEC estimator as expected.

Table 1: Mean Absolute Error for DGP w/o Random Coefficients

# Consumers (zero-share)	TSLS: Infeasible (1)	TSLS: Ad-Hoc (2)	EZ-MPEC (3)
500 (0.133)	0.094	0.317	0.231
750 (0.110)	0.091	0.289	0.186
1000 (0.095)	0.088	0.260	0.185
2000 (0.068)	0.091	0.191	0.166
3000 (0.054)	0.084	0.167	0.154
4000 (0.046)	0.088	0.173	0.150
5000 (0.040)	0.092	0.150	0.158

Notes. Results based on 1,000 Monte Carlo simulations. TSLS: Infeasible and TSLS: Ad-Hoc denote two-stage least squares estimators using the population-level and estimated market shares, respectively. TSLS: Ad-Hoc is computed using only those product-market combinations with positive estimated market shares. Parentheses state the average fraction of observations with zero-valued estimated market shares.

In a second set of simulations, we consider a data-generating process based on the random coefficient logit model. In particular, the consumer-specific demand parameters β_i are uncorrelated and generated as $\beta_i^0 \sim N(0, 0.25)$, $\beta_i^w \sim N((1\,1\,\text{-}1)^\top, 0.25 I_3)$, and $\beta_i^p \sim N(-3, 1)$. In this random coefficient logit setting without random coefficients, demand inversion with population-level market shares motivates estimation via a nested-fixed point estimation as in Berry et al. (1995) or via the MPEC estimator of Dubé et al. (2012) given in (4). Both approaches target identical estimands but the MPEC estimator has computational and numerical advantages (Dubé et al., 2012). Similar to the first simulation, we compare EZ-MPEC to both the infeasible MPEC estimator using the population-level market shares as well as to the feasible MPEC estimator based on estimated market shares that removes product-market combinations with zero-valued market shares. For all estimators, we evaluate the integral in (2) with 200 Monte Carlo draws. Because we do not focus on numerical integration here, we use the same 200 draws in data generation and estimation.

Table 2 presents the results for 1,000 simulations. As before, we focus on the median absolute error (MAE) of the mean demand parameter corresponding to the endogenous price variable ($E[\beta_i^p] = -3$). The share of zero-valued market shares in column (4) is the share of the sample dropped from the data when computing the feasible MPEC estimator based on estimated market shares (MPEC: Ad-Hoc) in column (2).

The infeasible MPEC estimator that uses the population-level market shares reported in column (1) of Table 2 has an MAE of approximately 0.14. In markets with a small number of consumers between 500-1000, the ad-hoc MPEC has MAEs between approximately 0.263-0.312, highlighting again the relative importance of the sampling uncertainty associated with market shares. In contrast to the setting without random coefficients, the EZ-MPEC estimator does not improve over the ad-hoc MPEC estimator for these small markets. When the number of consumers increases, however, the MAE of the EZ-MPEC estimator decreases at faster rate. We expect that this is due to the fast contraction of the Binomial quantiles as the number of consumers grow, which are used for constructing the feasible set in estimation, relative to the reduction incidental parameter and selection bias that the ad-hoc MPEC estimator suffers from.

Table 2: Mean Absolute Error for DGP w/ Random Coefficients

# Consumers (zero-share)	MPEC: Infeasible (1)	MPEC: Ad-Hoc (2)	EZ-MPEC (3)
500 (0.091)	0.146	0.312	0.350
750 (0.072)	0.143	0.296	0.298
1000 (0.061)	0.148	0.263	0.250
2000 (0.040)	0.151	0.222	0.205
3000 (0.031)	0.142	0.213	0.189
4000 (0.026)	0.146	0.199	0.178
5000 (0.023)	0.134	0.199	0.171

Notes. Results based on 1,000 Monte Carlo simulations. MPEC: Infeasible and MPEC: Ad-Hoc denote MPEC estimators using the population-level and estimated market shares, respectively. MPEC: Ad-Hoc is computed using only those product-market combinations with positive estimated market shares. Parentheses state the average share of observations with zero-valued estimated market shares.

6 Conclusion

This paper proposes a new EZ-MPEC estimator for demand estimation in settings with endogenous prices and estimated market shares. The estimator is constructed by generalizing the constrained optimization formulation of Dubé et al. (2012) for the random coefficient logit model of Berry et al. (1995) using probabilistic bounds on the sampling error of market shares. We show that the estimator is consistent as the number of markets grow large $T \to \infty$ and the number of consumers per market n grows at appropriate rate such that $\log(T)/n \to 0$. Under analogous conditions, we further provide confidence intervals that contain the true demand parameters at pre-specified confidence level.

Two Monte Carlo simulations illustrate the importance of estimation error in market shares and showcase that the incidental parameter and selection problem that conventional estimators suffer from can be substantial. In these settings, application of the proposed EZ-MPEC estimator can lead to meaningful improvements.

References

Berry, Steven, Levinsohn, James, and Pakes, Ariel (1995). Automobile Prices in Market Equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Berry, Steve, Linton, Oliver B, and Pakes, Ariel (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654.

Berry, Steven T (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Dubé, Jean-Pierre, Fox, Jeremy T, and Su, Che-Lin (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267.

Dubé, Jean-Pierre, Hortaçsu, Ali, and Joo, Joonhwi (2021). Random-coefficients logit demand estimation with zero-valued market shares. *Marketing Science*.

Freyberger, Joachim (2015). Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics*, 185(1):162–181.

Gandhi, Amit, Lu, Zhentong, and Shi, Xiaoxia (2020). Estimating Demand for Differentiated Products with Zeroes in Market Share Data. *Available at SSRN 3503565*.

Hortaçsu, Ali, Natan, Olivia R, Parsley, Hayden, Schwieg, Timothy, and Williams, Kevin R (2021). Incorporating search and sales information in demand estimation. Technical report, National Bureau of Economic Research.

Quan, Thomas W and Williams, Kevin R (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics*, 49(4):877–913.

Tebaldi, Pietro, Torgovitsky, Alexander, and Yang, Hanbin (2019). Nonparametric estimates of demand in the california health insurance exchange. Technical report, National Bureau of Economic Research.

Vershynin, Roman (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

A Appendix

A.1 Proof of Proposition 1

We begin by establishing the following auxiliary result.

Lemma 1. *Fix* $n \in \mathbb{N}_{++}$ *and* $\tau \in (0,1)$. *Let* $(Z_i)_{i\in\{1,\ldots,n\}}$ *be a sequence of random variables such that*

1. $(Z_i)_{i\in\{1,\ldots,n\}}$ *is a family of independent random variables,*
2. *for all* i, $\operatorname{supp} Z_i \subset [0,1]$.

Then

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mathbb{E}[Z_i])\right| \geq \sqrt{\frac{\log\left(\frac{2}{\tau}\right)}{2n}}\right) \leq \tau.$$

Proof of Lemma 1. Theorem 2.2.6 in Vershynin (2018) with $M_i = 1/n$, $m_i = 0$ implies that for any $t > 0$,

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mathbb{E}[Z_i]) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}\left(\frac{1}{n}-0\right)^2}\right) = \exp\left(-2t^2 n\right).$$

Then,

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mathbb{E}[Z_i])\right| \geq t\right) \\
=&\Pr\left(\left\{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mathbb{E}[Z_i]) \geq t\right\} \bigcup \left\{\frac{1}{n}\sum_{i=1}^{n}(-Z_i - \mathbb{E}[-Z_i]) \geq t\right\}\right) \\
\leq&2\exp\left(-2t^2 n\right)
\end{aligned}$$

where the last inequality applies a union bound. Finally, the right hand side is equal to τ if $t = \sqrt{\frac{\log\left(\frac{2}{\tau}\right)}{2n}}$. □

Proof of Proposition 1. Let $C^j_{n,T}(X,\xi;\theta,\tau)$ denote either

$$\left[\pi_j(X,\xi;\theta) - \sqrt{\frac{\log\left(\frac{2}{\tau}\right)}{2n}}, \pi_j(X,\xi;\theta) + \sqrt{\frac{\log\left(\frac{2}{\tau}\right)}{2n}}\right],$$

or

$$\left[\frac{1}{n}F^{-1}_{\text{Bin}}\left(\frac{\tau}{2}, \pi_j(X,\xi;\theta), n\right), \frac{1}{n}F^{-1}_{\text{Bin}}\left(1-\frac{\tau}{2}, \pi_j(X,\xi;\theta), n\right)\right].$$

Note that by Assumption 1-3, $E[\hat{S}^{(n)}_{jt}] = \pi_j(X_t,\xi_t;\theta_0)$, for all j and t. By Lemma 1 and the definition of quantiles, it follows that in either case

$$\Pr\left(\hat{S}_{jt} \in C^j_{n,T}(X_t,\xi_t;\theta_0,\tau)\right) \geq 1-\tau. \tag{7}$$

We then have

$$\begin{aligned}
\Pr\left(\forall j,t:\ \hat{S}_{j,t} \in C^j_{n,T}(X_t,\xi_t;\theta_0,\tau)\right) &= \Pr\left(\bigcap_{t=1}^{T}\bigcap_{j=1}^{J}\left\{\hat{S}_{jt} \in C^j_{n,T}(X_t,\xi_t;\theta_0,\tau)\right\}\right) \\
&\overset{[1]}{=} \prod_{t=1}^{T}\Pr\left(\bigcap_{j=1}^{J}\left\{\hat{S}_{jt} \in C^j_{n,T}(X_t,\xi_t;\theta_0,\tau)\right\}\right) \\
&= \prod_{t=1}^{T}\left[1-\Pr\left(\bigcup_{j=1}^{J}\left\{\hat{S}_{jt} \notin C^j_{n,T}(X_t,\xi_t;\theta_0,\tau)\right\}\right)\right] \\
&\overset{[2]}{\geq} \prod_{t=1}^{T}\left[1-\sum_{j=1}^{J}\Pr\left(\hat{S}_{jt} \notin C^j_{n,T}(X_t,\xi_t;\theta_0,\tau)\right)\right] \\
&\overset{[3]}{\geq} \left[1-J\tau\right]^T,
\end{aligned}$$

where [1] follows from Assumption 4, [2] follows from the union bound, and [3] follows from inequality (7) whenever $J\tau \leq 1$. Finally, setting the right hand side equal to $1-\alpha$ and solving for τ, we have $\frac{1-\sqrt[T]{1-\alpha}}{J} \geq \tau$. □

A.2 Proof of Theorem 1

Proof. The proof proceeds as follows. First, we state an equivalent formulation of the EZ-MPEC estimator using the demand inversion of Berry (1994). Second, we show that consistency of the estimator is implied by consistency in the latent demand shocks. Finally, we show consistency in the latent demand shocks.

Remark 3. *The second step of the proof, showing that consistency of the demand estimator is is implied by consistency in the latent demand shocks, relies heavily on the proof of Freyberger (2015), who shows consistency when population-level market shares are replaced by their estimates. We adapt the proof to allow for indeterminacy due to the set-constraints but follow the same arguments elsewhere. Further, when proving consistency in the latent demand shocks, we leverage a lemma that follows from the proof of Freyberger (2015).*

A.2.1 An Equivalent EZ-MPEC

By Berry et al. (1995), for any realization of the product characteristics and a given value of the demand parameters θ, there exists a bijective map between sample shares and latent demand shocks ξ. Let this map be denoted by $\xi(\cdot)$. For notational convenience, let $\xi_t(s;\theta) \equiv \xi(X_t, s; \theta)$ and similarly $\pi_t(\xi;\theta) \equiv \pi(X_t, s; \theta)$.

Define for any $\alpha \in (0,1)$

$$\mathcal{S}^t_{n,T}(\alpha) = \{s_t \in (0,1)^J : s_t \in C^t_{n,T}(\alpha),\ \|s_t\|_1 < 1\}$$
$$\mathcal{S}_{n,T}(\alpha) = \{\mathbf{s}_T \in (0,1)^{J\times T} : s_t \in \mathcal{S}^t_{n,T}(\alpha), \forall t = 1,\dots,T\},$$

where

$$C^t_{n,T}(\alpha) \equiv \left[\hat{S}^{(n)}_t - \delta_{n,T}(\alpha),\ \hat{S}^{(n)}_t + \delta_{n,T}(\alpha)\right]$$
$$\delta_{n,T}(\alpha) \equiv \sqrt{\frac{\log\left(\frac{2JT}{\alpha}\right)}{2n}}$$

where $\hat{S}^{(n)}_t$ is the vector of estimated market shares in market t with n consumers.

Let G_T denote the objective function in a sample of size T defined by

$$G_T(\theta, \mathbf{s}_T) = \left(\frac{1}{T} \sum_{t=1} Z_t^\top \xi_t(s_t;\theta) \right)^\top W_T \left(\frac{1}{T} \sum_{t=1} Z_t^\top \xi_t(s_t;\theta) \right).$$

The EZ-MPEC estimator in (5) is then equivalent to

$$\begin{aligned} \min_{\theta\in\Theta,\, \mathbf{s}_T \in (0,1)^{J\times T}} \quad & G_T(\theta, \mathbf{s}_T) \\ \text{s.t.} \quad & \mathbf{s}_T \in \mathcal{S}_{n,T}(\alpha), \end{aligned} \tag{8}$$

A.2.2 Consistency implied by Consistency in Latent Demand Shocks

First, we show that any estimator $\check{\theta}$ is consistent if

$$\|G_T(\check{\theta}, \mathbf{S}_T)\| = \inf_{\theta\in\Theta} \|G_T(\theta, \mathbf{S}_T)\| + o_p(1). \tag{9}$$

Fix $\delta > 0$. Note that

$$\begin{aligned} \Pr\left(\|\check{\theta} - \theta_0\| \geq \delta\right) =& \Pr\left(\|\check{\theta} - \theta_0\| \geq \delta,\ \|G_T(\check{\theta}, \mathbf{S}_T) - G_T(\theta_0, \mathbf{S}_T)\| \geq C(\delta)\right) \\ &+ \Pr\left(\|\check{\theta} - \theta_0\| \geq \delta,\ \|G_T(\check{\theta}, \mathbf{S}_T) - G_T(\theta_0, \mathbf{S}_T)\| < C(\delta)\right) \\ \leq& \Pr\left(\|G_T(\check{\theta}, \mathbf{S}_T) - G_T(\theta_0, \mathbf{S}_T)\| \geq C(\delta)\right) \\ &+ \Pr\left(\inf_{\theta\notin N_{\theta_0}(\delta)} \|G_T(\theta, \mathbf{S}_T) - G_T(\theta_0, \mathbf{S}_T)\| < C(\delta)\right), \end{aligned}$$

where by Assumption 8

$$\lim_{T\to\infty} \Pr\left(\inf_{\theta\notin N_{\theta_0}(\delta)} \|G_T(\theta, \mathbf{S}_T) - G_T(\theta_0, \mathbf{S}_T)\| < C(\delta)\right) = 0.$$

For the first term, it holds that

$$\begin{aligned} \|G_T(\check{\theta}, \mathbf{S}_T) - G_T(\theta_0, \mathbf{S}_T)\| &\leq \|G_T(\check{\theta}, \mathbf{S}_T)\| + \|G_T(\theta_0, \mathbf{S}_T)\| \\ &\stackrel{[1]}{=} \|G_T(\theta_0, \mathbf{S}_T)\| + \inf_{\theta\in\Theta} \|G_T(\theta, \mathbf{S}_T)\| + o_p(1) \\ &\leq 2\|G_T(\theta_0, \mathbf{S}_T)\| + o_p(1), \end{aligned}$$

where [1] applies Equation (9). Further, by the discussion in Appendix C of Freyberger (2015), assumption 4, 5 and 6, imply that the support of ξ_t is bounded. Using in addition 7, Kolmogorov's law of large numbers gives $\|G_T(\theta_0, \mathbf{S}_T)\| = o_p(1)$. Combining, we thus have

$$\Pr\left(\|\check{\theta} - \theta_0\| \geq \delta\right) = o_p(1)$$

for any $\check{\theta}$ such that (9).

Next, we show that if

$$\sup_{\theta\in\Theta}\left\| \left[\inf_{\tilde{\mathbf{s}}\in\mathcal{S}_{n,T}} G_T(\theta, \tilde{\mathbf{s}})\right] - G_T(\theta, \mathbf{S}_T)\right\| = o_p(1) \tag{10}$$

then (9) holds.

Take any $(\theta_T)_{T=1}^{\infty}$ with $\theta_T \in \Theta$ and note

$$\begin{aligned}
\left\|\left\|\inf_{\tilde{\mathbf{s}}\in\mathcal{S}_{n,T}} G_T(\theta_T, \tilde{\mathbf{s}})\right\| - \left\|G_T(\theta_T, \mathbf{S}_T)\right\|\right\| &\leq \left\| \left[\inf_{\tilde{\mathbf{s}}\in\mathcal{S}_{n,T}} G_T(\theta_T, \tilde{\mathbf{s}})\right] - G_T(\theta_T, \mathbf{S}_T)\right\| \\
&\leq \sup_{\theta\in\Theta}\left\| \left[\inf_{\tilde{\mathbf{s}}\in\mathcal{S}_{n,T}} G_T(\theta, \tilde{\mathbf{s}})\right] - G_T(\theta, \mathbf{S}_T)\right\| \\
&= o_p(1),
\end{aligned} \tag{11}$$

where the last equation applies Equation (10).

Now define

$$\tilde{\theta} = \arg\min_{\theta\in\Theta} \|G_T(\theta, \mathbf{S}_T)\|$$

and

$$\hat{\theta} \in \arg\inf_{\theta\in\Theta,\, \tilde{\mathbf{s}}\in\mathcal{S}_{n,T}} G_T(\theta, \tilde{\mathbf{s}}).$$

Then

$$
\begin{aligned}
0 \leq & \|G_T(\hat{\theta}, \mathbf{S}_T)\| - \inf_{\theta \in \Theta} \|G_T(\theta, \mathbf{S}_T)\| \\
= & \|G_T(\hat{\theta}, \mathbf{S}_T)\| - \|G_T(\tilde{\theta}, \mathbf{S}_T)\| \\
= & \|G_T(\hat{\theta}, \mathbf{S}_T)\| - \left\| \inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} G_T(\hat{\theta}, \tilde{\mathbf{s}}) \right\| + \left\| \inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} G_T(\hat{\theta}, \tilde{\mathbf{s}}) \right\| - \|G_T(\tilde{\theta}, \mathbf{S}_T)\| \\
\overset{[1]}{=} & \left\| \inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} G_T(\hat{\theta}, \tilde{\mathbf{s}}) \right\| - \|G_T(\tilde{\theta}, \mathbf{S}_T)\| + o_p(1) \\
\overset{[2]}{\leq} & \left\| \inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} G_T(\tilde{\theta}, \tilde{\mathbf{s}}) \right\| - \|G_T(\tilde{\theta}, \mathbf{S}_T)\| + o_p(1) \\
\overset{[3]}{=} & o_p(1),
\end{aligned}
$$

where [1] and [3] apply Equation (11), and [2] uses the definition of $\hat{\theta}$. Combining, we thus have that (9) holds.

By the above, it thus suffices to show that (10) holds. Let Z be the $JT \times d_Z$ matrix of instruments. By the Cauchy-Schwarz inequality

$$
\begin{aligned}
\left\| \left[\inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} G_T(\theta, \tilde{\mathbf{s}}) \right] - G_T(\theta, \mathbf{S}_T) \right\|^2 &= \frac{1}{T^2} \left\| Z^\top \left(\left[\inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} \xi_T(\tilde{\mathbf{s}}; \theta) \right] - \xi_T(\mathbf{S}_T; \theta) \right) \right\|^2 \\
&\leq \frac{1}{T} \|Z^\top Z\| \times \frac{1}{T} \left\| \left[\inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} \xi_T(\tilde{\mathbf{s}}; \theta) \right] - \xi_T(\mathbf{S}_T; \theta) \right\|^2.
\end{aligned}
$$

Since $\frac{1}{T}\|Z^\top Z\| = O_p(1)$ by Assumption 7, it suffices to prove that the second term is $o_p(1)$. For this purpose, note further that

$$
\begin{aligned}
\sup_{\theta \in \Theta} \frac{1}{T} \left\| \left[\inf_{\tilde{\mathbf{s}} \in \mathcal{S}_{n,T}} \xi_T(\tilde{\mathbf{s}}; \theta) \right] - \xi_T(\mathbf{S}_T; \theta) \right\|^2 &= \sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \left\| \left[\inf_{\tilde{\mathbf{s}} \in \mathcal{S}^t_{n,T}} \xi_t(\tilde{\mathbf{s}}_t; \theta) \right] - \xi_T(\mathbf{S}_T; \theta) \right\|^2 \\
&\leq \sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \sup_{\tilde{\mathbf{s}} \in \mathcal{S}^t_{n,T}} \left\| \xi_t(\tilde{\mathbf{s}}_t; \theta) - \xi_T(\mathbf{S}_T; \theta) \right\|^2 \qquad (12) \\
&\leq \sup_{\theta \in \Theta} \max_{t \in \{1,\ldots,T\}} \sup_{\tilde{\mathbf{s}} \in \mathcal{S}^t_{n,T}} \left\| \xi_t(\tilde{\mathbf{s}}_t; \theta) - \xi_T(\mathbf{S}_T; \theta) \right\|^2,
\end{aligned}
$$

so that we may consider the final expression.

A.2.3 Consistency of Latent Demand Shocks

For $\alpha \in (0,1)$, let $\mathcal{H}_{n,T}(\alpha)$ denote the the event that all Hoeffding bounds on the sampling error hold in the sample – that is,

$$\mathcal{H}_{n,T}(\alpha) \equiv \left\{\pi(X_t, \xi_t; \theta_0) \in C^t_{n,T}(\alpha),\ \forall t = 1, \ldots, T\right\}.$$

Now notice that by the definition of $\pi(\cdot)$ and $\xi(\cdot)$, it holds for any $C(\epsilon) > 0$ that

$$\begin{aligned}
&\Pr\left(\sup_{\theta\in\Theta} \max_{t\in\{1,\ldots,T\}} \sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}} \left\|\pi_t(\xi_t(\tilde{\mathbf{s}}_t;\theta);\theta) - \pi_t(\xi_t(\mathbf{S}_t;\theta);\theta)\right\| \geq C(\epsilon)\right) \\
&\overset{[1]}{=} \Pr\left(\max_{t\in\{1,\ldots,T\}} \sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}} \left\|\tilde{\mathbf{s}}_t - \mathbf{S}_t\right\| \geq C(\epsilon)\right) \\
&\overset{[2]}{=} \Pr\left(\max_{t\in\{1,\ldots,T\}} \sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}} \left\|\tilde{\mathbf{s}}_t - \mathbf{S}_t\right\|^2 \geq C(\epsilon)\middle|\mathcal{H}_{n,T}(\alpha)\right) \Pr\left(\mathcal{H}_{n,T}(\alpha)\right) \\
&\qquad + \Pr\left(\max_{t\in\{1,\ldots,T\}} \sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}} \left\|\tilde{\mathbf{s}}_t - \mathbf{S}_t\right\|^2 \geq C(\epsilon)\middle|(\mathcal{H}_{n,T}(\alpha))^c\right) \Pr\left((\mathcal{H}_{n,T}(\alpha))^c\right) \qquad (13)\\
&\overset{[3]}{\leq} \Pr\left(\max_{t\in\{1,\ldots,T\}} \sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}} \left\|\tilde{\mathbf{s}}_t - \mathbf{S}_t\right\|^2 \geq C(\epsilon)\middle|\mathcal{H}_{n,T}(\alpha)\right) + \Pr\left((\mathcal{H}_{n,T}(\alpha))^c\right) \\
&\overset{[4]}{\leq} \Pr\left(\max_{t\in\{1,\ldots,T\}} \sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}} \left\|\tilde{\mathbf{s}}_t - \mathbf{S}_t\right\|^2 \geq C(\epsilon)\middle|\mathcal{H}_{n,T}(\alpha)\right) + \alpha \\
&\overset{[5]}{\leq} \Pr\left(2\delta_{n,T}(\alpha) \geq C(\epsilon)|\mathcal{H}_{n,T}(\alpha)\right) + \alpha \\
&\overset{[6]}{\leq} \mathbb{1}\{2\delta_{n,T}(\alpha) \geq C(\epsilon)\} + \alpha
\end{aligned}$$

where [1] uses that the bounds $\mathcal{S}^t_{n,T}$ based on Hoeffding's inequality do not depend on θ, [2] follows from the law of total probability, [3] uses that probabilities are bounded by one, [4] follows from Proposition 1 which implies $\Pr\left(\mathcal{H}_{n,T}\right) \geq 1-\alpha$, [5] follows from the definition of $\mathcal{S}^t_{n,T}$, and [6] follows since $\delta_{n,T}(\alpha)$ is non-random. Choosing $\alpha = \alpha_{n,T} = o_p(1)$, it then follows from the definition of $\delta_{n,T}(\alpha)$ that the term converges to zero whenever $\log(T) = o_p(n)$.

This implies that (12) is $o_p(1)$. To see this, suppose by way of contradiction that for some

$\delta > 0$

$$\sup_{\theta\in\Theta}\max_{t\in\{1,\ldots,T\}}\sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}}\left\|\xi_t(\tilde{\mathbf{s}}_t;\theta)-\xi_T(\mathbf{S}_T;\theta)\right\|^2 > \delta.$$

Then by Lemma 2, which we state below, there exists $C(\delta) > 0$ such that

$$\sup_{\theta\in\Theta}\max_{t\in\{1,\ldots,T\}}\sup_{\tilde{\mathbf{s}}\in\mathcal{S}^t_{n,T}}\left\|\pi_t(\xi_t(\tilde{\mathbf{s}}_t;\theta);\theta)-\pi_t(\xi_t(\mathbf{S}_t;\theta);\theta)\right\| > C(\delta),$$

which contradicts that (13) is $o_p(1)$ whenever $\log(T) = o_p(n)$. Hence, we have the desired result. □

A.2.4 A Useful Lemma

Lemma 2. *Under the assumptions of Theorem 1, it holds that* $\forall\delta > 0$, $\exists C(\delta) > 0$ *such that for all* $t = 1,\ldots,T$,

$$\Pr\left(\inf_{\theta\in\Theta}\inf_{\tilde{\xi}\notin\mathcal{N}_{\xi_t}(\delta)}\left\|\pi_t(\tilde{\xi};\theta)-\pi_t(\xi_t;\theta)\right\| > C(\delta)\right) = 1.$$

Remark 4. *The lemma follows from the proof of Theorem 1 in Freyberger (2015) with only minor adaptations. We include it here for completeness.*

Proof. Take $\tilde{\xi} : \|\tilde{\xi}-\xi_t\| \geq J\delta$. Without loss of generality, take $|\tilde{\xi}_1-\xi_{1t}| \geq |\tilde{\xi}_j-\xi_{jt}|, \forall j = 1,\ldots,J$. We further take $\tilde{\xi}_1-\xi_{1t} \geq \delta$, but the arguments for $\tilde{\xi}_1-\xi_{1t} \leq -\delta$ are analogous.

Let $\tilde{\xi}+\delta$ denote element-wise addition. For all $\delta > 0$, it holds that

$$\exp(\delta)\pi_{1t}(\xi_t;\theta) > \pi_{1t}(\xi_t+\delta;\theta) > \pi_{1t}(\xi_t;\theta), \quad \text{and} \quad \pi_{1t}(\tilde{\xi};\theta) \geq \pi_{1t}(\xi_t+\delta;\theta). \tag{14}$$

Further, by assumptions 1, 5, 6, and 4, there exists $\gamma > 0$ and a compact set $\mathcal{V} \subset \mathbb{R}^{dx}$ such that $\int_{\mathcal{V}} dF(\beta;\theta) > p > 0$ and with probability 1,

$$\gamma < \Pr(Y_i = j|p_t, X_t, \xi_t, \beta) < 1-\gamma, \tag{15}$$

for all $j = 0, 1, \ldots, J$, and $\beta \in \mathcal{V}$, where

$$\Pr(Y_i = j | p, X, \xi; \beta) = \frac{\exp\left(V(p_j, X_j; \beta) + \xi_j\right)}{1 + \sum_{l=1}^{J} \exp\left(V(p_l, X_l; \beta) + \xi_l\right)}.$$

For notational convenience, let $v_{jt}(\xi;\beta) = \Pr(Y_i = j | p_t, X_t, \xi; \beta)$

Next, define $\delta_0 = \min\{\delta, -\frac{1}{4}\log(1-\gamma)\}$. We have

$$\begin{aligned}
&\int v_{1t}(\tilde{\xi};\beta) dF(\beta;\theta) - \int v_{1t}(\xi_t;\beta) dF(\beta;\theta) \\
&\geq \int_{\mathcal{V}} v_{1t}(\tilde{\xi};\beta) dF(\beta;\theta) - \int_{\mathcal{V}} v_{1t}(\xi_t;\beta) dF(\beta;\theta) \\
&\geq \int_{\mathcal{V}} v_{1t}(\xi_t + \delta_0;\beta) dF(\beta;\theta) - \int_{\mathcal{V}} v_{1t}(\xi_t;\beta) dF(\beta;\theta).
\end{aligned}$$

By the mean value theorem, it then holds that for some $\tilde{\delta} \in (0, \delta_0)$

$$\begin{aligned}
&\int_{\mathcal{V}} v_{1t}(\xi_t + \delta_0;\beta) dF(\beta;\theta) - \int_{\mathcal{V}} v_{1t}(\xi_t;\beta) dF(\beta;\theta) \\
&= \left(\int_{\mathcal{V}} v_{1t}(\xi_t + \tilde{\delta};\beta) dF(\beta;\theta) - \int_{\mathcal{V}} v_{1t}(\xi_t + \tilde{\delta};\beta) \left(\sum_{k=1}^{J} v_{kt}(\xi_t + \tilde{\delta};\beta) \right) dF(\beta;\theta) \right) \delta_0, \\
&\overset{[1]}{\geq} \left(\int_{\mathcal{V}} v_{1t}(\xi_t;\beta) dF(\beta;\theta) - \exp(2\tilde{\delta}) \int_{\mathcal{V}} v_{1t}(\xi_t;\beta) \left(\sum_{k=1}^{J} v_{kt}(\xi_t;\beta) \right) dF(\beta;\theta) \right) \delta_0 \\
&\overset{[2]}{\geq} \int_{\mathcal{V}} v_{1t}(\xi_t;\beta) dF(\beta;\theta) \left(1 - \exp(2\tilde{\delta})(1-\gamma) \right) \delta_0 \\
&\overset{[3]}{\geq} \int_{\mathcal{V}} dF(\beta;\theta) \gamma \left(1 - \exp(2\tilde{\delta})(1-\gamma) \right) \delta_0 \\
&\overset{[4]}{\geq} p\gamma \left(1 - \exp(2\tilde{\delta})(1-\gamma) \right) \delta_0,
\end{aligned}$$

where [1] follows from (14), [2] and [3] follows from (15), and [3] uses that $\int_{\mathcal{V}} dF(\beta;\theta) > p > 0$. The last term is greater than zero and only depends on δ. Hence, we can take $C(\delta) = p\gamma \left(1 - \exp(2\tilde{\delta})(1-\gamma) \right) \delta_0$ which completes the proof. □

A.3 Proof of Theorem 2

Proof. In addition to the test statistic $\hat{G}_T(\alpha)$ defined in (6) using the sampling bounds used in estimation, we define two additional test statistics. The first is based on *multinomial* quantile bounds $\tilde{C}_{n,T}(X,\xi;\theta,\alpha)$ with the property that

$$\Pr\left(\tilde{\mathcal{H}}_{n,T}(\alpha)\right) = \alpha, \tag{16}$$

where $\tilde{\mathcal{H}}_{n,T}(\alpha)$ denotes the the event that all multinhatomial bounds on the sampling error hold in the sample – that is,

$$\tilde{\mathcal{H}}_{n,T}(\alpha) \equiv \left\{\hat{S}_t^{(n)} \in \tilde{C}_{n,T}(X,\xi_t;\theta_0,\alpha),\ \forall t=1,\ldots,T\right\}.$$

In particular, define

$$\begin{aligned} \tilde{G}_T(\alpha) \equiv \min_{\{\tilde{\xi}_t\}_{t=1}^T} \quad & T\left(\frac{1}{T}\sum_{t=1}^T Z_t^\top \tilde{\xi}_t\right)^\top \left(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t^\top Z_t Z_t^\top \hat{\xi}_t\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^T Z_t^\top \tilde{\xi}_t\right) \\ \text{s.t.} \quad & \hat{S}_t^{(n)} \in \tilde{C}_{n,T}(X,\tilde{\xi}_t;\theta_0,\alpha),\ \forall t\in\{1,\ldots,T\}. \end{aligned} \tag{17}$$

Note that by construction,

$$\tilde{C}_{n,T}(X,\xi_t;\theta_0,\alpha) \subset C_{n,T}(X,\xi_t;\theta_0,\alpha),\ \forall t\in\{1,\ldots,T\},$$

where the RHS are the bounds specified in Proposition 1 that we consider in estimation, so that

$$\Pr\left(\hat{G}_T(\alpha) \leq \tilde{G}_T(\alpha)\right) = 1. \tag{18}$$

Further, define the infeasible GMM test statistic

$$G_T \equiv T\left(\frac{1}{T}\sum_{t=1}^T Z_t^\top \xi_t\right)^\top \left(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t^\top Z_t Z_t^\top \hat{\xi}_t\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^T Z_t^\top \xi_t\right),$$

where ξ_t is the set of true latent demand shocks. Note then that under H_0,

$$\Pr\left(\tilde{G}_T(\alpha) \leq G_T(\alpha) \middle| \tilde{\mathcal{H}}_{n,T}(\alpha)\right) = 1. \tag{19}$$

Then, let $c_K^{1-\tau}$ denote the $1-\tau$th quantile of a χ^2 distribution with K degrees of freedom

$$\begin{aligned}
E\left[\mathbb{1}\{\hat{G}_T(\alpha) > c_K^{1-\tau}\}\right] &\overset{[1]}{\leq} E\left[\mathbb{1}\{\tilde{G}_T(\alpha) > c_K^{1-\tau}\}\right] \\
&= E\left[\mathbb{1}\{\tilde{G}_T(\alpha) > c_K^{1-\tau}\} \middle| \tilde{\mathcal{H}}_{n,T}(\alpha)\right] \Pr\left(\tilde{\mathcal{H}}_{n,T}(\alpha)\right) \\
&\quad + E\left[\mathbb{1}\{\tilde{G}_T(\alpha) > c_K^{1-\tau}\} \middle| (\tilde{\mathcal{H}}_{n,T}(\alpha))^c\right] \Pr\left((\tilde{\mathcal{H}}_{n,T}(\alpha))^c\right) \\
&\overset{[2]}{\leq} E\left[\mathbb{1}\{\tilde{G}_T(\alpha) > c_K^{1-\tau}\} \middle| \tilde{\mathcal{H}}_{n,T}(\alpha)\right] (1-\alpha) + \alpha \\
&\overset{[3]}{\leq} E\left[\mathbb{1}\{G_T > c_K^{1-\tau}\} \middle| \tilde{\mathcal{H}}_{n,T}(\alpha)\right] (1-\alpha) + \alpha \\
&\overset{[4]}{=} E\left[\mathbb{1}\{G_T > c_K^{1-\tau}\}\right] (1-\alpha) + \alpha,
\end{aligned} \tag{20}$$

where [1] follows from (18), [2] follows from (16), and [3] follows from (19). To show step [4], we can show that

$$\{(Z_t, \xi_t)\}_{t=1}^T \perp\!\!\!\perp \tilde{\mathcal{H}}_{n,T}(\alpha),$$

since G_T is a function of only $\{(Z_t, \xi_t)\}_{t=1}^T$. To do so, let A_T be any set on $\mathbb{R}^{T\times J\times(K+1)}$ and consider

$$\begin{aligned}
E\left[\mathbb{1}_{A_T}\left(\{(Z_t, \xi_t\}_{t=1}^T\right) \mathbb{1}\{\tilde{\mathcal{H}}_{n,T}(\alpha)\}\right] &= E\left[\mathbb{1}_{A_T}\left(\{(Z_t, \xi_t\}_{t=1}^T\right) E\left[\mathbb{1}\{\tilde{\mathcal{H}}_{n,T}(\alpha)\} \middle| \{(Z_t, \xi_t, S_t\}_{t=1}^T\right]\right] \\
&= E\left[\mathbb{1}_{A_T}\left(\{(Z_t, \xi_t\}_{t=1}^T\right) E\left[\mathbb{1}\{\tilde{\mathcal{H}}_{n,T}(\alpha)\} \middle| \{S_t\}_{t=1}^T\right]\right] \\
&= E\left[\mathbb{1}_{A_T}\left(\{(Z_t, \xi_t\}_{t=1}^T\right) (1-\alpha)\right] \\
&= E\left[\mathbb{1}_{A_T}\left(\{(Z_t, \xi_t\}_{t=1}^T\right)\right] E\left[\mathbb{1}\{\tilde{\mathcal{H}}_{n,T}(\alpha)\}\right].
\end{aligned}$$

Freyberger (2015) shows that under assumptions 1, 5, and 6, the latent demand shocks ξ_t are bounded. Then, by assumptions 2, 4, and 7, it holds by the central limit theorem that $\sqrt{T}\frac{1}{T}\sum_{t=1}^T Z_t^\top \xi_t \overset{d}{\to} N(0, \Sigma)$, where ξ_t are the true latent demand shocks. Hence by

Theorem 1 and the continuous mapping theorem

$$G_T \equiv T\left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right)^\top \left(\frac{1}{T}\sum_{t=1}^{T} \hat{\xi}_t^\top Z_t Z_t^\top \hat{\xi}_t\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} Z_t^\top \xi_t\right) \xrightarrow{d} \chi^2(K), \tag{21}$$

where $\chi^2(K)$ is a χ^2 distribution with K degrees of freedom. Combining (20) and (21) and using that $\alpha_{n,T} = o_p(1)$, we have under H_0,

$$\limsup_{t\to\infty} E\left[\mathbb{1}\{\hat{G}_T(\alpha) > c_K^{1-\tau}\}\right] \leq \limsup_{T\to\infty} E\left[\mathbb{1}\{G_T > c_K^{1-\tau}\}\right](1-\alpha_{n,T}) + \alpha_{n,T} = \tau.$$

□

A.4 Nonparametric Demand Estimation

While the zero-share problem is particularly salient for the random coefficients logit model, the underlying problem of errors in estimated choice probabilities exists in many frameworks for demand estimation. To illustrate this, we extend the nonparametric demand estimation framework proposed by Tebaldi et al. (2019) to a finite number of consumers. For computational tractability, this model requires linearity of any additional constraints. We therefore focus on using Hoeffding's inequality.

In the simpler framework with exogenous prices, the linear program is[9]

$$\max_{\phi} / \min_{\phi} c'\phi \qquad \begin{cases} -\phi_k & \leq 0, \\ \sum_k \phi_k & = 1, \\ \sum_{k \in S(j,m)} \phi_k & = \hat{s}_{j,m} \quad \forall j, m. \end{cases} \tag{22}$$

where $S(j,m)$ is the set of indices of elements of the Minimum Relevant Partition whose union is the market share for product j in market m. $S(j,m)$ is a deterministic function of prices and known to the researcher. The only random variable here is $\hat{s}_{j,m}$, the estimated market share. We can apply Hoeffding's inequality to obtain simultaneous confidence intervals for all products in all markets. This leads to

$$\max_{\phi} / \min_{\phi} c'\phi \qquad \begin{cases} -\phi_k & \leq 0, \\ \sum_k \phi_k & = 1, \\ \left| \sum_{k \in S(j,m)} \phi_k - \hat{s}_{j,m} \right| & \leq B_{(n_t)_t,\alpha} \quad \forall j, m. \end{cases} \tag{23}$$

Note that the absolute value in the concentration constraints can be written as linear constraints. Hence our modification (23) of (22) maintains computational tractability since efficient solvers for linear programs are available.

One interesting feature of (23) and its generalization to endogenous prices proposed by Tebaldi et al. (2019) is that there often does not exist a feasible solution in applications.

[9] The goal is to bound the counterfactual from both sides, hence the max min notation.

This can be interpreted as a rejection of the econometric model by (22), notably the quasi-linearity of prices or the time-market-homogeneity of latent utility draws. However, the reason why the model appears rejected could be the finiteness of consumers. While we do not have access to the data used in Tebaldi et al. (2019), we can replicate this phenomenon in simulations.

A.5 Implementation details

This section outlines implementation details of the EZ-MPEC estimator using Binomial quantiles and supplies expressions for the Jacobian and Hessians used in estimation. All programs are implemented in Matlab using Knitro, with code readily available upon request.

A.5.1 Setup

Note that using the bounds derived in Proposition 1, we can write the corresponding EZ-MPEC estimator as

$$\begin{aligned}
\min_{(\theta,\xi,\eta)} \quad & \eta^\top W \eta \\
\text{s.t.} \quad & \frac{1}{n} F_{\text{Bin}}^{-1}\left(\frac{1-\sqrt[T]{1-\alpha}}{J}, P_{jt}(\theta), n \right) - \hat{S}_{j,t} \leq 0, \ \forall j,t, \\
& \hat{S}_{j,t} - \frac{1}{n} F_{\text{Bin}}^{-1}\left(1 - \frac{1-\sqrt[T]{1-\alpha}}{J}, P_{jt}(\theta), n \right) \leq 0, \ \forall j,t, \\
& \eta = \frac{1}{T} \sum_{t=1}^{T} Z_t^\top \xi_t
\end{aligned}$$

where $P_{jt}(\theta) \equiv \Pr(Y_i = j | p_t, X_t, \xi_t, \theta)$ as given in equation (2). The task is now to obtain the Jacobian of the objective function and the constraints, as well as the Hessian of the corresponding Lagrangian. Fortunately, the program is similar to Dubé et al. (2012), which derive first and second order derivatives of the objective, the moment equality, as well as the CCPs $P_{jt}(\theta)$. We thus focus on the derivatives of the Binomial quantile function with respect to the CCPs $P_{jt}(\theta)$.

Since the Binomial quantile function is not continuously differentiable, we approximate the Binomial distribution function with a function that has a continuously differentiable inverse. We consider

$$\hat{F}(r, P_{jt}(\theta), n) \equiv \sum_{i=0}^{n} \frac{1}{1+\exp(-2m(r-i))} \binom{n}{i} P_{jt}(\theta)^i \left(1 - P_{jt}(\theta)\right)^{n-i}, \tag{24}$$

where $m \geq 0$ is a hyperparameter.[10]

[10] To account for the approximation, we can further shift the bounds by a constant. This would not affect the derivatives.

A.5.2 Inverse function rules

Although the approximation (24) has an inverse in its first argument that is continuously differentiable in its second argument, it does generally have a closed form expression. We thus leverage simple results akin to the inverse function theorem.

In particular, we have

$$\begin{aligned} & \hat{F}\left(\hat{F}^{-1}(\alpha,p,n),p,n\right) = \alpha \\ \Rightarrow \qquad & \frac{\partial}{\partial}\hat{F}\left(\hat{F}^{-1}(\alpha,p,n),p,n\right) = 0, \end{aligned}$$

and hence by application of the chain rule and the inverse function theorem

$$\begin{aligned} 0 &= \frac{\partial}{\partial}\hat{F}\left(\hat{F}^{-1}(\alpha,p,n),p,n\right) \\ &= \frac{\partial \hat{F}(r,p,n)}{\partial r}\bigg|_{r=\hat{F}^{-1}(\alpha,p,n)} \frac{\partial \hat{F}^{-1}(\alpha,p,n)}{\partial p} + \frac{\partial \hat{F}(\hat{F}^{-1}(\alpha,p,n),p,n)}{\partial p} \\ \Rightarrow \qquad \frac{\partial \hat{F}^{-1}(\alpha,p,n)}{\partial p} &= \frac{-\frac{\partial \hat{F}(\hat{F}^{-1}(\alpha,p,n),p,n)}{\partial p}}{\frac{\partial \hat{F}(r,p,n)}{\partial r}\Big|_{r=\hat{F}^{-1}(\alpha,p,n)}}. \end{aligned} \tag{25}$$

For the second order derivative, we simplify notation and let subscripts denote partial derivatives with respect to the indexed argument.

Then

$$
\begin{aligned}
\frac{\partial^2 \hat{F}^{-1}(\alpha,p,n)}{\partial^2 p} &= \hat{F}_{22}^{-1}(\alpha,p,n) = \frac{\partial}{\partial p}\left(\hat{F}_2^{-1}(\alpha,p,n)\right) = \frac{\partial}{\partial p}\left(\frac{-\hat{F}_2(\hat{F}^{-1}(\alpha,p,n),p,n)}{\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)}\right) \\
&= \frac{\hat{F}_2(\hat{F}^{-1}(\alpha,p,n),p,n)}{\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)^2}\frac{\partial}{\partial p}\left(\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)\right) \\
&\quad - \frac{1}{\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)}\frac{\partial}{\partial p}\left(\hat{F}_2(\hat{F}^{-1}(\alpha,p,n),p,n)\right) \\
&= \frac{1}{\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)}\Bigg(\hat{F}_2^{-1}(\alpha,p,n)\frac{\partial}{\partial p}\Big(\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)\Big) \\
&\quad - \frac{\partial}{\partial p}\Big(\hat{F}_2(\hat{F}^{-1}(\alpha,p,n),p,n)\Big)\Bigg),
\end{aligned}
\tag{26}
$$

where

$$
\begin{aligned}
\frac{\partial}{\partial p}\Big(\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)\Big) &= \hat{F}_{11}(\hat{F}^{-1}(\alpha,p,n),p,n)\hat{F}_2^{-1}(\alpha,p,n) + \hat{F}_{12}(\hat{F}^{-1}(\alpha,p,n),p,n), \\
\frac{\partial}{\partial p}\Big(\hat{F}_2(\hat{F}^{-1}(\alpha,p,n),p,n)\Big) &= \hat{F}_{21}(\hat{F}^{-1}(\alpha,p,n),p,n)\hat{F}_2^{-1}(\alpha,p,n) + \hat{F}_{22}(\hat{F}^{-1}(\alpha,p,n),p,n).
\end{aligned}
$$

Note that by Schwarz's theorem, $\hat{F}_{12}(x,p,n) = \hat{F}_{21}(x,p,n)$. Hence, Equation (26) simplifies to

$$
\begin{aligned}
\frac{\partial^2 \hat{F}^{-1}(\alpha,p,n)}{\partial^2 p} = \frac{1}{\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)}\Bigg(&\hat{F}_{11}(\hat{F}^{-1}(\alpha,p,n),p,n)\left(\hat{F}_2^{-1}(\alpha,p,n)\right)^2 \\
&- \hat{F}_{22}(\hat{F}^{-1}(\alpha,p,n),p,n)\Bigg).
\end{aligned}
\tag{27}
$$

A.5.3 Jacobian

Using Equation (25) and the chain rule, we have

$$
\frac{\partial \hat{F}^{-1}(\alpha,P_{j,t}(\theta),n)}{\partial \theta} = \left(\frac{-\left.\frac{\partial \hat{F}(\hat{F}^{-1}(\alpha,P_{j,t}(\theta),n),p,n)}{\partial p}\right|_{p=P_{j,t}(\theta)}}{\left.\frac{\partial \hat{F}(r,p,n)}{\partial r}\right|_{r=\hat{F}^{-1}(\alpha,P_{j,t}(\theta),n)}}\right)\frac{\partial P_{j,t}(\theta)}{\partial \theta},
\tag{28}
$$

where $\frac{\partial P_{j,t}(\theta)}{\partial \theta}$ is the same as in Dubé et al. (2012), and

$$\frac{\partial \hat{F}(r,p,n)}{\partial p} = \sum_{i=0}^{n} \frac{\frac{i}{p} - \frac{n-i}{1-p}}{1+\exp(-2m(r-i))} \binom{n}{i} p^i (1-p)^{n-i},$$

and

$$\frac{\partial \hat{F}(r,p,n)}{\partial r} = 2m \sum_{i=0}^{n} \frac{\exp(-2m(r-i))}{(1+\exp(-2m(r-i)))^2} \binom{n}{i} p^i (1-p)^{n-i}.$$

A.5.4 Hessian

Using the chain rule, we have

$$\begin{aligned}\frac{\partial^2 \hat{F}^{-1}(\alpha, P_{j,t}(\theta), n)}{\partial \theta \partial \theta^\top} &= \frac{\partial}{\partial \theta^\top}\left(\frac{\partial \hat{F}^{-1}(\alpha, P_{j,t}(\theta), n)}{\partial \theta}\right) \\ &= \left.\frac{\partial^2 \hat{F}^{-1}(\alpha,p,n)}{\partial p^2}\right|_{p=P_{j,t}(\theta)} \frac{\partial P_{j,t}(\theta)}{\partial \theta}\frac{\partial P_{j,t}(\theta)}{\partial \theta^\top} + \left.\frac{\partial \hat{F}^{-1}(\alpha,p,n)}{\partial p}\right|_{p=P_{j,t}(\theta)} \frac{\partial P_{j,t}(\theta)}{\partial \theta \partial \theta^\top},\end{aligned}$$

where $\frac{\partial P_{j,t}(\theta)}{\partial \theta}$ and $\frac{\partial P_{j,t}(\theta)}{\partial \theta \partial \theta^\top}$ are the same as in Dubé et al. (2012), and $\left.\frac{\partial \hat{F}^{-1}(\alpha,p,n)}{\partial p}\right|_{p=P_{j,t}(\theta)}$ is given in equation (28). Finally, using Equation (27), we have

$$\begin{aligned}\frac{\partial^2 \hat{F}^{-1}(\alpha,p,n)}{\partial p^2} = \frac{1}{\hat{F}_1(\hat{F}^{-1}(\alpha,p,n),p,n)}\Big(&\hat{F}_{11}(\hat{F}^{-1}(\alpha,p,n),p,n)\left(\hat{F}_2^{-1}(\alpha,p,n)\right)^2 \\ &- \hat{F}_{22}(\hat{F}^{-1}(\alpha,p,n),p,n)\Big),\end{aligned}$$

where

$$\hat{F}_{22}(r,p,n) = \sum_{i=0}^{n} \frac{\frac{i(i-1)}{p^2} - 2\frac{i(n-i)}{\Pr(1-p)} + \frac{(n-i)(n-i-1)}{(1-p)^2}}{1+\exp(-2m(r-i))} \binom{n}{i} p^i (1-p)^{n-i},$$

and

$$\hat{F}_{11}(r,p,n) = -2m \sum_{i=0}^{n} \frac{\exp(-2m(r-i))}{(1+\exp(-2m(r-i)))^2}\left(1 - 2\frac{\exp(-2m(r-i))}{1+\exp(-2m(r-i))}\right)\binom{n}{i} p^i (1-p)^{n-i},$$

and $\hat{F}_1(r,p,n)$ and $\hat{F}_2^{-1}(\alpha,p,n)$ as before.