Filling Out the Instrument Set in Mixed Logit Demand Systems for Aggregate Data*

Charles J. Romeo
US Department of Justice
600 E Street, NW–Suite 10000
Washington, DC 20530

charles.romeo@usdoj.gov
(202) 307-5860

January 2010

*The Antitrust Division encourages independent research by its economists. The views expressed herein are entirely those of the author and are not purported to reflect those of the US Department of Justice.

Abstract

The random parameters logit model for aggregate data introduced by Berry, Levinsohn, and Pakes (1995) has been a driving force in empirical industrial organization for more than a decade. While these models are identified in theory, identification problems often occur in practice. In this paper we introduce a new set of readily available instruments that have the potential to substantially improve numerical performance in a variety of contexts. We use a set of endogenous price simulations to demonstrate that they are valid, and we use a real data illustration to demonstrate that they improve the numerical properties of the GMM objective function. In addition, we develop a metric that decomposes the explanatory power of the model into the proportion of market share variation that is explained by mean utility and that which is explained by the heterogeneity specification.

1. Introduction

Analysis of random parameters logit demand systems, as formulated and estimated using the methods developed by Berry (1994) and Berry, Levinsohn, and Pakes (1995) (henceforth BLP) have been a driving force behind the development of differentiated products demand analysis in empirical industrial organization and marketing science for the past 15 years. While these models are identified in theory,[1] in practice, these models often have multiple minima, and numerical problems often occur in optimization. The literature is relatively mum on such problems as researchers generally focus on presenting their best results, not on the myriad of problems they faced in reaching those results.[2]

In this paper we delve into the relationship between the choice of instruments and model structure on the properties of the objective function. These relationships have been examined by other researchers in linear models and in nonlinear GMM models, but generally in the context of weak identification due to weak instruments.[3] We argue that the problem in the mixed logit model is different: it is weak identification due to missing instruments. We offer two contributions to the understanding and improvement of identification in the context of the mixed logit demand model. First, we develop a metric that divides share variation into that which is explained by mean utility and that which is explained by deviations from the mean, and intuit how this metric responds to changes in model structure and the explanatory power of the instruments. In general, the proportion of share variation that is explained by mean utility increases with the strength of the instrument set and decreases with the richness of the random

---

[1]Recent papers by Berry and Haile (2009), Bajari, Fox, Kim, and Ryan (2009), and Fox and Gandhi (2009) all prove nonparametric identification for different variants of the model. The model in Berry and Haile nests random coefficient models used in the applied literature as it contains product specific unobservables, an additive extreme value error, and allows for endogenous prices. A key motivation for all three papers is to provide confidence to applied researchers that identification in mixed logit models does not hinge on assumptions on the distribution of random coefficients.

[2]Recent papers by Knittel and Metaxaglou (2008), and Dube, Fox, and Su (2008) are exceptions. Knittell and Metaxaglou discuss numerical problems that are driven by hill climbing algorithms, and Dube, et. al. discuss problems of convergence of the contraction mapping in the inner loop of the BLP algorithm and offer an alternative estimator that avoids it use.

[3]See eg., Stock and Wright (2000), and the survey by Stock, Wright, and Yogo (2002).

parameter specification. Second, we introduce a new set of instruments that to date have been missing from mixed logit demand models. We show that these instruments are valid first using theoretical arguments, and then using a simulation exercise. Then, in an illustration using Nielsen scanner data, we show that weak identification can occur in mixed logit demand systems even when the instruments are strong. But, introducing these heretofore missing instruments smooths the objective function producing steeply peaked objectives with a single minimum in all our model specifications.

For certain classes of problems then, we offer a simple way to the improve numerical performance: put the means of included demographics in the instrument set. Mean demographics enter expected utility and that provides the suggestion at least that they should be valid instruments. However, the Berry (1994)/BLP estimator with the contraction mapping nested within each optimization step, makes this issue more complicated. We demonstrate, that in the estimator, mean utility is *net* of the effect of included demographics and therefore they are valid instruments. Most importantly, including mean demographics among the instruments can greatly improve estimator performance.

The classes of problems in which mean demographics may prove to be important are ones where firms vary prices from store-to-store, from county-to-county, or from city-to-city at least partially in response to demographic differences. Marketing and economic applications that deal with grocery products are key candidates for substantial numerical improvements with the introduction of these instruments. Alternatively, products whose prices are set nationally, (e.g., online sales) will not gain from the use of these instruments, nor will products whose prices may be set locally, but for which data are only available at the national level (BLP's automobile data).

Mean demographics are likely to be correlated with willingness-to-pay and it is this correlation that potentially makes them important instruments in the mixed logit model. Price level shifts across markets are generally difficult to explain using instruments that are typically employed. In many product markets many of the same products are available in each geographic market leaving exogenous product characteristics unable to explain these shifts. By the same reasoning, instruments developed using the exchangeability arguments in Pakes (1994) and in BLP which are particular aggregations of exogenous characteristics are also likely to be unable to explain price level shifts. Nevo's (2001) use of average regional prices as instruments helps capture price level shifts. But, as he discusses, these instruments are not valid in the logit model if there are regional demand shocks. Unexplained city-specific valuation differences that are due to demographic differences are a likely source of demand shocks. Including mean demographics in the regression model, as he does in one of his logit specifications, controls for these shocks and makes it more likely that average regional prices are valid instruments.

Section 2 develops the mixed logit demand system for aggregate data. In Section 3, we deconstruct the BLP estimator to explain precisely why mean demographics are valid instruments. In addition, we develop a metric that decomposes market shares into two components: that which is explained by mean utility and that which is explained by the random coefficients. We then provide insights as to how these components will vary when mean demographics are included among the instruments. Section 4 presents two illustrations. The first is an endogenous price simulation study of the mixed logit model. We formulate prices as Bertrand-Nash markups plus marginal cost in order to mimic supply models most often used in the literature. We compare simulations with and without mean demographics in the instrument set. Results indicate that incorporating mean demographics in the instrument set yields consistent estimates that lower root mean squared error. The second uses Nielsen beer data. Here, we compare results of three random coefficient specifications each of with is estimated 25 times using three different instrument sets both with and without mean demographics to examine the effect on model statistics, on number of minima, on parameter estimates, and on random parameter functions. We report the following findings: models including mean demographics experience exhibit a single minimum in all our model and instrument specifications, and are more robust to instrument set changes. Section 5 contains conclusions.

2. Demand Model

We represent the conditional indirect utility of a consumer in market m at time t with preferences (a_{im},ν_{im},ε_{ijmt}) from the purchase of one unit of the jth product as

$$u_{ijmt} = x_{jmt}\beta + p_{jmt}\alpha_{im} + \xi_{jmt} + \varepsilon_{ijmt},$$

$$\alpha_{im} = \bar{\alpha} + \Gamma a_{im} + \upsilon\nu_{im}, \qquad \textbf{(1)}$$

$$i = 1,\ldots,I, \quad j = 0,\ldots,J_{mt}, \quad m = 1,\ldots,M, \quad t = 1,\ldots,T.$$

x_{jmt} and p_{jmt} are observed product characteristics and prices respectively. The ξ_{jmt} represent product characteristics that are observed by market participants, but unobserved by the econometrician. Assuming prices are set strategically, each p_{jmt} will be correlated with all J+1 ξ_{jmt} in each market-time period m,t. The ε_{ijmt} are i.i.d. type 1 extreme value errors.

The second equation in (1) is a hierarchical regression of α_{im} on D x 1 vector of demographics a_{im}, with standard normal error vector ν_{im}. Γ and υ are unknown parameters: Γ is a 1 x D vector and υ is a scalar. In this section, we limit attention to a model with a single random coefficient on price as that is sufficient for our purposes.

Normalize u_{0jmt} to zero and assume that each consumer maximizes utility by purchasing one unit of product j at time t if and only if $u_{ijmt} \geq u_{irmt}$, $r = 0,\ldots,J$. Then the set of consumers choosing product j in market-time period m,t is represented by

$$A_{jmt} = \{(a_{im},\nu_{im},\varepsilon_{ijmt}): u_{ijmt} \geq u_{irmt}, \ r = 0,\ldots,J\}.$$

Assuming that there are no ties, the market share of good j is given as

$$\begin{aligned} s_j(x_{mt},p_{mt},\xi_{mt};\psi) &= \int\!\!\int\!\!\int dP_0(a,\nu,\varepsilon|m) \\ &= \int\!\!\int\!\!\int dP_0(a|m)dP_0(\nu|m)dP_0(\varepsilon), \end{aligned} \qquad (2)$$

where ψ is the vector of utility function parameters, P_0 is the population distribution for individual level unobservables, and where the second line results from the assumption of mutual independence of a, ν, and ε.

Given that we have endowed ε with an i.i.d. type 1 extreme value distribution, the outer integral in (2) has a logit distribution as its analytical solution. Solving this integral yields an expression for market shares of the form

$$s_j(x_{mt}, p_{mt}, \xi_{mt}; \psi) = \int\int \frac{\exp(\delta_{jmt} + \mu_{ijmt})}{1 + \sum_r \exp(\delta_{rmt} + \mu_{irmt})} dP_0(a|m) dP_0(\nu|m)$$

$$= \int\int f_j(x_{mt}, p_{mt}, \xi_{mt}, a_{im}, \nu_{im}; \psi) dP_0(a|m) dP_0(\nu|m), \tag{3}$$

where $\delta_{jmt} = x_{jmt}\beta + p_{jmt}\bar{\alpha} + \xi$ and $\mu_{ijmt} = \Gamma a_{im} + \upsilon\nu_{im}$. Below we specify a as a discrete distribution of market level demographics, and we endow ν with a standard normal distribution. Under these specifications, the inner integral is replaced by a summation and the outer integral does not have an analytical solution.

To show that the means of demographic characteristics a_{im} enter the utility function we focus on the contribution of price to utility: $(\bar{\alpha} + \Gamma a_{im} + \upsilon\nu_{im})p_{jmt}$. Taking the expectation over over the distributions of $a|m$ and ν yields $(\bar{a} + \Gamma E[a|m])p_{jmt}$, which is the contribution of price to $E[u_{jmt}|m]$. Since the means of included demographics enter expected utility their correlations with price are observed. If the utility function was regression model then it would follow directly that they are valid instruments. Since the regression model uses the share functions, and the estimator contains an inner and an outer loop, the issue is more nuanced and will be discussed in detail below.

Mean demographics may be correlated with both the marginal cost and markup or willingness-to-pay components of price. If, for example, markets with higher income have higher labor costs then menu costs will be higher in these markets. Additionally, higher incomes may make consumers less price sensitive; they may be less willing to incur search costs to obtain a lower price. The market level means of included demographics then may be correlated with markups thereby providing correlations with parts of price movements that are missing from instruments correlated with marginal cost.

3. Deconstructing the GMM Estimator

In this section we lay out the steps involved in evaluating the GMM objective function at a given set of values for the nonlinear parameters. We use this deconstruction of a single function evaluation for two purposes. First, we show that mean demographics are in fact valid instruments in the context of the estimation algorithm. Second, we provide insights as to the impact of including these instruments on the estimates and on the characteristics of the objective function. To aid us in this discussion we develop a metric for the proportion of market shares explained by δ, μ, and $\mu + \xi$. Stepping back, we note here that the many of the insights regarding the inclusion of mean demographics in the instrument set apply more broadly to any valid relatively strong instruments, especially ones able to explain cross market price shifts.

Collect the linear and nonlinear parameters into $\theta = (\beta, \bar{\alpha})$ and $\lambda = (\Gamma, \upsilon)$ respectively. Assume that we have available a set of instruments z_{mt} and include among them $\bar{a}_m$, the market level means for included demographics. Make the following two assumptions regarding ξ_{mt} condition on z_{mt}.

A1. $E[\xi_{jmt}|z_{mt}] = 0$; and

A2. $Var[\xi_{jmt}|z_{mt}]$ is finite for almost every z_{jmt}.

Given a value of λ, say λ_0, there are four steps to evaluating the GMM objective function.

Step 1. Contraction Mapping: Simulate a solution to (3) for each j in each market-time period using a sample of $\{a_{i|m}, \nu_{i|m}\}_{i=1}^{I}$ for each m to obtain a consistent estimate of aggregate demand for each product j:

$$s_j(x_{mt}, p_{mt}, \xi_{mt}, P_I; \theta, \lambda_0) = I^{-1} \sum_i f_j(x_{mt}, p_{mt}, \xi_{mt}, a_{im}, \nu_{im}; \theta, \lambda_0). \tag{4}$$

δ_{mt} is then chosen as the solution to the vector of differences between observed and estimated aggregate product demand $\ln(s_{mt}) - \ln s(x_{mt}, p_{mt}, \xi_{mt}, P_I; \theta, \lambda_0)$ simultaneously for all J using the iterative procedure

$$\delta_{mt}^{(\ell+1)}(\lambda_0) = \delta_{mt}^{(\ell)}(\lambda_0) + \ln(s_{mt}) - \ln s(x_{mt}, p_{mt}, \delta_{mt}^{(\ell)}(\lambda_0), P_I),$$

yielding $\hat{\delta}_{mt}(\lambda_0)$ for each m and t.

Step 2. Form the estimator of θ:

Collect matrices x and p into r: $r = [x,p]$, and form the linear regression $\hat{\delta}(\lambda_0) = r\theta + \xi$.

Formulate $\hat{\theta}(\lambda_0)$ as

$$\hat{\theta}(\lambda_0) = (r'zWz'r)^{-1}r'zWz'\hat{\delta}(\lambda_0).$$

Step 3. Formulate the regression residuals:

$$\hat{\xi}(\lambda_0) = \hat{\delta}(\lambda_0) - r\hat{\theta}(\lambda_0).$$

Step 4. Evaluate the concentrated objective function:

$$\sum_{m,t} \hat{\xi}_{mt}(\lambda_0)' z_{mt} W_{mt} z_{mt}' \hat{\xi}_{mt}(\lambda_0),$$

where weighting matrix W is any symmetric positive definite matrix.

In Step 1, the chosen value of $\delta(\lambda_0)$ is net of the effects of included demographics. Equation (3) provides this insight; fixing λ fixes the μ_{ijmt}, leaving δ to explain that portion of the difference between observed and model shares that are not explained by the demographics. Since δ is net of the effects of included demographics so is ξ, and the $\bar{a}_m$ are valid instruments in the estimation of $\theta(\lambda_0)$ in Step 2. By this reasoning we are allowed to include instruments that likely capture part of consumer willingness-to-pay in each market even though these variables do not enter matrix r. Moreover, it is likely that mean demographics are studied by companies in setting prices and so they may have substantial explanatory power.

Assuming that companies factor mean demographics into price setting, including them in the instrument set should produce two effects on the model. First, mean utility will likely explain more of the variation in shares. To see this, suppose initially that the linear regression in Step 2 has little explanatory power. This will lead to large residuals in Step 3, and minimizing the weighted sum of squared residuals in Step 4 may necessitate updating λ to increase the portion of share variation that is explained by the μ_{ijmt}. In repeating Step 1, the variation in the δ's will get shrunk to better conform with the explanatory power of the linear regression thereby reducing the residuals. As the explanatory power of the linear regression increases, the reverse happens: more variation in shares gets captured by the systematic part of the linear regression and there is less residual variation for the nonlinear part of the model to explain.

Second, the objective function is smoother exhibiting fewer minima with mean demographics included in the instrument set than without. To see why this is, consider that any specific functional form for μ will generally not be able to explain all of the variation in market shares. Specifications having a random coefficient on price only and no or few demographics will likely be able to explain less variation than specifications having random coefficients on one or more product characteristics in addition to price and with a richer set of demographics. Remaining variation in shares will be absorbed by the δs. To the extent that this variation cannot be explained by the linear regression in Step 2, it will become part of ξ. The more variation there is in the ξs, the more likely that different (ξ,λ) combinations will produce multiple minima.

To give more precision to this argument we develop metrics for the proportion of s_{jmt} that is explained by δ, by μ, and by $\mu + \xi$. Noting that

$$s_{0mt} = s_0(x_{mt}, p_{mt}, \delta_{mt}(\lambda), P_I) = \sum_i \frac{1}{1 + \sum_k \exp(\delta_{kmt} + \mu_{ikmt})}$$

we define a set of metrics $S^{\omega} = (\sum_{mt} J_{mt})^{-1} \sum_{jmt} \sum_i \exp(\omega_{ijmt})^* s_{0mt}/s_{jmt}$ equal to the average proportion of the variation in s_{jmt} that is explained by ω, where $\omega = \{\delta, \mu, \mu + \xi\}$. Given that the contraction mapping sets $s(x_{mt}, p_{mt}, \delta_{mt}(\lambda), P_I) = s_{mt}$, this implies that the value of model shares are the same regardless of the specific parameterization of the model.[4] Only the division of the variation in shares into mean and heterogeneity model components differs from one formulation to the next. Hence, for a given dataset, the S^{ω} provide comparable measures of the source of explanatory

---

[4]This is true to the 10^{-14} tolerance we use in the contraction mapping.

power across models.

Suppose that at $\lambda = \lambda^*, S^{\mu(\lambda^*)}$ is at a local minimum, that is, the objective function in Step (4) is minimized at λ^* and the average proportion of the variation in s_{jmt} explained by μ is $S^{\mu(\lambda^*)}$. Further suppose that at λ^*, $\mu(\lambda^*)$ cannot explain all of the cross market variation in shares. Some cross market variation is subsumed into δ and is then explained by the regression, $r\theta$, or captured in $\xi(\lambda^*)$. Since both μ and ξ are functions of λ, the extent to which cross market variation is captured in the residual, shifts in λ allows these functions to trade off where this variation resides. Specifically, there may be $\lambda' \neq \lambda^*$ for which $S^{\mu(\lambda')}$ is also a local optimum and where $S^{\mu(\lambda')+\xi(\lambda')} \lesseqgtr S^{\mu(\lambda^*)+\xi(\lambda^*)}$ indicating that at these other optima more or less of the variation in shares may be explained by the heterogeneity terms and the regression residuals. The more cross market variation that gets captured in ξ, the more likely it is that the objective function will exhibit multiple minima.

Incorporating mean demographics into the instrument set increases S^{δ} and, in particular, it increases the amount of cross market variation explained by the linear regression. Hence, there is less cross market variation for μ to explain and less that gets captured in ξ, so there are fewer minima.

An alternative, less precise but more intuitive explanation is that the unconcentrated objective function is a quadratic equation in the linear parameters conditional on λ_0, the current draw of the nonlinear parameters. It follows that the objective is globally concave in θ given λ. However, it is not globally concave in λ. By increasing S^{δ} and improving the explanatory power of the linear regression, including mean demographics increases the weight given to the globally concave portion of the model, resulting in a smoother objective function. The beer illustration below shows that the increase in S^{δ} can be substantial when mean demographics are included in the instruments.

The multiple minima pathology we observe is one of the pathologies that can result from weak instruments. Nelson and Startz (1990) used a linear instrumental variables model to demonstrate that multiple modalities can occur and standard asymptotics can provide poor approximations to parameter distributions when instruments are weak.[5] Our problem, however, is not one of weak instruments, it is one of missing instruments. As we report in our beer illustration, partial R^2s and F-Statistics for our auxiliary price regressions indicate that our instruments are strong even without mean demographics included. But, multiple minima still occur in alsmot every model and instrument specification.

4. Illustrations

i. A simulation study with endogenous prices

To evaluate the effect of incorporating mean demographics in the instrument set we simulated data from a model where endogeneity in prices results from Bertrand markups entering the price function. We specified the utility function as

$$u_{ijm} = \beta_0 + \beta_1 x_{jm} + \alpha_{im} p_{jm} + \xi_{jm} + \varepsilon_{ijm}$$

$$\alpha_{im} = \bar{\alpha} + \gamma a_{im} + \upsilon \nu_{im}$$

$$i = 1,\ldots,I, \quad j = 0,\ldots,J, \quad m = 1,\ldots,M,$$

$$\{\beta_0, \beta_1, \bar{\alpha}, \gamma, \upsilon\} = \{2, 1, -2.5, 0.5, 0.25\},$$

and the price function as

$$p_{jm} = [diag\{\partial s_{jm}/\partial p_{jm}\}]^{-1} s_{jm} + \theta_0 + \theta_1 c_{jm} + \eta_{jm},$$

$$\{\theta_0, \theta_1\} = \{1, 1\},$$

where $[diag\{\partial s_{jm}/\partial p_{jm}\}]^{-1} s_{jm}$ are Bertrand markups and the $diag\{.\}$ functional indicates that the specification restricts each firm to own a single product in each market. To the extent that it's useful to provide context to simulated data, we construct the inputs so as to mimic a model in which the data were aggregated to the brand level; the single product characteristic, x, was designed to act like a share weighted composite characteristic that evolves slowly across markets

[5]Stock, Wright, and Yogo (2002) provide a survey of the literature.

in response to differences in cross-market product shares, and marginal cost, c, also evolves slowly across markets reflecting possibly differential transportation and menu costs. In addition, the single demographic, a, was drawn from a log normal distribution to have the character of an income variable. The specifications used for a, x, and c are

$$a_{im} \sim \exp(N_i(\mu_m, \sigma_m^2)), \text{ where } \mu_m \sim N(0.5,1) \text{ and } \sigma_m^2 \sim N(0.5,0.25),$$
$$x_{jm} = N_j(1,2) + N_{jm}(0,0.2),$$
$$c_{jm} = N_j(0,4) + N_{jm}(0,0.01).$$

ξ_{jm}, ν_{jm}, and η_{jm} are mutually independent and are each specified as having iid standard normal distributions, and ε_{ijm} is given an iid Type 1 extreme value distribution. ε is integrated out and then the model is solved for the Jx1 vectors of prices and shares p_m and s_m respectively for each m.

As in mixed logit models studied in the literature, price endogeneity results from correlations between prices and unobserved quality ξ, and is induced by ξ's presence in the Bertrand markups. The effect of demographics on prices also come through the markup term. We construct two instrument matrices. Both include an intercept, x, and c. In addition, we use the exchangeability of the products in the demand system to formulate $\bar{x}_{m,-j} = (J-1)^{-1}\sum_r (x_{rm} - x_{jm})$, and include $\bar{x}_{m,-j}$ and it's square in both matrices. Our final instrument is a vector of mean demographics $\bar{a}_m = \exp(\mu_m + \sigma_m^2/2)$, which only gets included in one instrument matrix. Specifically, we represent with $z_{-\bar{a}} = [1, x_{jm}, c_{jm}, \bar{x}_{m,-j}, \bar{x}^2_{m,-j}]$, and $z = [z_{-\bar{a}}, \bar{a}]$ respectively, instrument matrices that exclude and include mean demographics.

Table 1 contains two columns of mean price coefficient and root mean squared error (RMSE) estimates for simulations using $z_{-\bar{a}}$, and z. Each column contains the results from36 simulations using all permutations of three choices each of (I,J,M): I = {10,50,100}, J = {3,10,25}, and M = {48,100,200,400}. Each estimate is based on 25 runs. We limited runs to 25 per simulation in the interest of computational feasibility. Following Dube, Fox, and Su (2008), we retained the set of draws of individual types for all 25 runs of a simulation to remove the contribution of simulation error from the estimates. (Table is incomplete)

The following points can be gleaned from the table. First, and most importantly, $\bar{a}$ is a valid instrument. The true price coefficient, $\bar{\alpha}$ = -2.5. Reported means in both columns are generally biased toward zero for small M and J, but this bias diminishes quickly as J and M increase when z is used. Not so when $z_{-\bar{a}}$ is used. In this case the bias shrinks very slowly. Moreover, a comparison of the RMSE in the two columns shows that it is almost uniformly smaller when z is used, and the differences are substantial. The RMSE is often 20 to 50 percent smaller when mean demographics are included indicating that they are an important source of information. Results for unreported estimates of γ and υ show similar patterns of behavior across the two columns. When z is used, RMSE is generally much smaller thereby indicating cross market variation contained in mean demographics to be important in sorting out mean and heterogeneity price effects.

ii. Beer Data Illustration

We estimate models with three different random parameter specifications. The focus is on the impact of mean demographics and other instruments on model statistics (the number of minima, S^{δ}, generalized R^2, and computation time), on mean utility parameter estimates for endogenous variables, and on heterogeneity captured by the μ_{ijm}. This section proceeds in five parts. A short description of the data is presented first, followed by discussion of the construction of the outside share, and the results of evaluations of various instrument sets. Part four includes estimates of three different model specifications using three different instruments each of which both excludes and includes mean demographics. All model specifications are estimated 25 times from random normal starting values in order to search out multiple minima. This section concludes with a short discussion of the effect of incorporating mean demographics on policy analysis.

a) the data

For this illustration we used 16 weeks of Nielsen grocery store scanner data on beer pricing and sales from 37 Nielsen markets. For product characteristics we include variables for promotions (feature, display, feature and display, and discount), and dummy variables for package size (6, 12, 18, 24, and 30 packs), container type (bottles, longnecks, and cans), beer type (light, regular, and ice), package type (freezer packs, non-freezer packs), and brand (Busch, Bud Select, Bud Light, Coors, Coors Light, Corona, Heineken, Keystone, Labatts, Michelob, Michelob Ultra, Miller Genuine Draft, Miller High Life, Miller Lite, Milwaukee's Best, Natural). The preceding list contains all the major beer brands in terms of sales in these 37 markets. In the time period of these data, all of these brands were owned by one of five major brewers: Anheuser-Busch, Coors, Gruopo-Modelo, Heineken, and Miller. Together they accounted for more than 83 percent of the total reported quantity sales. The data also includes hundreds of smaller brands in each market-week. We define five aggregate "brands" (Craft, Import, Premium, Sub-Premium, Super Premium) and create share weighted aggregates of the product characteristics for these brands. With these aggregations in place, the data contain an average of 136 products in each of 37*16 = 592 market-time periods for a total of 80,861 observations.

Two other characteristics are included that may affect the utility of beer consumption. The first is a dummy for "event" which equals one in weeks that include either a major holiday, or the super bowl. Second, we include dummy variables equal to one for Anheuser-Busch, Coors, or Miller products in markets that include breweries for one or more of these brewers. These variables are included to allow for a hometown differential in the brand effect.

The promotional variables–discount, feature, display, and feature and display–record for each brand the percentage of stores, weighted by All Commodity Volume (%ACV),[6] in which these brands are on sale. Evidence gathered from instrumental variable tests indicates that all of these promotional variables are endogenous. There are likely to be two sources of endogeneity. First, ξ likely includes unobserved advertising that is correlated with observed promotions. Second, since these variables are only observed at the brand level they are really only proxies for the product level promotion variables. Any measurement error in these proxies gets absorbed into the ξ.

b) constructing the outside share

We assume that the market is all beer products, and given this we use three pieces of

[6] All Commodity Volume is total sales of all products in each store.

information in constructing the outside share. First, data from the 2008 edition of The U.S. Beer Market indicates that in 2007 each person, 21 and older, consumed an average 0.25308 cases of beer per week. Multiplying this figure by the 21 and over population in each market provides a first estimate of total market sales from which we subtract observed consumption recorded in our data to estimate outside beer consumption. Doing this yielded an average outside share of 95 percent.

The value of the outside share is that it brings the impact of outside competitors into the model. In the case of beer, however, this large outside share likely overstates the effect of outside competition. In large part outside competition results from the same brewers products being sold through other venues (convenience stores, mass merchandisers, and restaurants). Allowing such a large outside share would understate the control and coordination that brewers likely exert in pricing across venues. In most localities, brewers license distributors who have exclusive territories and who serve all venues. We use this information about the structure of the beer market to reduce the outside share down to 25 percent of our first pass calculation and then renormalize shares yielding an overall average outside share of roughly 59 percent.[7]

The third piece of information that we use in setting the outside share is the fact that beer consumption is cyclical. Consumption follows a sinusoid that peaks in the summer and troughs in the winter. The amplitude of peaks and troughs is greater the further north one goes, and are nearly flat for cities along southern tier of the US. Tampa's sine wave is an outlier among our 37 markets in that it peaks in the winter. We use this information to allow total market consumption to follow sine waves that we estimate separately for each market. For each market, we set per person consumption equal to the US average for the week in 2007 that includes the vernal equinox (the inflection point in most of the estimated sine waves), and allow total consumption to shift in accordance with our estimated sine waves.

---

[7] Our choice of 25 percent was simply a rough adjustment designed to keep to outside share large, while accounting for coordination of brewer/distributor pricing across venues. Reducing the size of the outside share makes demand for inside products less elastic. In one test, using the larger outside share increased the median own-elasticity by nearly 14 percent, while aggregate demand for inside products became 74 percent more elastic.

c) choosing and evaluating the strength of instruments

The focus of this section is to motivate the sets of instruments that we use in addition to our baseline set of exogenous product characteristics and mean demographics. We introduce three sets of instruments. The first is based on the own-product exchangeability argument in BLP. The second and third have intuitive bases: one uses lagged intertemporal market share differences, the other captures the effectiveness of past promotions.

BLP formulate exchangeability arguments that apply to differentiated product demand systems in which the demand for a product does not depend on the ordering of rival products, having cost functions that depend only on own-product costs, and with a Nash-Bertrand pricing equilibrium. In this context, as discussed in BLP, the demand and cost functions are fully exchangeable, and the pricing function is partially exchangeable:

(i) it is exchangeable in the order of the competing firms,

(ii) for a given competitor, exchangeable in the order of that competitor's products, and

(iii) for a given product, exchangeable in the order of other products marketed by the same firm.

The value of exchangeability is that it restricts the basis of the set of optimal instruments. In a Nash-Bertrand context, prices respond explicitly to changes in characteristics and prices of other products owned by the same firm, and implicitly to changes in characteristics and prices of products owned by competing firms. Hence, the set of optimal instruments includes characteristic information from all competing firms. As discussed by Newey (1990) one can use a polynomial to approximate the set of optimal instruments. However, without exchangeability, the basis of the polynomial grows with sample size: in a market with J products each of which has K observed characteristics, the unrestricted basis of a first order polynomial is JK. Pakes (1994) shows that exchangeability imposes restrictions on the polynomial basis. For example, (ii) implies that for a given competitor, the sum across products for each characteristic of that competitor's products forms the polynomial basis, while (i) further restricts this basis by implying that we can also sum across competitors. In addition (iii) implies that own-product characteristics and the sum of characteristics across all other products owned by the same firm contribute to the restricted polynomial basis. As such, the first order terms of this restricted polynomial grow only in K.

We lose exchangeability types (i) and (ii) if the competition cannot be characterized by a unique Nash-Bertrand equilibrium. Research on the beer industry raises questions as to the reasonableness of this assumption relative to Stackelberg type leader-follower behavior or other forms of coordinated behavior. Tremblay and Tremblay (2005) offer a long history of statements by Anheuser-Busch executives to the media as evidence that it plays the role of industry price leader. Econometric evidence, however, has been less compelling. Using data on the U.K. brewing industry, Pinske and Slade (2004) cannot reject Nash-Bertrand pricing, and Slade (2004) does not find evidence to support coordinated effects. Rojas (2008) finds that Stackelberg behavior may explain the U. S. data better, but that the estimated differences between two Stackelberg variants that he tests and Nash-Bertrand are not likely to differ statistically. This evidence, while admittedly weak, does raise concerns about the validity of the exchangeability arguments related to the products of other firms. More specifically, if one knows the form of coordination, knows the identity of the participants to a coordination arrangement, and knows the degree of coordination, then (i) and (ii) still may be satisfied. The pricing function is exchangeable in:

(ia′) the order of firms with which a firm coordinates assuming the degree of coordination between firms is the same;

(ib′) the order of firms with which a firm competes;

(ii′) for a given competitor, exchangeable in the order of that firm's products if the degree of coordination with that firm is the same among all products.

Suppose two firms tacitly collude in the pricing of all their products. Then, assuming their identity is known and coordination is perfect, it is as though we have a new Nash-Bertrand equilibrium with the products of two firms treated as though they were produced by a single firm. Alternatively, if the identity of the colluders is unknown or the degree of coordination across products is unknown or variable then the conditions required for exchangeability cannot be satisfied. For example, suppose in the context of the beer industry, that Anheuser-Busch/InBev (AB/InBev) and MillerCoors tacitly coordinate on pricing, but that coordination is imperfect, and varies by product and market. It may be that AB/InBev leads with its Budweiser and Bud Light products in markets where it has the dominant market share, and that MillerCoors leads with Miller Genuine Draft, Miller Lite, and Coors Light in markets where its share dominates. In both cases, the market leader might only respond to price moves by the competitors it deems most important.

Complications such as these make the conditions for exchangeability across competitors impossible to satisfy with any confidence. The alternatives are to use competitor characteristics in an unrestricted fashion, or to search out different instruments. We chose the latter path. Hence, in this illustration we formulate instruments using the product characteristics of other own-firm products[8] based on exchangeability argument (iii), and rely on intuition to fill out our instrument set.

We introduce two additional instrument sets. The first focuses on the promotion variables. We recognize that retailers decide how much of a discount to pass through to consumers and they contribute to decisions on which products to feature and/or display. Retailers are likely to use the success of past promotion programs as a decision guide. We capture this as

$$\Delta_{sw} = \frac{1}{4}\sum_{\ell=t-k-4}^{t-k} \frac{s_{jm\ell} - s_{jm,\ell-1}}{w_{jm\ell} - w_{jm,\ell-1}}$$

where w is used to index feature, display, and discount promotions. Δ_{sw} measures the average one week change in share of product j relative to the one week change in promotional level of j.[9] Effective past promotions will produce relatively high values of Δ_{sw}, ineffective ones will produce low or even negative values. In order to make Δ_{sw} predetermined and to reflect retailer decision processes, we set k = 4, 52, implying that the retailer reviews her experience from one

[8]we construct own-brand package size/container type instruments, such that, in each market-week, for each brand and package size by both bottles and cans, we exclude the current observation and formulate the share of that brand's products that are the same package size/container type.

[9]As noted above, we do not observe promotional levels by product, but only by brand. Hence, $w_{jmt} = w_{j'mt}$, for all $j,j' \in B$, where B indexes a brand.

month ago, and from one year ago.

The second additional instrument set also has an intuitive basis. One often hears on business reports or reads in the business pages how "same store sales" compare from month-to-month or year-to-year. This suggests that in setting prices, companies pay attention to intertemporal share changes. We form four share difference variables for each product in each market-time period: two one month differences, and two one year differences. Both the month and year differences are calculated using single week differences, and the difference of four week average shares. Suppressing *j* and *m* subscripts, share differences are calculated as follows:

$$\Delta s_{t-5,t-9} = s_{t-5} - s_{t-9}$$

$$\Delta s_{t-5,t-57} = s_{t-5} - s_{t-57}$$

$$\Delta \bar{s}_{month} = \frac{1}{4}\sum_{\ell=t-8}^{t-5} s_\ell - \frac{1}{4}\sum_{\ell=t-12}^{t-9} s_\ell$$

$$\Delta \bar{s}_{year} = \frac{1}{4}\sum_{\ell=t-8}^{t-5} s_\ell - \frac{1}{4}\sum_{\ell=t-60}^{t-57} s_\ell$$

We chose month and year differences based on single week and monthly average shares in an attempt to capture the full range of historical share information firms would review in updating prices. Lagging the differences has the practical effect of making it more likely that they will be valid instruments.

In using each instrument set, we form orthonormal polynomials of the own-brand, promotion effect, and share difference instrument sets and its square.[10] Our first step in using these instruments is to evaluate the relevance and strength of each set of instruments and mean demographics. We do this by evaluating partial R^2 measures developed by Shea (1997) and F-statistics based on these measures. Shea (1997) develops partial R^2s to evaluate instrument

---

[10]We have found that orthonormalizing the instrument polynomials and standardizing all data dramatically improves the condition numbers of the input matrices and substantially improves numerical performance. For example, the condition numbers for the instrument matrix in specification (iii) using instrument set (1) with mean demographics (see below) were cond(z) = 819 and cond($(z - \bar{z})/\sigma_z$) = 4.58. This degree of improvement was typical.

relevance in contexts with multiple endogenous variables. A simple example of how his measure works is given by considering the regression $Y_1 = Y_2\beta + \varepsilon$, where the vectors Y_1 and Y_2 are endogenous. Suppose now that two instruments vectors Z_1 and Z_2 are available and that Z_1 is highly correlated with both endogenous variables while Z_2 is uncorrelated with either endogenous variable. Reduced form regressions $Y_k = f(Z_1,Z_2)$, k = 1,2 would produce high R^2s even though β is unidentified. In contrast, Shea's measure reflects the fact that there are fewer relevant instruments than endogenous variables and produces a partial $R^2 = 0$ for both regressions.

Table 2 contains the results of these evaluations. All the partial R^2 estimates are from reduced form regressions of each of our five endogenous variables (price, discount, feature, display, and feature and display) on the baseline instrument set plus at least one additional instrument set. Columns 1 -3 include respectively results for regressions with the own-brand, share difference, and promotion effect instrument sets. The partial R^2 estimates indicate that our baseline instruments combined with each of these other instrument sets are relevant for all five endogenous variables. The F-statistics in the next block of the table, exclude the baseline instruments and examine the joint significance of the own-brand, share difference, and promotion effect instruments respectively in each column. These F-statistics are based on the partial R^2s in order to capture the explanatory power of each set of instruments for a given endogenous variable that is orthogonal to that instrument set's ability to explain the other endogenous variables. The results indicate that all three sets of instruments are strong for all five endogenous variables.

Above, we made the argument that mean demographics can better control for price shifts across markets than can exogenous product characteristics or the own-brand instruments: the set of products in each market is nearly identical yielding these instruments little power to explain cross-market price level differences. Columns 4 - 7 evaluate empirically the relevance and strength of mean demographics as additional instruments. The regressions in Columns 4 - 6, are the same as those in Columns 1 - 3 respectively, but with mean demographics as additional instruments, while in Column 7 all sets of instruments are included. Comparisons of the partial R^2 estimates in Columns 1 and 4, 2 and 5, and 3 and 6 indicate that mean demographics are relevant for all five endogenous variables as the partial R^2s increase substantially in all comparisons. The F-tests in Columns 4 - 6 evaluate the strength of mean demographics in regressions that include the baseline instruments and one other instrument set, while in Column 7 the F-test evaluates the strength of mean demographics in a regression that includes all the other sets of instruments. In all cases, mean demographics are shown to be strong instruments, thereby lending empirical support to the argument that they provide information that is orthogonal to that provided by all the other instruments.

d) Random coefficient model results

We gathered Current Population Survey (CPS) data on four different demographic variables for each of the 37 markets and tested three different random coefficient specifications.

Since CPS samples tend to be rather small, for each market we aggregated CPS data for April, August, and December 2007, and April 2008. Given the structure of household rotations in the CPS, we use four month intervals to minimize the number of households that contribute more than one response to our distributions.[11]

The four demographics we draw from these data are income, percent female, age, and percent Hispanic. From these four demographics we formulate six variables: income, income^2, income|female, age ranges 21 - 34, and 35-54, and percent Hispanic. Income is discretized into 10 adjacent intervals (in 000's): [0,10), [10,20), [20,30), [30,40), [40,50), [50,60), [60,75), [75,100), [100,∞). We specify a conditional income|female demographic to allow women to have a different response to beer pricing than men. We require this differential to operate through its effect on the income distribution in order to avoid the problem that there is likely to be very little variation in the percentage of women across markets making a solely female effect difficult to identify.

We estimate three models each having a different random coefficient specification:

i) random coefficient on price mixed with draws from income, income^2, income|female, and the two age variables;

[11] April, August, and December 2007 do not have any households in common, but approximately half of April 2008's households were in their first CPS rotation in April 2007, and will therefore make two contributions to our probability estimates.

ii) (i) plus a random coefficient on discount mixed with draws from income, income^2, income|female;

iii) (ii) plus a random coefficient on Corona mixed with draws from percent Hispanic.

We exclude the age variables from the discount mix because test runs showed that no systematic age-discount relationship in the data. In the third specification, we add a random coefficient on the Corona brand dummy and mix it with percent Hispanic in order to capture the greater affinity for Corona that Hispanics may have relative to other demographic groups. The progression of specifications provide an increasingly rich heterogeneity structure.

As detailed in Table 3, each specification is estimated using three different combinations of our sets of instruments both with and without mean demographics. All three instrument sets include the baseline instruments and the own-brand instruments. Specification (1) uses these alone, (2) is (1) plus share differences, (3) is (1) plus promotion effects. These same three instrument groupings are used both without and with mean demographics to form six combinations in total.

Table 3. Instrumental variable sets used in estimation of random coefficient logit specifications.

instrument set	without mean demographics	with mean demographics
(1)	baseline and own-brand	baseline and own-brand
(2)	(1) plus share differences	(1) plus share differences
(3)	(1) plus promotion effects	(1) plus promotion effects

[a]All specifications include the own-brand instruments and their squares in orthonormal polynomials. Specifications (2) and (3) including share differences or promotion instrument sets also include their squares in orthonormal polynomials.

In addition to being estimated with six different instrument sets, each random coefficients specification is estimated 25 times from random normal starting values in order to search for multiple minima. Varying the instrument set enables us to determine the effect of instruments on the structure of the objective. In addition, estimating different specifications with different instrument sets will expose patterns in the S^δ metric.

Finally, we note that varying the instrument sets has one other purpose. To the extent that a researcher is concerned that all her instrument sets are not exactly exogenous she would prefer to use fewer instrument sets rather than more. In the case of beer, products evolve slowly so exogeneity of the product characteristics is not likely to be an issue. Since the own-brand instruments are aggregations of the product characteristics, their exogeneity is also not likely to be at issue. However, the share difference and promotion effect instrument sets depend on being lagged long enough into the past to be predetermined, but not so long as to be irrelevant. This leaves these instruments fraught with some risk of remnant endogeneity.[12]

Table 4, parts a,b, and c, contain output statistics for the three random coefficients specifications. Each table reports the number of minima, estimation time, S^δ, and GR^2, Pesaran and Smith's (1994) generalized R^2 for regression models estimated using instrumental variables, for all three instrumental variable sets both without and with mean demographics.

Table 4a contains statistics for the model with a random coefficient only on price. The number of minima found are identified by the number of columns listed under each instrument set. Multiple minima are found in all three instrument specifications without mean demographics, while a single minimum is found in all three cases when mean demographics are included. This is yet another indication that mean demographics provide substantial identifying information. They also produce two other changes in the model statistics. S^δ, the proportion of share variation that is explained by mean utility is much lower in the instrument specifications without mean demographics. Roughly, between 1/4 and 3/4 of market shares can be explained by mean utility. Add mean demographics to the instrument set and S^δ increases to 99 percent. The random coefficients have virtually no explanatory power, mean utility explains everything. Since mean utility has both a systematic component and a residual, we use the GR^2 to further analyze how well we are explaining mean utility. With mean demographics, GR^2 is right around 0.40, and this is strictly greater than the GR^2 at the lowest minimums for all three instrument

[12]Nelson and Startz (1990) and Shea (1997) discuss the effects of remnant endogeneity of the instruments on the asymptotic distribution of model parameters.

specifications without mean demographics.[13]

Hence, the statistics from this first random coefficients specification indicate that while adding mean demographics produces an objective with a single optimum, it negates the value of the random coefficients specification. This suggests possibly that the random coefficients are only effective because of missing instruments: the random coefficients are capturing cross market variation in prices that is not controlled for by mean utility. This story evolves somewhat as we increase the flexibility of the random coefficients specification.

Compare the model statistics in 4a to those in Tables 4b and c. 4b adds a random coefficient on discount, while 4c adds random coefficients on both discount and Corona. One point of comparison is that the models generally have multiple minima when mean demographics are excluded from the instrument set. There is one exception in 4b. 25 runs using Instrument set (3) produce a single minima, even without mean demographics included. Turning attention to S^{δ}, the tables show that as the random coefficient specification becomes more flexible S^{δ} systematically decreases for estimations including mean demographics. In 4b, $S^{\delta} \in [0.613, 0.657]$, while in 4c $S^{\delta} \in [0.446, 0.516]$, a substantial decrease from 0.99 in Table 4a. As we will show when we turn to Tables 5a,b,c, the price demographic effects are not generally statistically significant when mean demographics are used, but the discount and Corona demographics are significant and this provides an explanation for the reduction in S^{δ} and corresponding increase in $S^{\mu} = 1 - S^{\delta}$.

S^{δ} for estimations excluding mean demographics behaves less systematically. It decreases from 4a to 4b as a more flexible random coefficients specification explains more share variation, but then increases somewhat in 4c. Also in 4c, S^{δ} is often smaller in estimations including mean demographics than in ones without. Alternatively, GR^2 moves in the opposite direction: in 4b GR^2 is larger for estimations including mean demographics, while in 4c it is smaller than the GR^2 at the lowest minimum in each column. My assessment of these comparisons is that without mean demographics in the instrument set the heterogeneity component of the model has to explain more and this leads to volatile behavior as the model

[13] When multiple minima are found, we use the global minimum from among the set uncovered when making comparisons between models with and without mean demographics.

specification evolves.

Two other points in these tables are worth noting. First, estimation time is generally much faster with mean demographics in the instrument set. The speed comparison varies with the specification, but speed gains are often between 25 and 50 percent, and this provides yet more evidence that adding these instruments produces a more steeply peaked objective. Second, while the overidentifying conditions were rejected for all specifications, increasing flexibility produced marked improvements in the objective function values from Table 4a to b to c for both sets of estimations.

Tables 5a,b,c contain estimates of all endogenous mean utility parameters, own-elasticity statistics, and random coefficients for each model specification and each instrument variable specification in Tables 4a,b,c respectively. Where estimation runs produced multiple minima, results are presented only for the lowest, possibly global, minimum.

Turning first to a comparison of the mean utility parameters across both estimations within a table across tables we make three points. First, note how the price parameters evolve. When mean demographics are excluded, instrument specification (1) always produces the estimates that are larger in absolute value than specifications (2) or (3). The differences in Tables 5a and b are statistically significant and seem to suggest that the share differences and promotion effects contain significant remnant endogeneity. However, once mean demographics are included, this pattern goes away suggesting a different explanation: share differences and promotion effects are explaining some of the cross market price differences that were not captured by the baseline and own-brand instruments alone, and that this significantly altered the price estimates.

Second, continuing to focus on the price coefficients, we note that with mean demographics included, the price coefficients always significantly smaller than the coefficients without mean demographics. With mean demographics, the coefficients also increase in absolute value as we move from Table 5a to b to c, while there is no systematic change in the estimates without mean demographics in place. In the estimates with mean demographics, there is a like increase the in the percentage of own-elasticities in the elastic range as we move from 5a to b to c. Together, these points indicate that the model is fitting the data better as it becomes more flexible.

Finally, compare the mean utility coefficients on discount, feature, display, and feature and display across specifications within each table. Note that without mean demographics among the instruments, these estimates make volatile swings as the instrument set is altered, while no such swings are evident with mean demographics in place.

The remaining blocks of each table show that the differences between estimations with and without mean demographics extend to the random parameters: estimations with mean demographics are much more similar across instrument specifications than estimations without.

The results in Table 5a indicate why S^{δ} explains 99 percent of share variation in models with a random coefficient only on price and with mean demographics. Other than υ_{price}, none of the random coefficients are significant. With mean demographics removed, the age 35 - 54 coefficient becomes significant in all specifications and income^2 is significant in specification (3). Subsets of the same coefficients are significant in the price random coefficients in the more flexible specifications in 5b and c. For the model with mean demographics, this indicates that individual heterogeneity is operating through the discount and discount and Corona random coefficients in 5b and c respectively. By contrast, in the estimations without mean demographics, none of the discount coefficients are significant in 5b while some discount and Corona coefficients are significant in 5c. Again, behavior of the results is erratic as the specification changes when mean demographics are excluded from the instrument set.

We use visuals to compare heterogeneity results both within and across tables. We graph the female price-income relationships, the female discount-income relations, and the Corona-Hispanic relationships in Figures 1 - 3. The income relationship graphs portray price and discount responsiveness using the relationship

$$\bar{\theta}_r * r + \gamma_{1r} * income + \gamma_2 r * income^2 + \gamma_{3r} * income|female, \quad r = \text{price, discount},$$

while the Corona-Hispanic graph portrays Corona responsiveness to changes in percent Hispanic using the relationship

$$\bar{\theta}_{Corona} * Corona + \gamma_{Corona} * (\%\text{Hispanic}),$$

where the $\overline{\theta}$ are mean utility parameters and the γ_{kr} are elements of Γ.

For ease of presentation, we designate instrument sets with and without mean demographics as z and z_{-a} and identify instrument sets as IV(1), IV(2), and IV(3). Figure 1 graphs the price heterogeneity results from Table 5a. Solid and dotted lines represent relationships with and without mean demographics in the instrument set respectively. Two points can be gleaned from this figure. First, with mean demographics the relationships are virtually identical across instrument specifications, while the relationships without mean demographics vary substantially. Second, the relationships with mean demographics are flat, thereby indicating that price sensitivity does not change with income. This is not surprising given that S^{δ} is 0.99 for these models. Alternatively, without mean demographics one relationship shows price sensitivity decreases substantially with increases in income and even turns positive at the highest income values, while the others show limited income heterogeneity.

These same general patterns are repeated for the price-income relationships in Figures 2 and 3. The relationships with mean demographics exhibit more income heterogeneity, however, given the lack of statistical significance of the coefficient estimates this exhibition of heterogeneity is not likely to be statistically significant.

The discount-income and Corona-Hispanic relationships exhibit similar patterns: the models with mean demographics show both flatter profiles and less variation among instrument specifications than the models without mean demographics. In these cases, however, the heterogeneity in the relationships with mean demographics are likely to be statistically significant, while the discount ones in Figure 2, without mean demographics are not.

e) thoughts on using the random parameters logit model for policy analysis

A policy analysis exercise is outside of the scope of this present paper[14], but the general point that with mean demographics in the instrument set the random parameters logit model is a more useful tool for policy analysis, seems worth a short discussion. First, having a single

[14] Such an exercise is included in a companion paper, Romeo (2010, in process).

minimum avoids issues raised by multiple minima. Does one formulate a weighted average of all available minima as in Sweeting (2009), and Gandhi (2007), work only with the global minimum among those minimum found as we have done here, or develop a different choice criteria. No matter how one proceeds, one runs the risk of being charged with "cherry picking" the minimum that supports one's position, or of having challenges raised based on other minima. A point that we excluded for sake of brevity is that for a given random coefficients specification and a given instrument set, there is substantial and statistically significant variation in results across minima. Moreover, finding up to seven minima in a run of 25 estimations leaves open the question as to whether there are still other minima to be found that might support one or another position. Second, the model with mean demographics is shown to be robust to changes in the instrument set thereby avoiding confounding issues in instrument choice. The volatile behavior of the model without mean demographics is certain to reduce the confidence policy makers place in any welfare analyses.

Clearly we have only conducted analysis in the context of a single simulation model and a single real data illustration, so it is yet to be shown whether the improvements demonstrated herein can be recreated in other datasets. Given that adding mean demographics to an instrument set is trivial, other researchers will be able to determine the effectiveness of this addition in short order.

5. Conclusions

Our sense is that finding that mean demographics are both valid and useful instruments is like found money: they have been overlooked in the literature to-date, are readily available, and are likely to be valuable in many contexts in marketing and economics. Our endogenous price simulation estimates demonstrate that they are valid instruments and that they improve estimator performance, decreasing RMSE relative to a model without them in 25 of 27 simulations. The beer data illustration shows that they improve identification, reshaping the objective by making it unimodal. This illustration also shows that including mean demographics reduces the sensitivity of model parameters and functions of interest to changes in the instrument set.

In conducting the exercises in this paper we limit attention to GMM estimation. A Gibbs sampler provides an alternative to searching out minima through repeated GMM estimation. Though it is outside the scope of the present study, the pseudo-likelihood approach of Romeo (2007) or the likelihood based approach of Jiang, Manchanda, and Rossi (2009) enables one to examine the entire parameter space, trace out the exact finite sample parameter distributions and shape of functions of interest.[15] The GMM objective or the closely associated posterior distribution are two such functions that will enable one to find all minima.

[15]Using the approach of Jiang, Manchanda, and Rossi (2009) requires that the objective has a single minimum as it uses a non-linear change-of-variables to define the likelihood. As noted by Berry (2003), "If there is the possibility that a given set of exogenous observable and unobservable variables could be associated with a different equilibrium set of prices and quantities, then there is no longer a one-to-one map between the unobservables and the endogenous prices (conditional on the exogenous observables and the demand errors) and so the change-of-variables necessary to define the likelihood is no longer correct."

References

Bajari, Patrick, Jeremy Fox, Kyoo il Kim, and Stephen Ryan, (2009), "The Random Coefficients Logit Model is Identified," Working Paper 14934, National Bureau of Economic Research.

Berry, Steven, (2003), "Comment on Bayesian Analysis of Simultaneous Demand and Supply," Quantitative Marketing and Economics, 1, 285-91.

_____ (1994), "Estimating Discrete-Choice models of Product Differentiation," RAND Journal of Economics, 25, 2, 242-62.

Berry, Steven, and Philip Haile, (2009), "Identification of Discrete Choice Demand from Market Level Data," mimeo, Yale University, Department of Economics.

Berry, S., J. Levinsohn, and A. Pakes, (1995), "Automobile Prices in Market Equilibrium," Econometrica, 63, 4, 841-90.

Dubé, Jean-Pierre, Jeremy Fox, and Che-Lin Su, (2008), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," Working Paper, Graduate School of Business, University of Chicago.

Fox, Jeremy, and Amit Gandhi, (2009), "Identifying Heterogeneity in Economic Choice Models," Working Paper 15147, National Bureau of Economic Research.

Gandhi, Amit, (2007), "The Stochastic Response Dynamic: A New Approach to Learning and Computing equilibrium in Continuous Games," Working Paper, Graduate School of Business, University of Chicago.

Jiang, Renna, Puneet Manchanda, and Peter Rossi, (2009), "Bayesian analysis of Random Coefficient Logit Models using Aggregate Data," Journal of Econometrics, 149, 136-48.

Knittel, Christopher, and Konstantinos Metaxoglou, (2008), "Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings," Working Paper 14080, National Bureau of Economic Research.

Nelson, Charles, and Richard Startz, (1990), "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," Econometrica, 58, 4, 967-76.

Nevo, A., (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," Econometrica, 69, 2, 307-42.

Newey, Whitney, (1990), "Efficient Instrumental Variables Estimation of Nonlinear Models," Econometrica, 58,4, 809-38.

Pakes, Ariel, (1996), "Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions," in Advances in Econometrics: Sixth World Congress, Vlo. II, edited by Christopher Sims, 171-260.

Pinske, Joris, and Margaret Slade, (2004), "Mergers, Brand Competition, and the Price of a Pint," European Economic Review, 48, 617-43.

Rojas, Christian, (2008), "Price Competition in the U.S. Brewing Industry," Journal of Industrial Economics, LVI, 1, 1-31.

Romeo, Charles, (2010, in process), "Incorporating Prior Information into a GMM Objective," internal mimeo, Economic Analysis Group, US Department of Justice.

_____, (2007), "A Gibbs Sampler for Mixed Logit Analysis of Differentiated Product Market using Aggregate Data," Comutational Economics, 29, 33-68.

Shea, John, (1997), "Instrument Relevance in Multivariate Linear Models: a Simple Measure," The Review of Economics and Statistics, 79, 2, 348-52.

Slade, Margaret, (2004), "Market Power and Joint Dominance in U.K. Brewing," Journal of Industrial Economics, LII, 1, 133-63.

Stock, James, Jonathan Wright, and Motohiro Yogo, (2002), "A Survey of Weak Instrument and Weak Identification in Generalized Method of Moments," Journal of Business and Economic Statistics, 20, 4, 518-29.

Stock, James, and Jonathan Wright, (2000), "GMM with Weak Identification," Econometrica, 68,5, 1055-96.

Sweeting, Andrew, (2009), "The Strategic Timing Incentives of Commercial Radio Stations: An Empirical Analysis Using Multiple Equilibria," RAND Journal of Economics, 40,4, 710-42.

The U.S. Beer Market: Impact Databank Review and Forecast, 2008 Edition, M. Shanken Communications, Inc., New York, NY.

Tremblay, Victor, and Carol Tremblay, (2005), The U.S. Brewing Industry: Data and Economic Analysis, MIT Press, Cambridge, Massachusetts.

Table 1. Means and root mean squared errors for price coefficients from simulation with endogenous prices: truth = -2.5. Output based on 25 runs per simulation.

J	M	I	instrument sets	
			$z_{-\bar{a}}$	z
3	48	10	mean RMSE	mean RMSE
		50	-1.426 (2.077)	-1.487 (1.678)
		100	-2.168 (1.700)	-2.399 (0.786)
			-1.856 (1.805)	-2.253 (0.962)
	100	10	-1.773 (1.882)	-2.400 (1.101)
		50	-2.383 (1.272)	-2.291 (0.667)
		100	-1.921 (1.931)	-2.308 (0.876)
	200	10	-2.138 (1.185)	-2.489 (0.483)
		50	-2.186 (1.663)	-2.291 (1.232)
		100	-2.065 (1.262)	-2.311 (1.150)
	400	10	-1.932 (1.261)	-2.287 (0.922)
		50	-2.000 (1.209)	-2.498 (1.359)
		100	-2.300 (1.636)	-2.517 (0.483)
10	48	10	-1.714 (2.050)	-1.667 (1.315)
		50	-2.064 (1.767)	-2.089 (0.841)
		100	-1.903 (1.512)	-1.901 (0.890)
	100	10	-2.242 (0.806)	-2.325 (0.836)
		50	-1.893 (1.311)	-2.307 (1.073)
		100	-2.355 (1.186)	-2.406 (0.590)
	200	10	-1.513 (2.432)	-2.427 (0.239)
		50	-2.064 (1.336)	-2.559 (0.162)
		100	-2.179 (0.823)	-2.358 (0.551)
	400	10	-1.960 (1.281)	-2.430 (0.207)
		50	-2.104 (1.088)	
		100	-2.349 (0.475)	-2.449 (0.192)
25	48	10	-2.065 (1.522)	-1.851 (1.191)
		50	-2.124 (1.336)	-2.237 (1.139)
		100	-2.623 (2.152)	-1.640 (1.579)
	100	10	-2.036 (1.118)	-2.365 (0.384)
		50	-1.865 (1.231)	-2.349 (0.352)
		100	-1.939 (1.176)	-2.249 (0.515)
	200	10	-1.982 (1.222)	-2.506 (-.127)
		50		-2.560 (0.162)
		100		
	400	10		-2.470 (0.191)
		50		
		100		

Table 2. Partial R^2 and F-statistics for first stage instrument variables regressions.

Instrument sets used:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
baseline	yes	yes	yes	yes	yes	yes	yes
own-brand	yes	no	no	yes	no	no	yes
share differences	no	yes	no	no	yes	no	yes
promotion effect	no	no	yes	no	no	yes	yes
mean demographics	no	no	no	yes	yes	yes	yes
1st stage partial R^2:							
price	0.506	0.486	0.487	0.531	0.505	0.504	0.539
discount	0.049	0.030	0.026	0.155	0.148	0.150	0.159
feature	0.033	0.022	0.019	0.078	0.072	0.069	0.086
display	0.103	0.081	0.064	0.255	0.245	0.241	0.262
feature and display	0.082	0.053	0.042	0.147	0.138	0.132	0.163
1st stage F-stats:							
price	209.48	65.23	40.11	242.05	107.80	30.45	77.78
discount	127.55	87.40	23.11	34.61	23.89	10.03	22.24
feature	79.74	62.46	17.99	52.91	73.37	8.40	72.47
display	217.41	228.01	22.85	167.12	84.36	2.57	91.09
feature and display	225.05	188.68	35.22	129.24	171.67	13.36	177.91
degrees of freedom	18	8	16	6	6	6	6
$P(F > f) = 0.05$	1.60	1.94	1.645	2.10	2.10	2.10	2.10
IV set evaluated in F-test:							
own-brand	yes	no	no	no	no	no	no
share differences	no	yes	no	no	no	no	no
promotion effect	no	no	yes	no	no	no	no
mean demographics	no	no	no	yes	yes	yes	yes

Table 4. Model statistics for the three specifications.

4a) Random coefficient only on price

variables	instrument set[a]		
	(1)	(2)	(3)
without mean demographics			
# of minima [# of times minima found][b]	5 [7,6,2,8,2]	3 [11,6,8]	4 [8,7,9,1]
median minutes to max	53	28	32
# of moment conditions	60	68	76
[range min objective value]	[565,642]	[1606,1762]	[1395,1670]
[S^{δ} at lowest, highest min]	[0.285, 0.516]	[0.410, 0.502]	[0.426, 0.797]
[GR^2 at lowest, highest min]	[0.329, 0.306]	[0.361, 0.407]	[0.388, 0.278]
with mean demographics			
# of minima [# of times minima found]	1 25	1 25	1 25
median minutes to max	19	17	17
# of moment conditions	65	73	81
mean min objective value	5152	5649	5425
mean S^{δ}	0.990	0.991	0.989
mean GR^2	0.403	0.405	0.402

[a](1) is baseline plus own-brand instruments; (2) is (1) plus share differences; (3) is (1) plus promotion effects.
[b]Minima ordered from lowest to highest.

4b) Random coefficient on price and discount

variables	instrument set[a]		
	(1)	(2)	(3)
without mean demographics			
# of minima [# of times minima found][b]	7 [16,4,1,1,1,1,1]	2 [7,18]	1 25
median minutes to max	133	78	109
# of moment conditions	64	72	80
[range min objective value]	[326,526]	[1114,1150]	1245
[S^δ at lowest, highest min]	[0.101, 0.467]	[0.382, 0.440]	0.421
[GR^2 at lowest, highest min]	[0.235, 0.155]	[0.229, 0.192]	0.273
with mean demographics			
# of minima [# of times minima found]	1 25	1 25	1 25
median minutes to max	43	35	34
# of moment conditions	69	77	85
mean min objective value	3934	4221	4316
mean S^δ	0.613	0.626	0.657
mean GR^2	0.316	0.325	0.333

[a](1) is baseline plus own-brand instruments; (2) is (1) plus share differences; (3) is (1) plus promotion effects.
[b]Minima ordered from lowest to highest.

4c) Random coefficient on price, discount, and Corona.

variables	instrument set[a]		
	(1)	(2)	(3)
without mean demographics			
# of minima [# of times minima found][b]	6 [16,3,1,1,1,3]	7 [10,3,1,6,2,2,1]	6 [12,1,3,5,2]
median minutes to max	152	100	101
# of moment conditions	66	74	82
[range min objective value]	[268,409]	[510,1134]	[678,1127]
[S^{δ} at lowest, highest min]	[0.481 0.173]	[0.485, 0.403]	[0.628, 0.261]
[GR^2 at lowest, highest min]	[0.579, 0.299]	[0.590, 0.255]	[0.676, 0.402]
with mean demographics			
# of minima [# of times minima found]	1 25	1 25	1 25
median minutes to max	98	81	89
# of moment conditions	73	68	76
mean min objective value	2226	2481	2765
mean S^{δ}	0.446	0.454	0.516
mean GR^2	0.432	0.445	0.467

[a](1) is baseline plus own-brand instruments; (2) is (1) plus share differences; (3) is (1) plus promotion effects.
[b]Minima ordered from lowest to highest.

Table 5a. Results for model with random coefficient only on price. Where multiple minima were found, estimates are presented for lowest minimum objective function.

variables	without mean demographics: instrument set #			with mean demographics: instrument set #		
	(1)	(2)	(3)	(1)	(2)	(3)
demand mean parameters and own-elasticities						
price	-16.311**	-11.843**	-11.728**	-5.798**	-5.644**	-5.569**
	(2.427)	(1.026)	(0.510)	(0.337)	(0.271)	(0.282)
discount	2.176**	1.286**	0.923	0.230**	0.348**	0.170**
	(0.339)	(0.251)	(0.197)	(0.051)	(0.038)	(0.043)
feature	1.133**	-0.054	-0.006	0.275**	0.431**	0.238**
	(0.418)	(0.136)	(0.164)	(0.086)	(0.052)	(0.056)
display	-1.599**	0.014	0.523	0.417**	0.339**	0.433**
	(0.231)	(0.075)	(0.103)	(0.042)	(0.034)	(0.044)
feature and display	-0.841	0.042	0.667	1.222**	0.696**	1.241**
	(0.543)	(0.122)	(0.151)	(0.090)	(0.052)	(0.076)
own-elasticities:						
% elastic	99.255	99.106	98.850	79.498	77.299	76.318
median	-4.427	-3.289	-3.360	-1.628	-1.585	-1.563
random coefficients on price (Γ,Υ)						
income	0.543	0.510	-0.672	-0.017	-0.029	-0.027
	(1.758)	(0.981)	(0.688)	(0.682)	(0.533)	(0.503)
$income^2$	-0.758	-0.764	6.206**	0.227	0.260	0.222
	(6.089)	(3.286)	(1.919)	(1.805)	(1.578)	(1.623)
income|female	-0.017	0.040	0.916†	0.037	0.048	0.036
	(2.740)	(1.143)	(0.537)	(0.868)	(0.727)	(0.715)
age 21 - 34	0.872	1.107	-0.253	0.116	0.103	0.117
	(3.240)	(0.728)	(0.881)	(0.915)	(0.675)	(0.803)
age 35 - 54	4.787**	2.801**	-1.891†	0.101	0.069	0.102
	(0.964)	(0.489)	(1.112)	(0.994)	(0.848)	(0.801)
υ_{price}	1.25e-9**	0.003	0.001*	2.51e-6*	9.59e-4*	6.82e-4*
	(2.22e-11)	(0.002)	(5.87e-4)	(1.05e-6)	(5.38e-4)	(3.41e-4)

(standard deviations in parentheses)

†Significant at the 10% level; *significant at the 5% level; **significant at the 1% level.

Table 5b. Results for model with random coefficients on price and discount. Where multiple minima were found, estimates are presented for lowest minimum objective function.

variables	without mean demographics: instrument set #			with mean demographics: instrument set #		
	(1)	(2)	(3)	(1)	(2)	(3)
demand mean parameters and own-elasticities						
price	-20.241**	-10.669**	-10.934**	-6.267**	-6.062**	-6.012**
	(3.794)	(0.649)	(0.599)	(0.542)	(0.419)	(0.444)
discount	2.400†	1.107**	0.995**	0.210	0.233**	0.142
	(1.413)	(0.243)	(0.316)	(0.189)	(0.155)	(0.172)
feature	1.751	-0.021	-0.201	0.482**	0.512**	0.328*
	(4.142)	(0.272)	(0.238)	(0.182)	(0.108)	(0.146)
display	-1.973*	0.196	-0.320	0.345	0.352**	0.391**
	(0.889)	(0.173)	(0.171)	(0.213)	(0.060)	(0.068)
feature and display	-1.629	0.518	1.070**	0.865**	0.686**	1.052**
	(5.620)	(0.219)	(0.171)	(0.145)	(0.070)	(0.119)
own-elasticities:						
% elastic	99.251	98.695	98.866	84.808	82.437	82.915
median	-5.379	-3.085	-3.166	-1.735	-1.681	-1.668
random coefficients on price (Γ,ϒ)						
income	2.243	0.286	0.090	-0.347	-0.319	-0.322
	(2.408)	(0.579)	(0.783)	(0.978)	(0.740)	(0.782)
income^2	-2.391	2.358	5.095**	1.452	1.348	1.356
	(12.601)	(2.971)	(1.214)	(2.872)	(2.096)	(2.310)
income|female	0.038	0.473	0.168	0.144	0.143	0.143
	(5.371)	(0.753)	(0.686)	(1.412)	(0.944)	(1.343)
age 21 - 34	0.748	0.064	0.008	-0.018	-0.003	0.014
	(14.579)	(1.797)	(2.596)	(1.108)	(0.754)	(0.852)
age 35 - 54	5.063†	-1.330**	-1.227	0.124	0.085	0.089
	(2.931)	(1.394)	(1.528)	(0.543)	(0.353)	(0.372)
υ_{price}	0.979	0.002	0.007	5.63e-4*	8.28e-4*	0.002†
	(4.879)	(0.003)	(0.006)	(3.35e-4)	(4.51e-4)	(0.001)
random coefficients on discount (Γ,ϒ)						
income	1.385	1.261	0.333	-1.071	-1.036**	-0.952*
	(4.991)	(0.616)	(0.581)	(1.149)	(0.308)	(0.484)
income^2	9.109	5.728	6.245	3.946	3.829**	3.400**
	(21.709)	(1.220)	(1.093)	(2.582)	(1.059)	(1.262)
income|female	2.821	2.197	0.977	0.357	0.375	0.317
	(6.091)	(0.675)	(0.604)	(1.075)	(0.448)	(0.812)
$\upsilon_{discount}$	1.419	0.264	0.098	0.311	0.301*	0.225*
	(5.590)	(0.261)	(0.126)	(0.413)	(0.138)	(0.093)

(standard deviations in parentheses)

†Significant at the 10% level; *significant at the 5% level; **significant at the 1% level.

Table 5c. Results for model with random coefficient on price, discount, and Corona. Where multiple minima were found, estimates are presented for lowest minimum objective function.

variables	without mean demographics: instrument set #			with mean demographics: instrument set #		
	(1)	(2)	(3)	(1)	(2)	(3)
demand mean parameters and own-elasticities						
price	-14.777	-13.692**	-11.828**	-7.820**	-7.698**	-7.519**
	(10.576)	(0.885)	(0.841)	(0.640)	(0.589)	(0.520)
discount	1.370	0.834*	0.511†	-0.186	-0.139	-0.103
	(2.475)	(0.408)	(0.267)	(0.290)	(0.216)	(0.187)
feature	2.085	0.018	0.179	0.074	0.057	-0.061
	(1.879)	(0.298)	(0.270)	(0.171)	(0.092)	(0.125)
display	0.075	1.199**	1.091**	0.875**	0.854	0.851
	(1.372)	(0.176)	(0.204)	(0.155)	(0.075)	(0.079)
feature and display	-2.831	-0.412†	0.533	0.679**	0.647**	0.931**
	(4.079)	(0.224)	(0.408)	(0.243)	(0.090)	(0.131)
own-elasticities:						
% elastic	99.247	99.219	99.022	93.691	93.536	93.360
median	-4.185	-3.916	-3.370	-2.147	-2.116	-2.074
random coefficients on price (Γ,ϒ)						
income	0.837	0.958	0.407	-0.518	-0.484	-0.485
	(4.303)	(0.906)	(0.803)	(1.110)	(0.788)	(0.647)
$income^2$	6.184	3.554	4.335**	1.934	2.436	1.877
	(17.364)	(2.226)	(1.482)	(3.569)	(3.569)	(1.794)
income|female	-0.210	0.069	-0.144	0.223	0.230	0.256
	(9.324)	(1.823)	(0.830)	(0.952)	(0.656)	(0.690)
age 21 - 34	0.231	1.128	0.858	-0.069	-0.061	-0.053
	(4.561)	(1.258)	(1.211)	(0.924)	(0.558)	(0.478)
age 35 - 54	1.353	1.422	1.088	0.220	0.234	0.169
	(10.584)	(1.590)	(1.002)	(0.661)	(0.573)	(0.493)
υ_{price}	0.390	0.621	0.246†	0.141	0.182*	0.138*
	(3.620)	(0.839)	(0.136)	(0.137)	(0.083)	(0.070)
random coefficients on discount (Γ,ϒ)						
income	1.555	1.591*	0.337	-1.743	-1.627**	-1.350**
	(5.166)	(0.636)	(0.591)	(0.666)	(0.426)	(0.403)
$income^2$	5.479	3.807**	4.184**	6.251**	5.920**	4.801**
	(13.678)	(1.094)	(1.311)	(1.583)	(1.231)	(1.194)
income|female	2.272	1.906*	0.972	0.565	0.553	0.256
	(4.414)	(0.947)	(0.744)	(0.810)	(0.396)	(0.690)
$\upsilon_{discount}$	0.001	0.012	0.009	0.602	0.597*	0.327*
	(0.003)	(0.019)	(0.009)	(0.560)	(0.295)	(0.160)
random coefficients on Corona (Γ,ϒ)						
% Hispanic	-9.696	-8.630**	-8.632**	-3.910**	-3.835**	-3.643**
	(9.488)	(0.701)	(0.546)	(0.330)	(0.281)	(0.254)
$\upsilon_{Hispanic}$	0.141	1.14e-5	-7.078**	0.011**	0.021**	0.004**
	(1.539)	2.63e-5	(0.826)	(0.004)	(0.006)	(0.001)

(standard deviations in parentheses)

†Significant at the 10% level; *significant at the 5% level; **significant at the 1% level.

GAUSS Mon Jan 04 18:24:01 2010

Figure 1. Female price response heterogeneity by income: random coefficient on price only

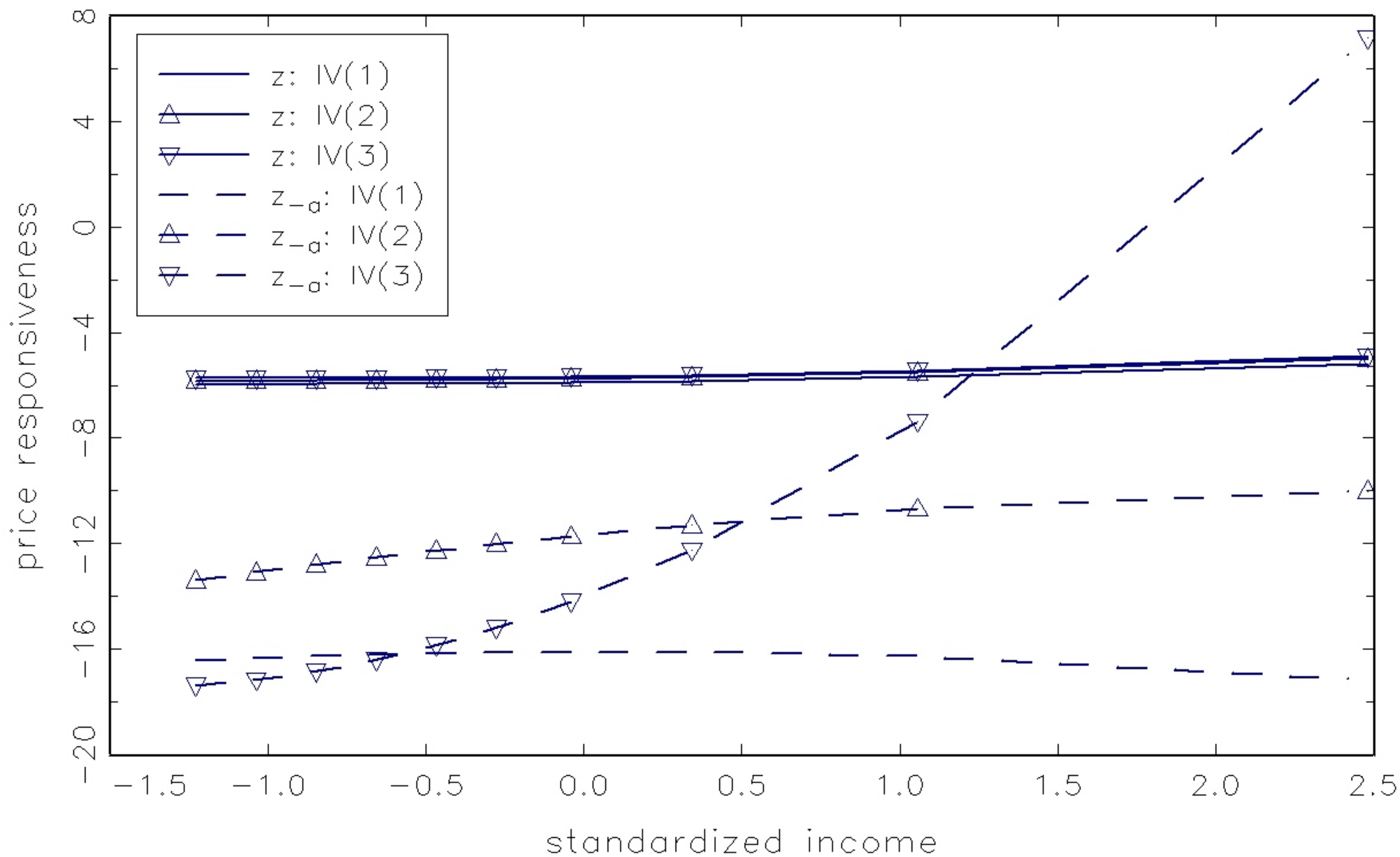


Figure 2. Female price and discount response heterogeneity by income: random coefficient on price and discount

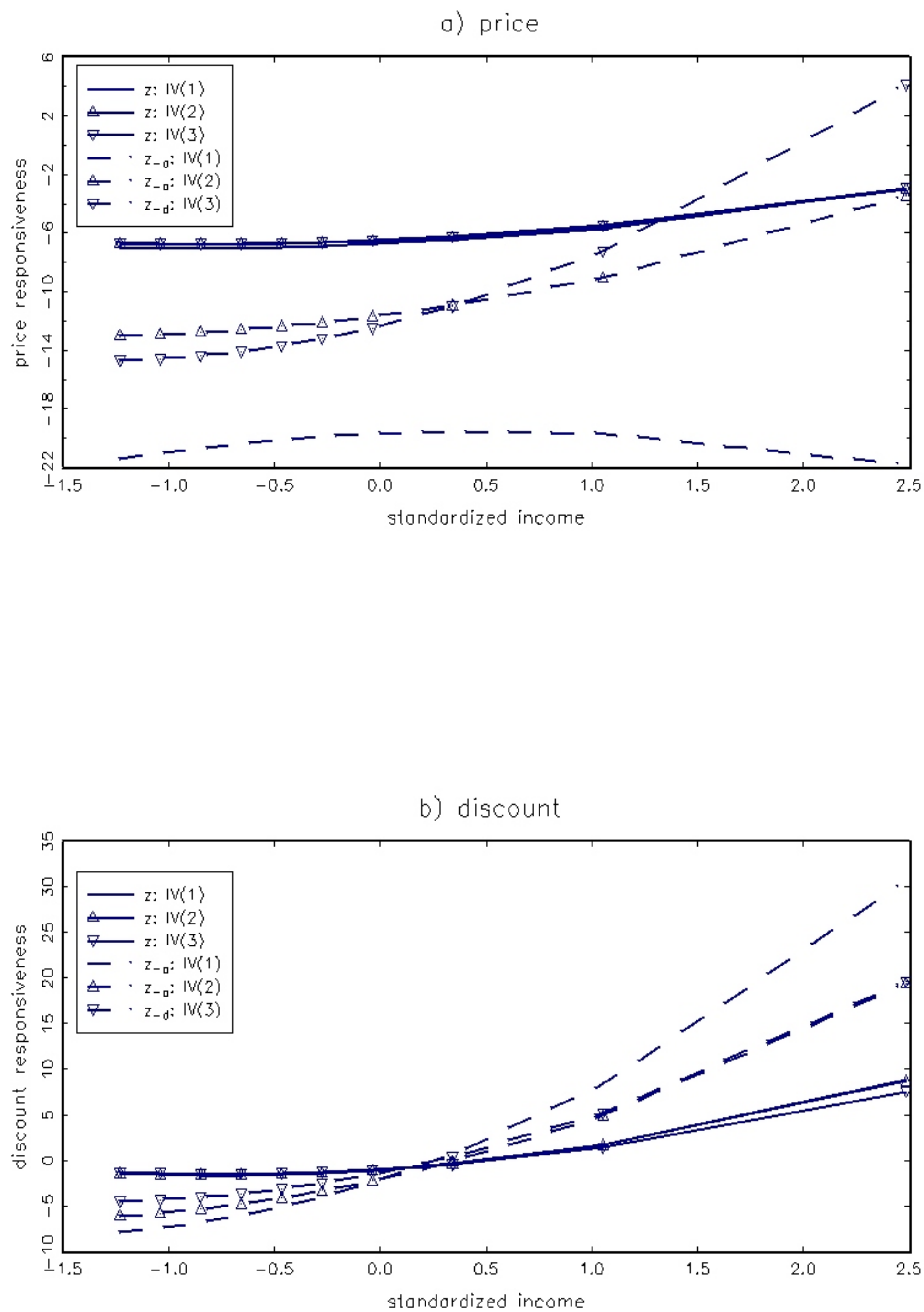
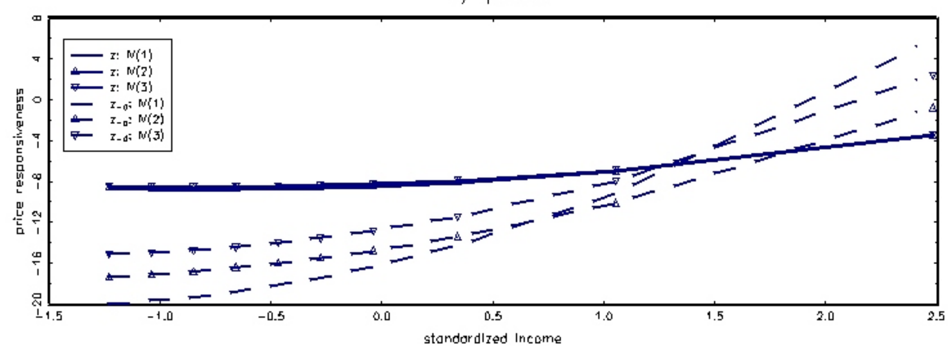
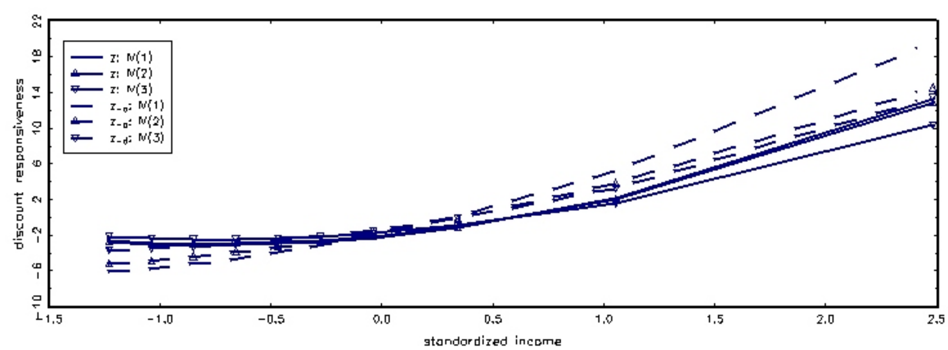


Figure 3. Female price and discount response heterogeneity by income, Corona heterogeneity by percent Hispanic: random coefficients on price, discount, and Corona

a) price



b) discount



c) Corona