Appendix: Beyond the endogeneity bias: the effect of unmeasured brand characteristics on household-level brand choice models

Pradeep Chintagunta* Jean-Pierre Dubé† Khim Yong Goh‡

February 2004

Abstract

In this appendix, we design a Monte Carlo simulation to explore the properties of the estimator discussed in the paper. First, we compare the theoretical assumptions of our approach with other popular likelihood-based approaches for household data. We then compare the empirical properties of these approaches in the context of a Monte Carlo simulation. We find that our approach provides robust results across numerous potential price-generation mechanisms. This is an important finding since alternative approaches compared require very specific assumptions about the price-generation mechanism. As a result, these alternative approaches are more vulnerable to specification biases.

We then focus on the proposed approach. We explore the importance of controlling for both endogeneity and heterogeneity in scanner data models. Using Monte Carlo simulations, we provide some evidence of biases that might arise if both sources of variation are not correctly accounted for. In particular, we find that ignoring the endogeneity bias not only biases mean effects, it also has an impact on the variance in heterogeneity. In general, we find that the degree of heterogeneity in price sensitivity is over-stated. This upward bias is amplified if price-insensitive consumers also systematically do not shop when UBCs are high.

1 Comparison of Approaches

1.1 Derivation of approaches

We now discuss full information versus limited information approaches to dealing with price endogeneity in the estimation of logit demand systems. We then contrast these approaches with our proposed instrumental variables procedure. In the end, we aim to show that the trade-off between full information versus alternative approaches is a matter of choosing between a more precise estimate versus a consistent estimate (e.g.

*Graduate School of Business, University of Chicago, pradeep.chintagunta@gsb.uchicago.edu.
†Graduate School of Business, University of Chicago, jdube@gsb.uchicago.edu.
‡Graduate School of Business, University of Chicago, kgoh@gsb.uchicago.edu.

asymptotically unbiased). Adding more structure helps reduce the variance in parameter estimates, but as we show below, it also increases the risk of specification error.

On the demand-side, consumer h derives the following conditional indirect utility from brand j at time t:

$$u_{hjt} = \alpha_j + \beta p_{jt} + \gamma F_{jt} + \lambda_{jt} + \varepsilon_{hjt},\ \varepsilon_{hjt}\tilde{}EV1$$

where p_{jt} is the price of brand j, F_{jt} is an indicator for a promotion for brand j , λ_{jt} denotes the impact of potentially unmeasured (to the researcher) product attributes and ε_{hjt} is an idiosyncratic taste shock. The coefficients $\Theta = (\alpha_1, ..., \alpha_J, \beta, \gamma)'$ are taste parameters to be estimated. An additional $J+1$ brand with $u_{h0t} = \varepsilon_{h0t}$ allows for an outside good (no-purchase option). The corresponding probability that household h chooses brand j at time t, P_{hjt}, is:

$$P_{hjt} = S_{jt}(\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta) \quad (1)$$

$$= \frac{\exp(\alpha_j + \beta p_{jt} + \gamma F_{jt} + \lambda_{jt})}{1 + \sum_{k=1}^{J} \exp(\alpha_k + \beta p_{kt} + \gamma F_{kt} + \lambda_{kt})} \quad (2)$$

which is identical to the share of brand j at time t in this case. Typical household choice data consists of a vector $Y_{ht} = (y_{h1t}, ..., y_{hJt})'$ denoting the brand chosen by household h at time t, where $y_{hjt} = \begin{cases} 1, & \text{if brand } j \text{ was chosen at time } t \\ 0, & \text{else} \end{cases}$. From the model, we derive the density of the vector Y_{ht} conditional on the marketing variables, p and F, and the unobserved attributes, λ:

$$f_y(Y_{ht}|\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta) = \prod_{j=1}^{J} P_{hjt}^{y_{hjt}}. \quad (3)$$

Typically, estimation is carried out by maxizing the corresponding likelihood function:

$$L(\Theta) = \prod_{t=1}^{T_h} \prod_{h=1}^{H} f_y(Y_{ht}|\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta) \quad (4)$$

and the terms λ_{jt} are assumed to be zero (i.e. are ignored). Alternatively, if one knew the density of λ, $f_\lambda(\lambda)$, one could integrate these terms out of the likelihood.

A concern arises if the terms λ, although unobserved to the researcher, do in fact influence choices and, moreover, if retail prices are set to some extent based on these demand-shifting terms (i.e. λ is observed by consumers and retailers). Essentially, this case is a data limitation for the researcher that results in $cov(\mathbf{p}_t,\lambda_t) \neq 0$. Past research has motivated such data limitations as unmeasured promotions, advertising, shelf space or stock-piling by consumers. Suppose, for instance, that shelf prices are generated according to the following rule:

$$p_{jt} = w_{jt} + MU(\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta) + \omega_{jt} \quad (5)$$

where w_{jt} is the wholesale price, $MU(\cdot)$ is a function denoting the mark-up and ω_{jt} captures additional marginal costs to the retailer that are unobserved by the researcher. Clearly one cannot naively integrate λ out of the likelihood in equation (4) as prices, by construction, vary with the level of λ. Many popular static analytical pricing models give rise to this type of pricing rule in equilibrium. For instance, the category profit-maximizing pricing problem gives rise to:

$$p_{jt} = w_{jt} - \frac{1}{\beta S_{0t}} + \omega_{jt} \tag{6}$$

and the individual brand profit-maxiziing pricing problem gives rise to:

$$p_{jt} = w_{jt} - \frac{1}{\beta \left(1 - S_{jt}\right)} + \omega_{jt}. \tag{7}$$

In both cases, changes in λ will shift demand, S_{jt}, and hence will shift prices. Thus, p is correlated with λ. Since prices depend on predicted demand, the density of prices will contain information about the demand parameters. Hence, the likelihood (4) must be modified to include this information:

$$L\left(\Theta\right) = \prod_{t=1}^{T_h} \int \prod_{h=1}^{H} f_y\left(Y_{ht} | \boldsymbol{\lambda}_t, \mathbf{p}_t; \Theta\right) f_p\left(\mathbf{p}_t | \boldsymbol{\lambda}_t; \Theta\right) f_\lambda\left(\lambda\right) d\lambda. \tag{8}$$

Evaluating the density of prices is tricky because (5) is an implicit function of prices. To derive the density of prices induced by the randomness in ω, we use the transformation of variables $r_{jt} = p_{jt} - w_{jt} - MU\left(\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta\right) - \omega_{jt}$, here $\boldsymbol{\omega}_t \sim N(0, \Omega_\omega)$ and hence $\mathbf{r}_t \sim N(p_t - w_t + MU\left(\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta\right), \Omega_\omega)$. Using the transformation-of-variabes theorem, we can derive the density of prices:

$$f_p\left(\mathbf{p}_t | \boldsymbol{\lambda}_t; \Theta, \Omega_\omega\right) = f_r\left(\mathbf{r}_t | \boldsymbol{\lambda}_t; \Theta, \Omega_\omega\right) |J_t\left(\Theta\right)|$$

where $J_t = \frac{dr}{dp}$ is a Jacobian. Estimation requires assuming a specific form of pricing conduct in order to derive the analytical forms for J_t and $MU\left(\cdot\right)$. We provide the example of the brand profit-maximizing model in the "examples" section below. Since this approach derives the price-generation mechanism "structurally" (i.e. from a specific economic model), we term this approach "Full Information Maximum Likelihood" (herertoafter FIML).

Similar structural approaches have been used by Villas-Boas and Zhao (2003) and, more recently, by Yang et al. (2003). The latter also account for consumer heterogeneity in demand. However, the inclusion of heterogeneity substantially complicates the evaluation of the Jacobian term, which they calculate numerically rather than analytically. Our experience (in the context of the simulations below) is that the use of numerical Jacobians creates a number of practical problems with the maximization of the likelihood. In general, if one is confident the pricing model is correct, then this is the most efficient method of estimating the demand parameters. However, in practice, this type of model may not provide a good representation of retail price variation over time. For instance, retail prices may often exhibit large temporary price cuts that are not reflected in the wholesale prices. These price cuts are essential for estimation as they provide the main source of price variation and, hence, variation in consumer choices that help identify the parameters. Very recent research has proposed dynamic models with consumer stock-piling to capture this type of sales timing (Erdem et al 2003). However, these dynamic models are computationally well outside the range of feasible structural models for the methods described above. A more severe problem with these full-information approaches is that, in the case of multi-product pricing, it is not possible to prove the uniqueness of the price equilibrium. In other words, these types of models typically suffer from multiple equilibria problems. Hence, the transformation-of-variables can not be performed and the Jacobian and the likelihood function itself are not well-defined. In such cases, the structural approach is simply infeasible[1].

An alternative approach is to work with a more agnostic "limited-information maximum likelihood" approach (heretoafter LIML), as in Villas-Boas and Winer (1999). While this method is less efficient than the full-information approach described above, it is far less susceptible to specification error. From (5) , we assume:

$$p_{jt} = \mu_j + \phi_j w_{jt} + \omega_{jt}$$

where μ_j and ϕ_j are parameters. One could interpret ω as $\omega_{jt} = MU(\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta) + \omega_{jt}$. Rather than take a stand on the precise analytical form of the mark-up term (and hence underlying economic behavior) $MU(\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Theta)$, we assume $\omega_t \sim f_\omega(\omega)$ and $cov(\boldsymbol{\omega}_t, \boldsymbol{\lambda}_t) \neq 0$. Note that this approach is entirely consistent with the structural approaches described above, but it makes no specific assumptions about pricing conduct. Now, the likelihood function is much simpler to construct as it does not require the evaluation of a Jacobian[2]:

$$L(\Theta) = \prod_{t=1}^{T_h} \int \prod_{h=1}^{H} f_y(Y_{ht}|\boldsymbol{\lambda}_t, \mathbf{p}_t; \Theta) f_p(\mathbf{p}_t|\boldsymbol{\lambda}_t; \Theta) f_\lambda(\lambda) d\lambda$$

where $f_p(\mathbf{p}_t|\boldsymbol{\lambda}_t; \Theta) = f_\omega(\boldsymbol{\omega}_t|\boldsymbol{\lambda}_t, \Omega_\omega)$ .This approach is termed limited information because the density of prices helps handle the endogeneity induced by λ, but it does not provide any additional information about the demand parameters, Θ. More specifically, the inclusion of the density of prices helps us integrate out λ correctly as prices vary with the level of λ. By relaxing the additional structure derived from a pricing conduct assumption, this approach also avoids the multiple equilibria problem.

The limited information approach clearly has the advantage over the full-information approach in that it requires far fewer assumptions about the economics generating prices. However, one must still make an assumption about $f_\omega(\omega|\boldsymbol{\lambda}_t, \Omega_\omega)$. For instance, assuming ω_t is distributed i.i.d. normal is still quite restrictive with respect to the implicit underlying pricing conduct. For instance, neither the category or brand profit models above would give rise to a normally-distributed error term of this sort. This approach also rules out any autocorrelation in λ. This would be a concerning assumption if λ reflected promotions, consumer inventories, advertising or other usual explanations used to motivate its inclusion, all of which would likely exhibit substantial time-dependence. The approach we propose below is still consistent with the models above, but it adds yet another level of flexibility in that it remains agnostic about the joint distribution $f_{\omega,\lambda}(\omega, \lambda)$, contemporaneously and across time, as well as the marginal distributions of ω and λ.

We now summarize our approach, which we refer to as the "instrumental variables" approach (heretoafter

[1] The multiple equilibria problem does not impede method-of-moments based estimators that have frequently been applied to aggregate data (e.g. Besanko et al. 1998 and 2003).

[2] The pricing equation is no longer an implicit function, so we can derive the density of prices analytically.

IV). At the likelihood stage, we estimate the parameters of the choice model:

$$P_{hjt} = S_{jt}\left(\boldsymbol{\lambda}_t, \mathbf{p}_t, \mathbf{F}_t; \Delta\right) = \frac{\exp\left(\delta_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt}\right)}$$

where $\Delta = (\delta_{11}, ..., \delta_{1T}, ..., \delta_{JT})$. Implicitly, $\delta_{jt} = \alpha_j + \beta p_{jt} + \gamma F_{jt} + \lambda_{jt}$ so these parameters absorb the impact of prices and unobserved attributes. Hence, we do not need to model the density of prices and we do not need to integrate λ out of the likelihood. Instead, we estimate Δ using:

$$L\left(\Delta\right) = \prod_{t=1}^{T_h} \prod_{h=1}^{H} \prod_{j=1}^{J} S_{jt}^{y_{hjt}}. \tag{9}$$

As a result, we do in fact propose a proper maximum likelihood estimator, in contrast with Goolsbee and Petrin (2002). We then obtain the parameters in Θ by running a second-stage regression:

$$\widehat{\delta}_{jt} = \alpha_j + \beta p_{jt} + \gamma F_{jt} + \lambda_{jt}$$

using the estimated vector, $\widehat{\delta}_{jt}$, as the dependent variable. To take into account the uncertainty around these estimates, we use a GLS procedure where the weights are simply the estimated covariance matrix $cov(\widehat{\Delta})$. At this stage, we need to be careful about the endogeneity of prices and their potential correlation with the error term, λ_{jt}. The econometric concern is a standard linear regression problem with correlation between the error term and a covariate. We resolve this problem by using an standard instrumental variables procedure where prices are projected onto instruments $Z_t = (F_{jt}, w_{jt})$ to partial out the component of prices that is uncorrelated with λ_{jt}. This approach has a number of advantages. First, we only assume λ is mean zero. The distribution of λ is estimated non-parametrically (implicity in the estimation of Δ) and, hence, the approach allows for an aribitrary form of dependence in λ. These estimates will be less efficient than a full information approach, but they will be consistent even in cases when the structural pricing models above are incorrect. The approach also does not require integrating out λ nor the evaluation of a Jacobian, vastly reducing the computational burden of the evaluation of the likelihood function.

1.2 Example

We now derive the likelihood for the case of the brand profit-maximization model, (7), where we will assume there are only 2 brands. First we derive the Jacobian. The j, k^{th} element of J_t is $\frac{-\beta\left(-\frac{dS_{jt}}{dp_{kt}}\right)}{[\beta(1-S_{jt})]^2} = -\frac{S_{jt}S_{kt}}{(1-S_{jt})^2}$ and the j, j^{th} element is $1 + \frac{S_{jt}(1-S_{jt})}{(1-S_{jt})^2}$, hence:$J_t = \begin{bmatrix} 1 + \frac{S_{1t}(1-S_{1t})}{(1-S_{1t})^2} & -\frac{S_{1t}S_{2t}}{(1-S_{1t})^2} \\ -\frac{S_{1t}S_{2t}}{(1-S_{2t})^2} & 1 + \frac{S_{2t}(1-S_{2t})}{(1-S_{2t})^2} \end{bmatrix}$. The corresponding likelihood function will have the form:

$$\begin{aligned} L\left(\Theta\right) &= \prod_{t=1}^{T_h} \int \prod_{h=1}^{H} \prod_{j=1}^{J} \left(\frac{\exp\left(\alpha_j + \beta p_{jt} + \gamma F_{jt} + \lambda_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\alpha_k + \beta p_{kt} + \gamma F_{kt} + \lambda_{kt}\right)} \right)^{y_{hjt}} \left(\frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left(-\frac{1}{2}\right) \mathbf{r}_t' \Sigma^{-1} \mathbf{r}_t \right) & (10) \\ & ... \left| \begin{bmatrix} 1 + \frac{S_{1t}(1-S_{1t})}{(1-S_{1t})^2} & -\frac{S_{1t}S_{2t}}{(1-S_{1t})^2} \\ -\frac{S_{1t}S_{2t}}{(1-S_{2t})^2} & 1 + \frac{S_{2t}(1-S_{2t})}{(1-S_{2t})^2} \end{bmatrix} \right| f_\lambda\left(\lambda\right) \partial\lambda. & (11) \end{aligned}$$

For estimation, the likelihood would need to be integrated over the distribution of λ. If the dimension is not too large, this task could be done numerically, via quadrature. Alternatively, this could be accomplished via Monte Carlo simulation.

1.3 Simulation Design

We now discuss the details for the creation of the simulated data sets. A total of 7 scenarios were analyzed to compare the empirical properties of the LIML approach, the FIML approach and the IV approaches. As a benchmark, we also estimate the standard conditional logit that ignores the error component λ entirely.

The 7 scenarios are as follows. The consumer demand is identical across all seven, only the price-generation mechanism and the assumptions about λ differ. The first 5 are generated from variations of the LIML model. The last 2 are generated from variations of the FIML model. In each of these seven cases, we replicated the data generation 30 times and, correspondingly, we estimate each of the models 30 times. In the results section below, we report the mean parameter estimates for each model and the mean absolute deviation from the true parameter values used to generate the data.

Notation

- brands $j = 0, 1, 2$ (0=no purchase)
- time $t = 1, ...T$, $T = 40$
- households $h = 1, ..., H$, $H = 150$

1.3.1 Demand

Assumptions:

- $u_{h1t} = 2 - 2p_{1t} + 0.3promotion_{1t} + \lambda_{1t} + \varepsilon_{h1t}$
- $u_{h2t} = 3 - 2p_{2t} + 0.3promotion_{2t} + \lambda_{2t} + \varepsilon_{h2t}$
- $u_{h0t} = \varepsilon_{h0t}$
- $\varepsilon_{htj} \sim Type\ I\ Extreme\ Value$
- $promotion_{jt} = I(\zeta_{jt} > .8), \zeta_{jt} \sim U(0, 1)$
- λ defined below

1.3.2 Scenario 1: LIML with correlation between λ and ω, and autocorrelation in ω

Data are generated from the LIML model, above, except that ω is drawn from a time-dependent distribution (auto-correlation). This specification is chosen to illustrate the performance of the LIML model when the conditional density is correctly-specified, but the joint-distribution is not (in this case the LIML approach will ignore the time dependence in the data).

Assumptions:

- $F_{\lambda,\omega}(\lambda_t, \omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0.2 & 0.08 \\ 0.01 & 0.05 & 0.08 & 0.2 \\ 0.2 & 0.08 & 1.5 & 0.5 \\ 0.08 & 0.2 & 0.5 & 1.5 \end{bmatrix}\right)$
- $corr(\omega_t, \omega_{t-1}) = 0.5$

1.3.3 Scenario 2: LIML with no correlation between λ and ω, and autocorrelation in ω

This is the same as scenario 1, except that it does not have any correlation between λ and ω. In principle, there is no endogeneity bias; however, there are still the additional demand-shifting error components λ_{jt}.

Assumptions:

- $F_{\lambda,\omega}(\lambda_t, \omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0 & 0 \\ 0.01 & 0.05 & 0 & 0 \\ 0 & 0 & 1.5 & 0.5 \\ 0 & 0 & 0.5 & 1.5 \end{bmatrix}\right)$
- $corr(\omega_t, \omega_{t-1}) = 0.5$

1.3.4 Scenario 3: LIML with correlation between λ and ω

This is the same as scenario 1, except that there is no auto-correlation in ω.

Assumptions:

- $F_{\lambda,\omega}(\lambda_t, \omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0.2 & 0.08 \\ 0.01 & 0.05 & 0.08 & 0.2 \\ 0.2 & 0.08 & 1.5 & 0.5 \\ 0.08 & 0.2 & 0.5 & 1.5 \end{bmatrix}\right)$
- $corr(\omega_t, \omega_{t-1}) = 0$

1.3.5 Scenario 4: LIML with no correlation between λ and ω

This is the same as scenario 3, except that there is no correlation between λ and ω, nor any auto-correlation in ω.

Assumptions:

- $F_{\lambda,\omega}(\lambda_t,\omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0 & 0 \\ 0.01 & 0.05 & 0 & 0 \\ 0 & 0 & 1.5 & 0.5 \\ 0 & 0 & 0.5 & 1.5 \end{bmatrix}\right)$
- $corr(\omega_t, \omega_{t-1}) = 0$

1.3.6 Scenario 5: LIML with correlation between λ and ω, prices follow a Markov Process

This is the same as scenario 3, except that there is no auto-correlation in ω and prices follow a Markov Process. The point of this scenario is to attempt to create a pricing policy that looks “realistic". Observed prices appear to follow a Markov-switching process between temporary sales and price cuts. Erdem, Imai and Keane (2003) find they could fit observed prices reasonably well using such a Markov-Process. In terms of our models, clearly both LIML and FIML will have the wrong density of prices in a week and, moreover, the wrong joint-distribution.

Assumptions:

- $F_{\lambda,\omega}(\lambda_t,\omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0.2 & 0.08 \\ 0.01 & 0.05 & 0.08 & 0.2 \\ 0.2 & 0.08 & 1.5 & 0.5 \\ 0.08 & 0.2 & 0.5 & 1.5 \end{bmatrix}\right)$
- $corr(\omega_t, \omega_{t-1}) = 0$
- $p_{tj} = \begin{cases} \mu_j + \phi_j w_{jt} + \omega_{jt}, & I_{(\kappa_{jt}>2.8)} \\ 0.6\left(\mu_j + \phi_j w_{jt} + \omega_{jt}\right), & 1 - I_{(\kappa_{jt}>2.8)} \end{cases}$
- $\kappa_{jt} = \begin{cases} \kappa_{jt-1} + \varpi_{jt}, & \kappa_{jt-1} + \varpi_{jt} \leq 2.8 \\ 0 & else \end{cases}, \varpi_{jt} \sim U(0,1)$

1.3.7 Scenario 6: FIML with correlation between λ and ω

Now we generate data from the FIML model. In addition to the Bertrand mark-up, we also allow for correlation between λ and ω. In principle, the advantage of the Bertrand approach is that it captures the correlation between price and λ parsimoniously through this mark-up. But, in this case we also add the correlation between λ and ω to create an additional source of endogeneity that would not be captured by the Bertrand mark-up.

Assumptions:

- $F_{\lambda,\omega}(\lambda_t,\omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0.2 & 0.08 \\ 0.01 & 0.05 & 0.08 & 0.2 \\ 0.2 & 0.08 & 1.5 & 0.5 \\ 0.08 & 0.2 & 0.5 & 1.5 \end{bmatrix}\right)$

1.3.8 Scenario 7: FIML with no correlation between λ and ω

This is the same as scenario 6, except we no-longer have any correlation between λ and ω. Hence, the FIML model described previously is the true model.

Assumptions:

- $F_{\lambda,\omega}(\lambda_t, \omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0 & 0 \\ 0.01 & 0.05 & 0 & 0 \\ 0 & 0 & 1.5 & 0.5 \\ 0 & 0 & 0.5 & 1.5 \end{bmatrix}\right)$

1.4 Simulation Results

We now report the findings from our Monte Carlo simulations in tables (1) to (7). In the tables, we report not only the demand-side parameters, we also report the covariance terms for λ and ω. These two terms correspond only to the LIML model as it requires them to resolve potential endogeneity (i.e. we do not estimate these terms for the other specifications). To conserve space, we do not report the supply-side coefficients from LIML or FIML, and we do not report the estimated weekly brand intercepts from IV. For those parameters reported, we include the mean point estimate across the 30 replications as well as the mean absolute deviation, MAD, from the true parameter value.

We begin with table (1), where the data are generated from the LIML model with autocorrelation. The conditional logit, as expected, gives biased parameter estimates in the anticipated direction. Mainly, price sensitivity is underestimated. Interestingly, while the LIML approach partially resolves this bias, it nevertheless still appears to underestimate the price response. This problem is probably due to the fact that the covariance terms appear to be underestimated, which in turn allows some of the bias in the price parameter to be retained. One explanation might be that the autocorrelation, which is not modeled in the LIML specification, might create additional inefficiency. However, it is surprising to observe that FIML seems to give more accurate results in terms of its lower MAD on intercepts and on the price parameter. Even though the FIML model is misspecified, it appears that it seem to have an easier time addressing the endogeneity problem. Intuitively, this is because the FIML model is more parsimonious. Instead of estimating the covariance terms between λ and ω, endogeneity is accounted for through the mark-up, which is a function of demand parameters. Finally, the IV approach clearly gives the most accurate results in terms of MAD. All point estimates are within a decimal place of the true parameter values.

In table (2) , we relax the autocorrelation condition. Nevertheless, our findings are still comparable to the previous case. LIML resolves some of the bias, but not all. In contrast, only the IV approach seems to give results within a decimal place of the true values.

In table table (3), we re-introduce the autocorrelation in ω, but we eliminate the correlation between λ and ω. In principle, there is no endogeneity problem built into this data. The logit model performs much better now, as expected. The remaining biases in the logit results are due to the fact that the logit does not account for the error components, λ, in demand. In this case, the LIML approach is able to recover the true parameters quite well. In contrast, the FIML approach does much worse. This result is expected since the FIML approach builds in endogeneity by construction. The only way for this specification issue to be offset would be with a price parameter of $-\infty$. Since the price parameter also enters demand, instead we simply get an overly-elastic demand estimate. Finally, the IV approach performs quite well. In fact, it seems to perform roughly as well as LIML.

In table (4), we remove autocorrelation. Now, LIML is the true model. Nevertheless, IV seems to perform quite comparably. In fact, for the price parameter, the IV approach seems to get closer to the true value on average.

In table table (5) , we allow prices to follow a Markov Process to capture the spirit of temporary price-cutting. In this case, the IV approach still gives results that are within a decimal place of the truth. Interestingly, the FIML results come reasonably close to the IV approach in terms of recovering the true parameters. In contrast, the LIML approach resolves some of the endogeneity biases; but FIML seems to perform relatively better.

In table (6), we shift our focus, generating our data from the FIML model. We also allow for additional correlation between λ and ω, so that the FIML model has the correct marginal density, but not the correct joint density. In this case, not only does the logit give biased results, it predicts the wrong sign on the price parameter. Also, the LIML approach resolves some of this bias, but it clearly does not entirely resolve the problem. FIML fairs much better; although the additional correlation between λ and ω introduces some specification bias. Recall that the FIML estimator does assumes all endogeneity can be recovered through the mark-up term. The advantage of this approach is parsimony. Finally, the IV approach, as always, gets within one decimal place of the true values.

In table (7), we remove the correlation between λ and ω. Now the FIML model is the true model. As such, it is not surprising that the FIML approach is the most accurate in terms of MAD. While this model outperforms the IV in this case, the IV approach still gives reasonably close estimates.

To summarize, the IV approach is robust across several different model specifications. The LIML and FIML approaches, by contrast, perform well when the data are generated from these models, but perform poorly in the presence of misspecification.

2 Heterogeneity and Endogeneity

Here we explore the role of the UBCs and the identification of heterogeneity. An advantage of household panel data is that we have within-household variation in choices, which can be used to identify heterogeneity in tastes, and across-household variation in choices within a week, which can be used to identify UBCs. In general, we are concerned about how failure to account for endogeneity in prices (i.e. correlation between prices and UBCs) might bias the heterogeneity parameters. In the previous section, we found that the endogeneity bias leads to understated price sensitivities. When data are generated from a population with heterogeneous price sensitivities, the bias in mean price effects could affect estimates of within-household deviations from these means. In addition, scanner panels are often unbalanced. In other words, consumers do not systematically shop in a focal chain each week. A potential concern is that some consumers' shopping decisions may correlate with UBCs. For instance, UBCs could be higher during summer, when some households leave for vacations. Alternatively, consumers may tend to shop in other stores when UBCs are low. In the context of a non-linear model, such as the random effects multinomial logit, it is not possible to derive these biases analytically. Hence, we explore these scenarios in the context of Monte Carlo simulation.

In designing the simulation, we first look at a balanced panel data set. In general, however, scanner panels are not balanced. A potential concern is that the manner in which households may enter or exit the panel (or the timing of shopping trips) may be systematically correlated with the distribution of UBCs. Hence, we also look at unbalanced panels. We begin with a case in which some randomly-selected households simply do not shop during high UBC weeks. Then, we look at the case in which the most price insensitive customers do not shop during high UBC weeks. Finally, we look at the case in which the most price sensitive customers do not shop during high UBC weeks. The goal of looking at the unbalanced panels is to see if any of these cases might amplify or, alternatively, mitigate some of the bias in the estimated heterogeneity.

2.1 Simulation Design

To address these concerns, we generate data from a heterogeneous logit demand system in the presence of price endogeneity. We consider 4 cases. First, we look at the case of a balanced panel (every household shops in every week). Then, we look at an unbalanced panel where a random subset of households do not shop during weeks when UBCs are high. Third, we look at an unbalanced panel where price sensitive shoppers tend not to shop during high-UBC weeks (i.e. there is positive correlation between UBCs and price sensitivities). Fourth, we look at an unbalanced panel where price insensitive shoppers tend not to shop during high-UBC weeks (i.e. there is negative correlation between price sensitivities and UBCs). To generate price endogeneity, we create data from the LIML specification (above). However, estimation will be carried out using our 2-step estimator.

To simplify, the analysis, we only include heterogeneity on the price sensitivity parameter. In general, we make the following assumptions for demand:

- $$\begin{bmatrix} \alpha_{1h} \\ \alpha_{2h} \\ \beta_h \\ \gamma_h \end{bmatrix} \tilde{} N \left(\begin{matrix} \alpha_1 \\ \alpha_2 \\ \beta \\ \gamma \end{matrix}, \Omega \right) = N \left(\begin{bmatrix} 2 \\ 3 \\ -2 \\ .3 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right)$$

- parameters to estimate: $\Theta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \\ \gamma \end{bmatrix}$, $\Sigma = chol(\Omega) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \Sigma_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
- $p_{jt} = \mu_j + \phi_j w_{jt} + \omega_{jt}$
- $F_{\lambda,\omega}(\lambda_t, \omega_t) \sim N\left(0, \begin{bmatrix} 0.05 & 0.01 & 0.2 & 0.08 \\ 0.01 & 0.05 & 0.08 & 0.2 \\ 0.2 & 0.08 & 1.5 & 0.5 \\ 0.08 & 0.2 & 0.5 & 1.5 \end{bmatrix}\right)$

2.1.1 Scenario 1:No Correlation Between UBC and price-sensitivity

Assumptions:

- All 150 households shop once per week for 40 weeks

2.1.2 Scenario 2:No Correlation Between UBC and price-sensitivity, unbalanced panel

Assumptions:

- A random subset of households does **not** shop when λ_{jt} is "high"
- random subset of households: randomly select 40 of the 150 households
- "high" λ_{jt}: $\lambda_{1t} > \lambda_1^{80^{th}\ percentile}$ or $\lambda_{2t} > \lambda_2^{80^{th}\ percentile}$

2.1.3 Scenario 3:Positive Correlation Between UBC and price-sensitivity

Assumptions:

- Households with "high" price sensitivity do **not** shop when λ_{jt} is "high"
- "high" price sensitivity: $|\beta_h| > \beta^{60^{th}\ percentile}$
- "high" λ_{jt}: $\lambda_{1t} > \lambda_1^{80^{th}\ percentile}$ or $\lambda_{2t} > \lambda_2^{80^{th}\ percentile}$

2.1.4 Scenario 4:Negative Correlation Between UBC and price-sensitivity

Assumptions:

- Households with "low" price sensitivity do **not** shop when λ_{jt} is "high"
- "low" price sensitivity: $|\beta_h| < \beta^{40^{th}\ percentile}$
- "high" λ_{jt}: $\lambda_{1t} > \lambda_1^{80^{th}\ percentile}$ or $\lambda_{2t} > \lambda_2^{80^{th}\ percentile}$

2.2 Simulation Results

We now discuss our findings. For each of the 4 cases described above, we generate 30 random datasets. For each of these 30 datasets, we estimate the heterogeneous direct MLE and the heterogeneous 2-step IV/MLE models. For each model, we report the mean parameter estimates across the 30 replications. We also report the mean absolute deviation from the true parameter values.

In table 8, we report the findings for the balanced panel. As expected, the mean effects are all estimated with bias in the case of direct MLE. However, the 2-step IV/MLE approach does reasonably well recovering the true parameter values. These results are consistent with our previous findings in the homogeneous case. Interestingly, the direct MLE approach also seems to over-estimate the variance in the price parameter. The MAD in this term relative to the true value is almost twice as high as the 2-step IV/MLE case.

In table 9, we report the findings for the unbalanced panel, where a random subset of households does not shop during high UBC weeks. Despite the unbalanced panel structure, the results look comparable to the balanced case.

In table 10, we report the findings for the unbalanced panel, where the most price insensitive households do not shop during high UBC weeks. This creates positive correlation between the UBCs and household-specific price parameters. This correlation creates an even larger upward bias in the estimated variance in price sensitivity for the direct MLE case, the MAD of which is now ten times the size of the 2-step MLE.

In table 11, we report the findings for the unbalanced panel, where the most price sensitive households do not shop during high UBC weeks. This creates negative correlation between the UBCs and household-specific price parameters. Interestingly, the variance in the price parameter is now much closer to the true values than for the previous cases. This result is perhaps not surprising given that the heterogeneity is over-estimated in the case of the balanced panel. Our results suggest that the negative correlation is in face off-setting some of this upward-bias.

To summarize, our simulation results indicate that even in the presence of heterogeneity, in addition to endogeneity, the proposed 2-step IV/MLE estimator performs well.

References

[1] Besanko, D., S. Gupta and D. Jain [1998], "Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework," *Management Science*, November.

[2] Besanko, D., J.-P. Dubé and S. Gupta [2003], "Competitive Price Discrimination Strategies in a Vertical Channel with Aggregate Data," *Management Science*, September.

[3] Erdem, T., S. Imai and M.P. Keane [2003], "A model of consumer brand and quantity choice dynamics under price uncertainty," *Quantitative Marketing and Economics*, 1(1).

[4] Goolsbee, A. and A. Petrin [2001], "Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," Working paper, The University of Chicago GSB.

[5] Villas-Boas, M. and R. Winer [1999], "Endogeneity in Brand Choice Models." *Management Science,* 45, 10, 1324-1338.

[6] Villas-Boas, M. and Y. Zhao [2003], "The Ketchup Marketplace: Retailer, Manufacturers, and Individual Consumers." forthcoming *Journal of Marketing Research*.

[7] Yang, S., Y. Chen and G. Allenby [2003], "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1(3).

	LIML with correlation between UBCs and supply shocks and autocorrelation								
		logit		LIML		FIML		IV	
	TRUE	param	MAD	param	MAD	param	MAD	param	MAD
α_1	2	0.2457	1.754	1.127	0.873	2.493	0.493	1.906	0.094
α_2	3	0.4917	2.508	1.934	1.066	3.655	0.655	3.058	0.058
price	-2	-0.6968	1.303	-1.400	0.601	-2.407	0.407	-2.013	0.013
promo	0.3	0.2752	0.025	0.299	0.001	0.251	0.049	0.333	0.033
$cov(\omega_{1t},\lambda_{1t})$	0.2			0.1507	0.049				
$cov(\omega_{1t},\lambda_{2t})$	0.08			0.0068	0.073				
$cov(\omega_{2t},\lambda_{1t})$	0.08			-0.0153	0.095				
$cov(\omega_{2t},\lambda_{2t})$	0.2			0.2666	0.067				

Table 1: Scenario 1

	LIML with NO correlation between UBCs and supply shocks and autocorrelation								
		logit		LIML		FIML		IV	
	TRUE	param	MAD	param	MAD	param	MAD	param	MAD
α_1	2	1.6734	0.327	1.893	0.107	3.246	1.246	1.863	0.137
α_2	3	2.4567	0.543	2.784	0.216	4.733	1.733	2.811	0.190
price	-2	-1.6708	0.329	-1.808	0.192	-2.807	0.807	-1.900	0.100
promo	0.3	0.2189	0.081	0.246	0.054	0.186	0.114	0.309	0.009
$cov(\omega_{1t},\lambda_{1t})$	0			-0.015	0.015				
$cov(\omega_{1t},\lambda_{2t})$	0			-0.0189	0.019				
$cov(\omega_{2t},\lambda_{1t})$	0			-0.0101	0.010				
$cov(\omega_{2t},\lambda_{2t})$	0			-0.0637	0.064				

Table 2: Scenario 3

	LIML with correlation between UBCs and supply shocks and NO autocorrelation								
		logit		LIML		FIML		IV	
	TRUE	param	MAD	param	MAD	param	MAD	param	MAD
α_1	2	0.0332	1.967	0.886	1.114	2.715	0.715	1.905	0.095
α_2	3	0.2496	2.750	1.550	1.450	4.033	1.033	3.051	0.051
price	-2	-0.5875	1.413	-1.215	0.785	-2.554	0.554	-2.024	0.024
promo	0.3	0.2812	0.019	0.304	0.004	0.458	0.158	0.346	0.046
$cov(\omega_{1t},\lambda_{1t})$	0.2			0.1863	0.014				
$cov(\omega_{1t},\lambda_{2t})$	0.08			0.13	0.050				
$cov(\omega_{2t},\lambda_{1t})$	0.08			0.15	0.070				
$cov(\omega_{2t},\lambda_{2t})$	0.2			0.3185	0.119				

Table 3: Scenario 2

	LIML with NO correlation between UBCs and supply shocks and NO autocorrelation								
		logit		LIML		FIML		IV	
	TRUE	param	MAD	param	MAD	param	MAD	param	MAD
α_1	2	1.6518	0.348	1.925	0.075	3.359	1.359	1.888	0.112
α_2	3	2.3599	0.640	2.683	0.317	4.827	1.827	2.755	0.245
price	-2	-1.6459	0.354	-1.858	0.142	-2.995	0.995	-1.914	0.086
promo	0.3	0.2051	0.095	0.213	0.087	0.404	0.104	0.281	0.019
$cov(\omega_{1t},\lambda_{1t})$	0			-0.0271	0.027				
$cov(\omega_{1t},\lambda_{2t})$	0			-0.0337	0.034				
$cov(\omega_{2t},\lambda_{1t})$	0			-0.0261	0.026				
$cov(\omega_{2t},\lambda_{2t})$	0			-0.0722	0.072				

Table 4: Scenario 4

	LIML with correlation between UBCs and supply shocks and Markovian prices								
		logit		LIML		FIML		IV	
	TRUE	param	MAD	param	MAD	param	MAD	param	MAD
α_1	2	0.3845	1.616	0.886	1.114	2.346	0.346	1.843	0.157
α_2	3	0.7196	2.280	1.550	1.450	3.475	0.475	2.936	0.064
price	-2	-0.7816	1.218	-1.215	0.785	-2.198	0.198	-1.945	0.055
promo	0.3	0.2836	0.016	0.304	0.004	0.195	0.105	0.358	0.058
$cov(\omega_{1t},\lambda_{1t})$	0.2			0.1863	0.014				
$cov(\omega_{1t},\lambda_{2t})$	0.08			0.13	0.050				
$cov(\omega_{2t},\lambda_{1t})$	0.08			0.15	0.070				
$cov(\omega_{2t},\lambda_{2t})$	0.2			0.3185	0.119				

Table 5: Scenario 5

	Bertrand with correlation between UBCs and supply shocks								
		logit		LIML		FIML		IV	
	TRUE	param	MAD	param	MAD	param	MAD	param	MAD
α_1	2	-1.6032	3.603	0.135	1.866	1.592	0.409	1.835	0.166
α_2	3	-1.434	4.434	0.843	2.158	2.522	0.478	2.940	0.061
price	-2	0.1265	2.127	-0.952	1.048	-1.795	0.205	-1.972	0.028
promo	0.3	0.2187	0.081	0.206	0.094	0.291	0.009	0.285	0.015
$cov(\omega_{1t},\lambda_{1t})$	0.2			0.2122	0.012				
$cov(\omega_{1t},\lambda_{2t})$	0.08			0.1522	0.072				
$cov(\omega_{2t},\lambda_{1t})$	0.08			0.2023	0.122				
$cov(\omega_{2t},\lambda_{2t})$	0.2			0.4639	0.264				

Table 6: Scenario 6

	Bertrand with NO correlation between UBCs and supply shocks								
		logit		LIML		FIML		IV	
	TRUE	param	MAD	param	MAD	param	MAD	param	MAD
α_1	2	-0.3147	2.315	1.014	0.986	1.877	0.123	1.697	0.303
α_2	3	0.0629	2.937	1.827	1.173	2.987	0.013	2.628	0.372
price	-2	-0.5495	1.451	-1.444	0.556	-1.964	0.036	-1.854	0.146
promo	0.3	0.2227	0.077	0.176	0.124	0.317	0.017	0.285	0.015
$cov(\omega_{1t},\lambda_{1t})$	0			0.0831	0.083				
$cov(\omega_{1t},\lambda_{2t})$	0			-0.0043	0.004				
$cov(\omega_{2t},\lambda_{1t})$	0			-0.0169	0.017				
$cov(\omega_{2t},\lambda_{2t})$	0			0.0065	0.007				

Table 7: Scenario 7

	TRUE	Direct		2-step	
		MLE		IV/MLE	
	param	param	MAD	param	MAD
int 1	2.00	0.1138	1.886	2.0633	0.063
int 2	3.00	0.2713	2.729	3.0753	0.075
price	-2.00	-0.5954	1.405	-2.05	0.050
promo	0.30	0.3537	0.054	0.4472	0.147
variance price	0.80	1.0212	0.221	0.9268	0.127

Table 8: No Correlation Between UBC and price-sensitivity, balanced panel

	TRUE	Direct		2-step	
		MLE		IV/MLE	
	param	param	MAD	param	MAD
int 1	2.00	0.1665	1.834	2.1972	0.197
int 2	3.00	0.3164	2.684	3.2505	0.251
price	-2.00	-0.528	1.472	-2.1262	0.126
promo	0.30	0.3478	0.048	0.4155	0.116
variance price	0.80	1.082	0.282	0.8984	0.098

Table 9: No Correlation Between UBC and price-sensitivity, unbalanced panel

	TRUE	Direct MLE		2-step IV/MLE	
	param	param	MAD	param	MAD
int 1	2.00	0.1069	1.893	1.1322	0.868
int 2	3.00	0.1488	2.851	1.7328	1.267
price	-2.00	-0.4803	1.520	-1.5311	0.469
promo	0.30	0.3265	0.027	0.234	0.066
variance price	0.80	1.3871	0.587	0.7451	0.055

Table 10: Positive Correlation Between UBC and price-sensitivity

	TRUE	Direct MLE		2-step IV/MLE	
	param	param	MAD	param	MAD
int 1	2.00	0.3056	1.694	1.4312	0.569
int 2	3.00	0.5079	2.492	2.3404	0.660
price	-2.00	-0.5233	1.477	-1.5385	0.462
promo	0.30	0.3531	0.053	0.3603	0.060
variance price	0.80	0.7439	0.056	0.8178	0.018

Table 11: Negative Correlation Between UBC and price-sensitivity