

DEPARTMENT OF ECONOMICS

**Improving the performance of random
coefficients demand models:
The role of optimal instruments**

Mathias Reynaert & Frank Verboven

UNIVERSITY OF ANTWERP
Faculty of Applied Economics



Stadscampus
Prinsstraat 13, B.226
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
<http://www.ua.ac.be/tew>

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ECONOMICS

Improving the performance of random coefficients demand models: The role of optimal instruments

Mathias Reynaert & Frank Verboven

RESEARCH PAPER 2012-011
JUNE 2012

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be

The papers can be also found at our website:
www.ua.ac.be/tew (research > working papers) &
www.repec.org/ (Research papers in economics - REPEC)

D/2012/1169/011

Improving the Performance of Random Coefficients Demand Models: the Role of Optimal Instruments

Mathias Reynaert and Frank Verboven*

June 2012

Abstract

We shed new light on the performance of Berry, Levinsohn and Pakes' (1995) GMM estimator of the aggregate random coefficient logit model. Based on an extensive Monte Carlo study, we show that the use of Chamberlain's (1987) optimal instruments overcomes most of the problems that have recently been documented with standard, non-optimal instruments. Optimal instruments reduce small sample bias, but prove even more powerful in increasing the estimator's efficiency and stability. Other recent methodological advances (MPEC, polynomial-based integration of the market shares) greatly improve computational speed, but they are only successful in terms of bias and efficiency when combined with optimal instruments.

**Mathias Reynaert*: University of Leuven, University of Antwerp and Ph.D. fellow of the Research Foundation Flanders (FWO) Email: Mathias.Reynaert@ua.ac.be or Mathias.Reynaert@econ.kuleuven.be *Frank Verboven*: University of Leuven and CEPR. Email: Frank.Verboven@econ.kuleuven.be. We would like to thank Benjamin Skrainka, Tim Armstrong, Geert Dhaene, Jeremy Fox, and Laura Grigolon for very useful comments and discussions. We made use of the code kindly provided online by Dubé, Fox and Su (MPEC) and by Heiss and Winschel (integration). We gratefully acknowledge financial support from University of Leuven Program Financing / Center of Excellence Grant, and from the Research Foundation - Flanders (FWO).

1 Introduction

Discrete choice models have a long tradition in empirical research. They were originally developed to analyze consumer choices with micro-level data (McFadden (1974)). In two important contributions, Berry (1994) and Berry, Levinsohn and Pakes (1995) (henceforth BLP) develop a random coefficients logit demand model that can be estimated with aggregate data on sales, prices and product characteristics. Since the random coefficients account for unobserved heterogeneity in consumer valuations of product characteristics, they create flexible substitution patterns between products. Since Nevo (2000, 2001), the aggregate random coefficients logit model has become increasingly popular in industrial organization, marketing, international trade, environmental economics and many other areas in economics and management.

BLP's random coefficients logit model generates a non-linear aggregate market share system. BLP show how to invert the system to solve for the product-specific unobservables and estimate the model using a generalized methods of moments (GMM) estimator. In recent years, several papers have documented numerical difficulties with BLP's approach, and attempted to formulate solutions, often based on Monte Carlo studies. Knittel and Metaxoglou (2012) focus on global convergence problems associated with the non-linearity of the model, including the role of starting values and optimization algorithms. Dubé, Fox and Su (2012) assess the performance of BLP's contraction mapping, which is a nested fixed point (NFP) algorithm to invert the market share system. As an alternative, they propose an approach called mathematical programming with equilibrium constraints (MPEC). This algorithm essentially eliminates the inner loop contraction mapping and instead minimizes the GMM objective function subject to the market share system as constraints. Skrainka and Judd (2011) focus on problems with pseudo Monte Carlo integration in the market share equations, and propose several numerical integration methods as alternatives. Skrainka (2012) gives an overview of computational problems and discusses small sample bias of the GMM estimator. Finally, Armstrong (2012) focuses on the instruments used in the GMM approach to account for price endogeneity. He shows that BLP's functions of characteristics across products may not be good price instruments in certain demand models, and suggests the use of more traditional (but less easily available) cost shifters as price instruments.

While this recent work has given interesting new insights, there are many open questions. In particular, the identification of the variances of the random coefficients proves difficult in both Monte Carlo studies and in applications, despite recent theoretical identification results (Berry, Gandhi and Haile (2011), Fox et al. (2012) and Fox and Gandhi (2012)). In this paper, we show that many of the difficulties relate to the use of inefficient instruments for estimating the variances of the random coefficients. Based on several Monte Carlo simulations, we document that Chamberlain's (1987) optimal instruments solve most of the problems reported in previous Monte Carlo studies. Chamberlain's (1987) optimal set of instruments consists of the expected value of the derivatives of

the structural error term (the product-specific unobservable) with respect to the parameter vector, evaluated at an initial estimate of the parameters. Intuitively, this is the most efficient set of instruments, out of an infinite set of possible functions of characteristics across products. BLP and most subsequent work used sums of characteristics of other products as instruments. Interestingly, BLP already implemented an approximation to Chamberlain’s optimal instruments in their follow-up application (see Berry, Levinsohn and Pakes (1999)). But to our knowledge it has only been applied in one other application (Goeree (2008)), and it has not been incorporated in the Monte Carlo studies that have recently documented difficulties with BLP’s GMM approach.

Our main results are in section 4, where we compare standard and optimal instruments in terms of bias and efficiency performance. We first consider the performance of the GMM estimator with a standard, non-optimal set of instruments. We take into account all advances that have been proposed in previous work: accurate numerical integration in the market share system, the MPEC algorithm instead of a NFP algorithm to invert the market share system, the use of cost shifters as price instruments, and careful checking of starting values. Despite all these precautionary measures, we find that the parameters are estimated with bias. But more importantly, the estimates are rather imprecise and unstable, with a high root mean squared error and a spike at zero for the variance of the random coefficient.

We then incorporate Chamberlain’s (1987) optimal instruments. Implementation requires an initial estimate of the parameters. We consider two approaches. First, we evaluate the instruments at the parameter estimates of a first-stage random coefficients logit model using standard, inefficient instruments (i.e. the parameters obtained in the previous paragraph). This is essentially the approach in Berry, Levinsohn and Pakes’ (1999) application (although they only consider an approximation). Second, we evaluate the instruments at the parameter estimates of a simple linear logit model, with a guess for the nonlinear parameters (the variances of the random coefficients). Both approaches eliminate the bias, but our most striking finding is that they drastically improve the efficiency and stability of the parameter estimates.

These results refer to a situation where strong instruments (“cost shifters”) are available to account for price endogeneity. But we also consider a situation where only weak instruments for price are available. Chamberlain’s (1987) optimal instruments still outperform standard instruments, with the strongest efficiency gains for the nonlinear parameter (the variance of the random coefficient). Finally, we consider bias as the sample size increases. With optimal instruments, the bias is decreasing in the sample size (number of products or number of markets). With standard instruments this is not the case, as documented earlier by Skrainka (2012).

We subsequently consider several extensions in section 5. We compare standard and optimal instruments in terms of their computational performance, and reconsider the above discussed recent advances by Dubé, Fox and Su (2012), Skrainka and Judd (2011) and Knittel and Metaxoglou (2012). First, with optimal instruments MPEC and NFP give identical estimates, but MPEC is

substantially faster. In contrast, with standard instruments MPEC and NFP give different results. Second, with optimal instruments, accurate polynomial-based integration (Sparse Grid quadrature) is faster and implies higher precision than pseudo Monte Carlo integration in the market shares. The opposite is true with standard instruments. Finally, under optimal instruments the estimates are much less sensitive to starting values than under standard instruments, the case documented by Knittel and Metaxoglou (2012). For example, the standard deviation of the estimates under 10 starting values decreases by a factor of 5 under optimal instruments. The remaining sensitivity to starting values reflects the inherent non-linearity of the GMM objective function as implied by the model, and warrants careful searching for a global optimum, even under optimal instruments.

In sum, the use of Chamberlain’s (1987) optimal instruments reduces bias and drastically improves the efficiency and stability of the parameter estimates. It also explains why earlier methodological advances still produce ambiguous findings. It is instructive to compare these results with other work on instruments in a GMM context, and with Bayesian work as an alternative to GMM.

First, Armstrong (2012) also focused on the role of instruments in aggregate random coefficients logit models. Armstrong (2012) only considers the instruments for the endogenous price variable (in a more general setting of imperfect competition). As mentioned above, he finds that cost-shifters perform better as price instruments than BLP’s proposed functions of characteristics across products. Armstrong (2012) does not, however, consider the instruments to identify the variances of the random coefficients. In fact, in his Monte Carlo simulations he sets these variances to their known values. In contrast, we focus precisely on the instruments that are required to identify the variances of the random coefficients. For these parameters, BLP’s proposed functions of product characteristics prove essential, in particular when Chamberlain’s (1987) optimal set of instruments is used. In applied work, researchers should thus search for good cost-side instruments to identify the price parameter (as stressed by Armstrong), and apply optimal instruments to identify the nonlinear parameters (the variances of the random coefficients).

Second, in simpler nested logit models researchers have already commonly used more refined functions of product characteristics than BLP’s original sums of characteristics of other products. Verboven (1996) proposed to use sums of characteristics of other products by nest (and subnest). Bresnahan, Stern and Trajtenberg (1997) apply similar instruments in their “principles of differentiation” GEV model. These instruments arise naturally in applications where the random coefficients (nesting parameters) refer to discrete characteristics (the nests). Since they typically result in fairly precise parameter estimates, they turn out to be good approximations to Chamberlain’s (1987) optimal instruments.

Third, Jiang, Manchanda and Rossi (2009) propose a Bayesian estimator to estimate the aggregate random coefficients model. We also implemented a Bayesian estimator in our Monte Carlo studies. We confirm that this approach also produces more precise estimates than GMM with non-optimal instruments (results available upon request). The Bayesian estimator is however con-

siderably slower, and it involves other trade-offs, since the Bayesian approach relies on stronger assumptions such as functional forms and supply side assumptions, as discussed in Berry (2003).

The outline of the paper is as follows. Section 2 discusses the random coefficients logit model for aggregate demand data, including the GMM estimator and the set of optimal instruments. Section 3 sets out the data-generating process for our Monte Carlo study and provides computational details for estimating the model. Section 4 presents our main results on the bias and efficiency with optimal instruments. Section 5 considers further extensions, covering the computational performance with optimal instruments. Section 6 summarizes and concludes with some cautionary warnings.

2 The Model

We first describe the random coefficients logit model for aggregate demand data. We then discuss marginal costs and the perfectly competitive market equilibrium. Next, we present the GMM estimator and the construction of the set of optimal instruments. Finally, we describe the simulated data-generating process.

2.1 Demand

There are T markets, indexed by $t = 1, \dots, T$. In each market t there are L_t potential consumers. Each consumer i chooses one alternative j , which is either the outside good, $j = 0$, or one of the J differentiated products, $j = 1, \dots, J$. Consumer i 's conditional indirect utility for the outside good is $u_{i0t} = \varepsilon_{i0t}$, and for products $j = 1, \dots, J$ it is:

$$u_{ijt} = x_{jt}\beta_i - \alpha p_{jt} + \xi_{jt} + \varepsilon_{ijt}, \quad (1)$$

where x_{jt} is a $1 \times K$ vector of observed product characteristics, p_{jt} is the price and ξ_{jt} is an unobserved product characteristic of product j in market t , unobserved by the researcher but observed by consumers and firms. The $K \times 1$ parameter vector β_i consists of random coefficients, capturing individual-specific valuations for the product characteristics, α is the marginal utility of income or price valuation (assumed to be equal for all consumers i), and ε_{ijt} is a remaining individual-specific valuation for product j . The random coefficient for characteristic k is given by $\beta_i^k = \beta^k + \sigma^k \nu_i^k$, where ν_i^k is a random variable with zero mean and unit variance, so that β^k represents the mean valuation for characteristic k and σ^k is its standard deviation across consumers. The individual-specific valuation for product j , ε_{ijt} , is an i.i.d. random variable with a type I extreme value distribution. We can write consumer i 's conditional indirect utility (1) as

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \sum_k x_{jt}^k \sigma^k \nu_i^k + \varepsilon_{ijt}. \quad (2)$$

Indirect utility can thus be decomposed into the sum of three terms: a mean utility term $\delta_{jt} \equiv x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ common to all consumers; an individual-specific utility term $\mu_{jt}(\nu_i) \equiv \sum_k x_{jt}^k \sigma^k \nu_i^k$; and an individual-specific utility term ε_{ijt} specific to each product j . If $\sigma^k = 0$ for all k , we obtain the standard logit model.

Each consumer i in market t chooses the alternative $j = 0, \dots, J$ that maximizes her utility. The predicted market share of product j in market t is the probability that product j yields the highest utility across all available products (including the outside good 0). This is given by the well-known logit choice probability, integrated over the individual-specific valuations for the continuous characteristics:

$$s_{jt}(\delta_t, \sigma) = \int \frac{\exp(\delta_{jt} + \mu_{jt}(\nu))}{1 + \sum_{l=1}^J \exp(\delta_{lt} + \mu_{lt}(\nu))} dP_\nu(\nu), \quad (3)$$

where δ_t is the $J \times 1$ mean utility vector in market t (dependent on the mean valuation parameters β and α), and σ is the vector of standard deviations around the mean valuations. In empirical work, the integrals are often approximated through m Monte Carlo draws of ν from the standard normal distribution:

$$s_{jt}(\delta_t, \sigma) \approx \frac{1}{m} \sum_{i=1}^m \frac{\exp(\delta_{jt} + \mu_{jt}(\nu_i))}{1 + \sum_{l=1}^J \exp(\delta_{lt} + \mu_{lt}(\nu_i))}. \quad (4)$$

An alternative approach uses more accurate polynomial-based integration such as Sparse Grid quadrature to approximate the integrals in (3):

$$s_{jt}(\delta_t, \sigma) \approx \sum_{i=1}^n \phi_i \frac{\exp(\delta_{jt} + \mu_{jt}(\nu_i))}{1 + \sum_{l=1}^J \exp(\delta_{lt} + \mu_{lt}(\nu_i))} \quad (5)$$

where n is the number of nodes for ν and ϕ_i is the weight associated with node ν_i . See Skrainka and Judd (2011) and Heiss and Winschel (2008) for a detailed discussion.

In our Monte Carlo analysis, we will base our main results on accurate Sparse Grid numerical integration. In an extension, we will also discuss the performance of more crude pseudo Monte Carlo integration of the market shares.

2.2 Costs and Market Equilibrium

Assume that the marginal costs of product j in market t are constant and given by

$$c_{jt} = x_{jt}\gamma_1 + w_{jt}\gamma_2 + \omega_{jt},$$

where x_{jt} is the above vector of product characteristics, which affects both utility and marginal cost, w_{jt} is a vector of other variables that only affect marginal cost, and ω_{jt} is an unobserved marginal cost component.

Most of the literature has adopted a model of imperfect competition, such as multiproduct

Bertrand competition as in BLP. We take a simpler approach and assume perfect competition, so that price equals marginal costs. In vector notation, the supply side in market t can then be described as

$$p_t = X_t\gamma_1 + W_t\gamma_2 + \omega_t. \quad (6)$$

The assumption of perfect competition has the advantage of simplicity and transparency. First, this gives a linear solution for the equilibrium price vector p_t , so we do not need to repeatedly solve a non-linear system of first-order conditions as under Bertrand pricing. Second, the competitive price solution (6) will sharply clarify the distinction between instruments to identify the price parameter (essentially cost shifters W_t , as stressed earlier by Armstrong (2012) in a more general setting of imperfect competition) and instruments to identify the variances of the random coefficients. Nevertheless, the supply side could in principle be extended to imperfect competition (such as Bertrand-Nash pricing).

To specify the demand side, we set the observed market share $s_{jt} = q_{jt}/L_t$ (aggregate quantity divided by number of potential consumers) equal to the predicted market share (3). In vector notation, the demand side in market t can then be described by the market share system

$$s_t = s_t(\delta_t, \sigma), \quad (7)$$

where

$$\delta_t \equiv X_t\beta - \alpha p_t + \xi_t. \quad (8)$$

2.3 GMM Estimator

We have a system of supply and demand equations (6) and (7), where ξ_t is the vector of demand unobservables and ω_t is the vector of cost unobservables in market t . Price enters as an endogenous variable in the demand system. It may be correlated with the demand error ξ_t , since it depends on ω_t , and ω_t and ξ_t may be correlated. BLP estimate the demand and supply side simultaneously, to increase the efficiency of the estimator. We follow much of the more recent literature and estimate the demand system (7) separately. We only use the supply side to generate additional instruments to account for the endogeneity of the price variable. These instruments are the cost shifters w_{jt} , which only enter marginal cost and not utility.

BLP and most of the subsequent literature estimated the demand system (7) using a nonlinear Generalized Methods of Moments (GMM) estimator, typically using a panel of markets $t = 1, \dots, T$. The main identification assumption is the conditional moment restriction

$$E [\xi_{jt}|X_t, w_{jt}] = 0,$$

i.e. the unobserved characteristic of product j is mean independent of the observed product char-

acteristics of all products and of the cost shifters of product j . With continuous characteristics X_t , the conditional moment restriction implies an infinite number of unconditional moment restrictions

$$E [\xi_{jt} z_{jt}] = 0, \quad (9)$$

where $z_{jt} = h_{jt}(X_t, w_{jt})$ are the instruments and $h_{jt}(X_t, w_{jt})$ is a vector-valued function of the product characteristics and exogenous cost shifters. For example, BLP include the own-product characteristics and the sums of the characteristics of other products (across all products and across products of the same firm of product j).

The vector of demand unobservables ξ_t enters highly non-linearly in the market share system (7). Berry (1994) shows how to solve the market share system analytically for ξ_t in simple examples, such as the logit model without random coefficients ($\sigma = 0$) or the nested logit model. In more general settings, Berry (1994) and BLP suggest to invert the market share system (7) numerically for δ_t (and therefore ξ_t). This gives

$$\delta_t = s_t^{-1}(s_t, \sigma) \equiv \delta_t(s_t, \sigma).$$

BLP show there is a contraction mapping to invert the market share system. Using (8), we can write the inverted market share system as

$$\xi_t = \delta_t(s_t, \sigma) - X_t \beta + \alpha p_t \equiv \xi_t(\theta)$$

where $\theta = (\beta, \alpha, \sigma)$ is the vector of demand parameters to be estimated.

BLP propose to apply GMM, based on the following minimization problem

$$\min_{\theta} \xi(\theta)' Z A Z' \xi(\theta),$$

where A is a weighting matrix and where the vectors and matrices are stacked over all markets t . BLP solve this minimization problem with a nested fixed point (NFP) algorithm, where the outer loop minimizes the objective function, and the inner loop solves a contraction mapping to obtain the inverted market share system $\delta_t(s_t, \sigma)$. The nonlinear minimization problem can be simplified after “substituting out” the first-order conditions with respect to the linear parameters β and α .

Dubé, Fox and Su (2012) propose an alternative formulation of the GMM problem. Instead of the NFP algorithm, they propose a mathematical program with equilibrium constraints (MPEC) approach. Their constrained minimization problem is

$$\min_{\theta, \xi} \xi' Z A Z' \xi$$

subject to $s(\delta, \sigma) = s$.

Su and Judd (2012) show that the NFP and MPEC algorithms give identical estimators. Dubé, Fox and Su (2012) characterize the numerical performance of NFP and MPEC in the context of the BLP model. They find that NFP requires a very tight convergence criterion for inverting the market share system (the inner loop) and that MPEC can be considerably faster.

In our Monte Carlo analysis, we will present our main results based on MPEC, and will provide a comparison with NFP in the extensions.

2.4 Instruments

Our primary interest is in the role of the instruments included in z_{jt} . We begin with three sets of instruments that have been used in most previous research, including the recent Monte Carlo studies on the performance of BLP's GMM estimator. The first set of instruments is $z_{jt}^1 = (x_{jt}, w_{jt})$, i.e. the set of observed product characteristics (that affect both demand and marginal cost) and an additional set of cost shifters (that do not directly affect demand). Armstrong (2012) shows that cost shifters can be more powerful to account for the endogeneity of price. The second set of instruments z_{jt}^2 adds polynomials to z_{jt}^1 (i.e. squares and interactions of x_{jt} and w_{jt}). Dubé, Fox and Su (2012) used this approach in their simulations. A third set of instruments z_{jt}^3 adds characteristics of other products as in BLP. More specifically, we add the sum of the characteristics of all other competitors to the first instrument set, so $z_{jt}^3 = (x_{jt}, w_{jt}, \sum_{k \neq j} x_{kt})$. This third set serves to assess to which extent BLP's instruments are useful in identifying the variances of the random coefficients σ (since Armstrong (2012) only focused on instruments to identify the price parameter α , and imposed σ to known values).

We next consider the performance of the set of optimal or efficient instruments, out of the infinite number of orthogonality conditions implied by (9). The set of optimal instruments results in an asymptotically efficient estimator. Amemiya (1977) obtained optimal instruments in nonlinear models, and Chamberlain (1987) finds that the optimal instruments attain the efficiency bound. See Arellano (2003) for an overview of optimal instruments in linear and nonlinear models. Berry, Levinsohn and Pakes (1999) propose an approximation to the optimal instruments for the random coefficients logit model. We will consider both their approximation and a more exact implementation of the optimal instruments. Chamberlain's (1987) optimal set of instruments is:

$$z_{jt} = E \left[\frac{\partial \xi_{jt}(\theta)}{\partial \theta'} \middle| X_t, w_{jt} \right].$$

The optimal set of instruments is therefore a vector of variables with the same dimension as the parameter vector $\theta = (\beta, \alpha, \sigma)$. Intuitively, these instruments exploit the functional forms behind

the model, in particular concerning the consumer heterogeneity that generates the market share system.

The optimal instruments for the linear parameters α and β are easy to interpret. In particular, we have

$$\begin{aligned} E \left[\frac{\partial \xi_{jt}(\theta)}{\partial \beta'} \middle| X_t, w_{jt} \right] &= E [x_{jt} | X_t, w_{jt}] = x_{jt} \\ E \left[\frac{\partial \xi_{jt}(\theta)}{\partial \alpha} \middle| X_t, w_{jt} \right] &= E [p_{jt} | X_t, w_{jt}] = x_{jt}\gamma_1 + w_{jt}\gamma_2. \end{aligned} \quad (10)$$

The optimal instrument for β is just x_{jt} , while the optimal instrument for α is the predicted price from a first-stage OLS regression on the linear competitive supply equation (6). Intuitively, the optimal instruments for the linear parameters are the same as those from the first stage in a 2SLS estimator. Note that these instruments do not depend on the demand parameters $\theta = (\beta, \alpha, \sigma)$, so they can be calculated without having to estimate the demand model in a first stage.

The optimal instruments for the nonlinear parameters σ are

$$E \left[\frac{\partial \xi_{jt}(\theta)}{\partial \sigma'} \middle| X_t, w_{jt} \right] = E \left[\frac{\partial \delta_{jt}(s_t, \sigma)}{\partial \sigma'} \middle| X_t, w_{jt} \right]. \quad (11)$$

These instruments are a non-linear function of the characteristics of all competing products. Since the expectation in (11) is a function of the true demand parameters $\theta = (\beta, \alpha, \sigma)$, the optimal instruments for σ are not feasible: they cannot be computed directly from the data and require a first stage estimate of the demand model. In the following we develop a parametric approach to calculate (11).

We first follow Berry, Levinsohn and Pakes' (1999) approximation. They replace the expected value of the derivatives in (11) by the derivatives evaluated at the expected value of the unobservables. More specifically, the procedure is as follows (in vector notation per market t):

1. Obtain an initial estimate $\hat{\theta} = (\hat{\beta}, \hat{\alpha}, \hat{\sigma})$ (for example, based on one of the three earlier inefficient sets).
2. Compute the predicted price $\hat{p}_t = X_t \hat{\gamma}_1 + W_t \hat{\gamma}_2$.
3. Compute the predicted mean utility $\hat{\delta}_t \equiv X_t \hat{\beta} - \hat{\alpha} \hat{p}_t$, and then the predicted market shares $\hat{s}_t = s_t(\hat{\delta}_t, \hat{\sigma})$.
4. Compute the Jacobian of the inverted market share system $\delta_t(\hat{s}_t, \sigma)$ evaluated at $\hat{\sigma}$:

$$\frac{\partial \delta_t(\hat{s}_t, \sigma)}{\partial \sigma'} \bigg|_{\sigma=\hat{\sigma}} \quad (12)$$

To compute the Jacobian of the mean utility with respect to σ , we differentiate the market share function (3) with respect to δ_t and σ' , and apply the implicit function theorem; see the appendix of Nevo (2000) for details. Note that this Jacobian is also used to optimize the GMM objective function and to compute the standard errors of the nonlinear parameters.

Next, as an alternative to the approximation of optimal instruments, we compute the exact expectation in (11). Since the structural error ξ_t enters the Jacobian nonlinearly, its distribution does not cancel out in the expectation. We can take this into account by the following procedure:

1. Obtain an initial estimate $\hat{\theta} = (\hat{\beta}, \hat{\alpha}, \hat{\sigma})$ and compute the density of the unobservable $\hat{\xi}_t$.
2. Compute the predicted price $\hat{p}_t = X_t \hat{\gamma}_1 + W_t \hat{\gamma}_2$.
3. Take K draws, $k = 1, \dots, K$, from the density of $\hat{\xi}_t$.¹ For each draw k , compute $\hat{\delta}_t^k \equiv X_t \hat{\beta} - \hat{\alpha} \hat{p}_t + \hat{\xi}_t^k$ and $\hat{s}_t^k = s_t(\hat{\delta}_t^k, \hat{\sigma})$.
4. For each draw k , compute the Jacobian of the inverted market share system $\delta_t(\hat{s}_t^k, \sigma)$ evaluated at $\hat{\sigma}$:

$$\left. \frac{\partial \delta_t^k(\hat{s}_t^k, \sigma)}{\partial \sigma'} \right|_{\sigma = \hat{\sigma}}$$

and compute the average across all K draws

$$\frac{1}{K} \sum_{k=1}^K \left. \frac{\partial \delta_t(\hat{s}_t^k, \sigma)}{\partial \sigma'} \right|_{\sigma = \hat{\sigma}}. \quad (13)$$

In our analysis below, we will use z_{jt}^4 and z_{jt}^5 to refer to the optimal instruments in, respectively, the approximate and “exact” approach. Note that the approximate and the exact approach will be identical only if $\partial \delta_t / \partial \sigma'$ is linear in $\hat{\xi}_t$, which is not generally the case.

Further intuition on the optimal instruments for σ can be obtained from the nested logit model, since this model has an analytical solution for the inverted market shares (although we do not consider this model in our Monte Carlo study). For the one-level nested logit model with nesting parameters σ_g for each group g , the inverted market share function is $\delta_{jt}(s_t, \sigma) = \ln s_{jt} / s_{0t} - \sigma_g \ln s_{j|g,t}$, where s_{0t} is the market share of the outside good, and $\ln s_{j|g,t}$ is the market share of product j within group g . We have $\partial \delta_{jt}(\hat{s}_t, \sigma) / \partial \sigma_g = -\ln \hat{s}_{j|g,t}$, so the optimal instrument for σ_g is the log of the predicted market share of the product within the group, either evaluated at $\hat{\xi}_t = 0$ (approximate approach) or averaged over the empirical density of $\hat{\xi}_t$ (exact approach). Note that the two-stage least squares estimator also uses the log of the predicted market shares within

¹We take k draws from the normal approximation to the distribution of $\hat{\xi}_t$, where the standard deviation $\hat{\sigma}_\xi$ is computed from the first stage. An alternative would be to sample directly from the empirical distribution $\hat{\xi}_t$. In case the distribution of ξ_t is not known, sampling from the empirical distribution is more appropriate. In our Monte Carlo simulations both approaches give nearly identical results.

the group as an instrument, but using a linear prediction instead of exploiting the nested logit functional forms.

To summarize, we first consider three sets of instruments based on standard approaches, z_{jt}^1 , z_{jt}^2 and z_{jt}^3 . We then consider two alternative sets of instruments, the approximate and exact implementation of Chamberlain’s (1987) optimal instruments, z_{jt}^4 and z_{jt}^5 . The optimal instruments also depend on the characteristics of all products, just like BLP’s original instruments (sums of characteristics of other products). But they do this in a way that better exploits the functional form implied by the model.

3 Monte Carlo Set-Up

3.1 Simulated Data-Generating Process

We consider a simple data-generating process. To demonstrate our main results on the performance of instruments, we construct 1000 different data sets for $T = 25$ markets, and $J = 10$ products. In some extensions we also consider the performance of instruments under varying sample sizes: $T = \{12, 25, 50\}$ with $J = 10$, and $J = \{5, 10, 20\}$ with $T = 25$. Each data set consists of the exogenous variables x_{jt} and w_{jt} and the endogenous variables s_{jt} and p_{jt} , as generated by the model and the demand and cost unobservables ξ_{jt} and ω_{jt} .

The model follows the assumptions set out in the previous subsections: random coefficients logit demand with competitive pricing. The vector of product characteristics that affects both utility and cost is $x_{jt} = (1, x_{jt}^1)$, where x_{jt}^1 is drawn from a uniform distribution $U(1, 2)$. The vector of additional characteristics that only shift cost is $w_{jt} = (w_{jt}^1, w_{jt}^2, x_{jt}^3)$, all independently drawn from a uniform distribution $U(0, 1)$. The unobserved demand and cost characteristics are drawn from a bivariate normal distribution

$$\begin{bmatrix} \xi_{jt} \\ \omega_{jt} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} \right).$$

To generate the endogenous price variable, we use the competitive price specification $p_{jt} = x_{jt}\gamma_1 + w_{jt}\gamma_2 + \omega_{jt}$, and set the cost parameters equal to $\gamma_1 = (0.7, 0.7)$ and $\gamma_2 = (3, 3, 3)$. The higher values for γ_2 ensure that the cost shifters w_{jt} have a strong impact on prices. They work as strong instruments for the endogenous price variable as in Armstrong (2012). In an extension we also consider the case where the cost shifters are weak instruments for price, by setting $\gamma_2 = (0.3, 0.3, 0.3)$.

To generate the endogenous market shares, we set the mean valuations of the exogenous product characteristics $x_{jt} = (1, x_{jt}^1)$ equal to $\beta = (2, 2)$ and their standard deviations to $\sigma = (0, 1)$. Hence, there is only consumer heterogeneity for the first product characteristic, and not for the constant.

The mean valuation for the endogenous product characteristic price p_{jt} is set equal to $\alpha = -2$. There is no consumer heterogeneity for the valuation of price in our main analysis, but we briefly consider this as an extension. The mass of consumers is $L_t = 1$ in each market t .

3.2 Computational Details

We minimize the GMM objective function using the Knitro 800 Interior\Direct algorithm in Matlab. Following Dubé, Fox and Su (2012), we use the MPEC algorithm with an analytic Jacobian and Hessian of both the objective function and the equilibrium constraints. For each of the 1,000 generated data sets, we estimate the model using 10 different starting values. So in practice we estimate the model 10,000 times (and we do this for five different instrument sets, and various data-generating processes). For each of the 1,000 generated datasets, we select the results based on the starting values that give the lowest value for the objective function. We approximate the market share integrals (3) using a Sparse Grid quadrature rule as given by (5), where ϕ_i are appropriate weights (see Heiss and Winschel (2008)). We use 9 nodes so that the unidimensional integral is exact for polynomials up to degree 17.

In our extensions (section 5), we compare the MPEC with the NFP algorithm (using a very tight convergence criterion for the contraction mapping, following Dubé, Fox and Su (2012)). Furthermore, we then also consider alternative approximations of the market share integrals using pseudo Monte Carlo (pMC) integration as given by (4).

4 Bias and Efficiency: Standard versus Optimal Instruments

We first compare the performance of standard and optimal instruments when cost-shifters are strong instruments for price (sections 4.1 and 4.2). We then consider a perhaps more typical situation where only weak instruments for price are available (section 4.3). Finally, we consider small and large sample performance of the optimal instruments (section 4.4).

4.1 Monte Carlo Results with Standard Instruments

As already discussed in section 2, we consider three “standard” instrument sets, similar to the instruments used in earlier Monte Carlo studies and most applied work. The first set of instruments is $z_{jt}^1 = (x_{jt}, w_{jt})$, where $x_{jt} = (1, x_{jt}^1)$ and $w_{jt} = (w_{jt}^1, w_{jt}^2, w_{jt}^3)$. The three cost shifters aim to identify the endogenous price effect α (as suggested by Armstrong (2012)) and the standard deviation of the random coefficient σ . The second set z_{jt}^2 adds polynomials of x_{jt} and w_{jt} (squares and interactions), resembling the instrument set of Dubé, Fox and Su (2012). Finally, the third set is $z_{jt}^3 = (x_{jt}, w_{jt}, \sum_{k \neq j} x_{kt})$, so it adds BLP’s sums of other product characteristics. As compared with the first instrument set, the cost shifters may prove more useful to identify α , whereas the

Table 1: Bias and Efficiency with Standard Instruments

	True	z_{jt}^1		z_{jt}^2		z_{jt}^3	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
β^0	2	0.295	1.739	-0.087	0.787	0.115	1.175
β^1	2	-0.291	1.584	0.032	0.617	-0.119	1.097
α	-2	-0.028	0.137	0.007	0.053	-0.009	0.075
σ^1	1	0.088	1.316	-0.100	0.549	0.011	0.817

Bias is the average parameter estimate minus the true parameter, over the 1000 generated data sets. RMSE is the root mean squared error. Estimates are based on the MPEC algorithm and Sparse Grid integration. The three “standard” instrument sets are $z_{jt}^1 = (x_{jt}, w_{jt})$; $z_{jt}^2 = z_{jt}^1$ plus polynomial squares and interactions of x_{jt} and w_{jt} ; and $z_{jt}^3 = (x_{jt}, w_{jt}, \sum_{k \neq j} x_{kt})$.

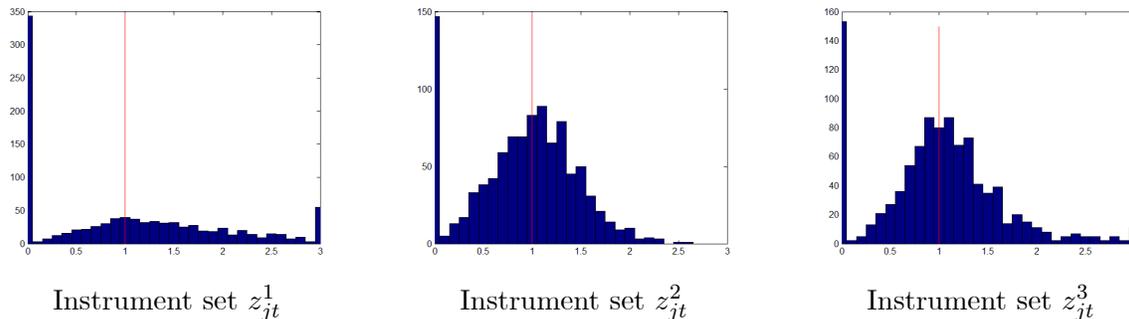
BLP instruments may be more useful to identify σ . For all three instrument sets, the cost shifters are strong instruments for price, i.e. they have a strong impact on marginal costs ($\gamma_2 = (3, 3, 3)$, as discussed in section 2.5).

Table 1 reports the bias and root mean squared error (RMSE) of all parameter estimates obtained from the Monte Carlo simulations. The bias is simply the average parameter estimate (over the 1,000 generated data sets) minus the true parameter value. Most parameters have moderate small sample bias. The first instrument set z_{jt}^1 gives the largest bias, roughly about 10% (e.g. bias of 0.088 for σ while its true value is $\sigma = 1$). The second and third instrument sets result in somewhat lower bias for most parameters, suggesting some identifying power of the additional instruments (polynomials of x_{jt} and w_{jt} in z_{jt}^2 and sums of characteristics $\sum_{k \neq j} x_{kt}$ in z_{jt}^3).

While standard instruments result in fairly moderate bias, the efficiency of the GMM estimator appears to be highly problematic. Table 1 shows that there is a very high RMSE for most parameters. Consider the first instrument set z_{jt}^1 . For the first product characteristic x_{jt}^1 , the mean valuation $\beta^1 = 2$ has a RMSE of 1.584, and the standard deviation $\sigma^1 = 1$ has a RMSE of 1.316. For the constant, the mean valuation $\beta^0 = 2$ has an equally large RMSE of 1.739. The only parameter with a reasonably low RMSE is α , the mean price valuation (RMSE of 0.137 while $\alpha = -2$). This is because the cost shifters are strong instruments and price is responsible for most variation in market shares in our data-generating process. The RMSE’s drop by a factor of more than 2 in the second instrument set z_{jt}^2 (with the polynomials) and by a factor of slightly less than 2 in the third instrument set (z_{jt}^3). This again suggests that the additional instruments have some extra identifying power. Yet for both instruments sets, the RMSE remains on average higher than half of the true parameter value (with the exception of the price parameter).

Figure 1 visualizes the problems with the efficiency of the GMM estimator under standard instruments, and also reveals a new problematic fact. The figure shows three histograms (one for each instrument set) for the taste parameter σ^1 , based on the 1000 generated data sets. Consistent

Figure 1: Histograms for $\hat{\sigma}^1$ with standard instruments as in Table 1



with Table 1, there is moderate bias (with a peak of the distribution around the true value $\sigma^1 = 1$) but a very high dispersion and fat tails, especially for the first instrument set z_{jt}^1 . Figure 1 also reveals another striking problem: there is a large spike in the distribution for the estimates of σ^1 around 0.² Estimates close to 0 occur in 30% of the cases with the first instrument set z_{jt}^1 , and in about 15% of the cases for the second and third instrument sets z_{jt}^2 and z_{jt}^3 .

To summarize, all three sets of “standard” instrument sets result in parameter estimates that are moderately biased, highly imprecise and unstable (in the sense of spikes in the distribution of σ^1 around 0). These findings are consistent with other recent studies that report sufficient detail on point estimates, bias or RMSE, such as Knittel and Metaxoglou (2012) and Skrainka (2012). We now turn to Chamberlain’s (1987) optimal instruments, and show that the efficiency and stability of the estimates drastically improves.

4.2 Monte Carlo Results with Optimal Instruments

As discussed in detail in section 2, Chamberlain’s (1987) optimal set of instruments consists of the expected value of the derivatives of the structural error term with respect to the parameter vector. Intuitively, the optimal instruments are as the third set of “standard” instruments, but where the sums of other product characteristics are replaced by either (12) for Berry, Levinsohn and Pakes’ (1999) approximation, or by (13) for our exact implementation. We label the approximate set of optimal instruments by z_{jt}^4 and the exact set by z_{jt}^5 .

In nonlinear models, the optimal set of instruments is not feasible since it depends on the parameters. We therefore need to perform a first stage to obtain an initial estimate of these parameters. We consider two approaches for the first stage. First, we obtain an initial estimate $(\hat{\beta}, \hat{\alpha}, \hat{\sigma})$ for the random coefficients logit model based on standard instruments, i.e. the earlier set z_{jt}^1 . Second, we obtain an initial estimate of the linear parameters $\hat{\beta}$ and $\hat{\alpha}$ from a simple logit model (with $\sigma = 0$)

²It is interesting to note that the spikes around 0 increase when we set the true σ_1 to lower values.

Table 2: Bias and Efficiency with Optimal Instruments

Parameter	True	z_{jt}^4		z_{jt}^5	
		Bias	RMSE	Bias	RMSE
β^0	2	0.000	0.466	-0.003	0.446
β^1	2	-0.003	0.328	0.001	0.278
α	-2	-0.0005	0.042	-0.0004	0.042
σ^1	1	-0.001	0.192	-0.003	0.142

Bias is the average parameter estimate minus the true parameter, over the 1000 generated data sets. RMSE is the root mean squared error. Estimates are based on the MPEC algorithm and Sparse Grid integration. The two instrument sets z_{jt}^4 and z_{jt}^5 are Chamberlain’s (1987) optimal instruments, i.e. x_{jt} for β , \hat{p}_{jt} for α , and (11) for σ , evaluated at first stage random coefficients logit estimates with standard instruments z_{jt}^1 . The first set z_{jt}^4 is an approximation of (11) evaluated at $\hat{\xi} = 0$, the second set z_{jt}^5 is an “exact implementation” using an estimated normal distribution for ξ .

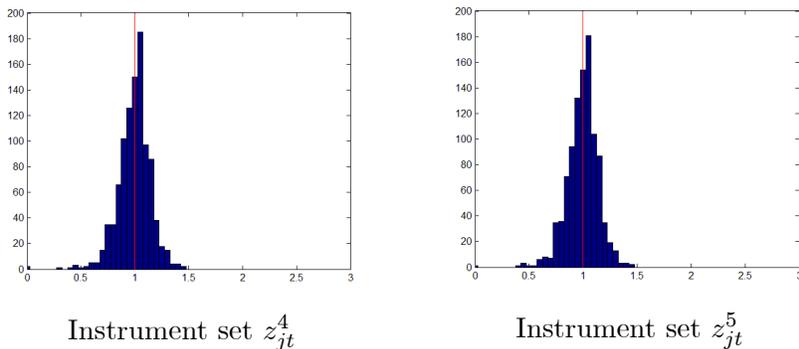
and use a guess for $\hat{\sigma}$. The second approach does not seem standard, but it avoids a computationally expensive first stage (which in any case leads to rather imprecise estimates).

First stage based on random coefficients logit Table 2 shows the bias and RMSE of all parameters for the approximate set (z_{jt}^4) and the exact set (z_{jt}^5) of instruments, evaluated at an initial estimate $(\hat{\beta}, \hat{\alpha}, \hat{\sigma})$ based on standard instruments z_{jt}^1 . The table shows several striking facts.

First, small sample bias decreases by a factor of more than 10, even when compared to the results from z_{jt}^2 , which gave the lowest bias among the three standard instrument sets in Table (1). Second, there is a spectacular increase in the efficiency of the estimates. The RMSE decreases by a factor of at least 2 for the linear parameters β . But the RMSE especially decreases for the nonlinear taste parameter $\sigma^1 = 1$. While the standard instruments gave a RMSE of 1.316 (z_{jt}^1), 0.549 (z_{jt}^2) and 0.817 (z_{jt}^3), the optimal instruments result in a RMSE of 0.192 (z_{jt}^4) and 0.142 (z_{jt}^5). The RMSE decreases only slightly for the price valuation α . This is because the price parameter was already estimated precisely, since the standard deviation for this characteristic was set to zero.

Figure 2 confirms these findings and gives interesting new insights. Compared with the histograms in Figure 1, it is immediately clear that the distribution of σ^1 has a much sharper peak. Furthermore, the spikes around $\sigma^1 = 0$ have almost completely disappeared, so the estimator becomes considerably more stable.

Figure 2: Histograms for $\hat{\sigma}^1$ with optimal instruments as in Table 2



Note that the exact instruments (z_{jt}^5) improve the efficiency compared to the approximate instruments (z_{jt}^4), but the gains are small. Furthermore, we also performed a third stage, i.e. re-estimated the model with optimal instruments based on the new estimates (results not reported). While such a third stage further reduces the RMSE, the efficiency improvements are very small. This was also true for further stages we considered. Hence, it appears a second stage is sufficient to implement optimal instruments, and further stages do not justify the additional computational burden.

To summarize, an optimal set of instruments drastically reduces parameter bias. But the gains in the efficiency and stability of the estimates are perhaps even more spectacular, in light of the problems with standard instruments reported earlier. The gains especially relate to the estimation of σ^1 , the standard deviation of the taste parameter. Intuitively, Chamberlain’s (1987) optimal instruments prove especially powerful to identify the nonlinear parameters, as they exploit the nonlinear functional forms of the model. To obtain these gains, it is sufficient to perform a second stage, and no further stages appear necessary. Nevertheless, a computationally costly first stage with inefficient instruments is required to obtain the optimal instruments. We now show how this may be avoided by implementing a much simpler linear first stage.

First stage based on simple logit We reconsider the approximate (z_{jt}^4) and the exact instrument set (z_{jt}^5), but now evaluated at an initial estimate $(\hat{\beta}, \hat{\alpha})$ of the simple logit model (with $\sigma^1 = 0$) and with a guess for $\hat{\sigma}^1$, i.e. the absolute value from a draw of the standard normal distribution.³ The optimal instruments for the linear parameters β and α obviously remain the same, since they do not depend on the demand parameters. From (10), they are just equal to x_{jt} and \hat{p}_{jt} . The optimal instrument for σ^1 will be different, but it will still be a function of the product

³We also considered alternative initial valuations for $\hat{\sigma}$, i.e. $\hat{\sigma} \approx 0$ and $\hat{\sigma} = \hat{\beta}$. This gave similar findings, though sometimes an additional stage was needed to obtain the full gains in precision.

characteristics (as BLP’s original sums of characteristics). These instruments avoid estimating a full random coefficients model in a first stage, which is computationally demanding and in any case yields imprecise parameter estimates.

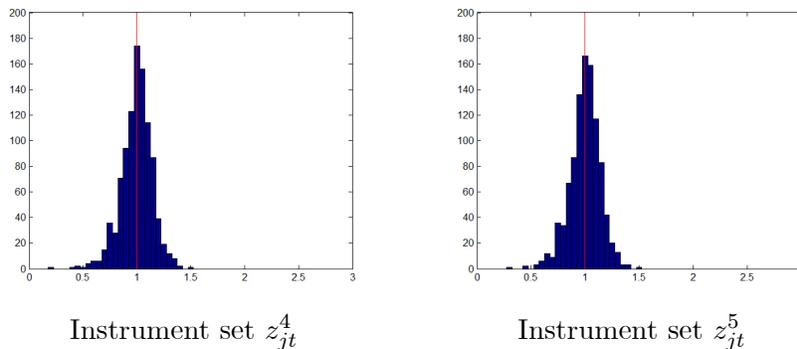
Table 3: Bias and Efficiency with Optimal Instruments from Linear First Stage Logit

Parameter	True	z_{jt}^4		z_{jt}^5	
		Bias	RMSE	Bias	RMSE
β^0	2	-0.003	0.447	-0.003	0.445
β^1	2	0.000	0.280	0.001	0.278
α	-2	0.000	0.042	0.000	0.042
σ^1	1	-0.002	0.143	-0.002	0.139

Bias is the average parameter estimate minus the true parameter, over the 1000 generated data sets. RMSE is the root mean squared error. Estimates are based on the MPEC algorithm and Sparse Grid integration. The two instrument sets are optimal instruments as in 2, but now evaluated a first stage simple logit estimates for the linear parameters, and an initial guess for the nonlinear parameter σ^1 .

Table 3 and Figure 3 show the results. Quite remarkably, the bias and RMSE is almost identical as those presented in Table 2 and Figure 2. Optimal instruments based on a first-stage simple logit and initial guess for $\hat{\sigma}^1$ apparently give the same performance as optimal instruments based on a computationally expensive first-stage random coefficients logit. The reason for the comparable performance is that the first-stage estimates of the random coefficients model were rather imprecise and contained a nontrivial mass of σ^1 at zero. Note also that the exact instrument set z_{jt}^5 again performs better in terms of RMSE than the approximate instrument set z_{jt}^4 , but the difference is almost negligible.

Figure 3: Histograms for $\hat{\sigma}^1$ with standard instruments as in Table 3



In sum, these findings suggest one may avoid estimating a computationally demanding first-stage random coefficients logit. It is sufficient to only estimate a simple first-stage logit with an initial guess for σ^1 . In practice, researchers could set their guess for σ^1 at a comparable order of magnitude as the estimate of $\hat{\beta}_1$.

4.3 Weak Instruments for Price

Up to now, we have assumed that the cost shifters have a strong impact on marginal costs $\gamma_2 = (3, 3, 3)$, so they serve as strong instruments for the endogenous price variable. In practice, it may be difficult to find product-specific cost shifters as price instruments, as stressed by BLP and the subsequent literature. It is therefore of interest to consider the performance of Chamberlain’s (1987) optimal instruments when good cost shifters are not available. We incorporate this by considering a situation where the cost shifters only have a weak impact on marginal costs, $\gamma_2 = (0.3, 0.3, 0.3)$. We keep all other parameters to generate the 1000 data sets the same as before.

Table 4 compares the performance of a standard set of instruments z_{jt}^1 with that of an optimal set z_{jt}^4 . As expected, both sets of instruments result in more small sample bias and higher RMSE than their twin counterparts of Table 1 and Table 3. Nevertheless, the optimal instruments still outperform the standard instruments. In particular, the RMSE drops considerably, especially for the constant β^0 (from 3.834 to 1.500) and the nonlinear taste parameter σ^1 (from 2.185 to 0.760). The bias of these two parameters also decreases substantially (though it increases somewhat for the other parameters β^1 and α).⁴ Finally, there is again a large spike around 0 for σ^1 with standard instruments. With optimal instruments the spike is much smaller (histogram not reported).

Table 4: Bias and Efficiency with Weak Price Instruments

Parameter	True	z_{jt}^1		z_{jt}^4	
		Bias	RMSE	Bias	RMSE
β^0	2	1.052	3.834	-0.032	1.500
β^1	2	0.041	0.664	-0.111	0.404
α	-2	0.098	0.533	0.165	0.441
σ^1	1	0.457	2.185	0.033	0.760

Bias and RMSE over 1000 generated data sets. Estimates are based on the MPEC algorithm and Sparse Grid integration. The results are parallel to Table 1 (for z_{jt}^1) and Table 3 (for z_{jt}^4), except that the cost shifters are weak instruments for price ($\gamma_2 = (0.3, 0.3, 0.3)$).

To conclude, when the cost shifters are only weak instruments for price, the optimal instruments

⁴Note however that the bias for α is twice as large when we treat price as an exogenous variable (i.e. when we do not use the weak cost shifters as price instruments).

still outperform the standard instruments, especially in terms of efficiency and stability. Nevertheless, the precision is much lower than when cost shifters are strong instruments for price. This is consistent with Armstrong (2012)'s conclusion on the importance of strong cost-side instruments to correct for endogeneity bias (although he considers imperfect competition and does not attempt to estimate σ^1).⁵

4.4 Small and Large Sample Performance

We finally ask how bias and precision change as the sample increases, either because of more products J or a larger number of markets T . As noted by Stock, Wright and Yogo (2002), the performance of an estimator does not improve under weak instruments. We therefore return to the case where cost shifters are strong instruments for the endogenous price variable, $\gamma_2 = (3, 3, 3)$.

Table 5 considers small and large sample performance based on the optimal instrument set z_{jt}^4 (as in Table 3). The top panel considers three sample sizes for the number of products, $J = \{5, 10, 20\}$ and fixes $T = 25$. The bottom panel considers three sample sizes for the number of markets, $T = \{12, 25, 50\}$ and fixes $J = 10$. Table 5 shows that the bias and precision improve both when J or when T increases. In particular, when the number of products J increases by a factor of 4 (from 5 to 20), the RMSE for all parameters tends to decrease by a factor of 2, a gain in the order of \sqrt{J} . Similarly, when the number of markets T increases by a factor of 4 (from 12 to 50), the RMSE also appears to decrease by a factor of 2, a gain in the order of \sqrt{T} .

⁵We also extended our analysis to allow for a random coefficient on the endogenous price variable, rather than on the exogenous characteristic x_{jt}^1 . We find that optimal instruments still reduce the RMSE (by more than half). The estimates are generally less precise than in the model with a random coefficient on x_{jt}^1 , because the endogeneity of prices carries over through the nonlinear part of utility.

Table 5: Small and Large Sample Performance with Optimal Instruments

		Number of Products					
$T = 25$		$J = 5$		$J = 10$		$J = 20$	
Parameter	True	Bias	RMSE	Bias	RMSE	Bias	RMSE
β^0	2	-0.019	0.647	-0.003	0.447	-0.016	0.315
β^1	2	0.015	0.446	0.000	0.280	0.008	0.191
α	-2	0.000	0.065	-0.000	0.042	0.001	0.031
σ^1	1	-0.023	0.223	-0.002	0.143	-0.009	0.124

		Number of Markets					
$J = 10$		$T = 12$		$T = 25$		$T = 50$	
Parameter	True	Bias	RMSE	Bias	RMSE	Bias	RMSE
β^0	2	-0.046	0.700	-0.003	0.447	-0.010	0.308
β^1	2	0.012	0.476	0.000	0.280	0.004	0.194
α	-2	0.004	0.065	-0.000	0.042	0.001	0.031
σ^1	1	-0.020	0.276	-0.002	0.143	-0.004	0.095

Bias and RMSE over 1000 generated data sets. Estimates are based on the MPEC algorithm and Sparse Grid integration. The instrument set is z_{jt}^4 . The top panel varies the number of products J , the bottom panel varies the number of markets T .

It is interesting to contrast these intuitive findings with the results in Skrainka (2012). He thoroughly considers small and large sample performance on a larger scale than we do. However, he uses standard instead of optimal instruments (variants of our set z_{jt}^1 and z_{jt}^3). With standard instruments, he finds that neither the bias nor the efficiency improves even when the sample size increases substantially, a finding that we also confirmed in our own calculations (not reported).

To conclude, the performance of the estimator improves with sample size under optimal instruments, but not under standard instruments. With a large sample, the benefits of optimal instruments thus become even more striking.

5 Computational Performance: Standard versus Optimal Instruments

Recent contributions by Dubé, Fox and Su (2012), Skrainka and Judd (2011) and Knittel and Metaxoglou (2012) have largely focused on the computational performance of the aggregate random coefficients logit model. In this section we reconsider the relevance of these three methodological contributions when optimal instruments are used. We first compare Dubé, Fox and Su's (2012) MPEC with BLP's NFP (the contraction mapping), under both standard and optimal instruments.

We then compare accurate polynomial-based integration (similar to Skrainka and Judd (2011)) with cruder pseudo Monte Carlo integration, again under both standard and optimal instruments. Finally, we discuss problems of local minima and solver convergence when optimizing a nonlinear objective function, as investigated by Knittel and Metaxoglou (2012).

5.1 Performance of MPEC and NFP

Su and Judd (2012) develop Mathematical Programming under Equilibrium Constraints (MPEC) as an alternative to commonly used nested fixed point (NFP) algorithms. In the context of the random coefficients logit model, Dubé, Fox and Su (2012) derive numerical properties of the NFP (contraction mapping) algorithm to invert market shares, and show how numerical error in the inner loop propagates to the outer loop GMM objective function. The MPEC algorithm avoids the inner loop, and may therefore perform both more precisely and faster. In their Monte Carlo study, Dubé, Fox and Su (2012) make use of standard instruments. It is therefore interesting to reconsider their results when optimal instruments are used. We also provide detail on the standard instruments, since Dubé, Fox and Su (2012) only report elasticities.

Table 6: Performance of MPEC and NFP algorithm

NFP					
CPU time:					
z_{jt}^1					
37.64					
z_{jt}^4					
90.63					
	True	Bias	RMSE	Bias	RMSE
β^0	2	0.250	1.597	-0.003	0.447
β^1	2	-0.241	1.380	0.000	0.280
α	-2	-0.025	0.121	0.000	0.042
σ^1	1	0.066	1.208	-0.002	0.143

MPEC					
CPU time:					
z_{jt}^1					
21.53					
z_{jt}^4					
9.31					
	True	Bias	RMSE	Bias	RMSE
β^0	2	0.295	1.739	-0.003	0.447
β^1	2	-0.291	1.584	0.000	0.280
α	-2	-0.028	0.137	0.000	0.042
σ^1	1	0.088	1.316	-0.002	0.143

Bias and RMSE over 1000 generated data sets. Estimates are based on Sparse Grid integration. The top panel uses the NFP algorithm (BLP's contraction mapping). The bottom panel uses MPEC, with results identical to Table 1 (for z_{jt}^1) and Table 3 (for z_{jt}^4).

Table 6 compares the performance of MPEC and NFP under both a standard instrument set z_{jt}^1 (similar to Dubé, Fox and Su (2012)) and an optimal instrument set z_{jt}^4 . (Recall that our above results were entirely based on MPEC, so the bottom panel duplicates results from Table 1 and Table 3.) Table 6 shows that MPEC indeed substantially outperforms NFP in terms of CPU time, both with standard and optimal instruments. But it is striking that the computational gains from MPEC are much higher under optimal instruments (a drop in average CPU time from 90.63 to 9.31) than under standard instruments (drop in average CPU time from 37.64 to 21.53).

While MPEC thus involves clear computational gains, it does not improve bias or efficiency (RMSE). It is interesting to see that bias and efficiency differ somewhat between NFP and MPEC under standard instruments, but they are exactly the same under our optimal instruments. This is reassuring since Dubé, Fox and Su (2012) derived that the two methods should in principle produce exactly the same results. Their result is thus confirmed with optimal instruments but not with standard instruments, even though we set the convergence criterion of contraction mapping NFP to a tight level of 10^{-12} .

We conclude that MPEC indeed brings substantial CPU time savings, especially under optimal instruments.⁶ It does not however improve the bias or the efficiency (at least under a tight convergence criterion for the inner loop).

5.2 Numerical Integration in the Market Share System

Most work since BLP has approximated the market share integrals (3) with pseudo Monte Carlo (pMC) integration. Skrainka and Judd (2011) consider the performance of more accurate polynomial-based integration methods; see also Heiss and Winschel (2008) for an analysis of different integration methods for likelihood problems. Up to now, we made use of accurate Sparse Grid quadrature. We now consider the performance of less accurate integration methods under our optimal instruments and standard instruments. More specifically, we first generate our 1000 data sets using a highly accurate Sparse Grid quadrature method that is exact to degree 69 to approximate the market share integrals.⁷ We then estimate the model with less accurate Sparse Grid quadratures (with 19, 9 and 7 nodes) and with pseudo Monte Carlo integration (with 200, 100 and 50 draws). We use optimal instruments, and will compare our findings with those of Skrainka and Judd (2011) who use standard instruments.

Table 7 shows the following results. First, polynomial-based integration yields lower bias and higher precision (lower RMSE) than pseudo Monte Carlo integration. For example, the RMSE of σ^1 is about 0.14 for any of the polynomial-based methods; there are no gains from increasing the

⁶This is certainly the case for small scale problems. In larger scale applications, MPEC may become more demanding in terms of memory usage.

⁷We can obtain an almost exact approximation because we have only one random coefficient. With multiple random coefficients, a higher dimensional integral needs to be computed, making the approximation less precise. See the above references for more discussion.

nodes from 7 to 19. In contrast, the RMSE of σ^1 is 0.249 under pseudo Monte Carlo integration with 200 draws, and it increases to 0.361 with 50 draws. Second, polynomial-based integration is also considerably faster. Average CPU time is 12.47 with 19 nodes and 6.43 with 7 nodes. In contrast, with pseudo Monte Carlo integration CPU time is 41.44 with 200 draws and 17.89 with 50 draws. In sum, pseudo Monte Carlo integration is a lot slower in computational terms and results in a higher bias and RMSE for all parameters. At least for one-dimensional integration, these findings indicate that polynomial-based methods are superior to pseudo Monte Carlo methods. Using standard instruments, Skrainka and Judd (2011) obtain similar findings in terms of speed, but they have no clear-cut conclusions regarding bias and RMSE.⁸

Table 7: Performance of pseudo Monte Carlo and Polynomial-based Integration

		Pseudo Monte Carlo Integration					
		# Draws=200		# Draws=100		# Draws=50	
CPU time		41.44		29.18		17.89	
Parameter	True	Bias	RMSE	Bias	RMSE	Bias	RMSE
β^0	2	-0.122	0.496	-0.205	0.536	-0.328	0.602
β^1	2	0.097	0.336	0.166	0.382	0.266	0.444
α	-2	0.005	0.043	0.010	0.044	0.015	0.047
σ^1	1	-0.089	0.249	-0.147	0.299	-0.236	0.361

		Polynomial-based Integration					
		# Nodes=19		# Nodes=9		# Nodes=7	
CPU time		12.47		8.00		6.43	
Parameter	True	Bias	RMSE	Bias	RMSE	Bias	RMSE
β^0	2	-0.005	0.449	-0.005	0.449	-0.006	0.447
β^1	2	0.002	0.280	0.002	0.281	0.003	0.278
α	-2	0.000	0.042	0.000	0.042	0.000	0.042
σ^1	1	-0.004	0.143	-0.004	0.143	-0.004	0.139

Bias and RMSE over 1000 generated data sets. Estimates are based on the MPEC algorithm. To generate the 1000 data sets, the market share integrals are approximated with a polynomial-based Sparse Grid quadrature with 35 nodes. To estimate the model, the market shares are approximated with pseudo Monte Carlo integration (top panel) or Sparse Grid quadrature (bottom panel), for varying degrees of accuracy.

The instrument set is z_{jt}^4 .

⁸We also investigated how numerical integration affects the standard errors of the parameter estimates. With standard instruments, we confirm Skrainka and Judd's (2011) curious finding that standard errors are on average lower under inaccurate pseudo Monte Carlo integration (with limited number of draws). In contrast, with optimal instruments we obtain an intuitive pattern: standard errors become higher on average when the quality of integration becomes cruder. Interestingly, with accurate integration the standard errors of the estimates are on average very close to the standard deviation of the estimates across the 1,000 generated datasets. With crude integration the standard errors are much higher.

To conclude, with optimal instruments polynomial-based integration outperforms pseudo Monte Carlo integration both in terms of mean squared error and computational speed. With optimal instruments, the standard deviations intuitively decrease as the market share is better approximated, while the opposite holds with standard instruments.

5.3 Local Minima of a Nonlinear Objective Function

Knittel and Metaxoglou (2012) extensively document that the GMM estimates of the random coefficients logit depend on the optimization algorithm, may converge to local minima and are therefore highly dependent on the starting values. See also Skrainka (2012) for a thorough review of the computational problems. As discussed in section 3, we minimize the GMM objective function using the Knitro 800 Interior\Direct algorithm with appropriate options. For each of the 1,000 generated data sets, we estimate the model using 10 different starting values and select the estimates that result in the lowest value for the objective function.

We have separately looked into more detail at the sensitivity of the results from different starting values. We confirm Knittel and Metaxoglou (2012)’s finding that the algorithm often converges to different point estimates and function values. This is the case for the standard instruments z_{jt}^1 reported in Table 1, but it holds to a much lesser extent for the optimal instruments z_{jt}^4 reported in Table 3. To see this, Table 8 shows the standard deviation of the estimates and function values under 10 different starting values, averaged over the first 50 generated datasets. These standard deviations would be zero if the objective function does not have multiple local optima. Interestingly, the standard deviations are about 5 times lower with optimal instruments than with standard instruments, showing the results are less sensitive to starting values with optimal instruments. Another striking finding relates to the convergence of the solver. With optimal instruments, the solver reported convergence to a (local) optimum in all 500 (50×10) cases. With standard instruments, the solver reported nonconvergence (“local infeasibility”) in 37 cases.⁹

⁹We discarded these outcomes in Table 8 and our earlier tables.

Table 8: Standard Deviations of Point Estimates from Different Starting Values

Parameter	z_{jt}^1 St. Dev.	z_{jt}^4 St. Dev.
β^0	0.117	0.028
β^1	0.140	0.024
α	0.007	0.001
σ^1	0.108	0.019
Function Value	0.029	1.07E-13

Mean of the standard deviations of point estimates obtained over 10 different starting values for 50 data sets. We generate data and estimate the model as in Table 1 and 3.

In sum, under optimal instruments the objective function shows considerably fewer local minima, so that the results become less sensitive to the choice of starting values. Nevertheless, there is still some variation over starting values. This just confirms the inherent non-linearity of the GMM objective function for the random coefficients model. Careful checking of different starting values and solver convergence therefore remains warranted regardless of the adopted instrument set.

6 Conclusion

We have shed new light on the performance of Berry, Levinsohn and Pakes' (1995) GMM estimator of the aggregate random coefficient logit model. Based on an extensive Monte Carlo study, we have shown that the use of Chamberlain's (1987) optimal instruments overcomes most of the problems that have recently been documented with standard, non-optimal instruments. Optimal instruments reduce small sample bias, but they prove even more powerful in increasing the estimator's efficiency (root mean squared error) and stability (avoiding spikes at zero for the standard deviation of the taste parameters). Other recent methodological advances (MPEC, polynomial-based integration of the market shares) greatly improve computational speed, but they are only successful in terms of bias and efficiency when combined with optimal instruments. Finally, with optimal instruments the GMM objective function is less flat resulting in less local minima, but careful searching for a global minimum still remains warranted.

These findings suggest that practitioners should apply optimal instruments when estimating aggregate random coefficients logit models. These instruments are generated from within the model, as they depend in a specific way on the product characteristics of the own and other products. Applying optimal instruments does not imply that researchers should not search for good cost-side instruments to estimate the endogenous price effects. As shown by Armstrong (2012), and confirmed in our setting, good cost-side instruments are important to obtain precise estimates of the price

parameter and indirectly also for the other parameters.

We conclude with some cautionary warnings. Optimal instruments guarantee a strong performance of the aggregate random coefficients logit model when the data-generating process is known and satisfies strong (but common) assumptions. It is not clear to which extent these assumptions are satisfied in real data sets, and whether optimal instruments still achieve the same powerful performance (in terms of bias, efficiency and computational speed). First, we assumed a data-generating process where the product-specific unobservable is uncorrelated with the observed product characteristics of all products. If this assumption is not satisfied, optimal instruments may need modification. See Crawford (2011), Akerberg and Crawford (2009) for recent progress on instruments when product characteristics are endogenous. Second, even if product characteristics are exogenous, they may be correlated across products, which may require modification of the optimal instruments. Third, we assumed the endogenous variables, prices and market shares, are generated based on known functional forms, from the random coefficients logit with a normal distribution for the random coefficients. In practice, the functional forms are not known so that the model may be misspecified. It therefore remains a very interesting topic for future research to assess the performance of random coefficients model with real data or more realistic simulated data.

References

- Akerberg, Daniel A. and Gregory S. Crawford**, “Estimating Price Elasticities in Differentiated Product Models with Endogenous Characteristics,” In Mimeo, University of Warwick 2009.
- Amemiya, Takeshi**, “The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model,” *Econometrica*, 1977, 45 (4), pp. 955–968.
- Arellano, Manuel**, *Panel Data Econometrics*, Oxford University Press, 2003.
- Armstrong, Timothy B.**, “Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply,” Technical Report, Stanford University 2012.
- Berry, Steven**, “Comment,” *Quantitative Marketing and Economics*, 2003, 1, 285–291.
- Berry, Steven T.**, “Estimating Discrete-Choice Models of Product Differentiation,” *The RAND Journal of Economics*, 1994, 25 (2), 242–262.
- , **Amit Gandhi, and Philip Haile**, “Connected Substitutes and Invertibility of Demand,” NBER Working Papers 17193, National Bureau of Economic Research, Inc July 2011.

- , **James Levinsohn**, and **Ariel Pakes**, “Automobile Prices in Market Equilibrium,” *Econometrica*, 1995, *63* (4), 841–890.
- , – , and – , “Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy,” *American Economic Review*, June 1999, *89* (3), 400–430.
- Bresnahan, Timothy F., Scott Stern, and Manuel Trajtenberg**, “Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s,” *The RAND Journal of Economics*, 1997, *28*, pp. S17–S44.
- Chamberlain, Gary**, “Asymptotic efficiency in estimation with conditional moment restrictions,” *Journal of Econometrics*, 1987, *34* (3), 305 – 334.
- Crawford, Gregory S.**, “Endogenous Product Choice: A Progress Report,” Technical Report, University of Warwick 2011.
- Dubé, Jean-Pierre H., Jeremy T. Fox, and Che-Lin Su**, “Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation,” *forthcoming in Econometrica*, 2012.
- Fox, Jeremy T. and Amit Gandhi**, “Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models,” Technical Report 2012.
- , **Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari**, “The random coefficients logit model is identified,” *Journal of Econometrics*, 2012, *166* (2), 204 – 212.
- Goeree, Michelle Sovinsky**, “Limited Information and Advertising in the U.S. Personal Computer Industry,” *Econometrica*, 09 2008, *76* (5), 1017–1074.
- Heiss, Florian and Viktor Winschel**, “Likelihood approximation by numerical integration on sparse grids,” *Journal of Econometrics*, May 2008, *144* (1), 62–80.
- Jiang, Renna, Puneet Manchanda, and Peter E. Rossi**, “Bayesian analysis of random coefficient logit models using aggregate data,” *Journal of Econometrics*, April 2009, *149* (2), 136–148.
- Knittel, Christopher R. and Konstantinos Metaxoglou**, “Estimation of Random Coefficient Demand Models: Two Empiricists’ Perspective,” In Mimeo 2012.
- McFadden, Daniel**, “The measurement of urban travel demand,” *Journal of Public Economics*, November 1974, *3* (4), 303–328.
- Nevo, Aviv**, “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand,” *Journal of Economics & Management Strategy*, December 2000, *9* (4), 513–548.

– , “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, March 2001, 69 (2), 307–42.

Skrainka, Benjamin S., “A Large Scale Study of the Small Sample Performance of Random Coefficient Models of Demand,” *SSRN eLibrary*, 2012.

– **and Kenneth L. Judd**, “High performance quadrature rules: how numerical integration affects a popular model of product differentiation,” CeMMAP working papers CWP03/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies Feb 2011.

Stock, James H, Jonathan H Wright, and Motohiro Yogo, “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business & Economic Statistics*, 2002, 20 (4), 518–529.

Su, Che-Lin and Kenneth L. Judd, “Constrained Optimization Approaches to Estimation of Structural Models,” *forthcoming in Econometrica*, 2012.

Verboven, Frank, “International Price Discrimination in the European Car Market,” *RAND Journal of Economics*, Summer 1996, 27 (2), 240–268.