Transportation Research Part B 84 (2016) 182–210

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Forecasting light-duty vehicle demand using alternative-specific constants for endogeneity correction versus calibration

C. Grace Haaf[a], W. Ross Morrow[b], Inês M.L. Azevedo[c], Elea McDonnell Feit[d], Jeremy J. Michalek[a,c,*]

[a] Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[b] Ford Motor Company, Ford Research and Innovation Center, 3200 Hillview, #200, Palo Alto, CA 94304, USA
[c] Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[d] Marketing, Drexel University, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 23 February 2015
Revised 10 November 2015
Accepted 19 November 2015
Available online 13 January 2016

Key words:
Automotive demand
Discrete choice model
Demand forecast
Forecast uncertainty

ABSTRACT

We investigate parameter recovery and forecast accuracy implications of incorporating alternative-specific constants (ASCs) in the utility functions of vehicle choice models. We compare two methods of incorporating ASCs: (1) a maximum likelihood estimator that computes ASCs post-hoc as calibration constants (MLE-C) and (2) a generalized method of moments estimator that uses instrumental variables (GMM-IV) to correct for price endogeneity. In a synthetic study we observe significant coefficient bias with MLE-C when the price-ASC correlation (endogeneity) is large. GMM-IV successfully mitigates this bias given valid instruments but exacerbates the bias given invalid instruments. Despite greater coefficient bias, MLE-C yields better forecasts than GMM-IV with valid instruments in most of the cases examined, including most cases where the price-ASC correlation present in the estimation data is absent in the prediction data. In a market study of U.S. midsize sedan sales from 2002 – 2006 the GMM-IV model predicts the 1-year-forward market better, but the MLE-C model predicts the 5-year-forward market better. Including an ASC in predictions by any of the methods proposed improves share forecasts, and assuming that the ASC of each new vehicle matches that of its closest competitor vehicle yields the best long term forecasts. We find evidence that the instruments most frequently used in the automotive demand literature may be invalid.

1. Introduction

Discrete choice models (DCMs) are used to interpret and forecast product demand in a variety of contexts, and a popular application is the new vehicle market. In particular, the automotive literature employs DCMs to understand drivers of purchase behavior (Allcott & Wozny, 2014; Copeland et al., 2011; Li et al., 2011; Dasgupta et al., 2007; Sudhir, 2001; Lave & Train, 1979) and predict future vehicle market shares (Greene et al., 2004; Greene et al., 2005; Duvall & Knipping, 2007; Balducci, 2008; Lin & Greene, 2010; U.S. Energy Information Administration, 2011). Alternative models of household vehicle choice can be used to forecast demand, but we focus exclusively on DCMs. DCM specifications include popular multinomial logit, nested logit, mixed

* Corresponding author. Tel.: +1 412 268 3765; fax: +1 412 268 2908.
E-mail addresses: grace.haaf@gmail.com (C.G. Haaf), wmorro13@ford.com (W.R. Morrow), iazevedo@cmu.edu (I.M.L. Azevedo), efeit@drexel.edu (E.M. Feit), jmichalek@cmu.edu (J.J. Michalek).

http://dx.doi.org/10.1016/j.trb.2015.11.012
0191-2615/

logit, and probit models (Train, 2009), as well as variants of these models, such as the generalized multinomial logit model (Fiebig et al., 2009).

DCMs of product purchases are generally estimated using either stated choice data or revealed preference data. Stated choice data can be obtained from choice-based conjoint experiments, for which the modeler selects a set of attributes related to product choice and designs hypothetical products for respondents to choose. These studies can avoid issues such as omitted variables, endogeneity, and multicollinearity. However, such studies typically rely on the respondent to make hypothetical choices that do not necessarily reflect real purchase choices in a market context. In contrast, revealed preference data (often aggregate market sales data) track real market purchases. Revealed preference studies have the limitation that buyers evaluate factors that are unobserved by the modeler or are difficult to represent mathematically (e.g. aesthetics). Also, attributes tend to be correlated among product alternatives in the marketplace (e.g. vehicles with luxury features routinely have higher prices), sometimes introducing multicollinearity and/or endogeneity issues, depending on whether the correlated attribute is observed by the modeler (Swait et al., 1994). Different sets of econometric assumptions are needed for the stated and revealed preference modeling approaches. We focus on models constructed from revealed preference (sales) data.

The number and nature of attributes considered by consumers in a vehicle purchase decision is sufficiently large and complex that any DCM posed will likely be missing information about some of the attributes that determine consumer choices. In order to address the utility not captured by explanatory variables, modelers frequently include an alternative-specific constant (ASC) in the utility function. For example, following Lave and Train (1979):

$$u_{ijt} = \mathbf{x}'_{jt}\boldsymbol{\beta}_i + \xi_{jt} + \varepsilon_{ijt} \tag{1}$$

where u_{ijt} is the utility consumer i derives from product j in market (year) t, $\mathbf{x}_{jt}$ is a vector of attributes specific to product j in market t, $\boldsymbol{\beta}_i$ is a vector of taste parameters for consumer i, ξ_{jt} is the ASC for product j in market t, and ε_{ijt} is an idiosyncratic error term treated as a random variable. Consumer-specific attributes like income or family size can also be included in the utility function, but are not included in this study since we use aggregate sales data where this information is not available.

An interpretation of the ASC introduced into the automotive demand context by Lave and Train (1979)[1] and popularized by Berry et al. (1995) is that it represents the mean cumulative effect of all product attributes that consumers use to evaluate a product but that are unknown to the researchers. Alternative terms for the ASC when it is used to represent omitted variables include the unobservable (Allcott & Wozny, 2014; Sudhir, 2001; Berry et al., 1995; Berry, 1994; Berry et al., 1999), the unobserved product characteristic or attribute (Berry et al., 2004; Beresteanu & Li, 2011), market-level disturbance (Petrin, 2002), and demand shock (Dubé et al., 2012; Knittel & Metaxoglou, 2012). However, the ASC need not necessarily be viewed as a representation of unobserved attributes but rather can be included as a purely mathematical construct to improve model fit (Greene et al., 2004; Greene et al., 2005), sometimes referred to as a calibration constant (Bunch et al., 2011). Indeed, in the literature, the treatment and interpretation of the ASC differs depending on whether the focus of the research is to *forecast* future vehicle demand shares (i.e. the "predictive" literature) or to measure the importance of attributes to consumers (i.e. the "explanatory" literature), especially as it pertains to willingness-to-pay and price elasticities of demand.

The predictive literature generally obtains ASCs by estimating coefficients in a model that excludes the ASCs and then "calibrating" the model post hoc by choosing values for the ASCs so that the modified model-predicted shares of the estimation data match observed shares. In contrast, the explanatory literature is primarily concerned with coefficient estimation and thus views it as imperative to address potential sources of coefficient inconsistency and bias — especially price endogeneity (Berry et al., 1995). Inconsistency arises if the ASC is correlated with an observed attribute, such as price. If the ASC is interpreted as a representation of aggregate utility from unobserved attributes, then it is plausible that observed and unobserved vehicle attributes (e.g. price and aesthetics) are correlated for markets in which prices are set by strategic firms, which would asymptotically bias[2] the coefficient of the observed attribute away from the true value. Though the true population taste parameters are unknowable for real data, researchers have demonstrated that for models estimated on actual market data the estimated price coefficient bias (measured as the difference between estimates when endogeneity is ignored versus when it is corrected for) is in the expected directions (Berry et al., 1995; Villas-Boas & Winer, 1999; Chintagunta, 2001). The explanatory literature implements estimation techniques that mitigate endogeneity bias — typically using instrumental variables (IVs) and estimating the ASC simultaneously with the coefficients.

There are drawbacks to mitigating coefficient bias with IVs. Model estimation is challenging in part because *valid* instruments are difficult to specify and impossible to verify, as demonstrated by Rossi (2014). Valid instruments require that they are correlated with the endogenous observed vehicle attribute(s), uncorrelated with the unobserved attribute(s), and do not affect the dependent model variable (market share) except through the observed attributes (Wooldridge, 2010). Instruments that do not meet these conditions are termed invalid. Because these properties are difficult to satisfy in many situations, instrument selection is somewhat subjective and ad hoc, and, as we show, the "wrong" choices can generate models that underperform those

[1] Lave and Train (1979) estimate a disaggregate model of vehicle choice in which all observed vehicle attributes are interacted with consumer attributes, e.g. income, so that the ASC is identical to the unobserved vehicle-specific utility. However, as in Train and Winston (2007), the ASC refers to the mean utility derived from both the observed and unobserved vehicle attributes. In the aggregate demand model context here the ASC refers only to the unobserved portion of utility.

[2] Methods that incorporate (valid) IVs result in estimators that are consistent (as the data sample size goes to infinity the expected value of the estimator converges to the true value of the parameters should they exist) but not unbiased in the sense that the sampling distribution of the estimator is centered on the true value of the parameters (also termed "finite sample bias"). In the literature discussed here "bias" is shorthand for "asymptotic bias" and is used interchangeably with "inconsistency." (Wooldridge, 2006)

that ignore endogeneity altogether. Specifically, Rossi shows that invalid instruments can lead to coefficient estimates exhibiting larger bias than those obtained *ignoring* endogeneity (Rossi, 2014). Furthermore, a common estimation technique used to incorporate IVs (the generalized method of moments, or GMM) is known to be inefficient, and a large number of instruments are needed to obtain statistically significant coefficient estimates (Wooldridge, 2010).

Even if valid instruments can be specified, asymptotically biased coefficients do not necessarily mean a model will predict poorly (Gigerenzer & Brighton, 2009). Particularly, biased models may predict future choices well in markets that have persistent patterns of endogeneity because bias may carry forward information about the connection between observed and omitted variables— a useful feature when the relationships that hold in the estimation data are also likely to hold in the future (Haaf et al., 2014). In the case of automotive demand modeling, a key source of endogeneity comes from strategic pricing by firms that observe and account for demand for vehicle aspects unobserved by the modeler. It is thus plausible that price-ASC correlation persists in future markets, absent dramatic changes in market structure, raising doubts as to whether correcting for endogeneity is needed for forecast purposes. Unbiased coefficients are often expected to predict better in the event of a structural market shift, for example, when a correlation with an unobserved variable in the estimation data changes in the prediction data. Whether or not unbiased coefficients successfully obtained from instrumental variables estimation will yield better or worse predictions than the biased MLE-C coefficients will be dependent upon the structure of past and future markets.

In this work, we aim to answer the following three questions:

(Q1) Should modelers address the potential endogeneity between price and omitted variables when forecasting new vehicle market shares?

For contexts in which the source of endogeneity is likely to persist — e.g. mature product markets with relatively stable consumer preferences like the automotive market — forecasting models may be well served by calibration constants free from the problems with specifying valid IVs. The main drawback of post-hoc ASC calibration is that it lacks the coherent theoretical grounding of the IV approach: the effect of missing attributes on choice is attributed to observed attributes when fitting the model, and the ASCs are added post hoc, outside a formal inferential framework, to match observed shares. Forcing model predictions to match observations *exactly* guarantees overfitting, which can degrade forecast quality. This suggests that use of ASCs as calibration constants in forecasts should be approached with caution and motivates our second research question:

(Q2) Can ASCs improve forecast accuracy?

Several studies implement methods to predict ASCs for out-of-sample alternatives (Greene et al., 2004; Berry et al., 2004; Bunch et al., 2011; Train & Winston, 2007), but neither the predictive nor explanatory literature suggests how to predict future values of the ASCs for products that do not appear in the estimation data. Furthermore, these prior studies model predicted ASCs as point estimates without characterizing the implications of uncertainty in the value of ASCs for future alternatives. We propose four methods for forecasting ASCs and evaluate the resulting accuracy and uncertainty of predicted shares in a synthetic study and in a case study using US sales of midsize sedan vehicles, when ASCs are included in the utility function in order to investigate our third research question:

(Q3) Should estimates of past ASCs be used in future share forecasts?

We are interested in the implications of estimation and prediction techniques employed in the literature on forecast accuracy and uncertainty. Much of the literature relies on these models to inform analysis of vehicle design and policy, and these types of analyses often require forecasting. We compare alternative methods of using past ASC estimates in future forecasts.

The remainder of this paper is structured as follows: Section 2 reviews the literature on DCMs that employ ASCs with a focus on the automotive demand literature. Section 3 describes model specifications and estimation techniques. Section 4 outlines the generation of a synthetic data set and the methods used to predict future ASCs as well as estimation and prediction results for the synthetic data set. Section 5 extends the analyses to a case study on US sales of midsize sedan for 2002-2006, 2007, and 2011. Section 6 discusses the studies' results and challenges posed by estimating a model with IVs. Section 7 addresses limitations, and Section 8 concludes.

2. Literature review

Table 1 compares automotive demand studies that use DCMs with ASCs and are estimated in a classical (as opposed to Bayesian) framework. The explanatory literature is composed of studies that introduce or evaluate model estimation techniques, estimate coefficients to describe consumer preferences or firm behavior, or test counterfactual policy scenarios, and they often report willingness-to-pay for or willingness-to-accept different vehicle attributes. For some of these purposes no predictions are made, but counterfactuals are simulated; such simulations are typically in-sample. For such analyses there is often no need to determine ASCs for new product entrants. (Berry et al. (2004) and Train and Winston (2007) are exceptions.) The forecast literature is composed of studies that forecast future market shares or test counterfactual scenarios. Reviewing the literature summarized in Table 1 a major methodological distinction between the two bodies of literature emerges. The explanatory literature estimates models by formal, econometric methods that often use IVs to obtain consistent coefficients, reducing or eliminating (asymptotic) coefficient bias and estimating ASCs concurrently with observed variable coefficients. In contrast, the forecast literature relies on expert opinion or historical coefficient estimates, calibrates ASCs outside the formal statistical framework, and does not take steps to correct for endogeneity.

Table 1
Automotive demand literature that includes an alternative-specific constant in the discrete choice model.

Study	Year	Specification[a,b]	Estimation[c]	Instruments[d]	Model purpose: Estimation technique or model proposal / investigation	Attribute valuation / market description	Counter-factual / policy evaluation	Forecast future market shares
Explanatory literature— ASC is estimated simultaneously with taste parameters								
Knittel and Metaxoglou (2012)	2013	Mixed logit	Two stage GMM	BLP	x			
Allcott and Wozny (2014)	2012	Nested logit	NFP + 2SLS	Fuel price[e]		x		
Dubé et al. (2012)	2012	Mixed logit	Two stage GMM, GMM MPEC	Synthetic[f]	x			
Klier and Linn (2012)	2012	Nested logit	2SLS	Engine atts.[g]			x	
Beresteanu and Li (2011)	2011	Mixed logit	Two stage GMM	Fuel price[e]			x	
Copeland et al. (2011)	2011	Mixed logit	Two stage GMM	BLP		x		
Li et al. (2011)	2011	Logit	NFP + 2SLS	BLP		x		
Frischknecht et al. (2010)	2010	Mixed logit	Two stage MLE	BLP	x			
Vance and Mehlin (2009)	2009	Nested logit	NFP + 2SLS	BLP			x	
Dasgupta et al. (2007)	2007	Nested logit	MLE	None		x		
Train and Winston (2007)	2007	Mixed logit	Two stage MLE	BLP		x		
Berry et al. (2004)	2004	Mixed logit	Two stage GMM[h], Expert elicitation	Price makeup /None	x			
Petrin (2002)	2002	Mixed logit	Two stage GMM	BLP	x			
Sudhir (2001)	2001	Mixed logit	Two stage GMM	BLP		x		
Berry et al. (1999)	1999	Mixed logit	Two stage GMM	BLP			x	
Berry et al. (1995)	1995	Mixed logit	Two stage GMM	BLP	x			
Lave and Train (1979)	1979	Logit	MLE	None		x		
Predictive literature— ASC is calibrated post-estimation of model taste parameters								
Whitefoot and Skerlos (2012)	2012	Logit	Literature informed	None			x	
Bunch et al. (2011)	2011	Nested logit	Canned software	None			x	
U.S. Energy Information Administration (2011)	2011	Nested logit	Expert elicitation	None				x
Greene et al. (2005)	2005	Nested logit	Expert elicitation	None				x
Greene et al. (2004)	2004	Nested logit	Expert elicitation	None				x
Unknown/other								
Choo and Mokhtarian (2004)	2004	Logit	Canned software	None		x		

[a] More than one model may be specified, model listed is the discrete choice model or study focus relevant to this work.
[b] Mixed logit models assume independent random coefficients in all studies listed, logit is multinomial logit.
[c] GMM = generalized method of moments, MLE = maximum likelihood estimation, 2SLS = two staged least squares, NFP = nested fixed point, IV = instrumental variable.
[d] "BLP" refers to the instruments used in Berry et al. (1995) or a similar variant.
[e] Allcott and Wozny (2014) use the vehicle's expected lifetime fuel costs (applicable when used vehicle sales are included in the model) and Beresteanu and Li (2011) use fuel costs in other Metropolitan Statistical Areas (MSAs).
[f] Synthetic data study, instruments are generated.
[g] Mean of engine characteristics of vehicles with same engine but of other vehicle class.
[h] First stage of GMM inverts shares to obtain mean utility, three methods for mean utility parameters: expert elicitation, IV regression, regression assuming no endogeneity.

While there are many reasons to include ASCs in DCMs, they do present a practical problem in forecasting. Typically, researchers will assume that estimated values of the ASCs for a particular vehicle model carry forward to future years, but when the forecast market includes new vehicle entrants, an assumption must be made about the value of the ASC for each entrant. We review new vehicle ASC forecasting methods from the literature and use these to inform our proposed methods in Sections 4 and 5. Four of the studies in Table 1 predict out-of-sample or new vehicle shares for which unknown ASCs must be generated or assumed: Berry et al. (2004) predict in-sample shares for a counterfactual scenario, but they introduce two new vehicles into the data set. For these entrants, the ASCs are generated by averaging the estimated ASCs of the respective brand and class of each entrant. Train and Winston (2007) forecast out-of-sample future market shares. They hold the ASCs of

the observed products constant from the calibrated value and, similarly to Berry et al. (2004), generate new product ASCs by averaging over the estimated ASCs of the same class. Greene et al. (2004) forecast out-of-sample shares for future markets in which the vehicle set is modified by introducing diesel and hybrid versions of vehicles in the estimation data set. They assume diesel and hybrid vehicles have identical ASCs to their conventional counterparts and hold the ASCs constant at the estimated values. Bunch et al. (2011) forecast out-of-sample future market shares under alternative policies and assume that the vehicle set is unchanged from the estimation data. Similarly to Greene et al. (2004), they hold future ASCs constant at the estimated values.

One of our model estimation techniques described in Section 3.2 is based on the two-stage generalized method of moments with instrumental variables (GMM-IV) implemented by Berry et al. (1995) that is commonly referred to as "BLP". Berry et al. (1995) develop a technique to estimate joint models of supply and demand that include ASCs and IVs in a nonlinear DCM framework. This canonical study informs many of the studies in Table 1 (Allcott & Wozny, 2014; Copeland et al., 2011; Li et al., 2011; Sudhir, 2001; Berry et al., 1999; Berry et al., 2004; Beresteanu & Li, 2011; Petrin, 2002; Dubé et al., 2012; Knittel & Metaxoglou, 2012; Train & Winston, 2007; Vance & Mehlin, 2009) and Knittel and Metaxoglou (2012) list BLP-type demand models in addition to the vehicle demand-focused studies included here. We isolate the demand side model and ignore the supply side as described in Nevo (2000). Several recent studies have focused on the difficulties of estimating models with the BLP method. Knittel and Metaxoglou (2012) find that for several combinations of algorithms and starting values the estimation routine results in specious convergence and convergence to inferior local minima; moreover they find that different local GMM-IV solutions have significantly different economic implications. Dubé et al. (2012) find that loose convergence tolerance criteria on BLP's nested fixed point iteration exacerbate coefficient bias and that reasonable estimation times often require a convergence criterion too loose for good estimates. They propose an alternative nonlinear programming formulation that we adopt here. Similarly Su and Judd (2012) propose an alternative formulation of the BLP-style estimation procedure for more general strategic market models to avoid nested fixed point iteration. We discuss some other difficulties vehicle demand researchers are likely to encounter in Appendix A.

No studies in Table 1 conduct a formal statistical test regarding the appropriateness of the IVs, but two of the studies qualitatively address it. Allcott and Wozny (2014) verify that their instruments should not be included as explanatory variables in the utility function, and Li et al. (2011) estimate a model that implies inelastic price elasticities, suggesting that their BLP-like instruments are invalid. Though not included in Table 1 because the model is a linear regression (as opposed to a DCM), Jenn et al. (2013) examine the effects of federal policies on hybrid vehicle sales and conduct a J Hansen test to verify that the model is not overspecified.

Most of the studies in Table 1 estimate models on aggregate data only. As exceptions, (Berry et al., 2004; Beresteanu & Li, 2011; Petrin, 2002; Bunch et al., 2011) combine aggregate and disaggregate data, and (Dasgupta et al., 2007; Lave & Train, 1979; Train & Winston, 2007; Choo & Mokhtarian, 2004) use disaggregate, household data only.

3. Methods

We are primarily interested in comparing the accuracy and uncertainty associated with forecasting using a random coefficients logit model with independent and normally distributed coefficients ("mixed logit" here for brevity) when potential endogeneity due to the correlation of the ASC with price is mitigated using IVs vs. when it is ignored. We estimate choice models using both maximum likelihood estimation with calibration (MLE-C) and GMM-IV methods, and we compare the out-of-sample market share predictions resulting from the two estimation methods under several proposed techniques for forecasting the new market ASCs.

3.1. Choice model

The assumed utility function is linear in coefficients and includes an ASC ξ:

$$u_{ijt} = \mathbf{x}'_{jt}\boldsymbol{\beta}_i + \xi_{jt} + \varepsilon_{ijt} \tag{1}$$

where u_{ijt} is the utility consumer i derives from product j in market (year) t, $\mathbf{x}_{jt}$ is a vector of attributes specific to product j in market t, $\boldsymbol{\beta}_i$ is a vector of taste parameters for consumer i, ξ_{jt} is the ASC for product j in market t, and ε_{ijt} is an idiosyncratic error term. Note that we use the terms "market" and "year" interchangeably, although they have different implications in some contexts. We assume the attribute vector $\mathbf{x}_{jt}$ for a given product is similar but not necessarily constant over time or across markets. For example, the price or weight of a Ford Focus may vary slightly from year to year, though it is still considered the same product with the same ASC[3].

We assume that the distribution of preference coefficients is constant in time (across markets) such that β is indexed only by i and not t. This is a standard assumption in the vehicle demand literature (see Axsen et al. (2009) for an exception). If the coefficients are assumed to be independently and normally distributed:

$$\boldsymbol{\beta}_i \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\beta\right) \tag{2}$$

[3] In the case study data the greatest year over year non-price attribute change is less than 15%.

where $\boldsymbol{\mu}$ is a $(K \times 1)$ vector and $\boldsymbol{\Sigma}_\beta$ is a $(K \times K)$ diagonal covariance matrix with $(K \times 1)$ diagonal element vector $\boldsymbol{\sigma}^2$, then the portion of the utility function dependent on product attributes can be expressed as the sum of the deterministic mean utility common to all consumers and the stochastic consumer-specific utility:

$$u_{ijt} = \left(\mathbf{x}'_{jt}\boldsymbol{\mu} + \xi_{jt}\right) + \left(\mathbf{x}'_{jt}(\boldsymbol{\sigma} \circ \boldsymbol{\nu}_i) + \varepsilon_{ijt}\right) \tag{3}$$

where $\boldsymbol{\nu}_i$ is a $(K \times 1)$ vector of independent standard normal random variables and the open circle $\circ$ is the Hadamard product (element-wise product). If ε_{ijt} is assumed to follow an independent and identically distributed (iid) extreme value type I distribution, the probability P_{ijt} of individual i selecting a product j in market t is then given by the logit probability:

$$P_{ijt} = \frac{\exp\left(\mathbf{x}'_{jt}\boldsymbol{\mu} + \xi_{jt} + \mathbf{x}'_{jt}(\boldsymbol{\sigma} \circ \boldsymbol{\nu}_i)\right)}{\sum_{k \in J_t} \exp\left(\mathbf{x}'_{kt}\boldsymbol{\mu} + \xi_{kt} + \mathbf{x}'_{kt}(\boldsymbol{\sigma} \circ \boldsymbol{\nu}_i)\right)} \tag{4}$$

where J is the set of distinct products observed across all markets and J_t is a subset of J containing the products that appear in market t (Train, 2009).[4] Our model considers only the set of consumers who purchase products, thus there is no outside good (option to not purchase any product). When an outside good is excluded, the choice probabilities (Eq. (4)) are invariant over uniform shifts in the ASCs. In order to enforce uniqueness of the ASCs we require throughout (and constrain in estimation) $\sum_{k \in J_t} \xi_{kt} = 0 \;\; \forall t$.

The share P_{jt} of product j in year t can be obtained by integrating over the consumer-specific stochastic utility:

$$P_{jt} = \int_{\mathbf{y}} \frac{\exp\left(\mathbf{x}'_{jt}\boldsymbol{\mu} + \xi_{jt} + \mathbf{x}'_{jt}(\boldsymbol{\sigma} \circ \boldsymbol{\nu}_i)\right)}{\sum_{k \in J_t} \exp\left(\mathbf{x}'_{kt}\boldsymbol{\mu} + \xi_{kt} + \mathbf{x}'_{kt}(\boldsymbol{\sigma} \circ \boldsymbol{\nu}_i)\right)} f_{\boldsymbol{\nu}}(\mathbf{y}) d\mathbf{y} \tag{5}$$

where $f_{\boldsymbol{\nu}}(\mathbf{y})$ is the multivariate standard normal probability density function and the integral is a K-dimensional integral. We approximate this integral using numerical integration with sample averages (Train, 2009). One hundred Halton draws of $\boldsymbol{\nu}$ are used and held constant throughout estimation. As discussed in Dubé et al. (2012), other methods are available, but they are not the focus of this work.

3.2. Estimation

The choice of estimator is determined almost entirely by the assumptions regarding the endogeneity of price and the ASC. If it is assumed that price is exogenous (and there is no model misspecification), then estimating the model in Eq. (5) *excluding* the ASC via MLE yields a consistent price coefficient estimate. This is a standard result in discrete choice model theory (Train, 2009) (though one of limited applicability, as several decades of econometric research has been motivated by the rejection of exogenous price, and no discrete choice model can reasonably be expected to be correctly specified for a complex market like the new vehicle market). Calibrating the ASCs post-hoc (MLE-C) represents a portion of the error term in the MLE model.

However, if it is assumed that price is endogenous, then GMM-IV is the estimation strategy preferred in the literature for incorporating IVs into the estimation routine in order to mitigate the potential price coefficient bias[5]. The presence of endogeneity (or assumption thereof) does not itself determine the estimator, but rather the techniques used to address endogeneity — in this case IVs — suggest one estimation method over the other. Note that it is possible that unobserved attributes are correlated with other observed attributes besides price, but we follow the prior literature here in considering only potential correlation with price.

3.2.1. Maximum likelihood with post hoc calibration

With the MLE-C approach, the likelihood at the estimated parameters L is defined as the probability of generating the observed data given the estimated parameter values:

$$L\left(\widehat{\boldsymbol{\beta}}|\mathbf{X}\right) = \prod_{t=1}^{T}\left(\prod_{k \in J_t}\left(P^-_{kt}\right)^{n_{kt}}\right) \tag{6}$$

where $\mathbf{X}$ is the $(V \times K)$ stacked matrix of transposed attribute vectors $\mathbf{x}_{jt}$ for all products in all markets, n_{kt} is the observed sales of product k in time t, T is the total number of markets, and P^-_{kt} is P_{kt} in Eq. (5) modified to exclude ξ_{jt}. The MLE estimator of the parameters $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\mu}}', \widehat{\boldsymbol{\sigma}}']'$ is the value of the $(2K \times 1)$ vector that maximizes L. The monotonic transformation $\ln(L)$ is typically used as the objective function for computational benefit. The ASCs are "calibrated" (see (Train, 2009) section 2.8) post-hoc by solving the system of equations:

$$\begin{aligned} &\ln\left(P_{jt}\left(\boldsymbol{\xi}_t|\mathbf{X}_t, \widehat{\boldsymbol{\beta}}\right)\right) = \ln\left(s_{jt}\right), \;\; \forall j \in J^-_t \\ &\sum_{k \in J_t} \xi_{kt} = 0, \quad \forall t \end{aligned} \tag{7}$$

[4] We follow the bulk of the literature in assuming that each consumer considers all vehicles in the market and makes a compensatory decision. Alternative approaches that model consideration sets are also possible (Morrow et al., 2012; Gilbride. & Allenby, 2004; Hauser & Wernerfelt, 1990), but we do not pursue them here.

[5] MLE methods can be used to incorporate IVs, however, GMM-IV is the most common estimation approach. Train and Winston (2007) is an exception. Park & Gupta (2011) propose a method of estimating observed coefficient parameters and the ASC simultaneously using MLE but do not use automotive data.

where $\boldsymbol{\xi}_t$ is the stacked vector of all ξ_{jt} in market t, $\mathbf{X}_t$ is the matrix of product attributes for all products in market t, s_{jt} is the observed share of product j in market t, P_{jt} includes the ASC as written in Eq. (5), and J_t^- is the set of products in market t excluding one[6]. We solve Eq. (7) using the popular fixed-point iteration routine originally proposed by BLP for computing ASCs that are globally convergent. For more detail on mixed logit models and MLE-C see Train (2009) chapters 3 and 6.

3.2.2. Generalized method of moments with instrumental variables

For GMM-IV, following BLP (Berry et al., 1995) we specify the moment conditions:

$$E\left[\xi_{jt}|\mathbf{z}_{jt}\right] = 0, \quad \forall j, t \tag{8}$$

where $\mathbf{z}_{jt}$ is an $(L \times 1)$ vector of instruments for product j in market t and $L \geq 2K$ is a necessary condition for identification (Wooldridge, 2006). The choice of these instruments is determined by the modeler and is specific to the data. We describe the instruments used for the synthetic data example and the case study in their respective sections.

Define $\widetilde{\boldsymbol{\beta}} = [\widetilde{\boldsymbol{\mu}}', \widetilde{\boldsymbol{\sigma}}']'$ and $\boldsymbol{\xi}$ as the stacked vector of ξ_{jt} for all products and markets. The GMM-IV estimator of the parameters is the value of the vector $[\widetilde{\boldsymbol{\beta}}', \widetilde{\boldsymbol{\xi}}']'$ that solves:

$$\begin{aligned} &\underset{\boldsymbol{\beta},\boldsymbol{\xi}}{\text{minimize}}\ (\mathbf{Z}'\boldsymbol{\xi})'\mathbf{W}(\mathbf{Z}'\boldsymbol{\xi})/T \\ &\text{subject to } \ln(P_{jt}(\boldsymbol{\beta}, \boldsymbol{\xi}_t)) = \ln(s_{jt}), \quad \forall j \in J_t^-, t \\ &\sum_{k \in J_t} \xi_{kt} = 0, \quad \forall t. \end{aligned} \tag{9}$$

where $\mathbf{Z}$ is the $(V \times L)$ matrix of stacked vectors for all $\mathbf{z}'_{jt}$, $\mathbf{W}$ is an $(L \times L)$ weighting matrix, T is the total number of markets, and P_{jt} is defined in Eq. (5). We set $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ as used by Dubé et al. (2012)[7]. Eq. (9) was not literally proposed by BLP (Berry et al., 1995), but it is a more recent interpretation of the model that has better statistical and computational properties (Dubé et al., 2012; Su &Judd, 2012). We further transform Eq. (9) to improve computational properties. See Appendix B for explicit formulation of the optimization problem provided to the solver. For more information on GMM estimation and IVs see Wooldridge (Wooldridge, 2006).

4. Simulation study

We begin our investigation of forecasting with ASCs using synthetic data sets. This allows us to evaluate the statistical properties of the predictions in a controlled setting for comparing model coefficients and predictions when the true parameters are known. We generate five years of estimation data and two years of prediction data representing 1-year- and 5-year-forward forecast horizons. Each year of data has 70 products with 40% year over year turnover (40% of products each year are newly introduced replacing 40% from the previous year). These parameters were chosen to closely mirror the structure of the real data in the case study. The estimation data is generated with low, base, and high price-ASC endogeneity and a mixed logit model is estimated by MLE-C and GMM-IV for each case using both valid and invalid instruments. The potential level of endogeneity and the validity of the instruments are both unknowns for real forecasting exercises, and we would like to know their possible implications. The prediction data is generated under two market futures — one for which the source of price-ASC endogeneity persists and one in which it does not. This is again unknown in forecasting. We compare the accuracy and uncertainty of predictions made by the MLE-C and GMM-IV estimated models.

4.1. Synthetic data generation

Mimicking data of a typical data set used to estimate automotive demand models, an initial market is assigned 70 products, each with a randomly generated price attribute p, a single non-price or "technology" attribute x (for simplicity), and an unobserved contribution to utility ξ that is correlated with prices. Instruments z are generated simultaneously with the price and technology attributes so that we are able to specify any arbitrary correlation between them.

The product attributes and instruments for year $t = 1$ are generated by drawing from a multivariate normal distribution:

$$\left[p_{jt}, x_{jt}, \boldsymbol{\xi}_{jt}, z_{jt}^{(1)}, z_{jt}^{(2)}, z_{jt}^{(3)}\right]' \sim N(0, \boldsymbol{\Sigma}_{\mathbf{x}}), \quad \forall j, t \tag{10}$$

Exogenous attributes (meaning those uncorrelated with the ASC) and functions of the instruments in the case of nonlinear models can also be used as instruments (Rossi, 2014). We adopt this approach and generate additional instruments:

$$\begin{aligned} \mathbf{z}_{jt}^* &= \left[z_{jt}^{(1)}, z_{jt}^{(2)}, z_{jt}^{(3)}\right] \\ \mathbf{z}_{jt} &= \left[\mathbf{z}_{jt}^*,\ \left(\mathbf{z}_{jt}^* \circ \mathbf{z}_{jt}^*\right),\ \left(\mathbf{z}_{jt}^* \circ \mathbf{z}_{jt}^* \circ \mathbf{z}_{jt}^*\right),\ \mathbf{z}_{jt}^* x_{jt},\ \mathbf{z}_{jt}^* x_{jt}^2,\ \left(\mathbf{z}_{jt}^* x_{jt}\right) \circ \left(\mathbf{z}_{jt}^* x_{jt}\right)\right] \end{aligned} \tag{11}$$

[6] The exclusion of one ASC from each market is for computational purposes. The selection of which ASC to exclude (set to zero) in each market is arbitrary.

[7] Other choices of $\mathbf{W}$ or more instruments can improve the statistical efficiency of the estimator (Wooldridge, 2006), though alternative formulations may be more computationally burdensome (Dubé et al., 2012). Nevo (2000) argues in favor of the less efficient but less statistically burdensome weighting matrix used here for BLP type models. For logit and nested logit models with homoskedastic errors $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ is the optimal weighting matrix (Nevo, 2000)

Increasing numbers of instruments improve the efficiency of the estimator so that coefficient estimates are more likely to be statistically significant for a given data set. We generate the additional instruments as in Eq. (11) rather than drawing from the distribution of Eq. (10) for numerical reasons discussed following Eq. (13).

Negative values of price are drawn under the specification in Eq. (10), but since only differences in utility between products (as opposed to absolute utility) affect the (mixed) logit probabilities when no outside good is included in the model, all prices could be trivially shifted upward by a constant so long as the population taste parameters are independent. We specify the data covariance matrix $\mathbf{\Sigma}_x$ to have the structure:

$$\Sigma_x = \begin{bmatrix} 1 & \rho_x & \rho_\xi & \rho_i & \rho_i & \rho_i \\ \rho_x & 1 & 0 & 0 & 0 & 0 \\ \rho_\xi & 0 & 1 & \rho_z & \rho_z & \rho_z \\ \rho_i & 0 & \rho_z & 1 & 0 & 0 \\ \rho_i & 0 & \rho_z & 0 & 1 & 0 \\ \rho_i & 0 & \rho_z & 0 & 0 & 1 \end{bmatrix} \tag{12}$$

where ρ_ξ is the price-ASC correlation that determines the presence and magnitude of price endogeneity, ρ_x is the correlation of price with the technology attribute, ρ_i is the correlation of price with each of the instruments, and ρ_z is the correlation of the ASC with each of the instruments ("IV-ASC correlation"). The covariance matrix $\mathbf{\Sigma}_x$ is by definition positive semi-definite and symmetric. We examine cases under the following correlations:

$$\begin{aligned} &\rho_\xi = 0.1 \textit{ or } 0.4 \textit{ or } 0.7 \\ &\rho_z = 0 \textit{ or } 0.4 \quad \rho_x = 0.1, \quad \rho_i = 0.4. \end{aligned} \tag{13}$$

The three levels of ρ_ξ represent the low, base, and high endogeneity cases respectively, and the two levels of ρ_z represent the valid and invalid instrument cases respectively (specifying ρ_z equal to 0.4 results in invalid instruments because valid instruments must be uncorrelated with the ASC).

Market shares are simulated for $t = 1$ according to Eq. (5) with the specified values of the taste parameters:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_p \\ \mu_x \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\sigma} = \begin{bmatrix} \sigma_p \\ \sigma_x \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{14}$$

where μ_p and σ_p are the population mean and standard deviation of the price coefficient and μ_x and σ_x are the population mean and standard deviation of the technology attribute coefficients.

In years $t = 2, 3, 4, 5$ we randomly select 28 (40%) of the products from the prior year to be replaced and generate new product attribute, instrument, and ASC values as described for the base year $t = 1$. Years $t = 1, 2, 3, 4, 5$ comprise the estimation data set and years $t = 6,11$ are the 1-year-forward and 5-year-forward prediction data sets, respectively. Note that the generated ASC for a given product is held constant over all estimation and prediction data, consistent with the idea that it is an aggregate measure of unobserved attributes (and assuming those unobserved attributes and the preferences for them do not change over time for the same vehicle model), but in our estimation procedures a year-specific ASC is estimated so that there are sufficient degrees of freedom to constrain predicted shares of the estimation data to equal observed shares exactly. The price and technology attribute for a given product are also held constant year over year.

4.2. Estimation

To assess performance across multiple data sets, we estimate independent mixed logit models for 125 synthetic data sets generated by the process described in Section 3 for each pairing of price-ASC and IV-ASC correlation levels. There are a total of six estimation cases (price correlation $\in$ {10% (low), 40% (base), 70% (high)} and instrument correlation $\in$ {0% (valid) and 40% (invalid)}). For a given estimation case, the expected value of the coefficients is approximated by averaging over the 125 estimates. A coefficient's bias is the deviation of its expected value (obtained here by estimating the model repeatedly for different simulated data sets) from the population parameters. The estimation routine successfully converged (exit flag of 0 using the Knitro solver for Matlab) for all 125 data sets for both the MLE-C and GMM-IV estimation methods.

Fig. 1. compares the distributions of estimates of the mean taste parameter price μ_p obtained by MLE-C estimation and GMM-IV estimation with valid and invalid instruments for the three levels of price-ASC endogeneity. The distributions represent the coefficients resulting from models fit to the 125 data sets excluding the least and greatest 2.5% of estimates of each coefficient (120 total coefficient estimates are included in each box plot). The expected values of the estimated coefficient values are indicated by *x*'s, and the coefficient values from Eq. (14) used to generate the data ("true coefficients") are represented by horizontal black lines. The distance between an *x* and a black line is the finite sample coefficient bias. The estimates of the mean taste parameter μ_x and the standard deviation of the taste parameters $\boldsymbol{\sigma}$ are presented in Appendix C.

The bias of the price coefficient increases with the degree of endogeneity for the MLE-C estimator. The valid GMM-IV estimator successfully corrects for endogeneity (Fig. 1) as expected, but the invalid GMM-IV estimator is more biased than the MLE-C estimator (this effect is also observed in a study by Andrews and Ebbes (2014)). The MLE-C estimates on average have a standard error on the order of 10-4, and the invalid GMM-IV estimates on average have standard error on the order of 1 (Fig. 1 shows the distribution of the expected value across multiple draws of the data – standard error of those estimates is not shown). So while

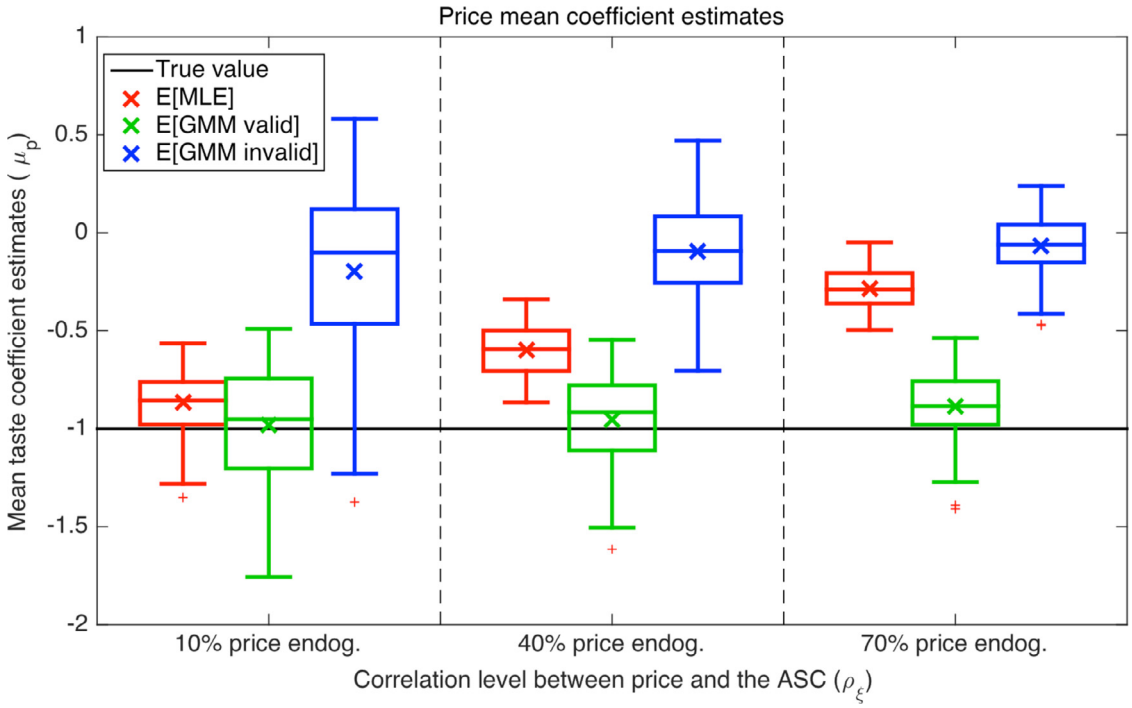


Fig. 1. The MLE-C estimator exhibits bias that increases with price correlation. The GMM-IV estimator mitigates the endogeneity bias when IVs are valid but exacerbates the bias when IVs are invalid.

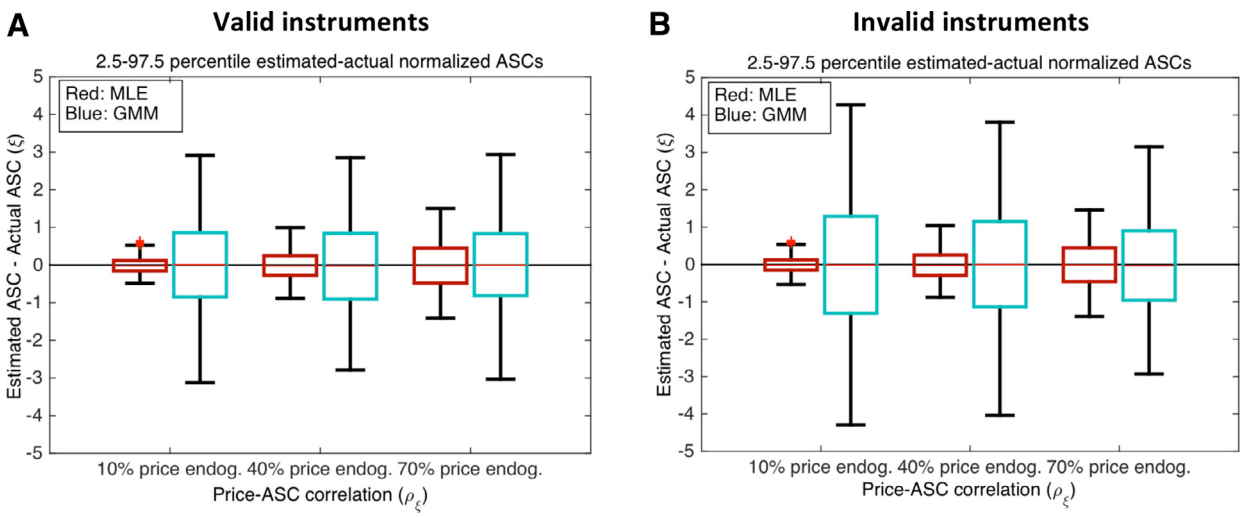


Fig. 2. The GMM-IV estimator exhibits greater ASC error variance than the MLE-C estimator at low levels of correlation, but the difference shrinks at greater price correlation levels when both valid (A) and invalid (B) instruments are used in GMM-IV estimation.

the biased MLE-C price parameter estimates are less biased than those of the GMM-IV estimates, the confidence intervals from MLE-C may be less likely to contain the truth (i.e. have worse coverage).

We are interested in the accuracy of the ASCs in addition to the accuracy of the observed attribute coefficients because the ASCs may be used in counterfactuals or for forecasting. Fig. 2 shows the distribution of error as the difference between the value of the estimated ASCs vs. true ASCs obtained from MLE-C and from GMM-IV estimation with valid vs. invalid instruments. The box plots in Fig. 2 include the ASC error from 125 estimated data sets for the valid instrument case excluding the least and greatest 2.5% of error differences. While both methods produce unbiased estimates of the ASC, the MLE estimator is more efficient,

Table 2
Description of ASC forecasting methods.

Method name	Method qualitative description	Method mathematical description
Products appearing in the estimation set (incumbents)		
Incumbent	For each product draw uniformly from the product's estimated ASCs	Draw ξ_{j6} or ξ_{j11} uniformly from $\{\xi_{j1=1}, \xi_{j2}, \xi_{j3}, \xi_{j4}, \xi_{j5}\}$
Products not appearing in estimation set (entrants)		
All (Method 1)	Draw uniformly from all estimated ASCs	Draw ξ_{j6} or ξ_{j11} uniformly from $\{\xi_{kt} \forall k \in J_t, t = 1, 2, 3, 4, 5\}$
Nearest neighbor (Method 2)	For each new product calculate the normalized vector distance of product observed attributes between the new product and each of the estimation data set products. Draw uniformly from the estimated ASCs of the observed product with the smallest vector distance ("nearest neighbor")	Draw ξ_{j6} or ξ_{j11} uniformly from $\{\xi_{k*1}, \xi_{k*2}, \xi_{k*3}, \xi_{k*4}, \xi_{k*5}\}$ where $k^* = argmin_{k \in J_t, t=1,2,3,4,5}(\mathbf{x}^*_{jt} - \mathbf{x}^*_{kt})$ and $\mathbf{x}^*_{jt}$ is the $(K \times 1)$ vector of normalized observed product attributes

meaning that for any individual data set, the MLE estimator is likely to be closer to the truth than the GMM-IV estimator[8]. The relative efficiency of the MLE estimator as compared to the GMM-IV estimator diminishes as the endogeneity increases, but it is still more efficient even at 70% price endogeneity. This is anticipated as MLE-C utilizes all of the variation in the data, whereas GMM-IV does not since the IVs are used to partition the data into that which is correlated with the unobserved attributes and that which is not (Rossi, 2014).

4.3. Prediction

The results from the previous section indicate that (1) the price coefficient estimated by MLE-C is inconsistent when price is correlated with the ASC; (2) GMM-IV with valid instruments corrects for endogeneity and provides better estimates for the price coefficient but with invalid instruments exacerbates endogeneity bias; and (3) there is more variation in the observed coefficient and ASC GMM-IV estimates. We examine the implications of these differences on forecast accuracy by comparing the predictions of the estimated models. If relationships that induce bias in the estimated coefficients persist in the prediction data, then models with biased coefficients need not predict any worse than models with unbiased coefficient estimates (Haaf et al., 2014). However, if there are fundamental shifts in the structure of the data, then models with biased coefficients may predict worse than models with unbiased coefficients.

We estimate the model in Eq. (5) on the estimation data set using MLE-C and GMM-IV. The models are then used to forecast product market shares one and five years forward (including new entrants in the prediction data that are not in the estimation data). The $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ parameters are assumed constant across markets, however, the $\widehat{\boldsymbol{\xi}}$ and $\widetilde{\boldsymbol{\xi}}$ parameters (the ASCs estimated by MLE-C and GMM-IV respectively) are market (year)-specific, and thus a method for determining their values in the prediction years is needed.

We ignore statistical uncertainty in the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ to focus on the larger source of uncertainty from forecasting values of $\boldsymbol{\xi}$ in future markets. Since the ASCs are unobserved, we treat the set of ASCs as a random vector and compare the distribution of share forecasts resulting from different realizations of the ASCs. We are interested in two questions:

1. Is there a particular method of generating prediction ASCs that is more robust than the others (in the case that the ASCs represent constant unobserved attributes)?
2. How do the models and ASC forecasting methods compare across various estimation and prediction data correlation structures?

4.3.1. Predicting ASCs

We propose two non-parametric distributions for generating ASCs in the prediction data, the second of which is based on prior literature (Berry et al., 2004). For "incumbent" products (those in the prediction data that also appear in the estimation data) there are historical estimates of the ASC that may provide information on potential future ASCs, if the ASCs do in fact represent unobserved attributes as they do in the synthetic data study. For new "entrant" products introduced in or before the prediction years but after the estimation data years, there are no such historical estimated ASCs. As a result, the ASC forecasting methods we propose differ for incumbents and entrants. Table 2 describes these methods. Method 2 would not necessarily be expected to predict ASCs well since we have explicitly generated data in which x_{jt} and ξ_{jt} are uncorrelated, so the technology attribute x should not contain any information about ξ. We include this method primarily for comparison to the case study with market data in Section 5.

[8] The slight differences in MLE-C ASC error distributions across the valid and invalid instrument cases are due to the difference in the data generation matrix $\boldsymbol{\Sigma}_x$ when $\xi_z = 0$ versus $\xi_z = 0.4$. When instruments are invalid there is less overall unexplained variation in the generated estimation data.

Table 3
BASE case results shown include the mean and std. dev. of the RAL of expected share across 125 data sets for MLE-C and GMM-IV models as well as the number of data sets for which the MLE-C or GMM-IV model had a greater respective RAL (# superior).

	1-year-forward			5-year-forward		
	None	All	Neigh.	None	All	Neigh.
Method:	0	1	2	0	1	2
10% price-ASC correlation (ρ_ξ)						
MLE-C	68%	85%	73%	68%	70%	53%
(Std. dev.)	*(7%)*	*(8%)*	*(12%)*	*(8%)*	*(9%)*	*(11%)*
# Superior	106	114	110	94	103	113
GMM-IV	65%	74%	59%	65%	65%	40%
(Std. dev.)	*(8%)*	*(12%)*	*(16%)*	*(9%)*	*(9%)*	*(14%)*
# Superior	19	11	15	31	22	12
Static	68%			42%		
No info	39%			39%		
40% price-ASC correlation (ρ_ξ)						
MLE-C	71%	86%	77%	71%	72%	55%
(Std. dev.)	*(7%)*	*(8%)*	*(10%)*	*(7%)*	*(7%)*	*(9%)*
# Superior	106	116	116	105	106	115
GMM-IV	66%	72%	60%	66%	66%	41%
(Std. dev.)	*(8%)*	*(14%)*	*(16%)*	*(7%)*	*(8%)*	*(13%)*
# Superior	19	9	9	20	19	10
Static	71%			48%		
No info	44%			45%		
70% price-ASC correlation (ρ_ξ)						
MLE-C	81%	91%	85%	80%	81%	69%
(Std. dev.)	*(5%)*	*(5%)*	*(6%)*	*(4%)*	*(4%)*	*(8%)*
# Superior	124	124	119	121	124	121
GMM-IV	72%	74%	68%	71%	71%	52%
(Std. dev.)	*(7%)*	*(14%)*	*(17%)*	*(8%)*	*(8%)*	*(15%)*
# Superior	1	1	6	4	1	4
Static	77%			56%		
No info	52%			54%		

4.3.2. Forecasting shares

We evolve each of the 125 estimation data sets once so that we have 125 pairs of estimation and prediction data sets. We make separate forecasts for 125 prediction data sets so that our prediction results are not sensitive to the random data generation process. For each of the 125 prediction data sets, we generate a Monte Carlo distribution of market share arising from the uncertainty of forecasting ASCs. One hundred vectors of ASCs for each prediction data set are drawn non-parametrically according to either method 1 or method 2 as described in Table 2. ASCs are necessarily drawn independently from one another since there is no available information about their potential correlation structure. For each draw of ASC we simulate a draw of share by integrating over taste heterogeneity as in Eq. (5). Thus for a given time frame (1-year- or 5-year-forward), estimation method (MLE-C or GMM-IV), and ASC forecasting method (method 1 or 2), 12,500 total share forecasts are made (125 prediction data sets $\times$ 100 ASC forecasts). We report the expected share and its standard deviation in each case.

4.3.3. Evaluating forecasts

We compare the accuracy of the predictions using the relative average likelihood (RAL). The RAL is a normalization (monotonic transformation) of the likelihood of the model share predictions L_P divided by the likelihood of an ideal model L_I that perfectly predicts the new shares:

$$\mathrm{RAL} = \frac{(L_P)^{1/N}}{(L_I)^{1/N}} \tag{15}$$

where N is the number of choices observed. When comparing two models on the same data set, the model with a larger RAL is more likely to generate the observed data. Using RAL instead of likelihood is important because markets that have more diffuse choice probabilities will necessarily have lower likelihoods of ideal prediction. RAL normalizes for this effect and can be interpreted as the fraction of the total possible explanatory power a model obtains. RAL is a monotonic transformation of the Kullback – Leibler (KL) divergence if observed shares are assumed equal to choice probabilities (Kullback & Leibler, 1951).

Tables 3 and 4 contain the mean of the RAL of expected shares across the 125 prediction data sets calculated as:

Table 4
MARKET SHIFT case results shown include the mean and std. dev. of the RAL of expected share across 125 data sets for MLE-C and GMM-IV models as well as the number of data sets for which the MLE-C or GMM-IV model had a greater respective RAL (# superior).

	1-year-forward			5-year-forward		
Method:	None 0	All 1	Neigh. 2	None 0	All 1	Neigh. 2
10% price-ASC correlation (ρ_ξ)						
MLE-C	69%	84%	74%	68%	69%	51%
(Std. dev.)	*(7%)*	*(8%)*	*(10%)*	*(7%)*	*(8%)*	*(10%)*
# Superior	102	112	110	87	87	116
GMM-IV	66%	75%	62%	66%	66%	39%
(Std. dev.)	*(8%)*	*(12%)*	*(15%)*	*(9%)*	*(9%)*	*(13%)*
# Superior	23	13	15	38	38	9
Static	66%			40%		
No info	38%			37%		
40% price-ASC correlation (ρ_ξ)						
MLE-C	68%	81%	72%	66%	67%	50%
(Std. dev.)	*(8%)*	*(9%)*	*(11%)*	*(7%)*	*(8%)*	*(11%)*
# Superior	88	95	116	68	76	114
GMM-IV	66%	73%	56%	65%	64%	38%
(Std. dev.)	*(9%)*	*(13%)*	*(16%)*	*(8%)*	*(9%)*	*(13%)*
# Superior	37	30	9	57	49	11
Static	62%			39%		
No info	40%			37%		
70% price-ASC correlation (ρ_ξ)						
MLE-C	67%	74%	68%	62%	62%	52%
(Std. dev.)	*(9%)*	*(11%)*	*(12%)*	*(8%)*	*(8%)*	*(10%)*
# Superior	85	56	108	40	34	112
GMM-IV	65%	73%	57%	65%	65%	41%
(Std. dev.)	*(9%)*	*(13%)*	*(17%)*	*(9%)*	*(9%)*	*(14%)*
# Superior	40	69	17	85	91	13
Static	58%			39%		
No info	42%			37%		

$$E_d\left[\mathrm{RAL}\left(E_a\left[\mathrm{P}\left(\boldsymbol{\beta}, \boldsymbol{\xi}\right)\right]\right)\right] = \frac{1}{D}\sum_{d=1}^{D}\mathrm{RAL}\left(\frac{1}{A}\sum_{a=1}^{A}\mathrm{P}\left(\boldsymbol{\beta}_d, \boldsymbol{\xi}_{ad}\right)\right) \tag{16}$$

where d indexes data sets, a indexes ASC draws, $D = 125$, and $A = 100$. Also included are the standard deviation of the RAL over the 125 data sets and a measure "# superior" that indicates the number of paired data sets for which the respective MLE-C or GMM-IV model had a greater RAL of expected share.

In addition to predictions from the models using ASC methods 1 and 2, the tables include predictions for which ASCs are assumed equal for either incumbents or entrants (method 0) during prediction (despite being estimated for GMM-IV estimation), a "static" model that holds shares of observed products constant from the last year of the estimation data and divides the remaining share equally among the new products, and a "no info" model that assumes all entrant and incumbent products have equal shares in the prediction years. Shaded boxes indicate the greatest RAL ("best" model) when comparing across models for a given estimation data price correlation and time horizon. For parsimony we refer to models estimated using MLE-C and GMM-IV as "MLE-C models" and "GMM-IV models" respectively, though MLE-C and GMM-IV are estimation techniques not models in and of themselves.

4.3.4. Tested cases

We test four prediction cases:

- Prediction case 1 ("Base", Table 3): the GMM-IV estimated model uses valid instruments and the price-ASC endogeneity present in the estimation data persists in the prediction data.
- Prediction case 2 ("Market shift", Table 4): the price correlation ρ_ξ is set to 0.1, 0.4, or 0.7 when generating the *estimation* data, but it is set to 0 when generating the *prediction* data so that price is no longer endogenous in the forecast years and the source of asymptotic coefficient bias disappears.
- Prediction case 3 ("Invalid", Appendix D): the GMM-IV estimated model uses invalid instruments and price-ASC endogeneity persists.
- Prediction case 4 ("Entrants", Appendix D): the shares are forecasted for all products, but shares of the incumbent products are included as a single lump sum in the RAL calculation.

In the following discussion, we remark upon mean RAL comparisons between models and between ASC prediction methods for a given model. Many of the paired RAL differences are statistically significant[9], even some as small as only 3%, but it is subjective as to whether or not this represents a practical difference in forecast accuracy between two sets. Our observations highlight the trends in the data rather than focus on the statistical significance of any individual pairing.

Is there a particular method of generating ASCs for forecasting that generates the best predictions?

Including an ASC by method 1 (drawing entrant ASCs from all product estimated ASCs) predicts best or near best in all cases tested. Method 0 and method 1 produce comparable results when examining entrants only as well as in a 5-year forecast, where most vehicles are entrants. Method 2 is always worse than excluding ASCs for the GMM-IV model regardless of time horizon, endogeneity level, or market correlation structure, as might be expected based on the structure of the synthetic data.

How do the models compare across various estimation and prediction data correlation structures?

Using RAL as a metric of comparison across the models, we find the following:

Base case: MLE-C is better on average than GMM-IV for any of the ASC forecasting methods, and MLE-C is typically better than GMM-IV at predicting any given data set (comparing "# superior").

Market shift case: MLE-C is better on average than GMM-IV except for the case of long term forecasts when high price-ASC correlation in the estimation data disappears in the prediction data.

Invalid case: If invalid instruments are used in the GMM-IV model, MLE-C predicts better on average than the best GMM-IV model.

Entrant case: MLE-C is better on average than GMM-IV.

Regardless of whether the price coefficient is biased, we find that the best attribute-based models are at least as good as the naïve (static) model and always better than random guessing (no info), except when predicting the short term market with a GMM-IV model estimated using invalid instruments.

The greatest discrepancy in prediction accuracy between the best MLE-C and GMM-IV models across the four prediction cases occurs when the GMM-IV model is estimated using invalid IVs. The MLE-C model is superior regardless of the level of endogeneity, and the penalty is steep at low endogeneity with an RAL difference of approximately 22%. Conversely, when the GMM-IV model does predict better than the MLE-C model in the high-endogeneity, long term market-shift case there is only a 2% lift in RAL from the best MLE-C to the best GMM-IV forecast. This illustrates the greatest drawback to using GMM-IV in a forecasting scenario; it risks making far worse predictions for limited and unlikely upside (especially when valid IVs are difficult to specify (Rossi, 2014)).

5. Empirical case study

We apply the same approach as described in Section 4 to a real automotive sales data set: A mixed logit model is estimated by MLE-C and GMM-IV on US consumer new midsize sedan purchase data from 2002 through 2006 and then used to predict market shares at the trim level for midsize sedans sold in the US during 2007 and 2011. The accuracy and uncertainty of the model forecasts are compared on the RAL measure.

5.1. Models & data

We define an independent random coefficient mixed logit model with a linear utility function including the covariates listed in Table 5 plus an ASC. The attributes capture vehicle price, operating cost, performance, size, and country of origin. This choice of covariates is loosely based on Haaf et al. (2014), who compare the market share forecasts of 9,000 possible logit model specifications informed by the vehicle demand literature. They find that predictive accuracy is relatively invariant to the specific form of the covariates (e.g. gallons/mile versus miles/gallon), so long as each covariate is included, and prediction accuracy increases with additional covariates. We omit some covariates that were included in the models tested in Haaf et al. (2014) due to the specific nature of the GMM-IV estimator. Dummies for A/C standard and automatic transmission are excluded because dummies increase the number of parameters to be estimated but cannot function as instruments, making it more difficult to meet the GMM-IV estimator requirement that there are more instruments than observed variable coefficients. We proxy dummies for brand (e.g. Honda, Ford, or Volkswagen) by dummies for producer firm geographic location (US, Europe, or Asia), reducing the number of dummies from 23 to 2 (a dummy for US is omitted for identification). This reduces the number of parameters to be estimated as well as results in a statistically significant price coefficient for the estimated model. Appendix D examines the case where the MLE-C model is estimated using brand dummies. Additional discussion of parameter selection is included in Appendix A.

Our data set uses vehicle attribute information from Ward's Automotive Group (2012) and MSRP and aggregate sales data from Polk (2012). We implicitly assume that all individuals who purchased a vehicle in this class considered only all of the other midsize sedans available in the same year and made a compensatory decision based on vehicle attributes. Our models consider

[9] The test for significance is a two-sample t-test for equal means.

Table 5
Coefficients for MLE-C and GMM-IV models estimated on 2002 – 2006 US midsize sedan new sales data.

	Estimated population mean taste (μ)		Estimated population taste heterogeneity (σ)	
	MLE-C	GMM-IV	MLE-C	GMM-IV
Price ($10,000)	−0.5***	−1.1***	0.0039*	0.0033
(Standard error)	(0.0)	(0.3)	(0.0022)	(2.1270)
Gallons/mile (gal./100-mi.)	−1.1***	−0.3	0.0033*	0.9536**
(Standard error)	(0.0)	(0.6)	(0.0018)	(0.4414)
Weight/horsepower (lbs/hp)	−0.1***	−0.2**	0.0003*	0.0000
(Standard error)	(0.0)	(0.1)	(0.0002)	(0.2762)
Length × width (100-ft^2)	3.8***	2.6	0.0085*	1.7229
(Standard error)	(0.0)	(1.8)	(0.0047)	(1.8169)
Europe (dummy)	−1.1***	−1.1**		
(Standard error)	(0.0)	(0.5)		
Asia (dummy)	0.3***	−0.1		
(Standard error)	(0.0)	(0.3)		

Note: US country dummy omitted for identification; zero estimates are zero to the precision shown but are not actually zero.
* = significant at the $\alpha = 0.10$ level
** = significant at the $\alpha = 0.05$ level
*** = significant at the $\alpha = 0.01$ level

only new midsize sedan buyers, thus there is no outside good (option to not purchase any midsize sedan). GMM-IV estimation of the model on the full vehicle market would require more years of data (markets) than are available to us[10].

The choice of instruments is non-trivial and a subject of much study in the econometrics literature (Wooldridge, 2006). By definition they must be correlated with the endogenous variable (in this case price) and uncorrelated with the error (ASC) (Wooldridge, 2006). We specify instruments similar to those of BLP (Berry et al., 1995): for a given vehicle, the IVs are the sum of each of the non-price attributes over all other vehicles in the market offered by the same firm as the given vehicle (excluding the given vehicle itself) and the sum of each of the attributes over all other vehicles in the market sold by the competitor firms. Berry et al. (1995) argue that these IVs are correlated with the vehicle price but uncorrelated with the ASC because firms are thought to fix vehicle design choices before fixing pricing choices, and the firms observe the ASCs (they are only unobservable to the researcher) of all vehicles and set prices of their vehicles to be competitive while accounting for the utility derived from the ASCs. They argue that the ASC of a given vehicle, however, is not expected to be affected by the non-price attributes of competitors. These instruments are predicated on the interpretation of the ASC as a representation of average preferences for aggregate unobserved vehicle attributes.

In our formulation of the BLP instruments we exclude attributes that enter as dummies since instruments must exhibit sufficient variation. We also add functions of these instruments to satisfy the requirement that $L \geq 2K$ and to increase the efficiency of our GMM-IV estimator (Dubé et al., 2012, Train & Winston, 2007). The instruments are defined as:

$$\begin{aligned}
&\mathbf{z}_{jt}^{(1)} = \mathbf{x}_{jt}^{*}, \quad \mathbf{z}_{jt}^{(2)} = \sum_{k \in F_{\mathrm{O}t} \setminus j} \mathbf{x}_{kt}^{*}, \quad \mathbf{z}_{jt}^{(3)} = \sum_{k \in F_{\mathrm{C}t}} \mathbf{x}_{kt}^{*} \\
&\mathbf{z}_{jt}^{(4)} = \sum_{k \in F_{\mathrm{O}t} \setminus j} \left(\mathbf{x}_{kt}^{*} \circ \mathbf{x}_{kt}^{*}\right), \quad \mathbf{z}_{jt}^{(5)} = \sum_{k \in F_{\mathrm{C}t}} \left(\mathbf{x}_{kt}^{*} \circ \mathbf{x}_{kt}^{*}\right) \\
&\mathbf{z}_{jt}^{\sim} = \left[\mathbf{z}_{jt}^{(1)}, \mathbf{z}_{jt}^{(2)}, \mathbf{z}_{jt}^{(3)}, \left(\mathbf{z}_{jt}^{(1)} \circ \mathbf{z}_{jt}^{(1)}\right), \mathbf{z}_{jt}^{(4)}, \mathbf{z}_{jt}^{(5)}, \left(\mathbf{z}_{jt}^{(1)} \circ \mathbf{z}_{jt}^{(2)}\right), \left(\mathbf{z}_{jt}^{(1)} \circ \mathbf{z}_{jt}^{(3)}\right)\right] \\
&\mathbf{z}_{jt} = \mathbf{z}_{jt}^{\sim} \circ \frac{1}{\text{column max}(\mathbf{z}_{jt}^{\sim})}, \quad \forall j, t
\end{aligned} \tag{17}$$

where $F_{\mathrm{O}t}$ is the set of vehicles made by the same firm in year t, $F_{\mathrm{C}t}$ is the set of vehicles made by competitor firms in year t, and $\mathbf{x}_{jt}^{*}$ is the vector of attributes excluding price and dummies. For numerical purposes we normalize each instrument by dividing by the maximum value of that instrument occurring over all vehicles in all panel years in order to prevent an ill-conditioned weighting matrix $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})$.

There are a total of six covariates: price, three exogenous non-dummy attributes, and two country dummies, yielding ten model parameters to be estimated (non-dummy covariates have mean and variance coefficient components) and 24 total instruments.

5.2. Coefficient estimates

The model coefficients estimated by MLE-C and GMM-IV are shown in Table 5. The standard errors presented for the MLE-C model are computed using the Hessian of the log-likelihood function at the solution to the MLE subproblem, which ignores

[10] We estimated several mixed logit model specifications on all 2004 – 2006 new vehicle sales (full market), but were unable to obtain a model with any statistically significant coefficients. 2002 – 2003 full market data was unavailable for the full market.

uncertainty in the ASCs, and are therefore likely smaller than one, would obtain from a method that accounts for uncertainty in the ASCs. Bootstrapping could be used to estimate standard errors that account for the ASCs, but we do not pursue that here. We note that robust standard errors are a problem for any misspecified model, and all models for automotive applications tend to be misspecified. The mean coefficient estimates are of the same sign (excepting the Asia location dummy) and same order of magnitude, though the only statistically significant difference between the two model mean coefficient estimates is on price[11]. The MLE-C estimated model coefficients indicate little taste heterogeneity in the population, but the GMM-IV estimated coefficients indicate larger taste heterogeneity. This suggests fundamentally different views of taste preference distribution for this data set. However, only the difference in the gallons/mile heterogeneity parameter is statistically significant between the two estimated models. Further, the difference in price coefficient is in the expected direction if price is indeed correlated with unobserved attributes that are desirable to the consumer: We would expect the MLE-C model to estimate a smaller (in absolute value) price coefficient because it cannot account for the portion of consumer willingness to pay for higher priced alternatives that is due to attributes correlated with price that are unobserved in the model. However, it is not possible to conclude which coefficients are more "correct" because with market data we lack access to the "true model" that generated the choices observed, and we cannot be certain that the BLP instruments are valid.

It is somewhat surprising that the MLE-C model suggests so little taste heterogeneity in the population, and we take several steps in order to test the validity of our estimation results. First, we use 100 Halston draws of $\boldsymbol{\nu}$ that are held constant throughout estimation in order to prevent simulation noise. Second, we perform multistart optimization with 100 mean taste parameter starting values drawn randomly from the interval [− 10,10] and heterogeneity parameters drawn randomly from the interval [0,10]. For 87/100 of the starting points the solver converged successfully and to the same objective function value and optimizer. Third, we check that the objective function is not flat in the neighborhood of the optimizer— the log likelihood is -4.52×10^7 at the optimizer versus -4.54×10^7 at values of $\boldsymbol{\sigma} = \mathbf{0.1}$. Fourth, the analytical Hessian is provided to the solver, and the eigenvalues of the Hessian calculated at the optimizer indicate that the convergence is not a specious result of derivatives that vanish near coefficient values of zero. Lastly, when the model is estimated on synthetic data with no price-ASC correlation, the routine successfully recovers the true heterogeneity coefficient values. Train and Winston (2007) are also unable to obtain statistically significant heterogeneity coefficient estimates when only the vehicle chosen by consumers (and not the specific choice set) is known. Berry et al. (2004) report algorithm convergence issues when only one market (year) of data is used and suggest that sufficient variation of choice set across markets may enable successful estimation. The limited variation in choice set for our data set may be the cause of the small heterogeneity coefficient estimates.

The standard errors for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are much larger for the GMM-IV estimated coefficients than for the MLE-C estimated coefficients. The GMM-IV is a less efficient estimator than MLE-C so we would expect the standard errors to be somewhat larger. That they are orders of magnitude larger is likely a result of the different estimation strategies for the ASC: There are an additional 339 parameters in the specification of the model estimated by GMM-IV over the model estimated by MLE since MLE-C ASC calibration is performed post hoc, after the estimation of the observed attribute coefficients. If the ASCs included in GMM-IV estimation are correlated with price, then the collinearity may increase the standard errors (Greene, 2003). If the ASCs are uncorrelated, mean zero error terms, then the MLE-C estimates are consistent and the standard errors are valid. However, since we do not believe that the ASCs are uncorrelated, additional steps would be required to calculate robust errors. Calculation of robust errors for MLE-C is not found in the literature, and development of a method is outside the scope of this work.

We treat the non-price attributes as exogenous in our model, since our focus is on price-ASC endogeneity (and this is common in the literature), but this is not likely to be true. If the non-price attributes are correlated with the ASC, then these coefficients will also be biased (and our instruments will be invalid). In automotive demand models this is virtually certain. Fuel economy, acceleration, weight, dimensions, and manufacturer geographic location are all likely to be correlated with aesthetics and other unobserved or unquantifiable vehicle features, thus there should be nontrivial correlations between observed features and the ASC.

5.3. Analyzing estimated ASCs

Fig. 3(A) contains histograms of ASCs estimated by the two models. The ASCs have a non-normal distribution for both of the models according to a Jarque – Bera test conducted at the $\alpha = 0.05$ significance level. Fig. 3(B) plots the MLE-C versus GMM-IV estimated ASCs. Points falling on the red diagonal line are estimated ASCs that are identical between the two models. There is dispersion of the points from the line, indicating that the models do not agree on the value of the ASCs, but there is no indication that one model consistently under or over estimates them as compared to the other.

In order to investigate the possible correlations among the ASCs and vehicle attributes we regress MLE-C and GMM-IV estimated ASCs on six sets of dependent variables. The estimated ASC is regressed on: (1) an intercept plus vehicle physical attributes (price, gallons/mile, weight/horsepower, and (length × width)), (2) brand dummies (e.g. Acura, Ford, etc.), and (3) a dummy variable for unique vehicles at the aggregate make-model level (a Toyota Camry and Toyota Camry Solara are both assigned a single ID representing a Toyota Camry). Though we do not include brand dummies in the estimated models of Table 5, we include them

[11] The asymptotic distribution of the difference of each mean coefficient is $N(\mu_{MLE} - \mu_{MLE}, s^2_{MLE} + s^2_{GMM})$ where s^2 is the square of the standard error of the mean parameter estimate. For each mean coefficient we test the null hypothesis H_0: $\mu_{\mathrm{MLE}} = \mu_{\mathrm{GMM}}$ against H_A: $\mu_{\mathrm{MLE}} \neq \mu_{\mathrm{GMM}}$. We reject H_0 at the $\alpha = 0.1$ level for price only; for all other mean coefficients we do not reject H_0.

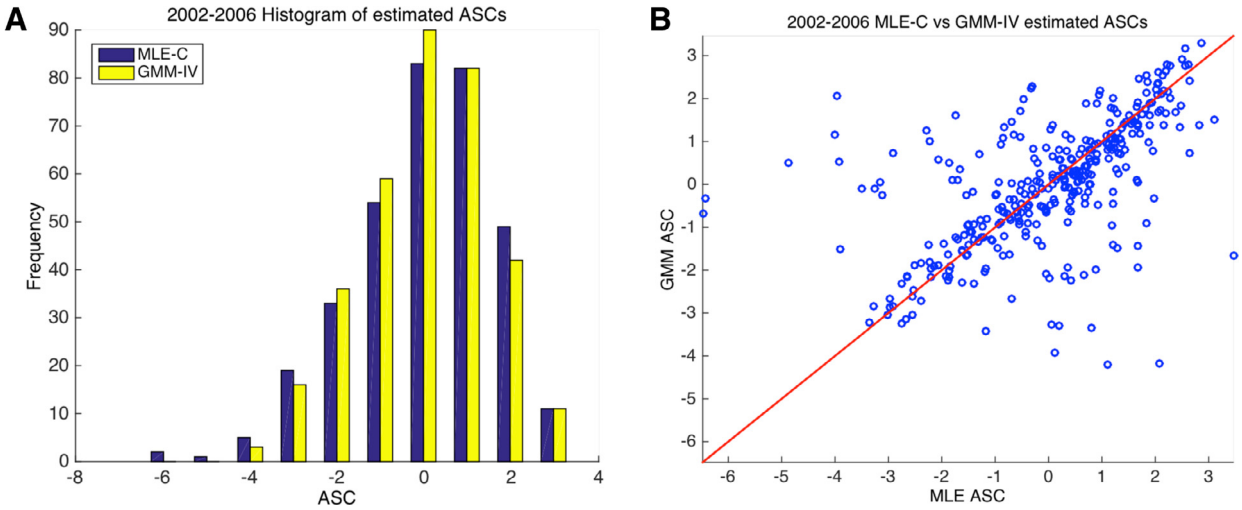


Fig. 3. The distribution of estimated ASCs for GMM-IV and MLE-C is non-normal and left skewed (A), but the MLE-C versus GMM-IV estimated ASCs (B) do not indicate that there is a systemic difference in their discrepancy.

in the ASC regressions as covariates since ASCs and brand are likely related. Additional results for an MLE-C estimated model that includes brand dummies are presented in Appendix E.

All three of these regressions yield at least 24% statistically significant coefficients for both MLE-C and GMM-IV estimated ASCs. This supports the use of observed vehicle characteristics in forecasting ASCs, particularly the methods used in this case study. Additionally, that the GMM-IV estimated ASCs are statistically significantly correlated with non-price vehicle attributes suggests that the BLP instruments (which are functions of non-price vehicle attributes) were, for our data, invalid.

Additional regressions, results and discussion are included in Appendix F.

5.4. Prediction

The coefficients estimated by MLE-C and GMM-IV are used to forecast 2007 and 2011 sales. In 2007, 33% of the midsize sedans were new ("entrants", 23 out of 68 did not appear in estimation data), and in 2011, 72% of the midsize sedans were new (34 out of 47 did not appear in estimation data)[12]. We test four methods of predicting new product ASCs. Methods 1 and 2 are the same as described in Table 2, where method 1 draws entrant ASCs from all estimated ASCs and method 2 draws entrant ASCs from estimated ASCs of the "nearest neighbor" vehicle. Additionally we introduce a third method in which entrant vehicle ASCs are drawn from the estimated ASCs of the same brand ("brand"), and a fourth method in which the ASCs are drawn from the estimated ASCs of other trim levels of the same vehicle make-model ("make-model' '). If the entrant brand or vehicle make-model is not observed in the estimation data set (e.g. Dodge Caliber), then the entrant vehicle ASCs are drawn from all estimated ASCs, as in method 1. Method 3 is similar to that used in Berry et al. (2004) and Train and Winston (2007). Both method 3 and method 4 are related to the interpretation of the ASC as a vehicle-specific fixed effect that represents aggregate unobserved attribute utility as opposed to a random error term.

Method 2 assumes that the unobserved attributes represented by the ASC are related to the vehicle's attributes and those of its competitor(s), which is an intuitive means for informing new product ASCs. However, this implies that the ASC is correlated with competitor non-price attributes, violating the assumption required for the BLP specification of instruments that for a given vehicle the sum of competitor attributes serving as instruments are uncorrelated with the ASC. Despite this issue we include results from prediction under this method because it is similar to an approach used in the literature and is tempting to researchers (Berry et al., 2004).

For each combination of the MLE-C and GMM-IV estimated models and four ASC generation methods, we draw 10,000 sets of ASCs for the predictive year vehicles. For each draw of ASCs we simulate a draw of shares by integrating over taste heterogeneity as in Eq. (5), generating a Monte Carlo simulation of predicted market shares. Forecast uncertainty is represented by the range of the 10,000 sets of share predictions, and our point estimate of share ("expected share") is obtained by averaging over the draws of share. We report the RAL of the expected share predictions in Table 6. The following comparisons of ASC and model estimation methods are based on this measure where greater values of RAL indicate more accurate predictions.

[12] The high turnover is a result of our designation as to what constitutes a "new vehicle". We consider make-models distinct at the trim level for different body styles (e.g. sedan versus wagon). As a result, several trims appear as "new" in the prediction data when only a subset of body styles is available in each year.

Table 6
RAL(E[share]) comparison of ASC forecasting methods for MLE-C and GMM-IV models estimated on 2002–2006 midsize sedan data and used to predict 2007 and 2011 midsize sedan market shares.

	1-year-forward (2007) forecasts					5-year-forward (2011) forecasts				
	No ASC (Meth. 0)	All (Meth. 1)	Near neighbor (Meth. 2)	Brand (Meth. 3)	Make-model (Meth. 4)	No ASC (Meth. 0)	All (Meth. 1)	Near neighbor (Meth. 2)	Brand (Meth. 3)	Make-model (Meth. 4)
RAL of expected share										
MLE-C	45%	66%	61%	63%	66%	38%	42%	60%	37%	41%
GMM-IV	34%	68%	64%	70%	70%	28%	33%	45%	34%	33%
Static	*68%*					*39%*				
No info	*32%*					*42%*				
RAL of expected share— ENTRANTS ONLY										
MLE-C	91%	91%	84%	88%	91%	49%	50%	71%	43%	49%
GMM-IV	86%	85%	81%	87%	88%	43%	43%	57%	43%	42%
Static	*87%*					*54%*				
No info	*87%*					*62%*				

Note: highlighted cells indicate the most accurate model and ASC forecasting method for a given time period and method of calculating RAL.

Two additional models are included in Table 6 for comparison purposes. The static model ("static") holds shares of incumbent vehicles constant from the last year of observation in the estimation data and divides the remaining share equally among the entrant vehicles. The no info model ("no info") assumes all vehicles have equal share in the prediction year.

Comparing the ASC generation methods in Table 6, method 4 forecasts best for both the MLE-C and GMM-IV methods for the 1-year-forward ("short term"), and method 2 forecasts best for the 5-year-forward ("long term") time horizons. There are two important comparisons in the 5-year-forward forecasts. First, the "no info" model predicts better than the "static" model. Second, the best MLE-C model predicts much better than the "no info" model, but the best GMM-IV model is only slightly better than the "no info" model. The first comparison suggests than the market has changed sufficiently in composition such that entrant products have significantly altered the share of the incumbent products. The second comparison suggests that the underlying relationships in the data, like the correlation of the ASC and price, persist and to some degree are successfully captured by the MLE-C coefficients in a way that facilitates better forecasts.

We estimate a model that includes brand dummies by MLE-C and present the results in Appendix E. The MLE-C model predictions with no forecast ASCs are greatly improved by the inclusion of brand dummies as expected (2011 RAL of 51% for no-ASC brand dummy model versus 38% for no-ASC model that does not include brand dummies). The brand dummy model is superior to the model that excludes brand dummies in estimation and predicts ASCs by the brand method (method 3). But the brand dummy model is inferior to the model that predicts ASCs by the nearest neighbor method. These comparisons suggest that the brand dummies capture much — but not all — of the explanatory power of the ASCs and that the brand method of forecasting ASCs is unable to recover it. The advantage of explicitly including brand dummies over aggregating the utility contribution with other unobservable characteristics is that brand is observed in future markets so that the predicted brand utility contribution is not susceptible to changes in correlation between brand and the aggregate ASC. The GMM-IV approach is unable to incorporate brand dummies due to the limitations described previously.

We suspect that the ASC is endogenous with the other vehicle physical attributes in addition to price. For the long term forecasts, method 2 is better than any of the other ASC prediction methods and better than omitting the ASC in prediction for both models. In the synthetic study, when, by design, the ASC was uncorrelated with the non-price attribute x, predicting the ASC by method 2 resulted in worse predictions than when no ASC was predicted by method 0. The regression of ASCs on vehicle physical characteristics also provides evidence for non-price attribute endogeneity with the ASC. This is in direct violation of the assumption for specification of valid instruments that non-dummy attributes and ASCs are uncorrelated, again implying that the instruments in this case study — used frequently (directly or with some variation) in vehicle demand literature (Copeland et al., 2011; Li et al., 2011; Sudhir, 2001; Berry et al., 1999; Petrin, 2002; Train & Winston, 2007; Vance & Mehlin, 2009) — are invalid.

Table 6 also contains the RAL for entrant products only. The shares are forecasted for all midsize sedans, but the shares of the incumbent products are included as a single lump sum in the RAL calculation. As for the whole market, method 4 is better in the short term, but method 2 is better in the long term.

6. Discussion

6.1. Lessons from the simulation study

Much of the econometrics literature on vehicle market modeling has worked to avoid biased coefficients. The GMM-IV approach mitigates coefficient bias in our synthetic data study when valid instruments are used; however, our synthetic data study also illustrates that (1) invalid instruments can exacerbate coefficient bias, and (2) correcting coefficient bias does not necessarily produce better forecasts. Invalidity of instruments is a significant issue for GMM-IV methods in practice because instrument

validity is impossible to verify. In our study, which used data similar in structure to the typical automotive demand data set, the greatest RAL penalty (relative to an MLE-C approach) for estimating a GMM-IV model with invalid instruments was about ten times larger than the greatest reward for specifying valid GMM-IV instruments. But even given valid instruments, so long as the underlying source of the endogeneity persists in the forward years, MLE with biased coefficients results in better forecasts than those made by GMM-IV's unbiased coefficients in the datasets considered in our study. And, perhaps surprisingly, even when the source of endogeneity present in the estimation data disappears in the prediction data, the MLE approach still can often produce better predictions than the GMM-IV approach with valid instruments.

6.2. Lessons from the empirical case study

We cannot evaluate coefficient bias in the empirical case study because we do not know the true coefficients[13]. However, the use of GMM-IV does result in a change to the price coefficient in the direction expected if price were correlated with unobserved attributes that are desired by consumers and if the resulting endogeneity bias were mitigated via the selected instruments.

For forecasting, when ASCs are not used to forecast, MLE-C outperforms GMM-IV. This might be expected because MLE-C captures some of the effect of un-modeled attributes that are correlated with price and projects the effect onto future vehicles based on their prices; in contrast, if GMM-IV corrects for endogeneity bias, GMM-IV forecasts that do not estimate ASCs of future vehicles effectively assume they have no valuable unobserved attributes. Thus the biased coefficients improve forecasts.

When ASCs are used in forecasting (using any of the methods examined), GMM-IV forecasts improve, and we observe the best GMM-IV forecast to be slightly better than the best MLE-C forecast in our 1 year forecast (though they are both comparable to the static model) while the best MLE-C forecast is substantially better for the 5-year forecast. The smaller differences in the 1-year forecast may be due to chance, and 1-year forward forecasts in markets planned out 3 – 5 years in advance aren't necessarily the best application for discrete choice models; other tools better fit underlying inertial trends (including inventory effects). For example, a simple static model that predicts shares in year n will be equal to those of year $n-1$ can outperform discrete choice models in 1-year forecasts (Table 6).

For counterfactual studies that alter observed market conditions and simulate choice outcomes, market competition, and measure related changes in economic measures such as market power, consumer welfare, fleet fuel economy, etc (Berry et al., 2004; Bunch et al., 2011; Train & Winston, 2007), the GMM-IV approach is more appropriate when the instruments are valid and the counterfactual change affects only price without affecting other unobserved attributes correlated with price (e.g.: short term effect of subsidies). When the counterfactual scenario is likely to involve fleet changes that might affect price as well as unobserved attributes that co-vary with price, MLE-C may predict better if such patterns are expected to follow those in the existing market; however, unobserved attributes may be expected to change in future markets as new features are developed.

When only forecast accuracy of entrants are evaluated, adding ASCs to the MLE-C and GMM-IV models does not meaningfully improve predictions in the short term and predicts worse than random guessing for the 5-year-forecast for all but the "nearest neighbor" method. The relative forecast improvement of the nearest neighbor method suggests correlations between the ASC and non-price observed attributes, which we (and the prior literature) do not instrument for. In fact, we assume no correlation between ASCs and non-price attributes when choosing non-price attributes as instruments. Validity of instruments can only be argued, not proven (Rossi, 2014), and our interpretation of ASCs as representations of unobserved attributes (consistent with both the explanatory and predictive bodies of literature) as well as the regressions of estimated ASCs on product characteristics in Section 5.3 suggests correlations with observations that conflict with IV assumptions.

For our data set and choice of instruments, which are popular in the automotive demand literature, the presumably-biased MLE-C coefficients were more successful at prediction overall than the GMM-IV coefficients. If the instruments are valid, then this suggests that endogeneity bias can *aid* predictions by implicitly partially capturing persistent unobserved effects. If the instruments are invalid, then prediction error in GMM-IV models may be exacerbated. Other choices of instruments may yield different results — predictions from a model estimated using truly valid or better instruments may be superior to a model that ignores endogeneity. There is an inherent risk trade-off: attempting to specify valid instruments risks degrading forecasts if the instruments are invalid.

7. Limitations

Our investigation examines only a portion of the factors affecting vehicle demand prediction uncertainty, and our empirical case study models have error resulting from misspecification and missing information (as do all models). We lack individual-level choice data with consumer covariates, such as demographics or usage variables (He et al., 2012), and are unable to quantify some key purchase drivers, such as aesthetics. We assumed that coefficient attributes are constant over time, i.e. that there are no changes in consumer valuation of attributes. The ASC is intended to capture commonly-held evaluations of unobserved attributes (in the GMM-IV estimated model) and/or other sources of error (in the MLE-C estimated model). However if we suspect that they are correlated with observed attributes other than price (and we do), our coefficient estimates for the observed attributes will be inconsistent.

We restrict our study of demand forecasting to random utility DCMs that treat consumers as observant rational expected utility maximizers with consistent preferences that fully consider every option in the relevant market. Several studies offer

[13] More precisely, because the model is necessarily misspecified there are no true coefficients.

Table 7
Comparison of MLE-C and GMM-IV properties and findings.

	MLE-C	GMM-IV
Coefficient estimates	Biased	Can reduce bias given valid IVs, but invalid IVs may exacerbate bias
Prediction	More accurate in almost all cases tested	Given valid IVs, can be better for markets with high correlation between price and unobserved attributes and anticipated dramatic change of that pattern in future markets
Computation	Concave nonlinear programming (NLP) when utility is linear, leading to global solution and fast estimation	Nonlinear equality constraints lead to local minima, infeasible regions lacking gradient information toward the feasible domain, and potentially multiple local minima; slow and requires multi start
ASC	Calibration constant does not necessarily represent missing attributes	Estimation parameter represents effect of missing attributes as long as IVs are valid

critiques or alternative treatments such as preferences that evolve over time (Axsen et al., 2009), incorporate cultural factors (Heffner et al., 2007; Axsen & Kurani, 2011), or are adapted to a specific choice situation (Min et al., 2014; Macdonald et al., 2009). Especially relevant to the data sources of this study, the Lucas critique warns against use of aggregated historical data to predict outcomes in counterfactual scenarios (Lucas, 1976).

We did not consider alternative estimation methods beyond MLE-C and GMM-IV. For example, we do not consider Bayesian methods, which estimate coefficients of the same model forms and are asymptotically equivalent to MLE (Train, 2009), and thus if our MLE estimates are reasonably good we should not expect to see significantly different results with Bayesian methods. Nor did we investigate other means of controlling for endogeneity (Frischknecht et al., 2010), alternative heterogeneity specifications, e.g. latent class models, mixed logit model with joint parameter distributions, mixture models, and generalized logit models that account for scale and coefficient heterogeneity (Fiebig et al., 2009). Finally, we also did not include in this study other econometric models that do not involve discrete choice and that may also be used for forecasting.

We assume in generating prediction ASCs that they are uncorrelated with one another, but this is a restrictive assumption. For GMM-IV estimation we specify a weighting matrix, $\mathbf{W}$ (Eq. (9)) that is less efficient than other candidates. An alternative choice of $\mathbf{W}$ (e.g. the inverse of the covariance matrix of the moments $\mathbf{W} = (\mathbf{Z}'\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{Z})^{-1}$) may lead to more accurate estimates for our finite data sample at increased computational cost, though Nevo (2000) suggests that this is not a primary concern.

8. Conclusion

Our synthetic simulation study shows that while a GMM-IV approach can reduce coefficient bias given valid instruments, (1) invalid instruments can exacerbate bias, and (2) correcting coefficient bias does not necessarily improve forecasts. For the automotive-relevant synthetic datasets we generated, the MLE-C forecasts were superior to GMM-IV forecasts in nearly all cases despite being theoretically inconsistent and having biased coefficients – even when the endogeneity patterns present in the estimation data are absent in the prediction data.

Our empirical market study showed that for mid-size sedan sales in the United States, when ASCs are not projected for future vehicles, MLE-C forecasts are superior to GMM-IV forecasts, due in part to the portion of the effect of unobserved features correlated with price that are implicitly captured in the biased price coefficient of the MLE-C model. When ASCs are projected for future vehicles, GMM-IV forecasts improve and may be better or worse than MLE-C forecasts. Because 5-year forecasts are best when ASCs are predicted using the nearest neighbor method, and because we find estimated ASCs to be correlated with non-price observed attributes, we argue that the BLP IVs used in most instrumented automotive demand studies appear to be invalid.

Table 7 summarizes the model characteristics observed in the synthetic and empirical case studies. For purely predictive purposes, our results show that, for the cases tested in this work, the drawbacks of GMM-IV (challenging model estimation, uncertain specification of valid instruments, and a less efficient estimator) may outweigh the expected benefits of potentially mitigating price endogeneity.

Acknowledgments

This work was completed with support from the National Science Foundation (grant #1064241), Ford Motor Company, Toyota Motor Corporation, and the center for Climate and Energy Decision Making (SES-0949710), through a cooperative agreement between the National Science Foundation and Carnegie Mellon University. Opinions expressed are those of the authors and not necessarily those of the sponsors.

Appendix A

The literature on GMM-IV methods has seen a recent focus on the computational and numerical challenges of estimating these types of models. Consistent with the studies discussed in Section 2 (Dubé et al., 2012; Knittel & Metaxoglou, 2012; Su & Judd, 2012), we encountered several difficulties that we discuss here in order to illuminate technical drawbacks of using GMM-IV. General estimator issues are documented in the marketing and econometrics literature, but we focus on the types of data sets vehicle demand researchers are likely to encounter.

Firstly, we found that the computation time required to estimate a model with GMM-IV was about 2.5 times greater than with MLE-C. The GMM-IV optimization problem has nonlinear and computationally expensive share constraints that must be solved at each iteration, as opposed to only once at the end of the optimization routine as in MLE-C.

Secondly, the GMM-IV estimator is statistically (as opposed to computationally) inefficient relative to the MLE estimator, and it is difficult to specify a sufficient number of instruments for real aggregate data sets of the size and form seen here such that at least some of the coefficients (particularly price) are statistically significant. At least as many instruments as observed coefficients must be used for identification, and we found that at least twice as many instruments as coefficients were needed to ensure a significant estimate of the price coefficient when estimated on our data set. Though any function of the exogenous variables and base instruments can be used, instruments that are too collinear cause numerical difficulties.

GMM-IV estimation was sensitive to the level of aggregation of the vehicle data. We were unable to obtain coefficient estimates when make-model sales were summed over trim-level variants.

Appendix B. Explicit formulation of the GMM-IV optimization problem

The optimization problem for GMM-IV estimation as stated in Eq. 9 is:

$$\begin{aligned}
&\underset{\boldsymbol{\beta},\boldsymbol{\xi}}{\text{minimize}}\left(\mathbf{Z}'\boldsymbol{\xi}\right)'\mathbf{W}\left(\mathbf{Z}'\boldsymbol{\xi}\right)/T\\
&\text{subject to } \ln\left(P_{jt}\left(\boldsymbol{\beta},\boldsymbol{\xi}_t\right)\right) = ln\left(s_{jt}\right), \quad \forall j \in J_t^-, t\\
&\sum_{k\in J_t}\xi_{kt} = 0, \quad \forall t
\end{aligned}$$

and we specify $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ as the weighting matrix. For execution, we transform the optimization problem so that the weighting matrix can be incorporated by singular value decomposition, rather than directly using Matlab's "inverse" function since it is less numerically stable, and we rewrite the objective as a simple inner product plus a linear constraint, which is more computationally efficient for the KNITRO solver. We obtain matrices **U, D**, and **V** from the singular value decomposition of **z** such that:

$$\mathbf{UDV}' = \mathbf{Z} \tag{18}$$

U is a $(V \times V)$ orthogonal matrix, **V** is a $(K \times K)$ orthogonal matrix, and **D** is $(V \times K)$ matrix composed of a stacked diagonal $(K \times K)$ matrix with positive entries and a $((V-K) \times K)$ matrix of zeros. We call the diagonal $(K \times K)$ upper matrix $\mathbf{D}*$. The product $\mathbf{Z}'\mathbf{Z}$ can be written:

$$\mathbf{Z}'\mathbf{Z} = \mathbf{VD}'\mathbf{U}'\mathbf{UDV}' = \mathbf{VD}'\mathbf{DV}' = \mathbf{V}(\mathbf{D}^*)^2\mathbf{V}' \tag{19}$$

so that $\mathbf{W} = (\mathbf{Z}'\,\mathbf{Z})^{-1}$ can be expressed:

$$\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1} = (\mathbf{V}(\mathbf{D}^*)^2\mathbf{V}')^{-1} = \mathbf{V}(\mathbf{D}^*)^{-2}\mathbf{V}' \tag{20}$$

We can now rewrite the objective function as a simple quadratic equation with the addition of a linear constraint:

$$\begin{aligned}
&\underset{\boldsymbol{\beta},\boldsymbol{\xi}}{\text{minimize}}\left(\mathbf{h}'\mathbf{h}\right)/T\\
&\text{subject to } \ln\left(P_{jt}\left(\boldsymbol{\beta},\boldsymbol{\xi}_t\right)\right) = ln\left(s_{jt}\right), \quad \forall j \in J_t^-, t\\
&\sum_{k\in J_t}\xi_{kt} = 0, \quad \forall t\\
&\mathbf{h} = (\mathbf{D}^*)^{-1}\mathbf{V}'\mathbf{Z}'\xi
\end{aligned} \tag{21}$$

where **h** is $(L \times 1)$. This objective function of Eq. (21) is equivalent to that of Eq. (9) as can be seen by substituting in the expression for **h** in the constraints and using the singular value decomposition.

Appendix C. Synthetic data estimation results

Fig. C.1.

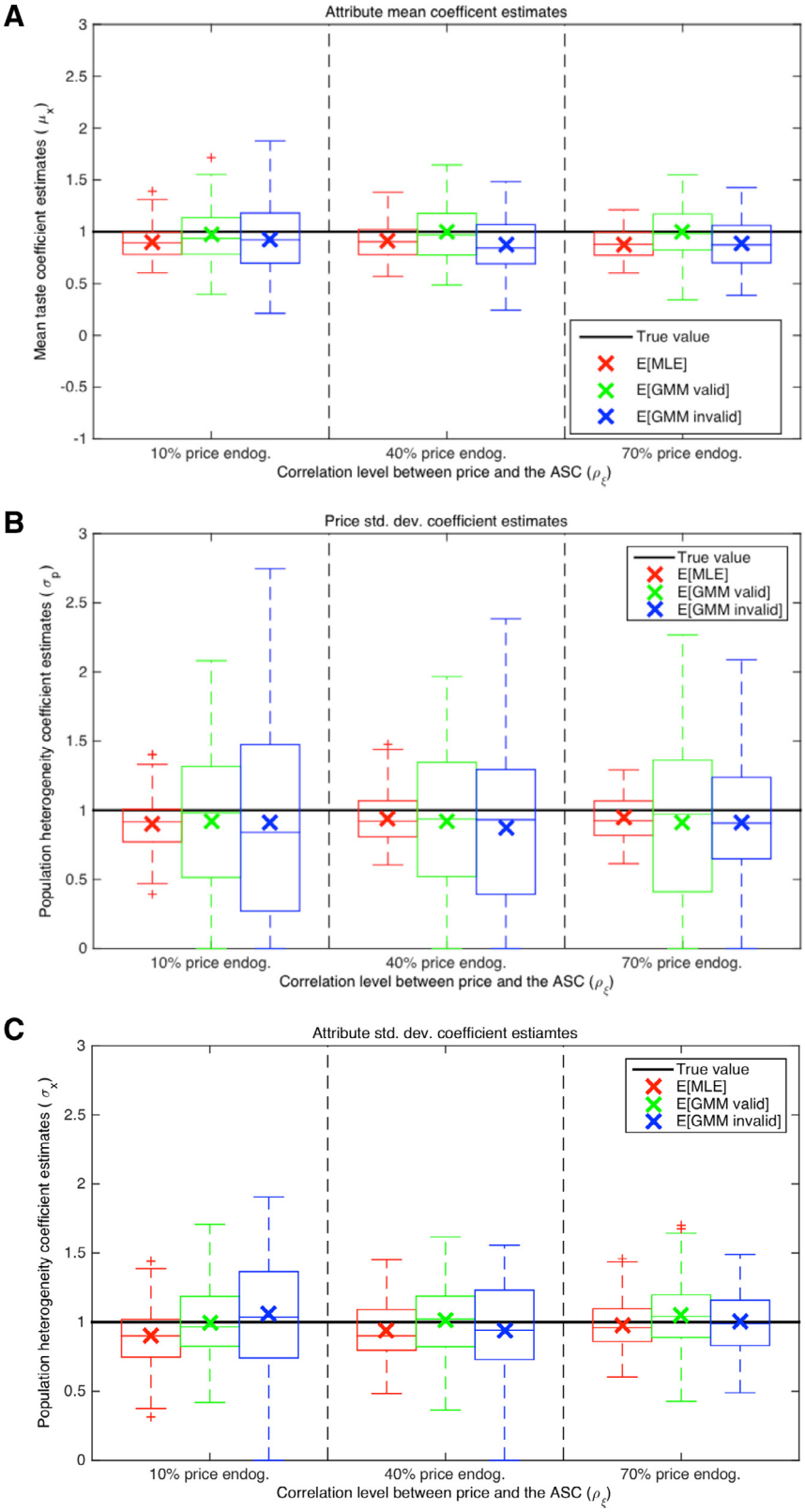


Fig. C.1. Estimates of population mean taste parameter μ_x (A) and population heterogeneous portion of taste parameters price σ_p (B) and technology σ_t (C).

Appendix D. Synthetic data prediction results

Tables D.1 and D.2.

Table D.1
INVALID case results shown include the mean and std. dev. of the RAL of expected share across 125 data sets for MLE-C and GMM-IV models as well as the number of data sets for which the MLE-C or GMM-IV model had a greater respective RAL (# superior).

	1-year-forward			5-year-forward		
	None	All	Neigh.	None	All	Neigh.
Method:	0	1	2	0	1	2
10% price-ASC correlation (ρ_ξ)						
MLE-C	67%	85%	74%	68%	70%	51%
(Std. dev.)	*(8%)*	*(8%)*	*(12%)*	*(8%)*	*(8%)*	*(11%)*
# Superior	119	123	121	116	121	119
GMM-IV	54%	63%	52%	55%	54%	32%
(Std. dev.)	*(13%)*	*(18%)*	*(20%)*	*(13%)*	*(12%)*	*(13%)*
# Superior	6	2	4	9	4	6
Static	68%			42%		
No info	38%			39%		
40% price-ASC correlation (ρ_ξ)						
MLE-C	72%	86%	78%	71%	72%	56%
(Std. dev.)	*(6%)*	*(6%)*	*(8%)*	*(6%)*	*(7%)*	*(8%)*
# Superior	120	121	119	114	117	121
GMM-IV	63%	67%	56%	62%	61%	38%
(Std. dev.)	*(10%)*	*(17%)*	*(19%)*	*(9%)*	*(9%)*	*(15%)*
# Superior	5	4	6	11	8	4
Static	72%			47%		
No info	44%			45%		
70% price-ASC correlation (ρ_ξ)						
MLE-C	81%	91%	85%	80%	82%	69%
(Std. dev.)	*(5%)*	*(4%)*	*(6%)*	*(4%)*	*(4%)*	*(7%)*
# Superior	116	124	113	113	117	122
GMM-IV	76%	81%	71%	76%	76%	53%
(Std. dev.)	*(7%)*	*(9%)*	*(14%)*	*(6%)*	*(7%)*	*(13%)*
# Superior	9	1	12	12	8	3
Static	76%			56%		
No info	52%			53%		

Note: Highlighted cells indicate the best model for a given time period and price-ASC correlation level.

Table D.2
ENTRANT ONLY case results shown include the mean and std. dev. of the RAL of expected share across 125 data sets for MLE-C and GMM-IV models as well as the number of data sets for which the MLE-C or GMM-IV model had a greater respective RAL (# superior).

	1-year-forward			5-year-forward		
	None	All	Neigh.	None	All	Neigh.
Method:	0	1	2	0	1	2
10% price-ASC correlation (ρ_ξ)						
MLE-C	85%	85%	74%	69%	70%	53%
(Std. dev.)	*0.54%*	*0.55%*	*1.49%*	*0.73%*	*0.78%*	*1.22%*
# Superior	87	103	103	92	101	113
GMM-IV	83%	80%	64%	66%	65%	40%
(Std. dev.)	*0.67%*	*0.77%*	*2.32%*	*0.78%*	*0.77%*	*1.84%*
# Superior	38	22	22	33	24	12
Static	68%			42%		
No info	69%			41%		
40% price-ASC correlation (ρ_ξ)						
MLE-C	86%	86%	77%	72%	72%	55%
(Std. dev.)	0.52%	0.55%	0.91%	0.48%	0.50%	0.83%
# Superior	81	107	114	102	104	115
GMM-IV	84%	81%	67%	68%	67%	41%
(Std. dev.)	0.67%	0.75%	1.87%	0.54%	0.55%	1.56%
# Superior	44	18	11	23	21	10
Static	72%			48%		
No info	73%			47%		
70% price-ASC correlation (ρ_ξ)						
MLE-C	91%	91%	85%	81%	81%	69%
(Std. dev.)	0.20%	0.21%	0.36%	0.20%	0.20%	0.62%
# Superior	112	122	118	120	121	120
GMM-IV	87%	82%	74%	72%	72%	52%
(Std. dev.)	0.37%	0.58%	1.68%	0.60%	0.56%	2.26%
# Superior	13	3	7	5	4	5
Static	78%			56%		
No info	77%			56%		

Note: Highlighted cells indicate the best model for a given time period and price-ASC correlation level.

Appendix E. MLE-C estimated model with brand dummies

Estimation results

Table E.1.

Prediction results

Table E.2 and Fig. E.1.

Appendix F. Estimated ASC regressions on observed vehicle attributes

We regress MLE-C and GMM-IV estimated ASCs on six sets of dependent variables in order to investigate the (possible) correlation between the ASCs and vehicle attributes. Table E.1 contains the number of coefficients that are statistically significant at the $\alpha = 0.05$ for six linear regression models. The estimated ASC is regressed on: (1) an intercept plus vehicle physical attributes (price, gallons/mile, weight/horsepower, and (length x width)), (2) geographic dummies for the US, Europe, and Asia,

Table E.1
MLE-C estimated parameters for a mixed logit model that includes brand dummies.

Coefficient	Mean taste parameter μ	Mean taste parameter (std. err.)	Heterogeneity taste parameter σ	Heterogeneity taste parameter (std. err.)
Price ($10,000)	−0.9***	(0.001)	0.4141***	(0.001)
Gallons/mile (gal./100-mi.)	0.1***	(0.002)	0.0027	(0.002)
Weight/HP (lbs/hp)	0.1***	(0.000)	0.0001	(0.000)
Len. × wid. (100-ft^2)	1.4***	(0.005)	1.1885***	(0.006)
Acura	−0.8***	(0.003)		
Cadillac	0.4***	(0.002)		
Chevrolet	−0.4***	(0.002)		
Chrysler	−1.6***	(0.002)		
Dodge	−1.2***	(0.003)		
Ford	0.0***	(0.002)		
Honda	0.5***	(0.002)		
Hyundai	−1.9***	(0.002)		
Infiniti	0.2***	(0.003)		
Kia	−2.6***	(0.003)		
Lincoln	−0.4***	(0.003)		
Mazda	−1.5***	(0.002)		
Mercury	−1.8***	(0.002)		
Mitsubishi	−1.8***	(0.003)		
Nissan	0.1***	(0.002)		
Oldsmobile	−1.2***	(0.011)		
Pontiac	−0.4***	(0.002)		
Saab	−1.9***	(0.005)		
Saturn	−2.9***	(0.004)		
Suzuki	−3.9***	(0.009)		
Toyota	0.0	(0.002)		
Volkswagen	−1.9***	(0.002)		
Volvo	−2.1***	(0.004)		

*** Coefficient is significant at the $\alpha = 0.01$ level.

Table E.2
RAL(E[share]) comparison of ASC forecasting methods for MLE-C models including brand dummies estimated on 2002 – 2006 midsize sedan sales and used to predict 2007 and 2011 midsize sedan market shares.

	1-year-forward (2007) forecasts					5-year-forward (2011) forecasts				
	No ASC (Meth. 0)	All (Meth. 1)	Near neighbor (Meth. 2)	Brand (Meth. 3)	Make-model (Meth. 4)	No ASC (Meth. 0)	All (Meth. 1)	Near neighbor (Meth. 2)	Brand (Meth. 3)	Make-model (Meth. 4)
All vehicles	54%	66%	65%	68%	67%	51%	45%	44%	48%	51%
Static	*68%*					*39%*				
No info	*32%*					*42%*				
Entrants only	79%	78%	77%	80%	79%	60%	57%	55%	60%	64%
Static	*87%*					*54%*				
No info	*87%*					*62%*				

Note: Highlighted cells indicate the most accurate model and ASC forecasting method for a given time period and means of calculating RAL

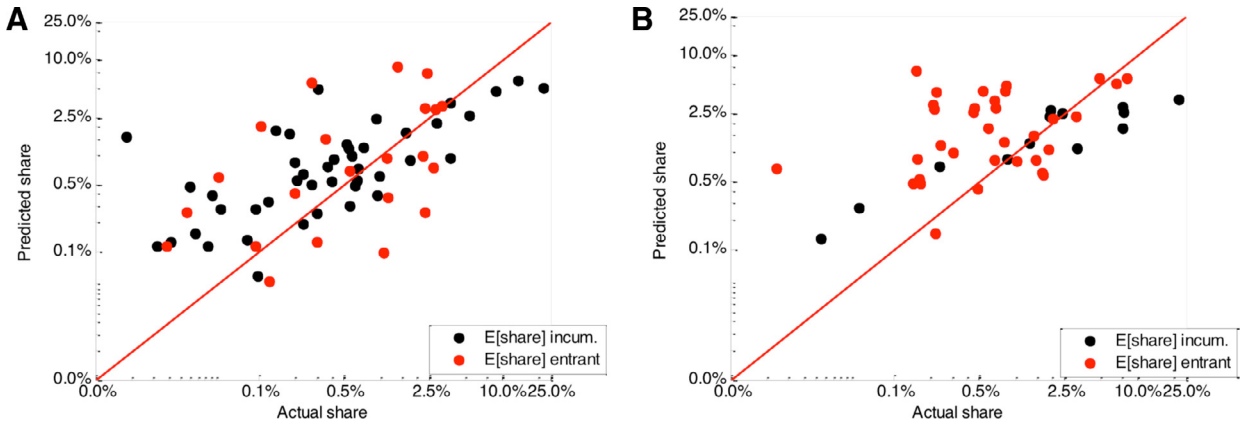


Fig. E.1. Actual versus predicted shares of midsize sedans predicted by a MLE-estimated mixed logit model that includes brand dummies and excludes ASCs in the prediction year for the 2007 (A) and 2011 (B) markets.

(3) the covariates of models (1) and (2) excluding the US dummy for identification, (4) brand dummies (e.g. Acura, Ford, etc.), (5) a dummy variable indicating each unique vehicle, and (6) a dummy variable for unique vehicles at the aggregate model level (a Toyota Camry and Toyota Camry Solara are both assigned a single ID representing a Toyota Camry). There are 339 total observations across five estimation data set years but only 153 unique vehicles since vehicles appear in multiple years. Though we do not include brand dummies in the estimated models of Table 5, we include them in the ASC regressions as covariates since ASCs and brand are likely related.

The number of statistically significant coefficients for a given regression is nearly identical between the MLE-C and GMM-IV estimated models. For both models, regressions 4 and 5 yielded statistically significant coefficients for ~1/3 of the covariates. GMM-IV estimated ASCs are statistically significantly correlated with non-price vehicle characteristics, suggesting that the BLP instruments were, for our data, invalid.

Table F.1.

Table F.1
Regression of MLE-C and GMM-IV estimated ASCs on select dependent variables.

Dependent variables (regression #)	Physical attributes (1)	Geographic dummies (2)	Physical attributes + geographic dummies (3)	Brand dummies (4)	Fixed effects (5)	Aggregate fixed effects (6)
Total covariates in regression	5	3	7	24	153	66
Number of statistically significant regression coefficients at the $\alpha = 0.05$ level						
MLE-C estimated ASCs	0	0	3	9	52	14
Significant coefficients	*Price, gal./mi, len. x wid.*		*gal./mi, len. x wid.*	*Not listed for brevity*	*Not listed for brevity*	*Not listed for brevity*
GMM-IV estimated ASCs	2	0	2	7	52	15
Significant coefficients	*Price, gal./mi.*		*price, gal./mi.*	*Not listed for brevity*	*Not listed for brevity*	*Not listed for brevity*

Note: A constant is excluded in regressions 2, 4, 5, and 6 for identification.

Appendix G. Case study prediction results

Tables G.1 and G.2 contain plots of the actual versus predicted shares for the 1-year-forward and 5-year-forward forecasts of the MLE-C and GMM-IV models using all four ASC generation methods. The error bar range for each of the forecasts covers the 2.5%–97.5% percentile of the simulated shares. Note that the ranges shown are independent of one another, meaning that the 2.5% percentile share shown for vehicle 1 may have occurred in a different draw of shares than the 2.5% percentile share shown for vehicle 2. Since shares of a given vehicle are related to the shares of all the other vehicles, the distributions would likely be tighter if the correlation were accounted for.

Tables G.1 and G.2.

Table G.1
Actual versus predicted 2.5–97.5% interval of shares of 1-year-forward predictions for the MLE-C and GMM-IV models using each of the ASC generation methods.

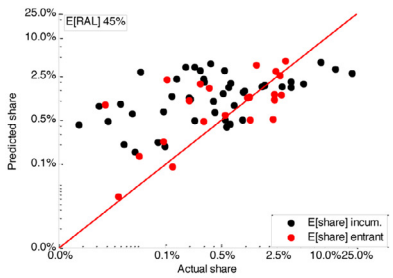
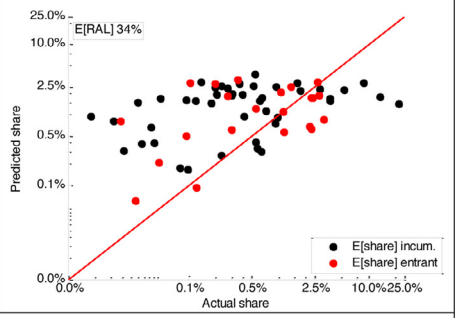
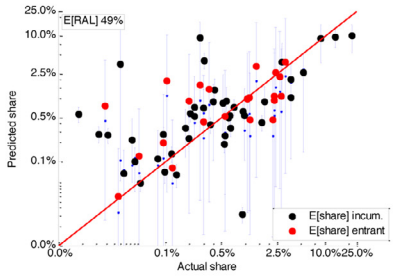
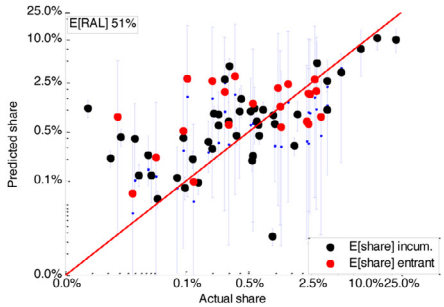
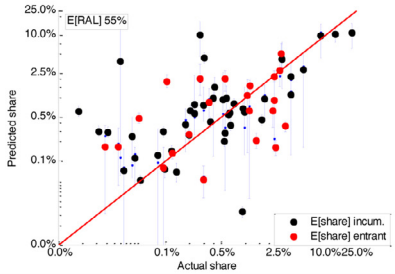
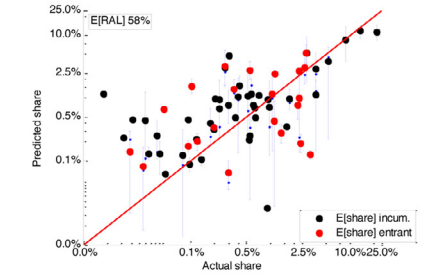
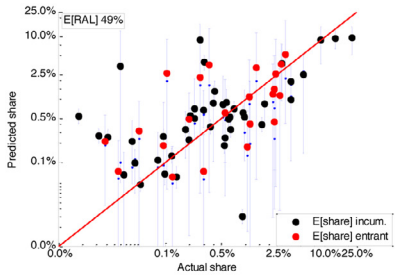
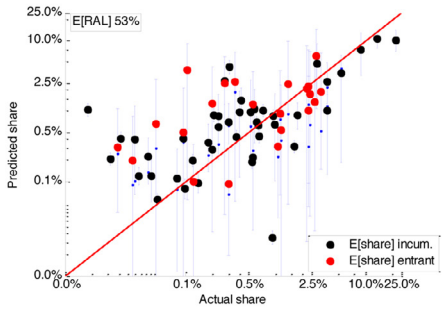
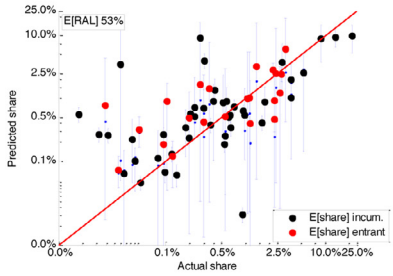
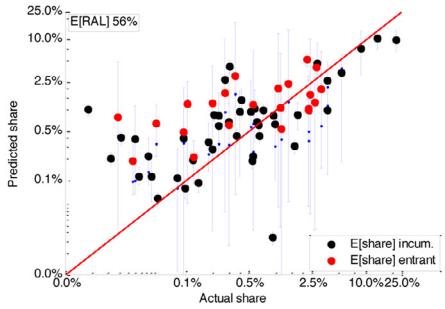
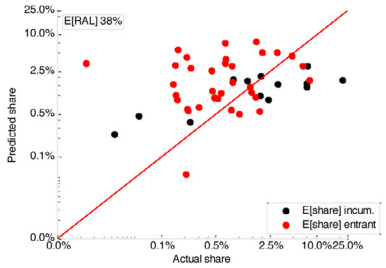
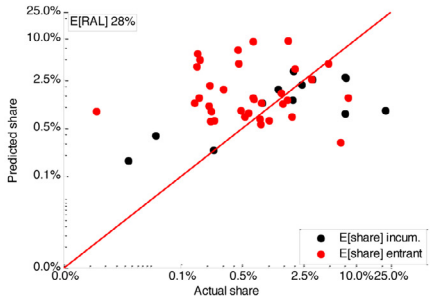
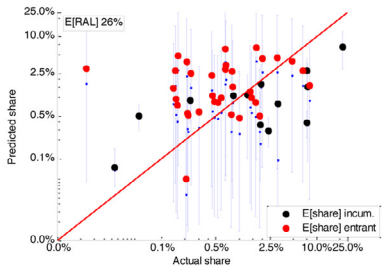
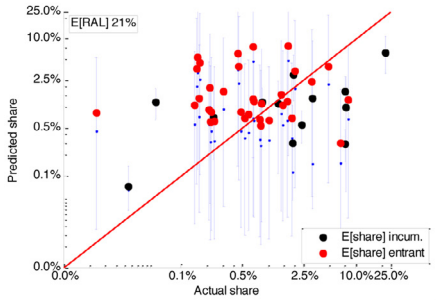
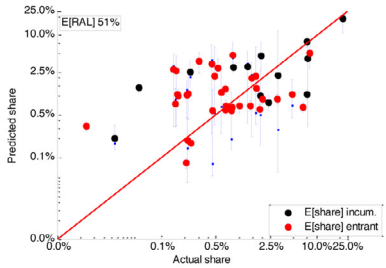
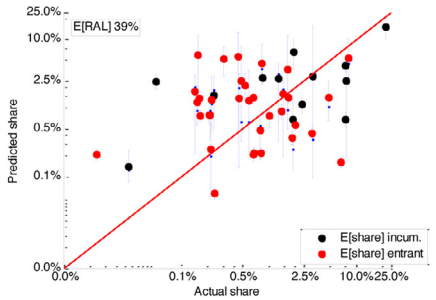
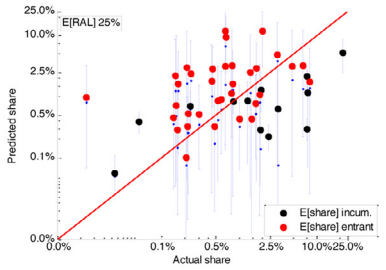
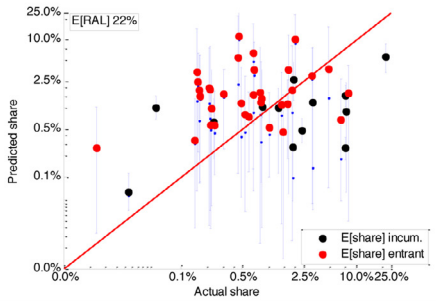
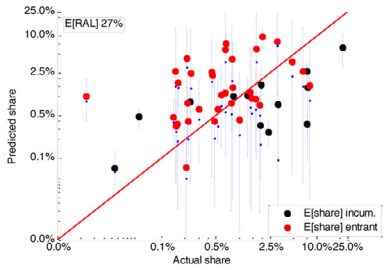
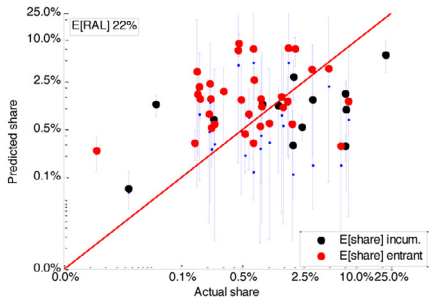
	MLE-C	GMM-IV
Actual v. predicted Method 0 (None)		
Actual versus predicted Method 1 (All)		
Actual versus predicted Method 2 (Nearest neighbor)		
Actual versus predicted Method 3 (Brand)		
Actual versus predicted Method 4 (Make-Model)		

Table G.2
Actual versus predicted 2.5–97.5% interval of shares of 5-year-forward predictions for the MLE-C and GMM-IV models using each of the ASC generation methods.

	MLE-C	GMM-IV
Actual versus predicted Method 0 (None)		
Actual versus predicted Method 1 (All)		
Actual versus predicted Method 2 (Nearest neighbor)		
Actual versus predicted Method 3 (Brand)		
Actual versus predicted Method 4 (Make-Model)		

References

Allcott, H., Wozny, N., 2014. Gasoline prices, fuel economy, and the energy paradox. Reviews of Economics and Statistics 96 (5), 779–795.

Andrews, R.L., Ebbes, P., 2014. Properties of instrumental variables estimation in logit-based demand models: Finite sample results. Journal of Modelling in Marketing Management.

Axsen, J., Kurani, K.S., 2011. Interpersonal influence in the early plug-in hybrid market: Observing social interactions with an exploratory multi-method approach. Transport Research Part D 16 (2), 150–159.

Axsen, J., Mountain., D.C., Jaccard, M., 2009. Combining stated and revealed choice research to simulate the neighbor effect: The case of hybrid-electric vehicles. Resources Energy Economics 31 (3), 221–238.

Balducci P. J., 2008, Plug-in hybrid electric vehicle market penetration scenarios, Richland, WA.

Beresteanu, A., Li, S., 2011. Gasoline prices, government support, and the demand for hybrid vehicles in the United States. International Economic Review 52 (1), 161–182.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.

Berry, S., Levinsohn, J., Pakes, A., 1999. Voluntary export restraints on automobiles: Evaluating a trade policy. American Economic Review 89 (3), 400–430.

Berry, S., Levinsohn, J., Pakes, A., 2004. Differentiated products demand systems from a combination of micro and macro data: The new car market. Journal of Political Economy 112 (1), 68–105.

Berry, S., 1994. Estimating discrete-choice models of product differentiation. RAND Journal of Economics 25 (2), 242–262.

Bunch D., Greene D., Lipman T., and Shaheen S., 2011, Potential design, implementation, and benefits of a feebate program for new passenger vehicles in California, Davis, CA.

Chintagunta, P.K., 2001. Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. Marketing Science 20 (4), 442–456.

Choo, S., Mokhtarian, P.L., 2004. What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. Transport Research Part A 38 (3), 201–222.

Copeland, A., Dunn, W., Hall, G., 2011. Inventories and the automobile market. RAND Journal of Economics 42 (1), 121–149.

Dasgupta, S., Siddarth, S., Silva-Risso, J., 2007. To lease or to buy? A structural model of a consumer's vehicle and contract choice decisions. Journal of Marketing Research 44 (3), 490–502.

Dubé, J.-P., Fox., J.T., Su, C.-L., 2012. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. Econometrica 80 (5), 2231–2267.

Duvall M., and Knipping E., 2007, Environmental assessment of plug-In hybrid electric vehicles, Palo Alto, CA.

Fiebig, D.G., Keane, M.P., Louviere, J., Wasi, N., 2009. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Marketing Science 29 (3), 393–421.

Frischknecht, B.D., Whitefoot, K., Papalambros, P.Y., 2010. On the suitability of econometric demand models in design for market systems. Journal of Mechanical Design 132 (12).

Gigerenzer, G., Brighton, H., 2009. Homo heuristicus: Why biased minds make better inferences. Topics in Cognitive Science 1 (1), 107–143.

Gilbride., T.J., Allenby, G.M., 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. Marketing Science 23 (3), 391–406.

Greene D. L., Duleep. K. G., and McManus W., 2004, Future potential of hybrid and diesel powertrains in the U.S. light-duty vehicle market, Oak Ridge, TN.

Greene, D.L., Patterson, P.D., Singh, M., Li, J., 2005. Feebates, rebates and gas-guzzler taxes: A study of incentives for increased fuel economy. Energy Policy 33 (6), 757–775.

Greene, W.H., 2003. Econometric Analysis. Prentice Hall, Upper Saddle River, NJ.

Haaf, C.G., Michalek, J.J., Morrow, W.R., Liu, Y., 2014. Sensitivity of vehicle market share predictions to discrete choice model specification. Journal of Mechanical Design 126 (12).

Hauser, J., Wernerfelt, B., 1990. An evaluation cost model of consideration sets. Journal of Consumer Research 16 (4), 393–408.

He, L., Chen, W., Hoyle, C., Yannou, B., 2012. Choice modeling for usage context-based design. Journal of Mechanical Design 134 (3), 031007.

Heffner, R.R., Kurani., K.S., Turrentine, T.S., 2007. Symbolism in California's early market for hybrid electric vehicles. Transport Research Part D 12 (6), 396–413.

Jenn, A., Azevedo, I., Ferreira, P., 2013. The impact of federal incentives on the adoption of hybrid electric vehicles in the United States. Energy Economics 40, 936–942.

Klier, T., Linn, J., 2012. New vehicle characteristics and the cost of the corporate average fuel economy standard. RAND Journal of Economics 43 (1), 186–213.

Knittel, C.R., Metaxoglou, K., 2012. Estimation of random coefficient demand models: Two empiricists' perspective. Review of Economics and Statistics.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. The Annals of Mathematical Statistics 22 (1), 79–86.

Lave, C., Train, K., 1979. A disaggregate model of auto-type choice. Transport Research Part A 13 (1), 1–9.

Li, S., Liu, Y., Zhang, J., 2011. Lose some, save some: Obesity, automobile demand, and gasoline consumption in the United States. Journal of Environmental Economics and Management 61 (1), 52–66.

Lin, Z., Greene, D., 2010. A plug-in hybrid consumer choice model with detailed market segmentation. In: Transportation Research Board 89th Annual Meeting. Washington, D.C.

Lucas, Robert E., 1976. Econometric policy evaluation: A critique. In: Carnegie-Rochester Conference Series on Public Policy. Elsevier Science Publishers, North-Holland, MI, pp. 19–46.

MacDonald, E.F., Gonzalez, R., Papalambros, P.Y. , 2009. Preference inconsistency in multidisciplinary design decision making. Journal of Mechanical Design 131 (3).

Min, J., Azevedo, I.L., Michalek, J., Bruine de Bruin, W., 2014. Labeling energy cost on light bulbs lowers implicit discount rates. Ecological Economics 97, 42–50.

Morrow, W.R., Long, M., MacDonald, E.F., 2012. Consider then choose models in decision-based design optimization. In: Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference. American Society of Mechanical Engineers, Chicago, IL, pp. 405–422.

Nevo, A., 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. Journal of Economic and Management Strategy 9 (4), 513–548.

Park, S., Gupta, S., 2011. Comparison of SML and GMM estimators for the random coefficient logit model using aggregate data. Empirical Economics 43 (3), 1353–1372.

Petrin, A., 2002. Quantifying the benefits of new products: The case of the minivan. Journal of Political Economy 110 (4), 705–729.

Polk, 2012, "Automotive Sales Data," IHS Polk Insight Yearly Vol. Data [Online]. Available: http://blog.polk.com/blog/new-vehicle-sales, last accessed 2014.

Rossi, P., 2014. Invited paper - Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. Marketing Science 33 (5), 655–672.

Su, C.-L., Judd, K.L., 2012. Constrained optimization approaches to estimation of structural models. Econometrica 80 (5), 2213–2230.

Sudhir, K., 2001. Competitive pricing behavior in the auto market: A structural analysis. Marketing Science 20 (3), 42–60.

Swait, J., Louviere, J., Williams, M., 1994. A sequential approach to exploiting the combined strengths of SP and RP data: Application to freight shipper choice. Transportation 21 (2), 135–152.

Train, K., Winston, C., 2007. Vehicle choice behavior and the declining market share of U.S. automakers. International Economic Review 48 (4), 1469–1497.

Train, K.E., 2009. Discrete Choice Models with Simulation. Cambridge University Press, New York, NY.

U.S. Energy Information Administration, 2011, Annual energy outlook 2011 with projections to 2035, Washington, D.C.

Vance C., and Mehlin M., 2009, Tax policy and CO2 emissions- An econometric analysis of the German automobile market, Essen, Germany.

Villas-Boas, J.M., Winer, R.S., 1999. Endogeneity in brand choice models. Management Science 45 (10), 1324–1338.
Ward's Automotive Group, 2012, "Ward's Yearly Vehicle Specifications," [Online]. Available: http://wardsauto.com/data-center. Last accessed 2014.
Whitefoot, K.S., Skerlos, S.J., 2012. Design incentives to increase vehicle size created from the u.s. footprint-based fuel economy standards. Energy Policy 41 (1), 402–411.
Wooldridge, J.M., 2006. Introductory Econometrics: A Modern Approach. Thomas Higher Education, Mason, OH.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, Massachusetts.