A Simple Solution to Invalid Inference in the Random Coefficients Logit Model*

Philipp Ketz†

May 19, 2017

We show that the standard approach to inference in the random coefficients logit model suffers from size distortions. The problem arises due to boundary issues and is aggravated when the model is parameterized with respect to standard deviations, which constitutes common practice. For example, in case of a single random coefficient, the asymptotic size of the nominal 95% confidence interval, which is obtained by inverting the two-sided t-test for the standard deviation, equals 83.65%. In seeming contradiction, we also find that standard errors can be unreasonably large. The proposed solution is to perform inference with respect to variances rather than standard deviations. This alleviates the problem of unreasonably large standard errors and, in combination with the estimator proposed in Ketz (2017a), allows the construction of confidence intervals that (uniformly) control asymptotic size.

Keywords: Random coefficients logit model, boundary, testing, asymptotic size.

*I thank the editor and two referees for helpful comments and suggestions that have considerably improved the paper. I am grateful to Frank Kleibergen, Adam McCloskey, Blaise Melly, and Eric Renault for helpful discussions and suggestions. I also thank Andrew Elzinga, Joachim Freyberger, Bruno Gasperini, Daniel Pollmann, and seminar participants at Brown University for valuable feedback.

†Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France. Email: philipp.ketz@psemail.eu.

1 Introduction

The random coefficients logit model, or BLP model (Berry, Levinsohn, and Pakes, 1995), is widely used in applied work, most prominently in the industrial organization and marketing literatures. The random coefficients in this model are typically assumed to be normally distributed; in its simplest form with only one random coefficient: $\beta \sim N(\mu, \sigma^2)$. The estimation procedure proposed by BLP minimizes a Generalized Method of Moments (GMM) objective function with respect to the mean, μ, and the standard deviation, σ, of the random coefficient. Inference on the parameters of the model then relies on the standard asymptotic normality result for GMM estimators. The fact that estimates for σ are often found to be small (see e.g., Nevo, 2001; Goeree, 2008), which is indicative of the true parameter being close to the boundary of the parameter space as σ cannot take on negative values, is generally ignored and statements about whether σ (or μ) is significantly different from zero are typically based on the two-sided t-test.[1]

In this paper, we show that this approach to inference is invalid, in the sense that the two-sided t-test for σ or, equivalently, the confidence interval (CI) based on it does not control asymptotic size (AsySz), in a uniform sense. (See Ketz (2017a) for corresponding results for μ.) Our characterization result implies that the AsySz of the nominal 95% CI equals 83.65% in case of a single random coefficient. In the presence of multiple random coefficients, the undercoverage of the CI gets worse. For example, in case of two random coefficients, the AsySz equals 80.74% (if an efficient GMM estimator is employed). The size distortion arises because of boundary effects on the asymptotic distribution of the estimator that are amplified by the fact that the Jacobian of the (sample and) population moment condition is of reduced rank at the boundary of the parameter space. Another consequence of the reduced rank Jacobian is that the standard error used in practice is unreasonably large when the estimate of σ is "practically" zero.[2]

As a solution to these inference problems, we propose that researchers perform inference with respect to a one-to-one transformation of the parameter vector, namely (μ, σ^2). This reparameterization alleviates the problem of a reduced rank Jacobian and, thus, the problem of unreasonably large standard errors.[3] However, the estimator of the transformed parameter vector is still subject to boundary effects and the CIs obtained by inverting the two-sided t-test based on μ and σ^2 can also suffer from undercoverage, as documented in Ketz (2017a) and Section 4, respectively. A simple solution for obtaining CIs that control asymptotic size is to invert the two-sided t-test that uses the asymptotically normal estimator (for (μ, σ^2)) suggested in Ketz (2017a). By construction, this

[1] See for example Petrin (2002) and Goeree (2008); other papers, such as Berry, Levinsohn, and Pakes (1995) and Nevo (2001), make statements about the significance of a parameter "at conventional significance level(s)" without specifying what test they rely on.

[2] That is, the estimate differs from zero only by an algorithm specific tolerance level.

[3] That is, upon reparameterization the Jacobian is no longer by *construction* of reduced rank, at the boundary. Whether it is, indeed, of full rank depends on the true data generating process.

estimator coincides with the constrained estimator when estimates are found to be in the interior of the parameter space such that, in that case, CIs based on the constrained estimator (for (μ, σ^2)) are valid.

The recent literature has highlighted the importance of drifting sequences of true parameters for determining the AsySz of tests and CIs, or, more generally, confidence sets when the asymptotic distribution of the underlying estimator is discontinuous in a parameter (see e.g., Andrews and Guggenberger, 2010b; Andrews and Cheng, 2012). In the context of the BLP model, the discontinuity arises at the boundary of the parameter space. We use the results in Andrews (2002) to derive the asymptotic distribution of the estimator of the transformed parameter vector under drifting sequences of true parameters that may drift towards the boundary. Since the results in Andrews (2002) require the Jacobian of the population moment condition to be of full rank, we cannot use them directly to obtain the asymptotic distribution of the estimator of the original parameter vector. Instead, we express the coverage probability of the CI for σ as a function of the (estimator of the) transformed parameter. Then, the characterization of AsySz is obtained by applying the results in Andrews and Cheng (2012) together with the aforementioned asymptotic distribution result.

A reduced rank Jacobian is a "first order condition for lack of identification," i.e., it is necessary but not sufficient for lack of identification (Sargan, 1983). The BLP model, when parameterized with respect to standard deviations, constitutes an example of a model where a reduced rank Jacobian does not imply lack of identification.[4] A closely related example is given by the random coefficients regression model (see e.g., Andrews, 1999). Cox and Hinkley (1974) (page 303) show that the score of the likelihood function with respect to the standard deviation of the random coefficient is identically zero at the boundary of the parameter space resulting in a reduced rank Hessian, which is akin to the problem of a reduced rank Jacobian encountered here. The authors also note that a reparameterization in terms of variances solves the problem. However, they do not analyze the consequences of using the "wrong" parameterization for inference. Rotnitzky, Cox, Bottai, and Robins (2000) provide a general theory for the asymptotic distribution of the Maximum Likelihood estimator and the Likelihood Ratio test for a wide class of Maximum Likelihood models which are identified but have a Hessian matrix whose rank is one below full rank. However, they do not analyze the asymptotic behavior of the commonly used t-test and do not discuss the problem of large standard errors. In the context of GMM, Dovonon and Renault (2013) analyze the asymptotic behavior of the J-statistic, used for testing overidentifying restrictions, when the model is locally identified, but the Jacobian is of reduced rank.[5]

[4] There are numerous examples of models in the literature where a reduced rank Jacobian implies lack of identification. See e.g., Staiger and Stock (1997) and Liu and Shao (2003) for early references.

[5] See also Ketz (2017c) for a related discussion.

This paper is not the first to consider asymptotic theory for the BLP model. Berry, Linton, and Pakes (2004) and Freyberger (2015) derive the asymptotic distribution of the estimator of the original parameter vector, (μ, σ), under a large number of products and a large number of markets, respectively.[6] Their results differ from the standard asymptotic normality result for GMM estimators through additional bias and variance terms that are due to sampling and simulation error. Our results are complementary, as they allow for the true parameter vector to be near the boundary of the parameter space. While we follow Freyberger (2015) in deriving asymptotic distribution theory under a large number of markets, the problem we address is inherent to the parameterization of the model and, therefore, also relevant under a large number of products. Therefore, we recommend that the reparameterization in terms of variances is also applied when asymptotic theory is based on a large number of products.

In addition, it turns out to be advantageous to estimate the model with respect to the transformed parameter vector: The algorithm is less prone to convergence failures and on average requires less iterations to achieve convergence. This finding contributes to the recent literature that concerns the numerical performance of the estimator for the BLP model. For example, Knittel and Metaxoglou (2014) illustrate the sensitivity of the estimation procedure with respect to different starting values and highlight the importance of the choice of the optimization algorithm. Dubé, Fox, and Su (2012a) show that the practice of loosening the tolerance level of the fixed point computation required in the original formulation of the estimation problem can lead to convergence problems such as non-convergence or convergence to local minima. They suggest an alternative formulation of the optimization problem, referred to as Mathematical Program with Equilibrium Constraints (MPEC), which does not require the fixed point computation and which they find to display speed advantages over the original so-called Nested Fixed Point (NFP) algorithm implemented with a tight tolerance level. Another issue was raised by Skrainka and Judd (2011), who find that the accuracy with which predicted market shares are computed greatly impacts the performance of the estimator. They find that Monte Carlo integration as commonly employed in practice performs poorly, whereas sparse grid integration is found to perform well.

In order to establish the relevance of this paper's findings for empirical work, we apply the proposed reparameterization to a series of published articles that use the BLP model, including Berry, Levinsohn, and Pakes (1995), Berry, Levinsohn, and Pakes (1999) and Reynaert and Verboven (2014). We find that in many cases the variance parameter is not significantly different from zero anymore upon reparameterization. This illustrates that the problem of size distortion can be of empirical relevance. The problem of unreasonably large standard errors is for example encountered in Neilson (2013). Upon reparameteri-

[6] Recently, Armstrong (2016) has shown that commonly employed instruments poorly identify the mean parameter of the random coefficient, μ, when the number of products is large.

zation, the standard error is much smaller and, consequently, the conclusion that there is little heterogeneity in consumer preferences (with respect to the corresponding product characteristic) has more support from the data than initially thought.

The outline of this paper is as follows. In Section 2, we derive the asymptotic distribution of GMM estimators under drifting sequences of true parameters. Section 3 introduces the BLP model and shows that the Jacobian of the population moment condition is of reduced rank at the boundary of the parameter space. In Section 4, we characterize the AsySz of the CIs based on σ and σ^2. Section 5 introduces the estimator proposed in Ketz (2017a). Section 6 provides a small Monte Carlo study showing that our asymptotic theory provides good finite-sample approximations and documenting the computational advantage of estimating the model with respect to the transformed parameter vector. The applications illustrating the relevance of our findings for empirical work are presented in Section 7. Section 8 concludes.

Throughout this paper, let "$\equiv$" denote "equals by definition". Also, let (a, b) denote $(a', b')'$ for any two column vectors, a and b. For any interval I, let $I^K \equiv I \times \cdots \times I$ with K copies.

2 Asymptotic distribution of GMM estimators under drifting sequences of true parameters

In this section, we reproduce the results in Andrews (2002) concerning the asymptotic distribution of Generalized Methods of Moments (GMM) estimators when the true parameter vector is at the boundary of the parameter space. The results are presented under slightly modified assumptions allowing for drifting sequences of true parameters as in Andrews and Cheng (2014a). Drifting sequences of true parameters are important in establishing (uniform) asymptotic size control of tests and confidence sets (CSs), as highlighted in the recent literature (see e.g., Andrews and Guggenberger, 2010b; Andrews and Cheng, 2012). The restrictions on the parameter space are motivated by the random coefficients logit model, which we introduce in Section 3. In what follows, we borrow notation from Andrews and Cheng (2012, 2014a).

The GMM objective function depends on the $(2K \times 1)$ dimensional parameter θ. It is given by a quadratic form of the sample moment $G_T(\theta)$, i.e.,

$$Q_T(\theta) = G_T(\theta)'\mathcal{W}_T(\theta)G_T(\theta)/2,$$

where $\mathcal{W}_T(\theta)$ denotes a weighting matrix. The dependence of $Q_T(\theta)$ on the data $\{W_t : t \leq T\}$, which may be i.i.d., independent and nonidentically distributed, or temporally dependent, is suppressed for notational ease. Define an estimator, $\hat{\theta}_T$, as any random variable that satisfies $\hat{\theta}_T \in \Theta$ and

$$Q_T(\hat{\theta}_T) = \min_{\theta \in \Theta} Q_T(\theta) + o_p(1/T), \tag{1}$$

where

$$\Theta = [-c, c]^K \times [0, c]^K,$$

for some $0 < c < \infty$, denotes the optimization parameter space.[7] The true parameter space, $\ddot{\Theta}$, is a strict subset of Θ. In particular, it takes the same form as Θ, but with $0 < \ddot{c} < c$. This ensures that "boundary effects" only occur at 0. Let $\theta = (\theta_1, \theta_2)$ such that the elements in θ_1 are unrestricted and the elements in θ_2 are restricted below by 0.

In most applications the distribution of the data and, thus, $Q_T(\theta)$ is not fully specified by the vector θ, but depends on an additional, commonly infinite-dimensional, parameter, ϕ. The parameter $\gamma = (\theta, \phi)$ is assumed to fully specify the distribution of the data and the corresponding true parameter space is assumed to be of the following form

$$\Gamma = \{\gamma = (\theta, \phi) : \theta \in \ddot{\Theta}, \omega \in \ddot{\Phi}(\theta)\},$$

where $\ddot{\Phi}(\theta) \subset \ddot{\Phi}\ \forall \theta \in \ddot{\Theta}$ for some compact metric space $\ddot{\Phi}$ with a metric that induces weak convergence of the bivariate distribution $(W_t, W_{t+t'})$ for all $t, t' \geq 1$.

In this paper, we are interested in obtaining the asymptotic size of confidence intervals (CIs) or, more generally, CSs that are obtained by inverting some test statistic. Let $\mathcal{T}_T(v)$ denote a test statistic for testing $H_0 : r(\theta) = v$. A nominal $1 - \alpha$ CS is given by

$$\mathrm{CS}_T = \{v : \mathcal{T}_T(v) \leq \mathrm{cv}_{1-\alpha}\},$$

where $\mathrm{cv}_{1-\alpha}$ denotes a fixed critical value.[8] The coverage probability of a CS for $r(\theta)$ is

$$P_\gamma(r(\theta) \in \mathrm{CS}_T) = P_\gamma(\mathcal{T}_T(r(\theta)) \leq \mathrm{cv}_{1-\alpha}).$$

The asymptotic size, which approximates finite sample size, is given by

$$\mathrm{AsySz} = \liminf_{T \to \infty} \inf_{\gamma \in \Gamma} P_\gamma(\mathcal{T}_T(r(\theta)) \leq \mathrm{cv}_{1-\alpha}).$$

As mentioned above, drifting sequences of true parameters, $\gamma_T = (\theta_T, \phi_T) = (\theta_{1,T}, \theta_{2,T}, \phi_T)$, are instrumental in determining AsySz. In this paper, the following sequences are crucial

$$\Gamma(\gamma_0, h) = \{\{\gamma_T \in \Gamma : T \geq 1\} : \gamma_T \to \gamma_0 \in \Gamma \text{ and } \sqrt{T}\theta_{2,T} \to h \in (\mathbb{R}_+ \cup \{\infty\})^K\},$$

[7]The use of a common end point c and the same dimension K is merely for notational ease. Similarly, the normalization to 0 is without loss of generality.

[8]Here, we consider a fixed critical value, as interest lies with standard Wald-type CSs.

where $\gamma_0 = (\theta_0, \phi_0) = (\theta_{1,0}, \theta_{2,0}, \phi_0)$. Throughout this paper, we use the terminology "under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$" to mean "when the true parameters are $\gamma_T \in \Gamma(\gamma_0, h)$ for any $\gamma_0 \in \Gamma$ with any $h \in (\mathbb{R}_+ \cup \{\infty\})^K$." Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ except $h = \infty^K$ we say that the true parameter vector is near the boundary.

Next, we state the assumptions underlying the asymptotic distribution results in this paper. They are slightly modified versions of Assumptions GMM1, GMM2, and GMM5 in Andrews and Cheng (2014a), as they do not allow for lack of identification in part of the parameter space, but allow θ_0 to be at the boundary as in Andrews (2002).

The first assumption ensures consistency of $\hat{\theta}_T$.

Assumption 1.

(i) Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, $\sup_{\theta \in \Theta} \|G_T(\theta) - G(\theta; \gamma_0)\| \to_p 0$ and $\sup_{\theta \in \Theta} \|\mathcal{W}_T(\theta) - \mathcal{W}(\theta; \gamma_0)\| \to_p 0$ for some nonrandom funtions $G(\theta; \gamma_0)$ and $\mathcal{W}(\theta; \gamma_0)$.

(ii) $G(\theta; \gamma_0) = 0$ if and only if $\theta = \theta_0$, $\forall \gamma_0 \in \Gamma$.

(iii) $G(\theta; \gamma_0)$ has continuous left/right (l/r) partial derivatives on Θ, denoted $G_\theta(\theta; \gamma_0)$, $\forall \gamma_0 \in \Gamma$.[9]

(iv) $\mathcal{W}(\theta; \gamma_0)$ is continuous in θ on Θ, $\forall \gamma_0 \in \Gamma$.

(v) $\mathcal{W} \equiv \mathcal{W}(\theta_0; \gamma_0)$ is nonsingular, $\forall \gamma_0 \in \Gamma$.

The next assumption ensures that the objective function is asymptotically well approximated by a quadratic function, see below.

Assumption 2. Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$,

$$\sup_{\theta \in \Theta : \|\theta - \theta_T\| \le \delta_T} \frac{\sqrt{T} \|G_T(\theta) - G(\theta; \gamma_0) - G_T(\theta_T) + G(\theta_T; \gamma_0)\|}{1 + \|\sqrt{T}(\theta - \theta_T)\|} = o_p(1)$$

for all constants $\delta_T \to 0$.

The following assumption is sufficient for Assumption 2 and can often be verified using a ULLN, see e.g., Andrews (1992).

Assumption 2*.

(i) $G_T(\theta)$ has continuous l/r partial derivatives on Θ $\forall T \ge 1$.

[9] A function is said to have left/right partial derivatives if it has partial derivatives at each interior point and left (right) partial derivatives at each boundary point (of Θ) that can be perturbed only to the left (right), see Section 3.3 in Andrews (1999).

(ii) Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$,

$$\sup_{\theta\in\Theta:\|\theta-\theta_T\|\le\delta_T} \left\| \frac{\partial}{\partial\theta'} G_T(\theta) - G_\theta(\theta, \gamma_0) \right\| = o_p(1)$$

for all constants $\delta_T \to 0$.

The last assumption concerns the Jacobian of the population moment condition and the asymptotic behavior of the scaled sample moment.

Assumption 3.

(i) $G_\theta \equiv G_\theta(\theta_0; \gamma_0)$ has full column rank, $\forall\gamma_0 \in \Gamma$.

(ii) Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, $\sqrt{T}G_T(\theta_T) \xrightarrow{d} Y \sim N(0, \Omega(\gamma_0))$ for some symmetric and positive definite matrix $\Omega(\gamma_0)$.

Next, we explain the intuition behind the results in Andrews (2002). Consider the following quadratic approximation of the GMM objective function

$$Q_T(\theta) = Q_T(\theta_T) + G_T(\theta_T)'\mathcal{W}G_\theta(\theta - \theta_T) + \frac{1}{2}(\theta - \theta_T)'(G_\theta'\mathcal{W}G_\theta)(\theta - \theta_T) + R_T(\theta). \quad (2)$$

Let

$$Z_T = -J^{-1}G_\theta'\mathcal{W}\sqrt{T}G_T(\theta_T),$$

where $J \equiv J(\gamma_0) \equiv G_\theta'\mathcal{W}G_\theta$. Then, equation (2) can be rewritten as

$$Q_T(\theta) = Q_T(\theta_T) - \frac{1}{2}Z_T'JZ_T + \frac{1}{2}q_T(\sqrt{T}(\theta - \theta_T)) + R_T(\theta), \quad (3)$$

where

$$q_T(\lambda) = (\lambda - Z_T)'J(\lambda - Z_T).$$

Given the above assumptions, the remainder, $R_T(\theta)$, is small enough such that the centered and scaled minimizer of $Q_T(\theta)$, $\sqrt{T}(\hat{\theta}_T - \theta_T)$, has the same asymptotic distribution as the minimizer of $q_T(\lambda)$.

The two determinants of the asymptotic distribution of the minimizer of $q_T(\lambda)$ are the asymptotic behavior of Z_T and $\lim_{T\to\infty}\sqrt{T}(\Theta - \theta_T)$, the limit of the centered and scaled parameter space. Given the above assumptions, we have that

$$Z_T \xrightarrow{d} Z \equiv N(0, V(\gamma_0)), \text{ where } V(\theta_0) \equiv J^{-1}G_\theta'\mathcal{W}\Omega(\gamma_0)\mathcal{W}G_\theta J^{-1}.$$

As formally stated in Proposition 1 below the asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_T)$, under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, is given by the distribution of

$$\hat{\lambda}_h = \arg\min_{\lambda\in\Lambda_h} q(\lambda), \quad (4)$$

where

$$q(\lambda) = (\lambda - Z)'J(\lambda - Z)$$

and

$$\Lambda_h \equiv (\mathbb{R} \cup \{\pm\infty\})^K \times [-h_1, \infty] \times \cdots \times [-h_K, \infty]. \tag{5}$$

Proposition 1. *Under* $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ *and Assumptions 1-3,* $\sqrt{T}(\hat{\theta}_T - \theta_T) \xrightarrow{d} \hat{\lambda}_h$*, where* $\hat{\lambda}_h$ *is defined in* (4) *with* Λ_h *given in* (5).

The proof of Proposition 1 is obtained by verifying the Assumptions in Ketz (2017a,b), see Appendix B for details.

Remark 1. When $h = \infty^K$, we have $\Lambda_h = (\mathbb{R} \cup \{\pm\infty\})^{2K}$ and $\hat{\lambda}_h = Z$. Note that $h = \infty^K$ allows for sequences of true parameters that drift towards the boundary, i.e., $\theta_{2,k,T} \to 0$ for some $k \in \{1, \ldots, K\}$, where $\theta_2 = (\theta_{2,1}, \ldots, \theta_{2,K})$, but at a rate slow enough such that the boundary does not impact the asymptotic distribution. Put differently, if the true parameter vector is "far enough" from the boundary, then we obtain the standard asymptotic normality result for GMM estimators. If, however, $h \neq \infty^K$, then the asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_T)$ is subject to boundary effects and given by the projection of Z onto Λ_h with respect to the norm $\|\lambda\| = (\lambda' J \lambda)^{1/2}$. The results in Section 6 of Andrews (1999) concerning the distribution of subvectors of $\sqrt{T}(\hat{\theta}_T - \theta_T)$ apply here as well with slight modifications, as Λ_h is a cone with (possibly) non-zero vertex. For example, if $K = 1$, there exists a simple closed form expression. Let $Z = (Z_1, Z_2)$ and let $J = \begin{bmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{bmatrix}$. Then, $\sqrt{T}(\hat{\theta}_{1,T} - \theta_{1,T}) \xrightarrow{d} Z_1 - J_{11}^{-1} J_{12} \min(0, Z_2 + h)$ and $\sqrt{T}(\hat{\theta}_{2,T} - \theta_{2,T}) \xrightarrow{d} \max(-h, Z_2)$. When an efficient weighting matrix is employed, i.e., $\mathcal{W} = \Omega(\gamma_0)^{-1}$, the asymptotic distribution of the GMM estimator simplifies. In particular, for $K = 1$, the asymptotic distribution of $\sqrt{T}(\hat{\theta}_{1,T} - \theta_{1,T})$ simplifies to $Z_1 - \frac{V_{12}}{V_{22}} \min(0, Z_2 + h)$, where $V(\theta_0) = \begin{bmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{bmatrix}$.

3 The random coefficients logit model

In this section, we introduce the random coefficients logit model, which we also refer to as the BLP model (Berry, Levinsohn, and Pakes, 1995). We assume that there exist T markets. In each market t $(t = 1, \ldots, T)$, there are N individuals, each of whom $(i = 1, \ldots, N)$ chooses one out of J products, which maximizes utility. The utility of product j $(j = 1, \ldots, J)$ in market t for individual i is assumed to be given by

$$u_{ijt} = x'_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}. \tag{6}$$

Here, x_{jt} denotes a $(K \times 1)$ dimensional vector of observed product characteristics, which in many applications include the price of the product. ξ_{jt} denotes an unobserved product characteristic, which may capture for example brand image. It is assumed known to the consumer, but unknown to the econometrician, and takes on the role of the error term. ε_{ijt} denotes an individual specific preference term, which is also assumed to be unobserved to the econometrician. β_i denotes a vector of random coefficients, which allows individuals to have heterogeneous preferences with respect to the different product characteristics.

It is commonly assumed that $\beta_i \sim N(\mu, \Sigma)$, where μ is a $(K \times 1)$ vector and Σ is a (positive semidefinite) $(K \times K)$ matrix. Typically, the model is further simplified by assuming that the random coefficients are mutually independent, i.e., Σ is assumed to be a diagonal matrix. Let $\sigma = (\sigma_1, \ldots, \sigma_K)'$ denote the vector of standard deviations, i.e., the square roots of the elements on the main diagonal. Then, the utility in (6) can be written as

$$u_{ijt} = \delta_{jt} + \sum_{k=1}^{K} x_{jt,k}\sigma_k v_{i,k} + \varepsilon_{ijt},$$

where $\delta_{jt} = x_{jt}'\mu + \xi_{jt}$ and $v_i \sim N(0, I_K)$. If, further, we assume that ε_{ijt} is extreme value type I distributed, then the model implied market share for product j in market t is given by

$$\mathrm{s}_j(\sigma, \delta_t, x_t) = \int \frac{e^{x_{jt}'\mu + \xi_{jt} + \sum_{k=1}^{K} x_{jt,k}\sigma_k v_k}}{1 + \sum_{l=1}^{J} e^{x_{lt}'\mu + \xi_{lt} + \sum_{k=1}^{K} x_{lt,k}\sigma_k v_k}} dF(v), \tag{7}$$

where $\delta_t = (\delta_{1t}, \ldots, \delta_{Jt})'$, $x_t = (x_{1t}, \ldots, x_{Jt})'$, and where $F(v)$ denotes the *cdf* of v. Let

$$\mathrm{s}(\sigma, \delta_t, x_t) = (\mathrm{s}_1(\sigma, \delta_t, x_t), \ldots, \mathrm{s}_J(\sigma, \delta_t, x_t))'$$

denote the vector of model implied market shares. Berry (1994) showed that for any vector of observed market shares, s_t, any given σ, and any x_t, there exists a unique vector $\delta(\sigma, s_t, x_t)$ such that

$$\mathrm{s}(\sigma, \delta(\sigma, s_t, x_t), x_t) = s_t.$$

Furthermore, there exists an inverse function, $\mathrm{s}^{-1}(\sigma, \cdot, x_t)$, such that $\delta(\sigma, s_t, x_t)$ is given by $\mathrm{s}^{-1}(\sigma, s_t, x_t)$. Let

$$\xi(\theta, x_t, s_t) = \delta(\sigma, s_t, x_t) - x_t'\mu,$$

where $\theta = (\mu, \sigma)$. In empirical applications, the existence of a $(J \times L)$ vector of instruments (with $L \geq 2K$), z_t, is assumed and the sample moment entering the GMM objective function is given by[10]

$$G_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} z_t' \xi(\theta, s_t, x_t), \tag{8}$$

where

$$\lim_{T\to\infty} G_T(\theta) = G(\theta, \gamma_0)$$

is assumed to uniquely identify θ_0, i.e., $G(\theta, \gamma_0) = 0 \Leftrightarrow \theta = \theta_0$. Inference on subvectors of θ is typically performed using Wald-type tests, such as the two-sided t-test, and is based on an asymptotic normality result. As evident from the results in Section 2, assuming asymptotic normality will provide a poor approximation to the finite sample distribution of the estimator when the true parameter vector is close to the boundary. As it turns out, the boundary effects are aggravated under the standard parameterization of the BLP model that uses $\theta = (\mu, \sigma)$. The reason is that the model suffers from first order lack of identification (see e.g., Sargan, 1983; Dovonon and Renault, 2013) at the boundary of the parameter space, i.e., Assumption 3(i), which is sufficient but not necessary for local identification, is violated if $\sigma_{k,0} = 0$ for some $k \in \{1, \ldots, K\}$. In fact, as shown in Appendix A,

$$\left.\frac{\partial \xi(\theta, s_t, x_t)}{\partial \sigma_k}\right|_{\sigma_k = 0} = 0,$$

which implies that

$$\left.\frac{\partial G_T(\theta)}{\partial \sigma_k}\right|_{\sigma_k = 0} = 0.$$

Put differently, the derivative of the sample moment and, consequently, the derivative of the population moment, the Jacobian, are of reduced rank at the boundary of the parameter space, regardless of the choice of instruments. This peculiarity can be avoided when we consider a reparameterization of the model in terms of variances, say $\theta^* = (\mu, \sigma^2)$.[11] With a slight abuse of notation, let

$$\mathrm{s}_j(\sigma^2, \delta_t, x_t) = \int \frac{e^{\delta_{jt} + \sum_{k=1}^{K} x_{jt,k}\sqrt{\sigma_k^2}\, v_k}}{1 + \sum_{l=1}^{J} e^{\delta_{lt} + \sum_{k=1}^{K} x_{lt,k}\sqrt{\sigma_k^2}\, v_k}} dF(v)$$

and define $\delta(\sigma^2, s_t, x_t)$, $\xi(\theta^*, s_t, x_t)$, $G_T(\theta^*)$ and $Q_T(\theta^*)$, with the same abuse of notation,

[10] Note that the objective function based on (8) can be minimized over $[-c, c]^{2K}$, since (7) is well defined for negative σ_k. However, the resulting estimator, which is obtained by taking the absolute value of the corresponding $\hat{\sigma}_T$, is equivalent to the estimator given in (1) and, therefore, the asymptotic theory derived in this paper also applies to this "alternative" estimator.

[11] The fact that the model does not suffer from lack of first order identification when parameterized with respect to σ^2, is equivalent to saying that the model is (or can be) second order identified at the boundary of the parameter space when parameterized with respect to σ, cf., equation (10). See Dovonon and Renault (2013) for a formal definition of second order identification.

accordingly. In Appendix A, it is shown that $\forall \sigma_k^2 > 0$

$$\frac{\partial \xi(\theta^*, s_t, x_t)}{\partial \sigma_k^2} = \frac{1}{2\sigma_k} \frac{\partial \xi(\theta, s_t, x_t)}{\partial \sigma_k}, \tag{9}$$

where $\sigma_k = \sqrt{\sigma_k^2}$. In addition, we have that

$$\lim_{\sigma_k^2 \to 0} \frac{\partial \xi(\theta^*, s_t, x_t)}{\partial \sigma_k^2} = \lim_{\sigma_k^2 \to 0} \frac{1}{2\sigma_k} \frac{\partial \xi(\theta, s_t, x_t)}{\partial \sigma_k} = \lim_{\sigma_k^2 \to 0} \frac{1}{2} \frac{\partial^2 \xi(\theta, s_t, x_t)}{\partial^2 \sigma_k}, \tag{10}$$

where the third equality follows by the rule of l'Hôpital. As shown in Appendix A, $\frac{\partial^2 \xi(\theta, s_t, x_t)(\theta)}{\partial^2 \sigma_k}$ is not identically equal to 0 when evaluated at $\sigma_k^2 = 0$, such that it is possible for Assumption 3(i) to be satisfied.

However, it is difficult to formulate low-level sufficient conditions (in terms of the joint distribution of z_t, x_t, and $\xi(\theta^*, s_t, x_t)$) due to the nonlinear nature of the model, cf. equations (9) and (10).[12] But we expect the approximation to the optimal instruments proposed in Berry, Levinsohn, and Pakes (1999) and Reynaert and Verboven (2014), which applies regardless of the parameterization of the model and which exploits the model inherent nonlinearities, to perform well in practice.

In order to be able to apply the results in Section 2, with θ^* in place of θ, we need to define the corresponding true parameter space, Γ.[13] For ease of exposition, we assume that $\{s_t, x_t, z_t\}_{t=1}^T$ are i.i.d. with distribution $\phi \in \ddot{\Phi}$, where $\ddot{\Phi}$ is a compact metric space with a metric that induces weak convergence.[14] Furthermore, for sake of concreteness, we consider a two-step estimator. In the first step, $\mathcal{W}_T = \frac{1}{T} \sum_{t=1}^T z_t' z_t$, yielding an estimator $\bar{\theta}_T^*$. In the second step,

$$\mathcal{W}_T = \frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\bar{\theta}_T^*, s_t, x_t) \xi_t'(\bar{\theta}_T^*, s_t, x_t) z_t,$$

where with a slight abuse of notation z_t may denote different instruments in the first and

[12] Berry, Linton, and Pakes (2004) and Freyberger (2015) (who analyze the asymptotic distribution of the GMM estimator in the BLP model) also directly assume the Jacobian to be of full rank, due to difficulty of formulating low-level sufficient conditions, see their Assumptions B2 and RC9, respectively.

[13] With slight abuse of notation, we continue to let $\ddot{\Theta}$ and Θ denote the true and optimization parameter spaces for θ^*.

[14] The i.i.d. assumption is not innocuous here, as it implies that J, the number of products, does not vary across markets, as for example allowed for in Freyberger (2015). While the results can easily be extended to allow for independent but not identically distributed data, we refrain from doing so for ease of exposition.

second step. Then, for any $\ddot{\theta}^* \in \ddot{\Theta}$, the true parameter space for ϕ is given by[15]

$$\ddot{\Phi}(\ddot{\theta}^*) = \{\phi \in \ddot{\Phi} : E_\phi z_t' \xi_t(\ddot{\theta}^*, s_t, x_t) = 0, E_\phi \sum_{i=1}^{4} M_i(s_t, x_t, z_t) \leq C$$
$$\lambda_{\min}(E_\phi z_t' z_t) \geq \epsilon, \lambda_{\min}(E_\phi z_t' \xi_t(\ddot{\theta}^*, s_t, x_t) \xi_t'(\ddot{\theta}^*, s_t, x_t) z_t) \geq \epsilon,$$
$$E_\phi z_t' \frac{\partial \xi(\theta^*, s_t, x_t)}{\partial \theta^{*\prime}} \text{ has full column rank } \forall \theta^* \in \Theta\} \tag{11}$$

for some constants $C < \infty$ and $\epsilon > 0$, where $M_1(s_t, x_t, z_t)$-$M_4(s_t, x_t, z_t)$ are defined in Appendix C. The verification of Assumptions 1-3 for the BLP model with $\ddot{\Phi}(\ddot{\theta}^*)$ given in (11) can also be found in Appendix C.

4 Asymptotic size

For expositional purposes, we focus on the scalar case and derive AsySz of the CI obtained by inverting the two-sided t-test based on σ, which constitutes common practice.[16,17] In addition, we derive AsySz of the CI obtained by inverting the two-sided t-test based on σ^2. From equation (9), it follows that the t-statistic used in practice is given by[18]

$$t_\sigma = \sqrt{T} \frac{\hat{\sigma}_{1,T} - \sigma_1}{\frac{\sqrt{\hat{V}_{K+1}(\hat{\theta}_T^*)}}{2\hat{\sigma}_{1,T}}}, \tag{12}$$

where $\hat{V}_{K+1}(\hat{\theta}^*) \equiv \hat{V}_{K+1,K+1}(\hat{\theta}^*)$ denotes the $K+1^{\text{th}}$ entry of the main diagonal of an estimator of $V(\gamma_0)$, the variance of the normal random variable underlying the asymptotic distribution of $\sqrt{T}(\hat{\theta}_T^* - \theta_T^*)$. Define $t_\sigma = 0$ when $\hat{\sigma}_{1,T} = 0$. Under $\gamma_T = (\theta_T^*, \phi_T)$, the coverage probability of the CI based on t_σ is given by

$$CP_{\sigma,T}(\gamma_T) = P_{\gamma_T}\left(\left|\sqrt{T}\frac{\hat{\sigma}_{1,T} - \sigma_{1,T}}{\frac{\sqrt{\hat{V}_{K+1}(\hat{\theta}_T^*)}}{2\hat{\sigma}_{1,T}}}\right| < z_{1-\alpha/2}\right).$$

[15]The moment conditions (involving $M_1(s_t, x_t, z_t)$-$M_4(s_t, x_t, z_t)$) are used to apply the uniform LLN and the CLT given in Andrews and Cheng (2014b). A sufficient condition is that $E_\phi \|z_t\|^{2+\epsilon} \leq C$, $\epsilon \leq s_{jt} \leq 1 - \epsilon$ for all $t = 1, \ldots, T$ and $j = 0, 1, \ldots, J$ with $s_{0t} = 1 - \sum_{j=1}^J s_{jt}$, and that x_t is in a compact set, because, then, $\xi(\theta^*, s_t, x_t)$ is bounded, see Appendix C in Freyberger (2015).

[16]We focus on the two-sided t-test, because it reflects common practice. However, from the exposition below, it is clear that a CI obtained by inverting a one-sided t-test will also suffer from undercoverage.

[17]From the results in Andrews and Guggenberger (2010b), it can be deduced that the CI obtained by inverting the two-sided t-test based on μ controls (asymptotic) size for $K = 1$, if $\mathcal{W} = \Omega(\gamma_0)^{-1}$. For large K, however, this no longer holds true, see Section 4.1.2 in Ketz (2017a).

[18]Without loss of generality, we consider the first entry of σ, σ_1.

Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, the asymptotic coverage probability equals (see Appendix C for details)

$$\mathrm{CP}_\sigma(\tilde{h}) \equiv 1 - P_{\gamma_0}\left(\hat{\lambda}_{h,K+1} > \frac{\left(\sqrt{h_1} + \sqrt{h_1 + 2 \cdot z_{1-\alpha/2}\sqrt{V_{K+1}(\gamma_0)}}\right)^2}{4} - h_1\right) - \mathbf{1}(h_1 - 2 \cdot z_{1-\alpha/2}\sqrt{V_{K+1}(\gamma_0)} > 0) \cdot$$
$$P_{\gamma_0}\left(\frac{\left(\sqrt{h_1} - \sqrt{h_1 - 2 \cdot z_{1-\alpha/2}\sqrt{V_{K+1}(\gamma_0)}}\right)^2}{4} - h_1 < \hat{\lambda}_{h,K+1} < \frac{\left(\sqrt{h_1} + \sqrt{h_1 - 2 \cdot z_{1-\alpha/2}\sqrt{V_{K+1}(\gamma_0)}}\right)^2}{4} - h_1\right), \tag{13}$$

where

$$\tilde{h} = (h, \gamma_0) \text{ and } \tilde{H} = \{\tilde{h} = (h, \gamma_0) : h \in (\mathbb{R}_+ \cup \{\infty\})^K, \gamma_0 \in \Gamma\}.$$

Equivalently, under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, the asymptotic coverage probability for the two-sided t-test based on σ^2, i.e.,

$$t_{\sigma^2} = \sqrt{T}\frac{\hat{\sigma}^2_{1,T} - \sigma^2_1}{\sqrt{\hat{V}_{K+1}(\hat{\theta}^*_T)}}, \tag{14}$$

is given by

$$\mathrm{CP}_{\sigma^2}(\tilde{h}) \equiv P_{\gamma_0}\left(\left|\hat{\lambda}_{h,K+1}\right| < z_{1-\alpha/2}\sqrt{V_{K+1}(\gamma_0)}\right). \tag{15}$$

Comparing equations (12) and (14), we can interpret the practice of relying on an asymptotic normality result for t_σ as applying the delta method to an asymptotic normality result for t_{σ^2}. Note that $\mathrm{CP}_\sigma(\tilde{h}) = \mathrm{CP}_{\sigma^2}(\tilde{h})$ when $h_1 = \infty$, which amounts to the delta method being applicable when $\sigma^2_{1,T}$ is "far enough" from the boundary, despite the (possible) lack of asymptotic normality of t_{σ^2}.

Proposition 2. *If the random coefficients logit model satisfies* (11)*, the nominal* $1-\alpha$*, two-sided CI based on* t_σ *satisfies* $AsySz_\sigma = \inf_{\tilde{h}\in\tilde{H}} CP_\sigma(\tilde{h})$ *and the nominal* $1-\alpha$*, two-sided CI based on* t_{σ^2} *satisfies* $AsySz_{\sigma^2} = \inf_{\tilde{h}\in\tilde{H}} CP_{\sigma^2}(\tilde{h})$*.*

The proof of Proposition 2 follows from (a slightly modified version of) Lemma 2.1 in Andrews and Cheng (2012), see Appendix C for details.

From (13), (15) and Proposition 1, it follows that AsySz_σ and $\mathrm{AsySz}_{\sigma^2}$, characterized in Proposition 2, depend on γ_0 only through $V(\gamma_0)$ and $J(\gamma_0)$. Ideally, we would like to obtain "numbers" for AsySz_σ and $\mathrm{AsySz}_{\sigma^2}$. However, the set of $V(\gamma_0)$ and $J(\gamma_0)$ admissible under (11) is rather abstract, preventing such computations. An exception occurs when $K = 1$, since then[19]

$$\sqrt{T}(\hat{\sigma}^2_{1,T} - \sigma^2_{1,T}) \xrightarrow{d} \hat{\lambda}_{h,2} = \max(-h_1, Z_2). \tag{16}$$

Put differently, AsySz does not depend on $J(\gamma_0)$ when $K = 1$. Furthermore, $V_2(\gamma_0)$ can be normalized to 1 without loss of generality (WLOG), such that AsySz only depends on h_1. We have $\text{AsySz}_\sigma = \inf_{h_1 \in [0,\infty]} \text{CP}_\sigma(h_1)$ and $\text{AsySz}_{\sigma^2} = \inf_{h_1 \in [0,\infty]} \text{CP}_{\sigma^2}(h_1)$.

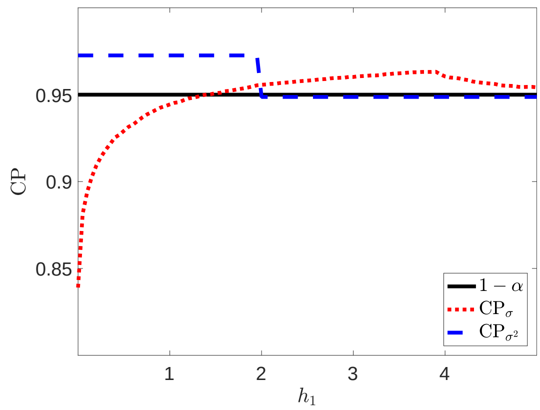

Figure 1: $\text{CP}_\sigma(h_1)$ and $\text{CP}_{\sigma^2}(h_1)$ as a function of h_1 - $K = 1$.

Figure 1 plots $\text{CP}_\sigma(h_1)$ and $\text{CP}_{\sigma^2}(h_1)$ as a function of h_1.[20] We see that $\text{AsySz}_\sigma = \text{CP}_\sigma(0) = 83.65\%$ and $\text{AsySz}_{\sigma^2} = \text{CP}_{\sigma^2}(\infty) = 95\%$ for $\alpha = 0.05$. The latter finding can equally be deduced from the results in Andrews and Guggenberger (2010a).

To shed some light on AsySz_σ and AsySz_{σ^2} for $K > 1$, we restrict our attention to efficient GMM estimators, in which case $J(\gamma_0) = V(\gamma_0)^{-1}$. Again, WLOG we can restrict the main diagonal of $V(\gamma_0)$ to a vector of 1s. If $K = 2$, the distribution of $\hat{\lambda}_{h,3}$ depends on h and $\rho \equiv V_{34}(\gamma_0)$ only. We have $\text{AsySz}_\sigma = \inf_{(h,\rho)\in[0,\infty]^2\times[1-\epsilon,1+\epsilon]} \text{CP}_\sigma(h,\rho)$ and $\text{AsySz}_{\sigma^2} = \inf_{(h,\rho)\in[0,\infty]^2\times[1-\epsilon,1+\epsilon]} \text{CP}_{\sigma^2}(h,\rho)$, assuming that (11) permits ρ to take on any value in $[-1+\epsilon, 1-\epsilon]$ for some $\epsilon > 0$.

The left panel of Figure 2 plots $\text{CP}_\sigma(0,\rho)$ and $\text{CP}_{\sigma^2}(0,\rho)$ as a function of ρ for $\alpha = 0.05$. We conclude, based on additional results not reported here, that $\text{AsySz}_\sigma = \text{CP}_\sigma(0, 0.55) \approx 80.74\%$ and $\text{AsySz}_{\sigma^2} = \text{CP}_{\sigma^2}(\infty^2, 0) = 95\%$.[21]

Computing AsySz requires a search over $\tilde{H}$, which amounts to a search over a $K + K(K-1)/2$ dimensional space, given by H and the set in which the off-diagonal elements

[19]By the Continuous Mapping Theorem, it follows that for $h < \infty$

$$T^{1/4}\hat{\sigma}_T \xrightarrow{d} \sqrt{\max(0, Z_2 + h)},$$

which explains why some researchers have found bimodal histograms for $\hat{\sigma}_T$ in Monte Carlo simulations, see for example Figure 1 in Reynaert and Verboven (2014) and their footnote 4. See also Figure 3 below.

[20]Figure 1 numerically evaluates $\text{CP}_\sigma(h_1)$ and $\text{CP}_{\sigma^2}(h_1)$ using 10,000 draws from $Z_2 \sim N(0,1)$.

[21]For any $\rho > 0$ we have $\text{CP}_\sigma(0,\rho) \leq \text{CP}_\sigma(h,\rho)$ for any $h \in [0,\infty]^2$, while for any $\rho \leq 0$ we have $\text{CP}_\sigma((0,0),\rho) \leq \text{CP}_\sigma((0,\infty),\rho) = 83.65\%$.

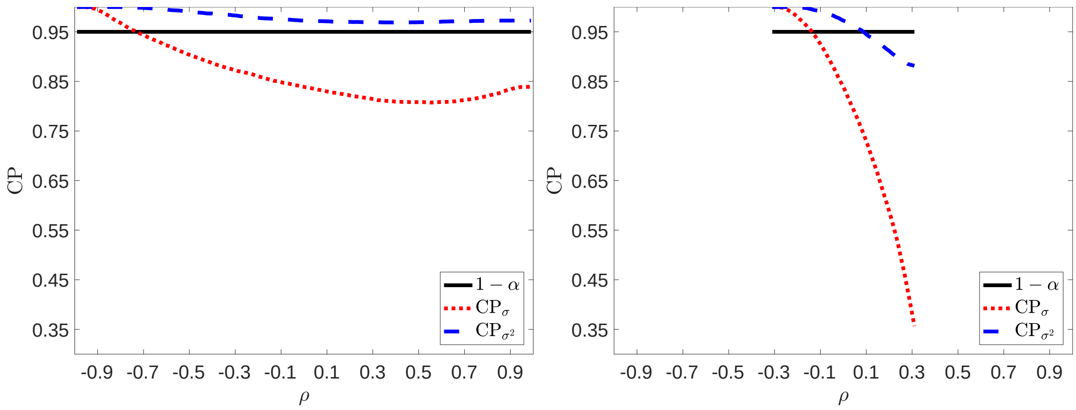

Figure 2: $\mathrm{CP}_{\sigma}(0, \rho)$ and $\mathrm{CP}_{\sigma^2}(0, \rho)$ as a function of ρ - $K = 2$ (left) and $K = 11$ (right).

of the block of $V(\gamma_0)$ corresponding to σ^2 lie. For $K > 2$, the search quickly becomes computationally infeasible. To get an idea of "how bad" things can get we consider a specific structure of $V(\gamma_0)$ that restricts the block of $V(\gamma_0)$ corresponding to σ^2 to be of the following form

$$\begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & 0 & \dots & 0 \\ \rho & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \rho & 0 & \dots & 0 & 1 \end{bmatrix}. \tag{17}$$

Assuming that (11) permits this structure, the right panel of Figure 2, which plots $\mathrm{CP}_{\sigma}(0, \rho)$ and $\mathrm{CP}_{\sigma^2}(0, \rho)$ as a function of ρ (for which (17) is invertible) for $K = 11$ and $\alpha = 0.05$, provides an upper bound for AsySz_{σ} and $\mathrm{AsySz}_{\sigma^2}$. We conclude that undercoverage of the CI based on t_{σ} becomes more severe for larger K and that also the CI based t_{σ^2} can suffer from undercoverage for large K.

5 A simple solution

While it is possible to construct CIs based on t_{σ} or t_{σ^2} that control asymptotic size (irrespective of the dimension of K), by choosing appropriate, possibly data-dependent, critical values (see e.g., Andrews and Cheng (2012) in a related context), we suggest an alternative approach, which uses standard critical values and is, thus, easy to implement. In particular, we suggest to construct CIs by inverting a two-sided t-test that is (also) based on σ^2 but, instead of $\hat{\theta}_T^*$, uses an estimator that is asymptotically normally distributed under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, say $\tilde{\theta}_T^*$, as suggested in Ketz (2017a). It is given by the minimizer of the sample analogue of the quadratic part of (2), i.e.,

$$\tilde{\theta}_T^* = \hat{\theta}_T^* - \left(\hat{G}_{\theta^*,T}' \mathcal{W}_T \hat{G}_{\theta^*,T}\right)^{-1} \hat{G}_{\theta^*,T}' \mathcal{W}_T G_T(\hat{\theta}_T^*), \tag{18}$$

where

$$\hat{G}_{\theta^*,T} = \frac{1}{T}\sum_{t=1}^{T} z_t' \frac{\partial \xi(\hat{\theta}_T^*, s_t, x_t)}{\partial \theta^{*\prime}}.$$

Since $\sqrt{T}(\tilde{\theta}_T^* - \theta_T^*) \xrightarrow{d} Z$ under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ (see Appendix B for details), it follows that the (two-sided) CI based on

$$t_{\sigma^2}^{\text{alt}} = \sqrt{T}\frac{\tilde{\sigma}_{1,T}^2 - \sigma_{1,T}^2}{\sqrt{\hat{V}_{K+1}(\hat{\theta}_T^*)}},$$

which is asymptotically pivotal, controls asymptotic size (uniformly). Note, however, that the two-sided CI based on $t_{\sigma^2}^{\text{alt}}$ is "unreasonable," i.e., it may be empty or unreasonably short with positive probability, cf. Müller and Norets (2016). While this possibility constitutes a valid concern, CIs based on $t_{\sigma^2}^{\text{alt}}$ have the advantage of being conceptually simple and easy to implement. In particular, if $\hat{\theta}_T^*$ is not at the boundary of the parameter space, then $\tilde{\theta}_T^* = \hat{\theta}_T^*$ (cf. equation (18)) and, consequently, $t_{\sigma^2}^{\text{alt}} = t_{\sigma^2}$, such that the CI that is obtained by inverting the two-sided t-test based on σ^2 that uses the constrained estimator, $\hat{\theta}^*$, is valid in that case. Alternatively, one could construct CIs that are never empty by inverting the test proposed in Ketz (2017a).

6 Monte Carlo

In this section, we perform a small Monte Carlo study to investigate the quality of the approximation provided by the asymptotic theory derived in this paper. For ease of reference, we use the same data generating process as Reynaert and Verboven (2014) (RV) with only minor modifications. We consider three product characteristics, $x_{jt,1}$ through $x_{jt,3}$. $x_{jt,1}$ can be thought to represent price and is modeled as being endogenous. In particular, $x_{jt,1}$ is generated as follows

$$x_{jt,1} = w_{jt}'\pi_1 + z_{jt}'\pi_2 + \zeta_{jt},$$

where $w_{jt} = (x_{jt,2}, x_{jt,3})'$ is the set of exogenous product characteristics and z_{jt} is 3×1 dimensional vector of cost shifters. This reflects the case of perfect competition, where there is no markup on price. The endogeneity of $x_{jt,1}$ arises, because the error terms, ξ_{jt} and ζ_{jt}, are generated according to

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

The two exogenous product characteristics, $x_{jt,2}$ and $x_{jt,3}$, are given by 1 and $U[1,2]$, respectively, where $U[a,b]$ denotes a uniform random variable with support on $[a,b]$. z_{jt} is generated as a vector of independent $U[0,1]$. RV generate data according to $\sigma_1 = \sigma_2 = 0$, while $\sigma_3 = 1$, but only estimate σ_3, i.e., $\sigma_1 = \sigma_2 = 0$ is assumed known. Here, we choose $\sigma \equiv \sigma_3 = 0$, while we continue to assume $\sigma_1 = \sigma_2 = 0$ known. The rest of the parameter values is chosen as follows: $\mu = (\mu_1, \mu_2, \mu_3)' = (-2, 2, 2)'$, $\pi_1 = (0.7, 0.7)'$, $\pi_2 = (3, 3, 3)'$. The sample size is given by $T = 25$ and $J = 10$.

The optimal instruments for the exogenous product characteristics, $x_{jt,2}$ and $x_{jt,3}$, are the product characteristics themselves. The optimal instrument for price, $x_{jt,1}$, is given by

$$w_{jt}'\hat{\pi}_1 + z_{jt}'\hat{\pi}_2,$$

where $\hat{\pi}_1$ and $\hat{\pi}_2$ denotes the ordinary least squares estimators from a regression of $x_{jt,1}$ on w_{jt} and z_{jt}. We implement a one-step estimator. We avoid the first-step of a two-step estimation procedure by evaluating the approximation to the optimal instruments for σ, suggested by RV, at a random guess of σ rather than at a first-step estimate.[22] This procedure is described in footnote 5 of RV and found to perform equally well in Monte Carlo simulations. Due to the homoskedastic nature of the data, the usual reason for implementing a two-step estimator does not apply.

The integral in equation (7) is approximated by sparse grid integration. The resulting approximation error is immaterial, because the main points of this paper remain valid as long as the distribution of v has mean zero, see Appendix A. Furthermore, we use the same number of knots (7) and the same weights for estimation and for generating true market shares as to not create any sampling error. In light of recent findings, we employ the mathematical program with equilibrium constraints (MPEC) formulation of the estimation problem as proposed by Dubé, Fox, and Su (2012a).

Column 1 of Table 1 below reports the average (over 1000 replications) of the point estimates and what we, with a slight abuse of terminology, refer to as standard errors of $\hat{\sigma}_T^2$ and $\hat{\sigma}_T$, namely $\sqrt{\frac{\hat{V}_{K+1}(\hat{\theta}_T^*)}{T}}$ and $\frac{\sqrt{\hat{V}_{K+1}(\hat{\theta}_T^*)}}{\sqrt{T}2\hat{\sigma}_T}$.[23] Column 2 reports the rejection frequency (RF) of the common Wald-type t-test for testing $H_0 : \sigma^2 = \sigma = 0$ at the 5% significance level. The rest of the table reports some quantiles (over 1000 replications) of the estimates and their standard errors. The motivation for looking at the quantiles of the standard errors stems from the fact that $\hat{\sigma}_T$ enters the denominator of $\frac{\sqrt{\hat{V}_{K+1}(\hat{\theta}_T^*)}}{\sqrt{T}2\hat{\sigma}_T}$ and that $\hat{\sigma}_T$ can be at the boundary, i.e., equal to zero, such that we expect the standard error for $\hat{\sigma}_T$ to behave very irregularly. Note that in practice computing standard errors does not cause a problem, because typically (constrained) minimization algorithms restrict $\hat{\sigma}_T$ to be strictly greater than zero, i.e., if $\hat{\sigma}_T$ is "at the boundary" it differs from 0 by an algorithm specific tolerance level.

Table 1 shows that the asymptotic results obtained in Section 4 provide good approxi-

[22] A random guess is obtained as $|N(0,1)|$.

[23] This constitutes an abuse of terminology because the asymptotic distribution of $\hat{\sigma}_T^2$ and $\hat{\sigma}_T$ are not normal when $\sigma^2 = \sigma = 0$, see equation (16) and footnote 19.

Table 1: Monte Carlo Results - $H_0 : \sigma^2 = \sigma = 0$

			Quantiles				
Estimator	Average	RF	0.05	0.25	0.5	0.75	0.95
$\hat{\sigma}_T^2$	0.079	0.023	0.000	0.000	0.010	0.115	0.328
SE($\hat{\sigma}_T^2$)	0.202		0.115	0.148	0.181	0.229	0.353
$\hat{\sigma}_T$	0.185	0.167	0.000	0.000	0.101	0.340	0.573
SE($\hat{\sigma}_T$)	10,807.094		0.161	0.269	0.874	11,485.022	49,797.439

mations. In particular, the finite-sample rejection frequencies of 2.3% and 16.7% are very close to the corresponding asymptotic rejection frequencies of 2.5% and 16.35%, given in Figure 1 for $h_1 = 0$, which corresponds to $H_0 : \sigma^2 = \sigma = 0$. Furthermore, the standard error for $\hat{\sigma}_T$ can be huge with the 95$^{\text{th}}$ quantile at about 50,000.

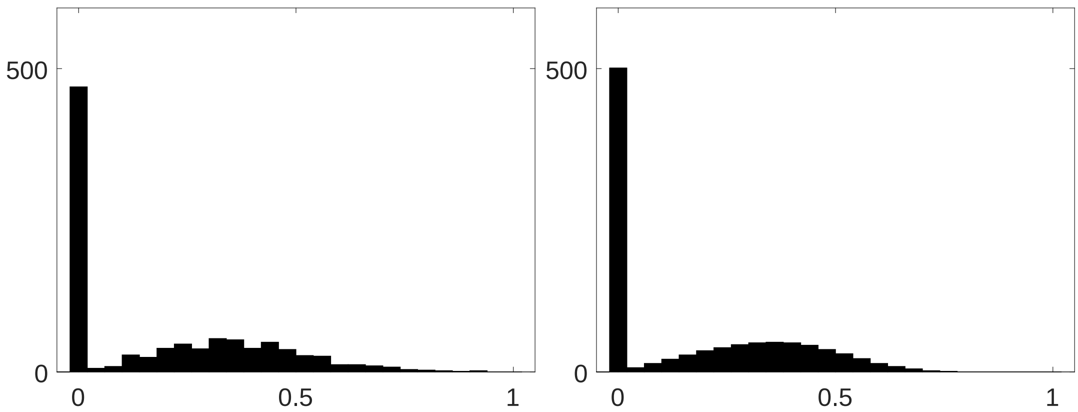

Figure 3: Histogram of finite-sample (left) and asymptotic (right) distribution of $\hat{\sigma}_T$.

While Table 1 shows that the finite sample rejection frequency of the two-sided t-test is well approximated by our asymptotic theory, it does not illustrate the nonstandard distribution of $\hat{\sigma}_T$. Therefore, Figure 3 plots the finite-sample (over 1000 replications) and the asymptotic distribution of $\hat{\sigma}_T$, see also footnote 19.

6.1 Computational efficiency

So far, we have been quite about how to obtain $\hat{\theta}_T^*$. Since θ^* is a one-to-one function of θ, $\hat{\theta}_T^*$ is easily obtained by estimating the model with respect to θ and by squaring the corresponding $\hat{\sigma}_{T,k}$ $\forall k \in \{1, \dots, K\}$, which might be tempting given that the publicly available code for estimating the random coefficients logit model uses θ. However, since typical optimization algorithms, such as variants of the Newton-Raphson method, rely on either closed form expressions or numerical approximations of the Jacobian and the Hessian of the minimization problem, which are functions of the first order derivative of the sample moment, we expect minimization with respect to θ^* to be more reliable.

Next, we present some Monte Carlo evidence of this computational advantage using the same data generating process as above. Estimation is performed with respect to θ and θ^* separately. In addition to allowing for only one random coefficient in estimation, we also perform estimation with respect to an additional variance parameter, namely on the constant. With a slight abuse of notation, these two cases are referred to as $K = 1$ and $K = 2$ in Table 2 below, which reports the average and the median number of iterations that the optimization algorithm needs in order to achieve convergence. For each cell, the total number of optimizations is 2000, 1000 Monte Carlo replications using 2 different starting values. In addition, Table 2 reports the number of times the algorithm did not converge (defined as having failed to find a local minimum after 100 iterations).

Table 2: Numerical performance

# iterations until convergence				
	θ		θ^*	
K	Average	Median	Average	Median
1	6.64	6	4.86	4
2	20.35	7	6.77	5
# convergence failures				
K	θ		θ^*	
1	5		2	
2	37		2	

Table 2 shows that on average the algorithm needs fewer iterations in order to achieve convergence when optimization is performed with respect to θ^*, 4.86 vs. 6.64. The difference in the average number of iterations is more pronounced when $K = 2$, 6.77 vs. 20.35. This is partly due to the high number of convergence failures that are encountered when $K = 2$. The median, which is not influenced by such convergence failures, indeed only indicates a minor speed advantage of the optimization with respect to θ^*. Nevertheless, the speed advantage is present and consistent across specifications, $K = 1, 2$.

In summary, we recommend optimization with respect to θ^*, because the expected number of iterations required to achieve convergence is lower and the algorithm is less prone to convergence failures. To facilitate implementation, Appendix A contains all the first and second order derivatives that are required to implement a modified MPEC algorithm and that differ from those presented in Dubé, Fox, and Su (2012b).

7 Applications

Comparing equation (12) with equation (14), we see that the standard error of $\hat{\sigma}^2_{T,1}$ is obtained by multiplying the standard error of $\hat{\sigma}_{T,1}$ by $2 \cdot \hat{\sigma}_{T,1}$. Therefore, a reparameterization can be performed without direct access to the data, as long as $\hat{\sigma}_{T,1}$ and its standard error are reported.

The problem of large standard errors can, for example, be observed in Neilson (2013), who reports an estimate of the standard deviation of the random coefficient on the "Quality" variable in (his) Table 4 of 0.001. The corresponding standard error is 0.7607, which is much larger than the other standard errors from the same estimation, which are in the order of 0.01. Upon reparameterization, the estimate of the variance of the random coefficient is 0.000001 with a standard error of 0.0015. His conclusion remains unaltered: There seems to be little or no heterogeneity with respect to the "Quality" variable. But his conclusion has, in fact, more support from the data than initially thought.

Next, we investigate in how far undercoverage of the CI based on t_σ may have influenced conclusions with respect to the presence of heterogeneity in previous work. In what follows, we reproduce the estimation results for a few published articles that use the BLP model and apply the reparameterization in terms of variances. This merely serves to illustrate the potential relevance of our findings and is not meant to question the validity of the findings in these papers. The following tables reproduce estimates of standard deviations and corresponding standard errors given in Berry, Levinsohn, and Pakes (1995) (BLP), Berry, Levinsohn, and Pakes (1999), and Reynaert and Verboven (2014). In addition, we also report estimates of σ^2 along with corresponding standard errors. For convenience, we also report the corresponding 95% CIs obtained by inverting the two-sided t-test. All three papers analyze the demand for cars. BLP and Berry, Levinsohn, and Pakes (1999) analyze the US car market from 1971 to 1990, while Reynaert and Verboven (2014) analyze the car market of nine European countries from 1998 to 2010. The product characteristics in BLP and Berry, Levinsohn, and Pakes (1999), which are interacted with a random coefficient in estimation, are horse power per weight (HP/Weight), a dummy for whether the car has air conditioning (Air), miles per gallon (MP$) and size (Size). The corresponding product characteristics in Reynaert and Verboven (2014) are price divided by income (Price/Inc.), both in local currency, horse power divided by weight (Hp/We.), a dummy variable to indicate if the car make is foreign (Foreign), size (Size), height (Height) and a measure of fuel efficiency (€/km). In all three papers, the estimates of σ and, thus, σ^2 are in the interior of the parameter space such that CIs based on t_{σ^2} are valid, cf. Section 5.

Table 3: Berry, Levinsohn, and Pakes (1995) - Table 4 (left panel)

Variable	σ	SE	95% CI	σ^2	SE	95% CI
Constant	3.612	1.485	[0.701,6.523]	13.047	10.728	[0,34.074]
HP/Weight	4.628	1.885	[0.933,8.323]	21.418	17.448	[0,55.616]
Air	1.818	1.695	[0,5.140]	3.305	6.163	[0,15.384]
MP$	1.050	0.272	[0.517,1.583]	1.103	0.571	[0,2.222]
Size	2.056	0.585	[0.909,3.203]	4.227	2.506	[0,9.139]

Table 3 reproduces part of (the left panel of) Table 4 in BLP. None of the CIs based on t_σ except for the one on Air include 0, while all CIs based on t_{σ^2} include 0.

Table 4: Berry, Levinsohn, and Pakes (1999) - Table 5

Variable	σ	SE	95% CI	σ^2	SE	95% CI
Constant	1.112	1.171	[0,3.407]	1.237	2.604	[0,6.341]
HP/Weight	0.167	4.652	[0,9.285]	0.028	1.554	[0,3.074]
Size	1.392	0.707	[0.006,2.778]	1.938	1.968	[0,5.795]
Air	0.377	0.886	[0,2.114]	0.142	0.668	[0,1.451]
MP$	0.416	0.132	[0.157,0.675]	0.173	0.110	[0,0.389]

Table 4 reproduces part of Table 5 in Berry, Levinsohn, and Pakes (1999). The CIs based on t_σ are indicative of heterogeneity in consumer tastes with respect to Size and MP$. The CIs based on t_{σ^2}, on the other hand, all include 0 and, thus, provide less evidence of heterogeneous preferences.

Table 5: Reynaert and Verboven (2014) - Table 6 Optimal instruments (ii)

Variable	σ	SE	95% CI	σ^2	SE	95% CI
Price/Inc.	0.524	0.168	[0.195,0.853]	0.274	0.176	[0,0.619]
Hp/We.	3.202	0.679	[1.871,4.533]	10.252	4.346	[1.734,18.770]
Foreign	0.718	0.513	[0,1.723]	0.515	0.736	[0,1.958]
Size	0.239	0.394	[0,1.011]	0.057	0.189	[0,0.427]
Height	0.104	0.030	[0.045,0.163]	0.011	0.006	[0,0.028]
€/km	2.103	4.715	[0,11.344]	4.424	19.835	[0,43.301]

Table 5 reproduces part of Table 6 in Reynaert and Verboven (2014). Here, the CIs based on t_σ exclude 0 for three out of six product characteristics, namely Price/Inc., Hp/We., and Height. The CIs based on t_{σ^2}, however, include 0 for all but one product characteristic, namely Hp/We.

8 Conclusion

In this paper, we show that the standard parameterization of the BLP model can lead to inference problems. On the one hand, CIs obtained by inverting the standard two-sided t-test can suffer from undercoverage, while, on the other hand, unreasonably large standard errors can be obtained. The proposed solution is to make inference with respect to a one-to-one transformation of the original parameter vector. This approach avoids the problem of unreasonably large standard errors and the resulting CIs are valid, as long as estimates are found to be in the interior. When estimates are found to be at the boundary, an alternative estimator, suggested in Ketz (2017a), can be used in the construction of

CIs. We also recommend that researchers estimate the BLP model with respect to the transformed parameter vector, given the computational advantages documented in our Monte Carlo study.

A Derivatives of sample moment

First, we provide the derivative of the sample moment, $G_T(\theta)$, with respect to θ. Then, we provide the derivative of $G_T(\theta^*)$ with respect to θ^*. The latter can be used to modify the code of Dubé, Fox, and Su (2012a) in order to estimate the model with respect to θ^*.

A.1 Derivative of $G_T(\theta)$ with respect to θ

Define

$$\mathcal{S}_{jt}(\sigma, v) = \frac{e^{\delta_{jt}+\sum_{k=1}^{K} x_{jt,k}\sigma_k v_k}}{1+\sum_{l=1}^{J} e^{\delta_{lt}+\sum_{k=1}^{K} x_{lt,k}\sigma_k v_k}}.$$

With slight abuse of notation, let δ_{jt} and ξ_{jt} denote the j^{th} entry of $\delta(\sigma, s_t, x_t)$ and $\xi(\theta, s_t, x_t)$, respectively. Then, we have

$$\frac{\partial \xi_{jt}}{\partial \mu_k} = -x_{jt,k}$$

and for $i, i' \geq 1$ and $k, k' \in \{1, \dots, K\}$

$$\frac{\partial^{i+i'} \xi_{jt}}{\partial^{i'} \mu_{k'} \partial^{i} \mu_k} = 0.$$

Also for $i, i' \geq 1$ and $k, k' \in \{1, \dots, K\}$

$$\left(\frac{\partial^{i+i'}}{\partial^{i} \mu_k \partial^{i'} \sigma_{k'}} \xi_{jt} = \right) \frac{\partial^{i+i'}}{\partial^{i'} \sigma_{k'} \partial^{i} \mu_k} \xi_{jt} = 0.$$

Furthermore, we have (by the implicit function theorem)

$$\frac{\partial \xi_t}{\partial \sigma_k} = \frac{\partial \delta_t}{\partial \sigma_k} = -\left(\frac{\partial \mathrm{s}}{\partial \delta_t'} \right)^{-1} \frac{\partial \mathrm{s}}{\partial \sigma_k}.$$

$\frac{\partial \mathrm{s}}{\partial \delta_t'}$ has typical elements $(j' \neq j)$

$$\frac{\partial \mathrm{s}_j}{\partial \delta_{jt}} = \int \mathcal{S}_{jt}(\sigma, v) \left(1 - \mathcal{S}_{jt}(\sigma, v)\right) dF(v),$$
$$\frac{\partial \mathrm{s}_j}{\partial \delta_{j't}} = -\int \mathcal{S}_{jt}(\sigma, v) \mathcal{S}_{j't}(\sigma, v) dF(v)$$

and $\frac{\partial \text{s}}{\partial \sigma_k}$ has typical element

$$\frac{\partial \text{s}_j}{\partial \sigma_k} = \int \mathcal{S}_{jt}(\sigma, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt}(\sigma, v) x_{mt,k} \right) v_k \; dF(v).$$

Note that $\mathcal{S}_{jt}(\sigma, v)$ does not depend on v_k when $\sigma_k = 0$ and therefore can be written as $\mathcal{S}_{jt}(\sigma, v_{-k})$ when $\sigma_k = 0$, where v_{-k} denotes v without v_k. Thus, evaluated at $\sigma_k = 0$, the typical element of $\frac{\partial \text{s}}{\partial \sigma_k}$ equals

$$\left. \frac{\partial \text{s}_j}{\partial \sigma_k} \right|_{\sigma_k = 0} = \int \mathcal{S}_{jt}(\sigma, v_{-k}) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt}(\sigma, v_{-k}) x_{mt,k} \right) \; dF(v_{-k}) \int v_k \; dF(v_k) = 0. \tag{19}$$

The above equation equals zero because v_k has mean zero. Evaluated at $\sigma_k = 0$, we therefore have

$$\left. \frac{\partial \xi_t}{\partial \sigma_k} \right|_{\sigma_k = 0} = 0$$

and, thus,

$$\left. \frac{\partial \xi_{jt}}{\partial \sigma_k} \right|_{\sigma_k = 0} = 0$$

for all $j \in \{1, \ldots, J\}$. Thus, $\frac{\partial}{\partial \sigma_k} G_T(\theta) = 0$ for every $T \in \mathbb{N}$ and therefore $\frac{\partial}{\partial \sigma_k} G(\theta, \gamma_0) = 0$ whenever $\sigma_k = 0$. Furthermore, for $k, k' \in \{1, \ldots, K\}$

$$\frac{\partial^2 \xi_t}{\partial \sigma_{k'} \partial \sigma_k} = \frac{\partial}{\partial \sigma_{k'}} \frac{\partial \xi_t}{\partial \sigma_k} = \left(\frac{\partial \text{s}}{\partial \delta_t'} \right)^{-1} \frac{\partial^2 \text{s}}{\partial \sigma_{k'} \partial \delta_t'} \left(\frac{\partial \text{s}}{\partial \delta_t'} \right)^{-1} \frac{\partial \text{s}}{\partial \sigma_k} - \left(\frac{\partial \text{s}}{\partial \delta_t'} \right)^{-1} \frac{\partial^2 \text{s}}{\partial \sigma_{k'} \partial \sigma_k}.$$

Next to $\frac{\partial \text{s}}{\partial \delta_t'}$ and $\frac{\partial \text{s}}{\partial \sigma_k}$, we encounter $\frac{\partial^2 \text{s}}{\partial \sigma_{k'} \partial \delta_t'}$ and $\frac{\partial^2 \text{s}}{\partial \sigma_{k'} \partial \sigma_k}$. $\frac{\partial^2 \text{s}}{\partial \sigma_{k'} \partial \delta_t'}$ has typical elements ($j' \neq j$)

$$\begin{aligned}
\frac{\partial^2 \text{s}_j}{\partial \sigma_{k'} \partial \delta_{jt}} &= \int \mathcal{S}_{jt}(\sigma, v)(1 - 2\mathcal{S}_{jt}(\sigma, v)) \left(x_{jt,k'} - \sum_m \mathcal{S}_{mt}(\sigma, v) x_{mt,k'} \right) v_{k'} \; dF(v), \\
\frac{\partial^2 \text{s}_j}{\partial \sigma_{k'} \partial \delta_{j't}} &= - \int \mathcal{S}_{jt}(\sigma, v) \mathcal{S}_{j't}(\sigma, v) \left(x_{jt,k'} + x_{j't,k'} - 2 \sum_m \mathcal{S}_{mt}(\sigma, v) x_{mt,k'} \right) v_{k'} \; dF(v).
\end{aligned}$$

For $k' = k$, $\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'} \partial \sigma_k}$ is given by

$$\begin{aligned}
\frac{\partial^2 \mathrm{s}_j}{\partial^2 \sigma_k} &= \int \mathcal{S}_{jt}(\sigma, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt,k}(\sigma, v) x_{mt,k} \right)^2 v_k^2 \; dF(v) \\
&+ \int \mathcal{S}_{jt}(\sigma, v) \left(- \sum_m \{ \; \mathcal{S}_{mt}(\sigma, v) \left[x_{mt,k} - \sum_n \mathcal{S}_{nt}(\sigma, v) x_{nt,k} \right] v_k \; \} \, x_{mt,k} \right) v_k \; dF(v) \\
&= \int \mathcal{S}_{jt}(\sigma, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt}(\sigma, v) x_{mt,k} \right)^2 v_k^2 \; dF(v) \\
&- \int \mathcal{S}_{jt}(\sigma, v) \left(\sum_m \mathcal{S}_{mt}(\sigma, v) \left[x_{mt,k} - \sum_n \mathcal{S}_{nt}(\sigma, v) x_{nt,k} \right] x_{mt,k} \right) v_k^2 \; dF(v).
\end{aligned}$$

This evaluated at $\sigma_k = 0$, is not equal to zero. The expression factorizes as in (19) above, but $\int v_k^2 \; dF(v_k) = 1 \neq 0$. For $k' \neq k$, $\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'} \partial \sigma_k}$ is given by

$$\begin{aligned}
\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'} \partial \sigma_k} &= \int \mathcal{S}_{jt}(\sigma, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt,k}(\sigma, v) x_{mt,k} \right) v_k \left(x_{jt,k'} - \sum_m \mathcal{S}_{mt,k'}(\sigma, v) x_{mt,k'} \right) v_{k'} \; dF(v) \\
&- \int \mathcal{S}_{jt}(\sigma, v) \left(\sum_m \{ \; \mathcal{S}_{mt}(\sigma, v) \left[x_{mt,k'} - \sum_n \mathcal{S}_{nt}(\sigma, v) x_{nt,k'} \right] v_{k'} \; \} \, x_{mt,k} \right) v_k \; dF(v),
\end{aligned}$$

which evaluated at $\sigma_k = 0$ equals zero, again by the same argument as in (19).

A.2 Derivative of $G_T(\theta^*)$ with respect to θ^*

In what follows, we only present derivatives that differ from those above and those given in Section 1.2.1 of Dubé, Fox, and Su (2012b). Let

$$\mathcal{S}_{jt}(\sigma^2, v) = \frac{e^{\delta_{jt} + \sum_{k=1}^{K} x_{jt,k} \sqrt{\sigma_k^2} v_k}}{1 + \sum_{l=1}^{J} e^{\delta_{lt} + \sum_{k=1}^{K} x_{lt,k} \sqrt{\sigma_k^2} v_k}},$$

such that with a slight abuse of notation $\mathcal{S}_{jt}(\sigma^2, v) = \mathcal{S}_{jt}(\sigma, v)$. $\frac{\partial \mathrm{s}}{\partial \sigma_k^2}$ has typical element

$$\frac{\partial \mathrm{s}_j}{\partial \sigma_k^2} = \frac{1}{2\sqrt{\sigma_k^2}} \int \mathcal{S}_{jt}(\sigma^2, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt}(\sigma^2, v) x_{mt,k} \right) v_k \; dF(v_k).$$

Therefore, $\frac{\partial \mathrm{s}_j}{\partial \sigma_k^2} = \frac{1}{2\sigma_k}\frac{\partial \mathrm{s}_j}{\partial \sigma_k}$. The elements of $\frac{\partial^2 \mathrm{s}}{\partial \sigma_{k'}^2 \partial \delta_t'}$ are given by (for $j' \neq j$)

$$\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_k^2 \partial \delta_{jt}} = \frac{1}{2\sqrt{\sigma_k^2}} \int \mathcal{S}_{jt}(\sigma^2, v)(1 - 2\mathcal{S}_{jt}(\sigma^2, v)) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt}(\sigma^2, v) x_{mt,k} \right) v_k \, dF(v),$$
$$\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_k^2 \partial \delta_{j't}} = -\frac{1}{2\sqrt{\sigma_k^2}} \int \mathcal{S}_{jt}(\sigma^2, v)\mathcal{S}_{j't}(\sigma^2, v) \left(x_{jt,k} + x_{j't,k} - 2\sum_m \mathcal{S}_{mt}(\sigma^2, v) x_{mt,k} \right) v_k \, dF(v).$$

Thus, $\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_k^2 \partial \delta_{jt}} = \frac{1}{2\sigma_k}\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_k \partial \delta_{jt}}$ and $\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_k^2 \partial \delta_{j't}} = \frac{1}{2\sigma_k}\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_k \partial \delta_{j't}}$. For $k' = k$, $\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'}^2 \partial \sigma_k^2}$ is given by

$$\begin{aligned}
\frac{\partial^2 \mathrm{s}_j}{\partial^2 \sigma_k^2} &= \frac{1}{4\sigma_k^2} \int \mathcal{S}_{jt}(\sigma^2, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt,k}(\sigma^2, v) x_{mt,k} \right)^2 v_k^2 \, dF(v) \\
&\quad - \frac{1}{4\sigma_k^2} \int \mathcal{S}_{jt}(\sigma^2, v) \left(\sum_m \mathcal{S}_{mt}(\sigma^2, v) \left[x_{mt,k} - \sum_n \mathcal{S}_{nt}(\sigma^2, v) x_{nt,k} \right] x_{mt,k} \right) v_k^2 \, dF(v) \\
&\quad - \frac{1}{4(\sigma_k^2)^{\frac{3}{2}}} \int \mathcal{S}_{jt}(\sigma^2, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt}(\sigma^2, v) x_{mt,k} \right) v_k \, dF(v_k).
\end{aligned}$$

Note that $\frac{\partial^2 \mathrm{s}_j}{\partial^2 \sigma_k^2} = \frac{1}{4\sigma_k^2}\frac{\partial^2 \mathrm{s}_j}{\partial^2 \sigma_k} - \frac{1}{4\sigma_k^3}\frac{\partial \mathrm{s}_j}{\partial \sigma_k}$. For $k' \neq k$, $\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'}^2 \partial \sigma_k^2}$ is given by

$$\begin{aligned}
\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'}^2 \partial \sigma_k^2} &= \frac{1}{4\sqrt{\sigma_k^2}\sqrt{\sigma_{k'}^2}} \int \mathcal{S}_{jt}(\sigma^2, v) \left(x_{jt,k} - \sum_m \mathcal{S}_{mt,k}(\sigma^2, v) x_{mt,k} \right) v_k \\
&\qquad \left(x_{jt,k'} - \sum_m \mathcal{S}_{mt,k'}(\sigma^2, v) x_{mt,k'} \right) v_{k'} \, dF(v) \\
&\quad - \frac{1}{4\sqrt{\sigma_k^2}\sqrt{\sigma_{k'}^2}} \int \mathcal{S}_{jt}(\sigma^2, v) \\
&\qquad \left(\sum_m \{ \, \mathcal{S}_{mt}(\sigma^2, v) \left[x_{mt,k'} - \sum_n \mathcal{S}_{nt}(\sigma^2, v) x_{nt,k'} \right] v_{k'} \, \} \, x_{mt,k} \right) v_k \, dF(v).
\end{aligned}$$

It follows that $\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'}^2 \partial \sigma_k^2} = \frac{1}{4\sigma_k \sigma_{k'}}\frac{\partial^2 \mathrm{s}_j}{\partial \sigma_{k'} \partial \sigma_k}$.

B Verification of assumptions in Ketz (2017a,b)

We first provide the proof of Proposition 1 and then details about the asymptotic normality result for $\sqrt{T}(\tilde{\theta}_T^* - \theta_T^*)$, referred to in Section 5 above.

Proof of Proposition 1. Proposition 1 in Ketz (2017b) obtains the same asymptotic distribution result as Proposition 1, but under general high-level assymptions, namely Assumptions 2 and 3 in Ketz (2017a) and Assumptions 5 and 6 in Ketz (2017b). Therefore, the proof proceeds by verifying these Assumptions. Assumptions 1-3 correspond to Assumptions GMM1, GMM2, and GMM5 in Andrews and Cheng (2014a). Assumptions 2 and 3 in Ketz (2017a) and Assumption 6 in Ketz (2017b) correspond to Assumptions D1-D3 in Andrews and Cheng (2012). By Lemma 10.3 in Andrews and Cheng (2014b), Assumptions GMM1, GMM2, and GMM5 in Andrews and Cheng (2014a) imply Assumptions D1-D3 in Andrews and Cheng (2012).[24] Furthermore, by Lemma 10.1 in Andrews and Cheng (2014b) and Lemma 3.1 in Andrews and Cheng (2012), Assumption GMM1 in Andrews and Cheng (2014a) implies Assumption 5 in Ketz (2017b). Therefore, Assumptions 1-3 imply 2 and 3 in Ketz (2017a) and Assumptions 5 and 6 in Ketz (2017b) with

$$DQ_n(\theta) = G'_{\theta^*}\mathcal{W}G_T(\theta^*)$$

and

$$D^2Q_n(\theta_n) = J(\gamma_0) = G'_{\theta^*}\mathcal{W}G_{\theta^*}.$$

□

The asymptotic normality result for $\sqrt{T}(\tilde{\theta}^*_T - \theta^*_T) \xrightarrow{d} Z$ under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ follows from Theorem 1 in Ketz (2017a), which applies under Assumptions 1-4 in Ketz (2017a). The proof of Proposition 1 shows that Assumptions 1-3 imply Assumptions 2 and 3 in Ketz (2017a) and Assumptions 5 and 6 in Ketz (2017b). Therefore, by Lemma 4 in Ketz (2017b), Assumption 1 in Ketz (2017a) holds. Then, it remains to show that Assumption 4 in Ketz (2017a) holds. By Assumptions 1 and 2, we have

$$\hat{G}'_{\theta^*,T}\mathcal{W}_T\hat{G}_{\theta^*,T} = G'_{\theta^*}\mathcal{W}G_{\theta^*} + o_p(1)$$

such that Assumption 4(ii) in Ketz (2017a) is satisfied. Similarly, we have

$$\hat{G}'_{\theta^*,T}\mathcal{W}_T = G'_{\theta^*}\mathcal{W} + o_p(1).$$

Therefore, it suffices to show that

$$\sup_{\theta^*\in\Theta:\|\sqrt{T}(\theta^*-\theta^*_T)\|\leq\epsilon} \|G_T(\theta^*) - G_T(\theta^*_T) + G'_{\theta^*}(\theta^* - \theta^*_T)\| = o_p(1/\sqrt{T}),$$

which holds by Assumption 2 together with $\|G_{\theta^*}(\theta^*_T;\gamma_0) - G_{\theta^*}(\theta^*_0;\gamma_0)\| = o(1)$ and

$$G(\theta^*;\gamma_0) = G(\theta^*_T;\gamma_0) + G_{\theta^*}(\theta^*_T;\gamma_0)(\theta^* - \theta^*_T) + o(\|\theta^* - \theta^*_T\|),$$

which holds by Theorem 6 in Andrews (1999) and Assumption 1(iii).

[24] Note that the proof of Lemma 10.3 in Andrews and Cheng (2014b) goes through when continuous differentiability is replaced by l/r continuous differentiability.

C Details for Section 4

Before verifying Assumptions 1-3 for the BLP model under the conditions in (11), we define $M_1(s_t, x_t, z_t)$-$M_4(s_t, x_t, z_t)$:

$$M_1(s_t, x_t, z_t) = \sup_{\theta^*} \|z'\xi(\theta^*, s, x)\|^{2+\epsilon}, M_2(s_t, x_t, z_t) = \sup_{\theta^*} \left\| z' \frac{\partial \xi(\theta^*, s, x)}{\partial \theta^{*\prime}} \right\|^{1+\epsilon},$$
$$M_3(s_t, x_t, z_t) = \sup_{\theta^*} \left\| \frac{\partial}{\partial \theta^{*\prime}} \mathrm{vec}\left(z' \frac{\partial \xi(\theta^*, s, x)}{\partial \theta^{*\prime}} \right) \right\|, \text{ and}$$
$$M_4(s_t, x_t, z_t) = \sup_{\theta^*} \left\| \frac{\partial}{\partial \theta^{*\prime}} \mathrm{vec}\left(z_t'\xi(\theta^*, s_t, x_t)\xi'(\theta^*, s_t, x_t) z_t \right) \right\|.$$

Instead of verifying Assumption 2, we verify Assumption 2^*. Assumption 2^*(i) follows from Appendix A. Assumption 2^*(ii) and Assumptions 1(i), (iii), (iv) hold by Lemma 12.2 in Andrews and Cheng (2014b) given the conditions in (11). Assumptions 1(ii) follows from a mean value expansion together with the conditions in (11). Assumptions 1(v) and 3(i) follow immediately from the conditions in (11). Lastly, Assumption Assumption 3(ii) follows by Lemma 12.3 in Andrews and Cheng (2014b), which applies under the conditions in (11).

Next, we derive equation (13). The two-sided t-test based on (12) rejects if

$$\sqrt{T}\frac{\hat{\sigma}_{1,T} - \sigma_{1,T}}{\frac{\sqrt{\hat{V}_{K+1}(\hat{\theta}_T^*)}}{2\hat{\sigma}_{1,T}}} < -z_{1-\alpha/2} \text{ or } \sqrt{T}\frac{\hat{\sigma}_{1,T} - \sigma_{1,T}}{\frac{\sqrt{\hat{V}_{K+1}(\hat{\theta}_T^*)}}{2\hat{\sigma}_{1,T}}} > z_{1-\alpha/2}.$$

It never rejects if $\hat{\sigma}_{1,T} = 0$. For $\hat{\sigma}_{1,T} > 0$, we can solve the resulting quadratic equations. The first rejection region is "active" only if $\hat{\sigma}_{1,T}^2 - 2 \cdot z_{1-\alpha/2}\sqrt{\frac{\hat{V}_{K+1}(\hat{\theta}_T^*)}{\sqrt{T}}} > 0$ and is given by

$$\frac{\sqrt{\hat{\sigma}_{1,T}^2} - \sqrt{\hat{\sigma}_{1,T}^2 - 2 \cdot z_{1-\alpha/2}\sqrt{\frac{\hat{V}_{K+1}(\hat{\theta}_T^*)}{\sqrt{T}}}}}{2} < \hat{\sigma}_{1,T} < \frac{\sqrt{\hat{\sigma}_{1,T}^2} + \sqrt{\hat{\sigma}_{1,T}^2 - 2 \cdot z_{1-\alpha/2}\sqrt{\frac{\hat{V}_{K+1}(\hat{\theta}_T^*)}{\sqrt{T}}}}}{2}.$$

The second rejection region is given by

$$\hat{\sigma}_{1,T} > \frac{\sqrt{\hat{\sigma}_{1,T}^2} + \sqrt{\hat{\sigma}_{1,T}^2 + 2 \cdot z_{1-\alpha/2}\sqrt{\frac{\hat{V}_{K+1}(\hat{\theta}_T^*)}{\sqrt{T}}}}}{2},$$

as we never reject for $\hat{\sigma}_{1,T} < 0$. Multipyling through by $T^{1/4}$, squaring, subtracting $\sqrt{T}\sigma_{1,T}^2$ gives equation (13) under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$.

Proof of Proposition 2. Lemma 2.1 in Andrews and Cheng (2012) takes a slightly different form from Proposition 2: AsySz $= \min\{\inf_{h\in H} \mathrm{CP}(h), \mathrm{CP}_\infty\}$, where h takes the role of our $\tilde{h}$. The presence of CP_∞ arises since Andrews and Cheng (2012) allow for lack of identification in parts of the parameter space, which also necessitates Assumptions

ACP(ii)-(iv). The proof of Proposition 2 is thus given by a slightly modified version of the proof of Lemma 2.1 under the following slightly modified version of Assumption ACP(i): For any $\gamma_T \in \Gamma(\gamma_0, h)$, $\mathrm{CP}_T(\gamma_T) \to \mathrm{CP}(\tilde{h})$ for some $\mathrm{CP}(\tilde{h}) \in [0,1]$, where $\tilde{h} = (h, \gamma_0) \in \tilde{H}$. Details are omitted for sake of brevity. (Here, $\mathrm{CP}_T(\gamma_T)$ and $\mathrm{CP}(\tilde{h})$ are placeholders for $\mathrm{CP}_{\sigma,T}(\gamma_T)$, $\mathrm{CP}_{\sigma^2,T}(\gamma_T)$ and $\mathrm{CP}_\sigma(\tilde{h})$, $\mathrm{CP}_{\sigma^2}(\tilde{h})$, respectively, where

$$CP_{\sigma^2,T}(\gamma_T) = P_{\gamma_T}\left(\left|\sqrt{T}\frac{\hat{\sigma}^2_{1,T} - \sigma^2_{1,T}}{\sqrt{\hat{V}_{K+1}(\hat{\theta}^*_T)}}\right| < z_{1-\alpha/2}\right).)$$

The modified version of Assumption ACP(i) is satisfied given equations (13) and (15). □

References

Andrews, D. W. (1992). Generic Uniform Convergence. *Econometric Theory 8*, 241–257.

Andrews, D. W. (1999). Estimation When a Parameter Is on a Boundary. *Econometrica 67*, 1341–1383.

Andrews, D. W. (2002). Generalized Method of Moments Estimation When a Parameter Is on a Boundary. *Journal of Business & Economic Statistics 20*, 530–544.

Andrews, D. W. and X. Cheng (2012). Estimation and Inference With Weak, Semi-Strong, and Strong Identification. *Econometrica 80*, 2153–2211.

Andrews, D. W. and X. Cheng (2014a). GMM Estimation and Uniform Subvector Inference With Possible Identification Failure. *Econometric Theory 30*, 287–333.

Andrews, D. W. and X. Cheng (2014b). Supplement to 'GMM Estimation and Uniform Subvector Inference With Possible Identification Failure'. *Econometric Theory 30*.

Andrews, D. W. and P. Guggenberger (2010a). Applications of subsampling, hybrid, and size-correction methods. *Journal of Econometrics 158*, 285–305.

Andrews, D. W. and P. Guggenberger (2010b). Asymptotic Size and a Problem With Subsampling and With the m out of n Bootstrap. *Econometric Theory 26*, 426.

Armstrong, T. B. (2016). Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply. *Econometrica 84*, 1961–1980.

Berry, S. (1994). Estimating Discrete Choice Models of Product Differentiation. *RAND Journal of Economics 25*, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*, 841–890.

Berry, S., J. Levinsohn, and A. Pakes (1999). Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy. *American Economic Review 89*, 400–430.

Berry, S., O. Linton, and A. Pakes (2004). Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems. *Review of Economic Studies 71*, 613–654.

Cox, D. and D. Hinkley (1974). *Theoretical Statistics*. London: Chapman and Hall.

Dovonon, P. and E. Renault (2013). Testing for Common Conditionally Heteroskedastic Factors. *Econometrica 81*, 2561–2586.

Dubé, J.-P., J. Fox, and C.-L. Su (2012a). Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation. *Econometrica 80*, 2231–2267.

Dubé, J.-P., J. Fox, and C.-L. Su (2012b). Supplememt to 'Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation'. *Econometrica 80*.

Freyberger, J. (2015). Asymptotic Theory for Differentiated Products Demand Models With Many Markets. *Journal of Econometrics 185*, 162–181.

Goeree, M. S. (2008). Limited Information and Advertising in the US Personal Computer Industry. *Econometrica 76*, 1017–1074.

Ketz, P. (2017a). Subvector Inference When the True Parameter Vector is near the Boundary. Working Paper.

Ketz, P. (2017b). Supplementary material for "Subvector Inference When the True Parameter Vector is near the Boundary".

Ketz, P. (2017c). Testing Overidentifying Restrictions When the True Parameter Vector Is near or at the Boundary of the Parameter Space. Working Paper.

Knittel, C. R. and K. Metaxoglou (2014). Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective. *Review of Economics and Statistics 96*, 34–59.

Liu, X. and Y. Shao (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics 31*, 807–832.

Müller, U. K. and A. Norets (2016). Credibility of Confidence Sets in Nonstandard Econometric Problems. *Econometrica 84*, 2183–2213.

Neilson, C. (2013). Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students. Working Paper.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica 69*, 307–342.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy 110*(4), 705–729.

Reynaert, M. and F. Verboven (2014). Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments. *Journal of Econometrics 179*, 83–98.

Rotnitzky, A., D. R. Cox, M. Bottai, and J. Robins (2000). Likelihood-Based Inference With Singular Information Matrix. *Bernoulli 6*, 243–284.

Sargan, J. D. (1983). Identification and lack of identification. *Econometrica 51*, 1605–1633.

Skrainka, B. and K. Judd (2011). High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation. Working Paper.

Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.