



Contents lists available at ScienceDirect

Journal of Econometrics

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

On asymptotic size distortions in the random coefficients logit model[☆]

Philipp Ketz

Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France



ARTICLE INFO

Article history:

Received 7 January 2016

Received in revised form 17 September 2018

Accepted 3 February 2019

Available online 15 May 2019

JEL classification:

C12

L13

Keywords:

Asymptotic size

Boundary

Lack of first-order identification

Size distortion

Random coefficients logit model

ABSTRACT

We show that, in the random coefficients logit model, standard inference procedures can suffer from asymptotic size distortions. The problem arises due to boundary issues and is aggravated by the standard parameterization of the model, in terms of standard deviations. For example, in case of a single random coefficient, the asymptotic size of the nominal 95% confidence interval obtained by inverting the two-sided t-test for the standard deviation equals 83.65%. In seeming contradiction, we also show that standard error estimates for the estimator of the standard deviation can be unreasonably large. This problem is alleviated if the model is reparameterized in terms of variances. Furthermore, a numerical evaluation of a conjectured lower bound suggests that the asymptotic size of the nominal 95% confidence interval obtained by inverting the two-sided t-test for variances (means) is within 0.5 percentage points of the nominal level as long as there are no more than five (four) random coefficients and as long as an optimal weighting matrix is employed.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The random coefficients logit model (Berry et al., 1995), or simply BLP model, is widely used in applied work, most prominently in the industrial organization and marketing literatures. The random coefficients are typically assumed to be independently normally distributed such that the model parameters are given by a K_1 -dimensional vector of means, μ , and a K_2 -dimensional vector of standard deviations, σ , with $K_2 \leq K_1$. The estimation procedure proposed by Berry et al. (1995) minimizes a Generalized Method of Moments (GMM) objective function with respect to $\theta = (\mu', \sigma')$ and inference relies on an asymptotic normality result for the thus obtained estimator. The fact that estimates for σ_{k_2} ($k_2 \in \{1, \dots, K_2\}$) are often found to be small (see e.g., Nevo, 2001; Goeree, 2008), which is indicative of the true parameter being near or at the boundary of the parameter space as σ_{k_2} cannot take on negative values, is generally ignored and statements about whether σ_{k_2} or μ_{k_1} ($k_1 \in \{1, \dots, K_1\}$) are significantly different from zero are typically based on the (symmetric) two-sided t-test.¹

In this paper, we show that this standard inference procedure can suffer from asymptotic size distortions. For example, the asymptotic size of the nominal 95% confidence interval (CI) obtained by inverting the two-sided t-test for σ_{k_2} is shown

[☆] A previous version of this paper was circulated under the title “A Simple Solution to Invalid Inference in the Random Coefficients Logit Model.” I thank the associate 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.

E-mail address: philipp.ketz@psemail.eu.

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

to equal 83.65% when $K_2 = 1$. The problem arises due to boundary effects on the asymptotic distribution of the underlying estimator that are amplified because the Jacobian of the sample moments and, thus, of the population moments is of reduced rank at the boundary of the parameter space, i.e., at $\sigma_{k_2} = 0$. In fact, the corresponding column of the Jacobian is identically equal to a vector of zeros at the boundary of the parameter space. This also implies that the typically employed estimator of – what is incorrectly referred to as – the standard error of the estimator of σ_{k_2} takes on unreasonably large values when the latter is (approximately) equal to zero.^{2,3} We show that the problem of a reduced rank Jacobian can be alleviated by considering a reparameterization of the model in terms of variances, σ^2 , in the sense that all columns of the Jacobian of the sample moments with respect to, say, $\theta^* = (\mu', \sigma^{2'})'$ are non-trivial functions of the data (for any θ^*) and will therefore, under appropriate conditions on the data generating process, be of full rank with probability approaching 1. Put simply, the reparameterization in terms of σ^2 alleviates the problem of unreasonably large standard error estimates or, hereinafter, standard errors.

However, it *a priori* does not ensure asymptotic size control of the resulting CIs, as the asymptotic distribution of the estimator of θ^* is still subject to boundary effects when the true parameter vector is near or at the boundary. From the results in Andrews and Guggenberger (2010a), it can be deduced that the nominal 95% CI obtained by inverting the two-sided t-test for $\sigma_{k_2}^2$ does control asymptotic size when $K_2 = 1$. Similarly, the results in Andrews and Guggenberger (2010b) imply that the nominal 95% CI obtained by inverting the two-sided t-test for μ_{k_1} controls asymptotic size when $K_2 = 1$, as long as an optimal weighting matrix is employed. The results in Ketz (2018a), however, imply that the latter result may not hold when $K_2 = 10$. To the best of our knowledge, no other results for $K_2 > 1$ are available in the literature to date. Here, we show that the results in Andrews and Guggenberger (2010b) extend to the nominal 95% CI obtained by inverting the two-sided t-test for $\sigma_{k_2}^2$ when $K_2 = 2$. In order to shed light on the asymptotic size of the nominal 95% CIs obtained by inverting the two-sided t-test for σ_{k_2} , $\sigma_{k_2}^2$, and μ_{k_1} for other values of K_2 , we rely on lower bounds that we, based on a partially corroborated conjecture, are able to numerically evaluate and that are also of independent interest, as they apply, for example, in other random coefficients models. The lower bounds suggest that the asymptotic size distortion of the nominal 95% CI obtained by inverting the two-sided t-test for σ_{k_2} increases with K_2 . Using empirical estimation results from Reynaert and Verboven (2014), we illustrate that asymptotic size distortions are, indeed, larger for $K_2 > 1$. Furthermore, the lower bounds suggest that, as long as an optimal weighting matrix is employed, the asymptotic size of the nominal 95% CI obtained by inverting the two-sided t-test for $\sigma_{k_2}^2$ (μ_{k_1}) is reasonably close to the nominal level if $K_2 \leq 5$ ($K_2 \leq 4$), with the asymptotic size distortion being less than 0.5 percentage points. For the construction of CIs that control asymptotic size regardless of the choice of the weighting matrix and regardless of the dimension of σ^2 , K_2 , one solution is to employ the *quasi* unconstrained estimator proposed in Ketz (2018a); see Section 5 for details.

The above results rely on a characterization result of asymptotic size (AsySz). The recent literature has highlighted the importance of drifting sequences of true parameters for characterizing AsySz of tests and CIs, or, more generally, confidence sets when the asymptotic distribution of the underlying test statistic 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, θ^* , under drifting sequences of true parameters that may drift towards the boundary. Since the results in Andrews (2002) require the Jacobian of the population moments to be of full rank, we cannot use them directly to obtain the asymptotic distribution of the estimator of the original parameter vector, θ . In order to characterize AsySz of the CI for σ_{k_2} , we, therefore, express its coverage probability as a function of the estimator of θ^* . Then, the characterization of AsySz for the different CIs is obtained by applying the results in Andrews et al. (2011) together with the aforementioned asymptotic distribution result for the estimator of θ^* .

A reduced rank Jacobian is a “first order condition for lack of identification” (Sargan, 1983), i.e., it is necessary but not sufficient for lack of identification. 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). In the context of this model, 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 et al. (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

² The term “standard error” is incorrect because the asymptotic distribution of the corresponding estimator is not normal when the true parameter vector is near or at the boundary.

³ To reconcile the possibility of unreasonably large standard error estimates with the lack of asymptotic size control (at the 5% significance level), note that they arise in different parts of the sample space. To see this, note that the two-sided t-test for testing, for example, $H_0 : \sigma_{k_2} = 0$ (asymptotically) overrejects because the 0.95-quantile of the asymptotic null distribution of the underlying statistic, given by $\max\{0, 2Z\}$ where $Z \sim N(0, 1)$ if $K_2 = 1$ —cf. equation (11), exceeds the corresponding critical value. Put differently, (over-) rejection occurs when the estimator of σ_{k_2} takes on large positive values.

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.

used t-test and do not discuss the problem of unreasonably large standard errors. In the context of GMM, [Dovonon and Renault \(2013\)](#) analyze the nonstandard asymptotic distribution of the J-statistic, used for testing overidentifying restrictions, when the model is locally identified but the Jacobian is of reduced rank. Recently, [Lee and Liao \(2018\)](#) have shown how to recover standard asymptotic distribution results for the J-statistic (and the underlying estimator) in a certain class of models, including the one considered in [Dovonon and Renault \(2013\)](#), where the “local identification failure” is of known form but a reparameterization as performed here is not available; see [Ketz \(2017\)](#) for a related discussion.

This paper is not the first to consider asymptotic theory for the BLP model. [Berry et al. \(2004\)](#) and [Freyberger \(2015\)](#) derive the asymptotic distribution of the estimator of the original parameter vector, θ , under a large number of products and a large number of markets, respectively.⁵ 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 or at 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 et al. \(2012a\)](#) 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 suggested reparameterization to a series of published articles that use the BLP model, including [Berry et al. \(1995\)](#), [Berry et al. \(1999\)](#) and [Reynaert and Verboven \(2014\)](#). Relying on a powerful test suggested by [Feldman and Cousins \(1998\)](#) that can be implemented using the *quasi* unconstrained estimator proposed in [Ketz \(2018a\)](#), we find less evidence of heterogeneity in consumer preferences than suggested by the two-sided t-test for σ_{k_2} , illustrating that size distortions can be empirically relevant. The problem of unreasonably large standard errors is, for example, encountered in [Neilson \(2013\)](#). Upon reparameterization, 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 that may drift towards the boundary. Section 3 introduces the BLP model and shows that the Jacobian of the sample moments is of reduced rank at the boundary of the parameter space. In Section 4, we derive the results concerning asymptotic size. Section 5 introduces the estimator proposed in [Ketz \(2018a\)](#) as well as the test suggested by [Feldman and Cousins \(1998\)](#). 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 results illustrating the relevance of our findings for empirical work are given in Section 7.

Throughout this paper, let “ \equiv ” denote “equals by definition”. Let $\mathbb{R} \equiv (-\infty, \infty)$, $\mathbb{R}_\infty \equiv \mathbb{R} \cup \{\pm\infty\}$, $\mathbb{R}_+ \equiv [0, \infty)$, and $\mathbb{R}_{+, \infty} \equiv [0, \infty) \cup \{\infty\}$. For any $a \in \mathbb{R}_\infty$, let $a_k = (a, \dots, a)'$ denote the k -dimensional vector whose entries are all equal to a . Also, let (a, b) denote $(a', b)'$ for any two column vectors, a and b , and let a_i denote the i th entry of a . For any set A , let $A^k \equiv A \times \dots \times A$ with $k \in \mathbb{N}$ copies. For any matrix A , we let $A_{i,j}$ denote the entry with row index i and column index j . For any square matrix, we also let A_i denote $A_{i,i}$ and, similarly, let $A_{i':i''}$ ($A_{i':i''}$) denote the submatrix of A with all rows and columns removed whose indices are not in $i' : i'' \equiv \{i', i' + 1, \dots, i'' - 1, i''\}$ ($i : i' \equiv \{i\} \cup i' : i''$). In addition, we let $\lambda_{\min}(A)$ denote the smallest eigenvalue of A . Lastly, \xrightarrow{p} and \xrightarrow{d} denote convergence in probability and distribution, respectively, while all limits are taken as “ $T \rightarrow \infty$ ”, unless otherwise noted.

⁵ 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.

2. Asymptotic distribution theory

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). 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 and terminology from Andrews and Cheng (2012, 2014a).

The GMM objective function depends on the K -dimensional parameter vector θ . It is given by a quadratic form of the L -dimensional vector of sample moments $G_T(\theta)$, i.e.,

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

where $\mathcal{W}_T(\theta)$ denotes a $L \times L$ symmetric 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

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

where

$$\Theta = [-c, c]^{K_1} \times [0, c]^{K_2}$$

for some $0 < c < \infty$ with $K_1 + K_2 = K$ and $K_1, K_2 \geq 0$, denotes the optimization parameter space. Here, the use of a common end point c is merely for notational ease. Similarly, the normalization to 0 is without loss of generality. The true parameter space, $\check{\Theta}$, is a strict subset of Θ . In particular, it takes the same form as Θ , but with $0 < \check{c} < c$. 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 the context of GMM, the distribution of the data is in general not fully specified by θ , but depends on an additional, commonly infinite-dimensional, parameter, say ϕ . The parameter $\gamma = (\theta, \phi)$ is assumed to fully specify the distribution of the data, say F_γ . In what follows, P_γ and E_γ denote the probability and expectation under F_γ , respectively. The true parameter space for γ is assumed to be compact and of the following form

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

where $\check{\Phi}(\theta) \subset \check{\Phi} \forall \theta \in \check{\Theta}$ for some compact metric space $\check{\Phi}$ with a metric that induces weak convergence of $(W_t, W_{t+t'})$ for all $t, t' \geq 1$.⁶

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

$$CS_T = \{v : \mathcal{T}_T(v) \leq cv_{1-\alpha}(v)\},$$

where $cv_{1-\alpha}(v)$ denotes the critical value that may depend on the null hypothesis. The coverage probability of CS_T for $r(\theta)$ is

$$P_\gamma(r(\theta) \in CS_T) = P_\gamma(\mathcal{T}_T(r(\theta)) \leq cv_{1-\alpha}(r(\theta))).$$

The asymptotic size, which approximates finite-sample size, is defined as

$$AsySz \equiv \liminf_{T \rightarrow \infty} \inf_{\gamma \in \Gamma} P_\gamma(\mathcal{T}_T(r(\theta)) \leq cv_{1-\alpha}(r(\theta))).$$

Asymptotic size distortion is defined as $\max\{1 - \alpha - AsySz, 0\}$. We say that a CS “controls” asymptotic size if its asymptotic size distortion equals zero, while a CS is said to “suffer from” asymptotic size distortion if the latter is positive. As mentioned above, drifting sequences of true parameters, $\gamma_T = (\theta_T, \phi_T)$, are instrumental in characterizing $AsySz$. Formally, we consider the following set of drifting sequences of true parameters

$$\Gamma(\gamma_0) = \{\{\gamma_T \in \Gamma : T \geq 1\} : \gamma_T \rightarrow \gamma_0 \in \Gamma\},$$

where $\gamma_0 = (\theta_0, \phi_0)$. Of particular interest in this paper is the following subset

$$\Gamma(\gamma_0, h) = \{\{\gamma_T\} \in \Gamma(\gamma_0) : \sqrt{T}\theta_{2,T} \rightarrow h \in \mathbb{R}_{+, \infty}^{K_2}\},$$

where $\theta_T = (\theta_{1,T}, \theta_{2,T})$. Throughout this paper, we use the terminology “under $\{\gamma_T\} \in \Gamma(\gamma_0)$ ” to mean “when the true parameters are $\{\gamma_T\} \in \Gamma(\gamma_0)$ for any $\gamma_0 \in \Gamma$ ”. Similarly, we use the terminology “under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ ” to mean “when

⁶ See also Section 2.1 and, in particular, footnote 21 in Andrews and Cheng (2012).

the true parameters are $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ for any $\gamma_0 \in \Gamma$ and any $h \in \mathbb{R}_{+, \infty}^{K_2}$. Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ except $h = \infty_{K_2}$ we say that the true parameter vector is near or at 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 in the spirit of Andrews (2002).

The first assumption ensures consistency of $\hat{\theta}_T$ under $\{\gamma_T\} \in \Gamma(\gamma_0)$.

Assumption 1.

- (i) Under $\{\gamma_T\} \in \Gamma(\gamma_0)$, $\sup_{\theta \in \Theta} \|G_T(\theta) - G(\theta; \gamma_0)\| \xrightarrow{p} 0$ and $\sup_{\theta \in \Theta} \|\mathcal{W}_T(\theta) - \mathcal{W}(\theta; \gamma_0)\| \xrightarrow{p} 0$ for some nonrandom functions $G(\theta; \gamma_0)$ 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) Each element of $G(\theta; \gamma_0)$ has continuous left/right (l/r) partial derivatives on $\Theta, \forall \gamma_0 \in \Gamma$, with the $L \times K$ matrix of l/r partial derivatives denoted $G_\theta(\theta; \gamma_0)$.⁷
- (iv) $\mathcal{W}(\theta; \gamma_0)$ is continuous in θ on Θ and nonsingular for all $\theta \in \Theta, \forall \gamma_0 \in \Gamma$.

The next assumption ensures that the objective function is asymptotically well approximated by a quadratic function, see below. Here and in what follows, “for all $\epsilon_T \rightarrow 0$ ” stands for “for all sequences of positive scalar constants $\{\epsilon_T : T \geq 1\}$ for which $\epsilon_T \rightarrow 0$ ”.

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

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

for all $\epsilon_T \rightarrow 0$.

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

Assumption 2*.

- (i) Each element of $G_T(\theta)$ has continuous l/r partial derivatives on $\Theta \forall T \geq 1$, with the $L \times K$ matrix of l/r partial derivatives denoted $\frac{\partial}{\partial \theta'} G_T(\theta)$.
- (ii) Under $\{\gamma_T\} \in \Gamma(\gamma_0)$,

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

for all $\epsilon_T \rightarrow 0$.

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

Assumption 3.

- (i) $G_\theta \equiv G_\theta(\theta_0; \gamma_0)$ has full column rank, $\forall \gamma_0 \in \Gamma$.
- (ii) Under $\{\gamma_T\} \in \Gamma(\gamma_0)$, $\sqrt{T} G_T(\theta_T) \xrightarrow{d} Y \sim N(0, \Omega(\gamma_0))$ 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 expansion of the GMM objective function under $\{\gamma_T\} \in \Gamma(\gamma_0)$

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

where $\mathcal{W} \equiv \mathcal{W}(\theta_0; \gamma_0)$. Let

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

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

$$Q_T(\theta) = Q_T(\theta_T) - \frac{1}{2T} \mathcal{Z}_T' \mathcal{J} \mathcal{Z}_T + \frac{1}{2T} q_T(\sqrt{T}(\theta - \theta_T)) + R_T(\theta),$$

⁷ A function is said to have left/right partial derivatives if it has partial derivatives at each interior point, partial derivatives at each boundary point with respect to coordinates that can be perturbed to the left and the right, and left (right) partial derivatives at each boundary point with respect to coordinates that can be perturbed only to the left (right), see Section 3.3 in Andrews (1999).

where

$$q_T(\lambda) = (\lambda - \mathcal{Z}_T)' \mathcal{J}(\lambda - \mathcal{Z}_T).$$

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 \mathcal{Z}_T and $\lim_{T \rightarrow \infty} \sqrt{T}(\Theta - \theta_T)$, the limit of the centered and scaled parameter space. Given the above assumptions, we have that under $\{\gamma_T\} \in \Gamma(\gamma_0)$

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

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

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

where

$$q(\lambda) = (\lambda - \mathcal{Z})' \mathcal{J}(\lambda - \mathcal{Z}) \text{ and } \Lambda_h \equiv \mathbb{R}^{K_1} \times [-h_1, \infty] \times \dots \times [-h_{K_2}, \infty].$$

Proposition 1. Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ and Assumptions 1–3, $\sqrt{T}(\hat{\theta}_T - \theta_T) \xrightarrow{d} \hat{\lambda}_h$.

Proposition 1 follows from Theorem 1 in Andrews (2002), which in turn is based on Theorem 3(b) in Andrews (1999), under suitable adjustments to accommodate drifting sequences of true parameters along the lines of Andrews and Cheng (2012, 2014a); see Appendix C for details.

Remark 1. When $h = \infty_{K_2}$, we have $\Lambda_h = \mathbb{R}^{K_1}$ and $\hat{\lambda}_h = \mathcal{Z}$. Note that $h = \infty_{K_2}$ allows for sequences of true parameters that drift towards the boundary, i.e., $\theta_{2,T,k} \rightarrow 0$ for some $k \in \{1, \dots, K_2\}$, where $\theta_{2,T} = (\theta_{2,T,1}, \dots, \theta_{2,T,K_2})'$, but at a rate slow enough such that the boundary does not impact the asymptotic distribution of the estimator. 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, $h \neq \infty_{K_2}$, then the asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_T)$ is subject to boundary effects and given by the projection of \mathcal{Z} onto Λ_h with respect to the norm $\|\lambda\| = (\lambda' \mathcal{J} \lambda)^{1/2}$. 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 $K_1 = K_2 = 1$, there exists a simple closed form expression: $\sqrt{T}(\hat{\theta}_{1,T} - \theta_{1,T}) \xrightarrow{d} \mathcal{Z}_1 + \frac{\mathcal{J}_{1,2}(\gamma_0)}{\mathcal{J}_1(\gamma_0)} \min\{0, \mathcal{Z}_2 + h\}$ and $\sqrt{T}(\hat{\theta}_{2,T} - \theta_{2,T}) \xrightarrow{d} \max\{-h, \mathcal{Z}_2\}$. When an optimal weighting matrix is employed, i.e., $\mathcal{W} = \Omega(\gamma_0)^{-1}$ such that $\mathcal{J}(\gamma_0) = V(\gamma_0)^{-1}$, the asymptotic distribution of the GMM estimator simplifies. In particular, for $K_1 = K_2 = 1$, the asymptotic distribution of $\sqrt{T}(\hat{\theta}_{1,T} - \theta_{1,T})$ simplifies to $\mathcal{Z}_1 - \frac{V_{1,2}(\gamma_0)}{V_2(\gamma_0)} \min\{0, \mathcal{Z}_2 + h\}$.

3. The random coefficients logit model

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

$$U_{ijt} = x'_{jt} \beta_i + \xi_{jt} + \varepsilon_{ijt}. \tag{3}$$

Here, x_{jt} denotes a K_1 -dimensional vector of observed product characteristics, which in many applications includes 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 by 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 β_i is independent across i and that $\beta_i \sim N(\mu, \Sigma)$, where μ is a K_1 -dimensional vector and Σ is a positive-semidefinite $K_1 \times K_1$ 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. Moreover, it is often assumed that some product characteristics (such as the constant) do not interact with a random coefficient. Put differently, some elements of the main diagonal of Σ may a priori be known to equal zero. Let $\sigma = (\sigma_1, \dots, \sigma_{K_2})'$ denote the vector of possibly nonzero standard deviations, i.e., the square roots of the possibly nonzero elements of the main diagonal of Σ , which we, without loss of generality, take to be the first $K_2 \leq K_1$ elements. Then, the utility in (3) can be written as

$$U_{ijt} = \delta_{jt} + \sum_{k=1}^{K_2} x_{jt,k} \sigma_k u_{i,k} + \varepsilon_{ijt},$$

where $\delta_{jt} = x'_{jt}\mu + \xi_{jt}$ and $u_i \sim u \sim N(0, I_{K_2})$. 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

$$s_j(\sigma, \delta_t, x_t) = \int \frac{e^{x'_{jt}\mu + \xi_{jt} + \sum_{k=1}^{K_2} x_{jt,k}\sigma_k u_k}}{1 + \sum_{l=1}^J e^{x'_{lt}\mu + \xi_{lt} + \sum_{k=1}^{K_2} x_{lt,k}\sigma_k u_k}} dF_u(u), \tag{4}$$

where $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})'$, $x_t = (x_{1t}, \dots, x_{Jt})'$, and where $F_u(u)$ denotes the cdf of u . Let

$$s(\sigma, \delta_t, x_t) = (s_1(\sigma, \delta_t, x_t), \dots, s_J(\sigma, \delta_t, x_t))'$$

denote the vector of model implied market shares. [Berry \(1994\)](#) shows 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

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

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

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

where $\theta = (\mu, \sigma)$. In empirical applications, the existence of a $J \times L$ matrix of instruments with $L \geq K$, z_t , is assumed such that $E_\gamma z'_t \xi(\theta, s_t, x_t) = 0$ and the sample moments entering the GMM objective function are given by⁸

$$G_T(\theta) = \frac{1}{T} \sum_{t=1}^T z'_t \xi(\theta, s_t, x_t). \tag{5}$$

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 near or at 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 lack of first-order 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. In fact, as shown in [Appendix A](#), for any $k \in \{1, \dots, K_2\}$

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

which implies that

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

Put differently, the Jacobian of the sample moments and, consequently, the Jacobian of the population moments 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 $\theta^* = (\mu, \sigma^2)$, where $\sigma^2 = (\sigma_1^2, \dots, \sigma_{K_2}^2)' = ((\sigma_1)^2, \dots, (\sigma_{K_2})^2)'$.⁹ With a slight abuse of notation, let

$$s_j(\sigma^2, \delta_t, x_t) = \int \frac{e^{\delta_{jt} + \sum_{k=1}^{K_2} x_{jt,k} \sqrt{\sigma_k^2} u_k}}{1 + \sum_{l=1}^J e^{\delta_{lt} + \sum_{k=1}^{K_2} x_{lt,k} \sqrt{\sigma_k^2} u_k}} dF_u(u) \tag{6}$$

and define $\delta(\sigma^2, s_t, x_t)$, $\xi(\theta^*, s_t, x_t)$, $G_T(\theta^*)$ and $Q_T(\theta^*)$ accordingly, with the same abuse of notation. In [Appendix A](#), it is shown that for $\sigma_k^2 > 0$ ($k \in \{1, \dots, K_2\}$)

$$\frac{\partial \xi(\theta^*, s_t, x_t)}{\partial \sigma_k^2} = \frac{1}{2\sigma_k} \frac{\partial \xi(\theta, s_t, x_t)}{\partial \sigma_k}. \tag{7}$$

In addition, we have that

$$\lim_{\sigma_k^2 \rightarrow 0} \frac{\partial \xi(\theta^*, s_t, x_t)}{\partial \sigma_k^2} = \lim_{\sigma_k^2 \rightarrow 0} \frac{1}{2\sigma_k} \frac{\partial \xi(\theta, s_t, x_t)}{\partial \sigma_k} = \lim_{\sigma_k^2 \rightarrow 0} \frac{1}{2} \frac{\partial^2 \xi(\theta, s_t, x_t)}{\partial^2 \sigma_k}, \tag{8}$$

⁸ Note that the objective function based on (5) can be minimized over $[-c, c]^K$, since (4) is well defined for negative σ_k ($k \in \{1, \dots, K_2\}$). However, the resulting estimator, which is obtained by taking the absolute value (element-by-element) 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.

⁹ 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 (8). See [Dovonon and Renault \(2013\)](#) for a formal definition of second-order identification.

where the second equality follows by the rule of l'Hôpital. As shown in Appendix A, $\frac{\partial^2 \xi(\theta, s_t, x_t)}{\partial^2 \sigma_k}$ is not identically equal to 0_J 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 s_t , x_t , and z_t) given the nonlinear nature of the model.¹⁰ But we expect the approximation to the optimal instruments proposed in Berry et al. (1999) and Reynaert and Verboven (2014), which applies regardless of the parameterization of the model and which exploits the model's 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 , i.e., with $\gamma = (\theta^*, \phi)$, we need to define the corresponding true parameter space, Γ . With a slight abuse of notation, we continue to let $\ddot{\Theta}$ and Θ denote the true and optimization parameter spaces for θ^* , respectively. For ease of exposition, we assume that $\{\xi_t, x_t, z_t\}_{t=1}^T$ are i.i.d. with distribution $\phi \in \ddot{\Phi}$.¹¹ Let E_ϕ denote the expectation under ϕ . Then, for any $\theta^* \in \ddot{\Theta}$ the true parameter space for ϕ is defined as¹²

$$\begin{aligned} \ddot{\Phi}(\theta^*) &= \{ \phi \in \ddot{\Phi} : E_\phi z_t' \xi_t = 0, \lambda_{\min}(E_\phi z_t' z_t) \geq \epsilon, \lambda_{\min}(E_\phi z_t' \xi_t \xi_t' z_t) \geq \epsilon, \\ &E_\gamma \sum_{i=1}^4 M_i(s_t, x_t, z_t) \leq C, E_\gamma z_t' \frac{\partial \xi(\hat{\theta}^*, s_t, x_t)}{\partial \theta^{*j}} \text{ has full column rank } \forall \theta^* \in \ddot{\Theta} \} \end{aligned} \tag{9}$$

for some constants $\epsilon > 0$ and $C < \infty$, where $M_1(s_t, x_t, z_t) - M_4(s_t, x_t, z_t)$ are defined in Appendix D. The verification of Assumptions 1–3 for the BLP model, with θ^* in place of θ and $\ddot{\Phi}(\theta^*)$ given in (9), is equally provided in Appendix D. In accordance with common practice, (9) covers the standard one-step estimator that uses $\mathcal{W}_T \equiv \mathcal{W}_T(\theta^*) = \frac{1}{T} \sum_{t=1}^T z_t' z_t$ as well as the standard two-step estimator that uses

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

evaluated at a first-step estimator (using e.g., $\mathcal{W}_T = \frac{1}{T} \sum_{t=1}^T z_t' z_t$), where, with a slight abuse of notation, z_t may denote different instruments in the first and the second step. Similarly, (9) implies that the standard “plug-in” estimator of $V(\gamma_0)$, say \hat{V} , is consistent under $\{\gamma_T\} \in \Gamma(\gamma_0)$; see Appendix B for details.

4. Asymptotic size

In this section, we analyze the consequences of ignoring the possible lack of asymptotic normality of the (centered and scaled) estimators of θ and θ^* when relying on standard Wald-type inference procedures. For expositional purposes, we focus on the scalar case and derive the asymptotic size of the CI obtained by inverting the (symmetric) two-sided t-test that is prevalent in applied work. Let $\hat{\theta}_T^* = (\hat{\mu}_T, \hat{\sigma}_T^2)$ denote the estimator of θ^* and let $\hat{\theta}_T = (\hat{\mu}_T, \hat{\sigma}_T)$ with $\hat{\sigma}_T = (\hat{\sigma}_{T,1}, \dots, \hat{\sigma}_{T,K_2})' = (\sqrt{\hat{\sigma}_{T,1}^2}, \dots, \sqrt{\hat{\sigma}_{T,K_2}^2})'$ denote the estimator of θ . Without loss of generality, we consider inference on the first element of each parameter vector, μ_1 , σ_1 , and σ_1^2 . From Eq. (7), it follows that the t-statistic for testing $H_0 : \sigma_1 = \sigma_{\text{null}}$ (i.e., $r(\theta) = \sigma_1$) that is typically used in practice is given by

$$t_{\sigma,T} \equiv t_{\sigma,T}(\sigma_{\text{null}}) \equiv \sqrt{T} \frac{\hat{\sigma}_{T,1} - \sigma_{\text{null}}}{\frac{\sqrt{\hat{V}_{K_1+1}}}{2\hat{\sigma}_{T,1}}} \text{ (with } t_{\sigma,T}(\sigma_{\text{null}}) \equiv 0 \text{ if } \hat{\sigma}_{T,1} = 0); \tag{10}$$

see Appendix B for details. Under $\gamma_T = (\theta_T^*, \phi_T)$, the coverage probability of the nominal $1 - \alpha$ CI based on $|t_{\sigma,T}|$ is given by $P_{\gamma_T} (|t_{\sigma,T}(\sigma_{T,1})| < z_{1-\alpha/2})$, i.e., $\text{cv}_{1-\alpha}(v) = z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of $N(0, 1)$. Under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, its asymptotic coverage probability equals

$$\text{CP}_\sigma(\tilde{h}) \equiv 1 - P \left(\hat{\lambda}_{h,K_1+1} > \frac{\left(\sqrt{h_1} + \sqrt{h_1 + 2 \cdot z_{1-\alpha/2} \sqrt{V_{K_1+1}(\gamma_0)}} \right)^2}{4} - h_1 \right) - \mathbf{1}(h_1 - 2 \cdot z_{1-\alpha/2} \sqrt{V_{K_1+1}(\gamma_0)} > 0) \times$$

¹⁰ Berry et al. (2004) and Freyberger (2015), who also 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.

¹¹ 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.

¹² 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 $t \in \{1, \dots, T\}$ and $j \in \{0, 1, \dots, J\}$ 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).

$$P \left(\frac{\left(\sqrt{h_1} - \sqrt{h_1 - 2 \cdot z_{1-\alpha/2} \sqrt{V_{K_1+1}(\gamma_0)}} \right)^2}{4} - h_1 < \hat{\lambda}_{h,K_1+1} < \frac{\left(\sqrt{h_1} + \sqrt{h_1 - 2 \cdot z_{1-\alpha/2} \sqrt{V_{K_1+1}(\gamma_0)}} \right)^2}{4} - h_1 \right), \tag{11}$$

where $\tilde{h} = (h, \gamma_0)$; see [Appendix D](#) for details. Equivalently, under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$, the asymptotic coverage probability of the nominal $1 - \alpha$ CI based on the absolute value of

$$t_{\sigma^2, T} \equiv t_{\sigma^2, T}(\sigma_{\text{null}}^2) \equiv \sqrt{T} \frac{\hat{\sigma}_{T,1}^2 - \sigma_{\text{null}}^2}{\sqrt{\hat{V}_{K_1+1}}} \left(t_{\mu, T} \equiv t_{\mu, T}(\mu_{\text{null}}) \equiv \sqrt{T} \frac{\hat{\mu}_{T,1} - \mu_{\text{null}}}{\sqrt{\hat{V}_1}} \right), \tag{12}$$

for testing $H_0 : \sigma_1^2 = \sigma_{\text{null}}^2$ ($H_0 : \mu_1 = \mu_{\text{null}}$) with $\text{cv}_{1-\alpha}(v) = z_{1-\alpha/2}$ is given by

$$\text{CP}_{\sigma^2}(\tilde{h}) \equiv P \left(\left| \hat{\lambda}_{h,K_1+1} \right| < z_{1-\alpha/2} \sqrt{V_{K_1+1}(\gamma_0)} \right) \left(\text{CP}_{\mu}(\tilde{h}) \equiv P \left(\left| \hat{\lambda}_{h,1} \right| < z_{1-\alpha/2} \sqrt{V_1(\gamma_0)} \right) \right). \tag{13}$$

Comparing equations (10) and (12), we can interpret the practice of relying on a normal approximation for $t_{\sigma, T}(\sigma_{T,1})$ under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ with $h_1 < \infty$, where $\sigma_{T,1} = \sqrt{\sigma_{T,1}^2}$, as a non-permissible application of the delta method based on a non-applicable asymptotic normality result for $\sqrt{T}(\hat{\sigma}_{T,1}^2 - \sigma_{T,1}^2)$. Note that $\text{CP}_{\sigma}(\tilde{h}) = \text{CP}_{\sigma^2}(\tilde{h})$ when $h_1 = \infty$, which amounts to the delta method being applicable when the true value of σ_1^2 is “far enough” from the boundary, despite the possible lack of asymptotic normality of $\sqrt{T}(\hat{\sigma}_{T,1}^2 - \sigma_{T,1}^2)$ under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$.

Let AsySz_{σ} , AsySz_{σ^2} , and AsySz_{μ} denote the AsySz of the nominal $1 - \alpha$ CIs based on $|t_{\sigma, T}|$, $|t_{\sigma^2, T}|$, and $|t_{\mu, T}|$, respectively, and define

$$\tilde{H} \equiv \{ \tilde{h} = (h, \gamma_0) \in \mathbb{R}_{+, \infty}^{K_2} \times \Gamma : \gamma_T \rightarrow \gamma_0 \text{ and } \sqrt{T}\sigma_T^2 \rightarrow h \text{ for some } \{\gamma_T \in \Gamma : T \geq 1\} \}.$$

Then, we have the following result.

Proposition 2. *If the random coefficients logit model satisfies the conditions in (9), then $\text{AsySz}_{\sigma} = \inf_{\tilde{h} \in \tilde{H}} \text{CP}_{\sigma}(\tilde{h})$, $\text{AsySz}_{\sigma^2} = \inf_{\tilde{h} \in \tilde{H}} \text{CP}_{\sigma^2}(\tilde{h})$, and $\text{AsySz}_{\mu} = \inf_{\tilde{h} \in \tilde{H}} \text{CP}_{\mu}(\tilde{h})$.*

The proof of [Proposition 2](#) is given in [Appendix D](#) and uses [Corollary 2.1\(b\)](#) in [Andrews et al. \(2011\)](#).

From [Eqs. \(11\) and \(13\)](#) and [Propositions 1 and 2](#) it follows that AsySz_{σ} , AsySz_{σ^2} , and AsySz_{μ} depend on γ_0 only through $V(\gamma_0)$ and $\mathcal{J}(\gamma_0)$. Ideally, we would like to obtain “numbers” for AsySz_{σ} , AsySz_{σ^2} , and AsySz_{μ} . However, the set of $V(\gamma_0)$ and $\mathcal{J}(\gamma_0)$ that is admissible over \tilde{H} given the conditions in (9) is difficult to characterize due to the nonlinear nature of the model, preventing such computations. An exception occurs when $K_2 = 1$, since then¹³

$$\sqrt{T}(\hat{\sigma}_T^2 - \sigma_T^2) \xrightarrow{d} \hat{\lambda}_{h,K_1+1} = \max\{-h, \mathcal{Z}_{K_1+1}\} \tag{14}$$

under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$. Since, in addition, $V_{K_1+1}(\gamma_0)$ can be normalized to 1 without loss of generality we have that AsySz_{σ}^1 and $\text{AsySz}_{\sigma^2}^1$ only depend on h , i.e., $\text{AsySz}_{\sigma}^1 = \inf_{h \in \mathbb{R}_{+, \infty}} \text{CP}_{\sigma}(h)$ and $\text{AsySz}_{\sigma^2}^1 = \inf_{h \in \mathbb{R}_{+, \infty}} \text{CP}_{\sigma^2}(h)$, where, here and in what follows, the superscript indicates the dimension of σ^2 , K_2 .

[Fig. 1](#) plots $\text{CP}_{\sigma}(h)$ and $\text{CP}_{\sigma^2}(h)$ as a function of h . Here and in what follows, α is set to 5% and all numerical evaluations are performed using 10,000 simulations. [Fig. 1](#) together with an extended grid search reveals that $\text{CP}_{\sigma}(h)$ and $\text{CP}_{\sigma^2}(h)$ attain their infimum at $h = 0$ and $h = \infty$, respectively.¹⁴ Given [equations \(11\) and \(13\)](#), an exact calculation is possible, which gives $\text{AsySz}_{\sigma}^1 = 83.65\%$ and $\text{AsySz}_{\sigma^2}^1 = 95\%$. Note that the latter finding can equally be deduced from the results in [Andrews and Guggenberger \(2010a\)](#).

To shed some light on $\text{AsySz}_{\sigma}^{K'+1}$, $\text{AsySz}_{\sigma^2}^{K'+1}$, and $\text{AsySz}_{\mu}^{K'}$ for $K' \geq 1$, we, in what follows, restrict our attention to efficient GMM estimators, i.e., we impose $\mathcal{J}(\gamma_0) = V(\gamma_0)^{-1}$, and, without loss of generality, restrict the main diagonal of $V(\gamma_0)$ to a vector of 1s.¹⁵ Note that, then, $\text{AsySz}_{\sigma}^{K'+1}$ and $\text{AsySz}_{\sigma^2}^{K'+1}$ ($\text{AsySz}_{\mu}^{K'}$) depend on $V(\gamma_0)$ only through $V_{K_1+1:K}(\gamma_0)$ ($V_{1:K_1+1:K}(\gamma_0)$). Let \mathcal{V}^R denote the set of symmetric and positive-definite ($R \times R$) matrices with 1s on the main diagonal. Then, we have $\text{AsySz}_{\sigma}^{K'+1} \geq \underline{\text{AsySz}}_{\sigma}^{K'+1} \equiv \inf_{(h, V) \in \mathbb{R}_{+, \infty}^{K'+1} \times \mathcal{V}^{K'+1}} \text{CP}_{\sigma}(h, V)$, $\text{AsySz}_{\sigma^2}^{K'+1} \geq \underline{\text{AsySz}}_{\sigma^2}^{K'+1} \equiv \inf_{(h, V) \in \mathbb{R}_{+, \infty}^{K'+1} \times \mathcal{V}^{K'+1}} \text{CP}_{\sigma^2}(h, V)$,

¹³ By the Continuous Mapping Theorem, it follows that for $h < \infty$

$$T^{1/4} \hat{\sigma}_T \xrightarrow{d} \sqrt{\max\{0, \mathcal{Z}_{K_1+1} + h\}},$$

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 [Fig. 2](#).

¹⁴ The grid search was performed over $h \in \{0, 0.05, \dots, 9.95, 10, 11, \dots, 19, 20, \infty\}$.

¹⁵ Asymptotic size distortions are expected to be significantly larger when the restriction $\mathcal{J}(\gamma_0) = V(\gamma_0)^{-1}$ is not imposed.

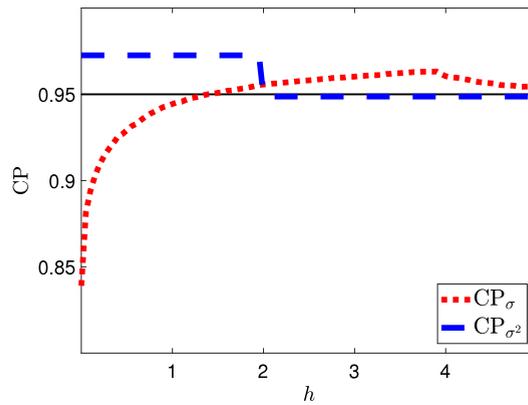


Fig. 1. $CP_{\sigma}(h)$ and $CP_{\sigma^2}(h)$ as a function of $h - K_2 = 1$.

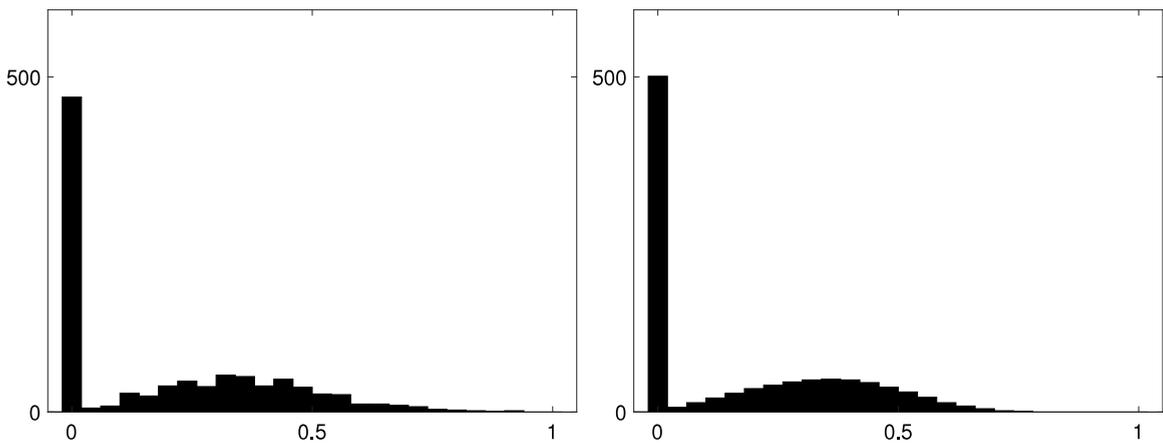


Fig. 2. Histograms of finite-sample (left) and asymptotic (right) distribution of $\hat{\sigma}_T$.

and $AsySz_{\mu}^{K'} \geq \underline{AsySz}_{\mu}^{K'} \equiv \inf_{(h,V) \in \mathbb{R}_{+, \infty}^{K'} \times \mathcal{V}^{K'+1}} CP_{\mu}(h, V)$, where V takes the place of the corresponding submatrix of $V(\gamma_0)$ in the definition of $CP(\cdot)$.¹⁶ From Eq. (13) and Theorem 5 in Andrews (1999), it follows that $AsySz_{\sigma^2}^{K'+1} = AsySz_{\mu}^{K'}$.

From the results in Andrews and Guggenberger (2010b), we have that $\underline{AsySz}_{\mu}^1 = CP_{\mu}(\infty, V) = 95\%$ for any $V \in \mathcal{V}^2$ and we conclude that $AsySz_{\sigma^2}^2 = AsySz_{\mu}^1 = 95\%$. The following Corollary summarizes the results obtained thus far.

Corollary 1. *Suppose the random coefficients logit model satisfies the conditions in (9). Then, $AsySz_{\sigma}^1 = 83.65\% < 95\%$ and $AsySz_{\sigma^2}^1 = 95\%$. If, in addition, an optimal weighting matrix is employed, then $AsySz_{\sigma^2}^2 = AsySz_{\mu}^1 = 95\%$.*

Corollary 1 shows that the CI based on $|t_{\sigma, T}|$ can suffer from asymptotic size distortion and that the CIs based on $|t_{\sigma^2, T}|$ and $|t_{\mu, T}|$ can control asymptotic size, under some conditions. In what follows, we analyze in how far these results generalize.

While it is still feasible to numerically evaluate $\underline{AsySz}_{\sigma}^2$, it is computationally prohibitive to numerically evaluate $\underline{AsySz}_{\sigma}^{K'+1}$ and $\underline{AsySz}_{\mu}^{K'}$ for $K' \geq 2$. We, therefore, rely on the following conjecture: $\underline{AsySz}_{\sigma}^{K'+1} = \underline{ConAsySz}_{\sigma}^{K'+1} \equiv \inf_{(h, \rho_1, \rho_2) \in \{0\}^{K'+1} \times \mathcal{P}} CP_{\sigma}(h, \rho_1, \rho_2)$ and $\underline{AsySz}_{\mu}^{K'} = \underline{ConAsySz}_{\mu}^{K'} \equiv \inf_{(h, \rho_1, \rho_2) \in \{0\}^{K'} \times \mathcal{P}} CP_{\mu}(h, \rho_1, \rho_2)$,¹⁷ where $\mathcal{P} = \{(\rho_1, \rho_2) :$

¹⁶ These lower bounds apply, for example, also in the random coefficients regression model (Andrews, 1999) when the likelihood function is correctly specified.

¹⁷ The grid searches to compute $\underline{ConAsySz}_{\sigma}^{K'+1}$ and $\underline{ConAsySz}_{\mu}^{K'}$ for $K' \geq 2$ were performed over $(\rho_1, \rho_2) \in \{-0.99, -0.98, \dots, 0.98, 0.99\}^2$ (subject to positive-definiteness of $V(\rho_1, \rho_2)$).

Table 1
Conjectured lower bounds on AsySz in % for $K_2 \geq 3$.

K_2	3	4	5	6	7	8	9	10
$\text{ConAsySz}_\sigma^{K_2}$	76.67	71.40	65.77	60.24	54.87	49.04	43.75	38.99
$\text{ConAsySz}_\sigma^{K_2}$	94.81	94.80	94.61	93.81	92.86	91.65	90.41	88.76

Table 2
 $\text{CP}_\sigma(0_6)$ in %.

Price/Inc.	Hp/We.	Foreign	Size	Height	€/km
74.56	76.43	69.79	87.59	88.44	78.37

$V(\rho_1, \rho_2) \in \mathcal{V}^{K'+1}$ and where

$$V(\rho_1, \rho_2) = \begin{bmatrix} 1 & \rho_1 & \rho_1 & \dots & \rho_1 \\ \rho_1 & 1 & \rho_2 & \dots & \rho_2 \\ \rho_1 & \rho_2 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_2 \\ \rho_1 & \rho_2 & \dots & \rho_2 & 1 \end{bmatrix}. \tag{15}$$

The conjecture is based on two observations. First, boundary effects on the distribution of $\hat{\lambda}_{h,K_1+1}$ ($\hat{\lambda}_{h,1}$) and, thereby, on the asymptotic coverage probability given in Eq. (11) ((13)) are most pronounced when $h = 0_{K'+1}$ ($h = 0_{K'}$). Second, the $\hat{\lambda}_{h,k}$ s for $k \in \{2, \dots, K' + 1\}$ ($k \in \{1, \dots, K'\}$) enter $\text{CP}_\sigma(\cdot)$ ($\text{CP}_\mu(\cdot)$) “symmetrically”. To corroborate our conjecture, we consider the case for $K' = 2$ in more detail. A first grid search reveals that $\text{ConAsySz}_\sigma^3 = \inf_{(h,V) \in \{0\}^3 \times \mathcal{V}^3} \text{CP}_\sigma(h, V)$ and $\text{ConAsySz}_\mu^2 = \inf_{(h,V) \in \{0\}^2 \times \mathcal{V}^3} \text{CP}_\mu(h, V)$, with the infimum uniquely attained at $V_{1,2} = V_{1,3} = 0.48$ (and $V_{2,3} = -0.18$) and $V_{1,2} = V_{1,3} = 0.05$ (and $V_{2,3} = -0.63$), respectively.¹⁸ A second grid search reveals that $\text{ConAsySz}_\sigma^3 = \inf_{(h,V) \in \mathbb{R}_{+, \infty}^3 \times \{V(0.48, -0.18)\}} \text{CP}_\sigma(h, V)$ and $\text{ConAsySz}_\mu^2 = \inf_{(h,V) \in \mathbb{R}_{+, \infty}^2 \times \{V(0.05, -0.63)\}} \text{CP}_\mu(h, V)$, where the infimum is uniquely attained at $h = 0_3$ and $h = 0_2$, respectively.¹⁹ The two grid searches combined show that ConAsySz_σ^3 and ConAsySz_μ^2 are, at least, “local” lower bounds. The conjectured lower bounds on $\text{AsySz}_\sigma^{K_2}$ and $\text{AsySz}_\mu^{K_2}$ for $K_2 \geq 3$ and, thus, on $\text{AsySz}_\mu^{K_2}$ for $K_2 \geq 2$ are reported in Table 1.²⁰

We already established that the nominal 95% CI based on $|\tilde{t}_\sigma|$ suffers from asymptotic size distortion when $K_2 = 1$, as $\text{AsySz}_\sigma^1 = 83.65\% < 95\%$. A grid search shows that $\text{AsySz}_\sigma^2 = 80.74\%$, with the infimum attained at $h = 0_2$ and $V_{1,2} = 0.55$.²¹ Assuming our conjecture is true, Table 1 shows that $\text{AsySz}_\sigma^{K_2}$ further decreases as K_2 increases. To investigate in how far this decline may carry over to $\text{AsySz}_\mu^{K_2}$, we rely on an empirical estimate of $V(\gamma_0)$. The estimate of $V(\gamma_0)$ is taken from the application of the random coefficients logit model to the European car market in Reynaert and Verboven (2014) and is part of the estimation results reported in the last two columns of their Table 6.²²

Table 2 shows $\text{CP}_\sigma(0_6)$ for $K_2 = 6$ taking the relevant submatrix of $V(\gamma_0)$ equal to its estimated counterpart; see Section 7 for more information on the variables. The results in Table 2 confirm an increase in asymptotic size distortions for $K_2 > 1$, with the minimum in Table 2, 69.79%, being considerably lower than $\text{AsySz}_\sigma^1 = 83.65\%$.

As mentioned above, the nominal 95% CI based on $|\tilde{t}_{\sigma,2}|$ ($|\tilde{t}_\mu|$) controls asymptotic size for $K_2 \in \{1, 2\}$ ($K_2 = 1$) as long as an optimal weighting matrix is employed. Assuming our conjecture is true, Table 1 shows that these results can, to some extent, be generalized. In particular, Table 1 shows that the asymptotic size distortion of the nominal 95% CI based on $|\tilde{t}_{\sigma,2}|$ ($|\tilde{t}_\mu|$) is less than 0.5 percentage points and, thus, reasonably small as long as $K_2 \leq 5$ ($K_2 \leq 4$). For larger values of K_2 , however, the CIs may suffer from (larger) asymptotic size distortions.

5. A general solution

The positive asymptotic size distortions of the CIs based on $|t_{\sigma,T}|$, $|t_{\sigma^2,T}|$, and $|t_{\mu,T}|$ result from the lack of asymptotic normality of the underlying estimators under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ except $h = \infty_{K_2}$. Due to the presence of the nuisance

¹⁸ The grid searches were performed over $(V_{1,2}, V_{1,3}, V_{2,3}) \in \{-0.99, -0.98, \dots, 0.98, 0.99\}^3$ (subject to positive-definiteness of V). Note that $\text{CP}_\sigma(h, V_{1,2}, V_{1,3}, V_{2,3}) = \text{CP}_\sigma(h, V_{1,3}, V_{1,2}, V_{2,3})$. As this relationship can be violated in simulations, we computed the average of the two simulated counterparts.

¹⁹ The grid searches were performed over $h \in \{0, 0.05, \dots, 9.95, 10, 11, \dots, 19, 20, \infty\}^{K_2}$ with $K_2 = 3$ and $K_2 = 2$, respectively.

²⁰ The “reference point” for the numbers in Table 1 is 94.87%, the value of the numerical evaluation of, for example, $\text{CP}_{\sigma^2}(\infty_{K_2}, V)$ for any $V \in \mathcal{V}^{K_2}$, whose theoretical value is known to equal 95%.

²¹ The grid search was performed over $h \times V_{1,2} \in \{0, 0.05, \dots, 9.95, 10, 11, \dots, 19, 20, \infty\}^2 \times \{-0.99, -0.98, \dots, 0.98, 0.99\}$.

²² I thank Mathias Reynaert and Frank Verboven for sharing their estimate of the asymptotic variance matrix with me.

parameter h , which is not consistently estimable (if $\|h\| \neq \infty_{K_2}$), it is difficult to construct tests based on $\hat{\theta}_T^*$ that control asymptotic size regardless of the dimension of σ^2 and regardless of whether $\mathcal{J}(\gamma_0) = V(\gamma_0)^{-1}$. One solution is to employ the *quasi* unconstrained estimator proposed in Ketz (2018a) whose asymptotic distribution does not depend on h ; note that an unconstrained estimator is not available in the random coefficients logit model as the model implied market share(s) given in Eq. (6) cannot be evaluated at negative values of σ_k^2 ($k \in \{1, \dots, K_2\}$). The *quasi* unconstrained estimator is given by the minimizer of the sample analogue of the quadratic approximation of Eq. (2), i.e.,

$$\tilde{\theta}_T^* = \hat{\theta}_T^* - \left(\hat{G}_{\theta^*}' \mathcal{W}_T \hat{G}_{\theta^*} \right)^{-1} \hat{G}_{\theta^*}' \mathcal{W}_T G_T(\hat{\theta}_T^*), \tag{16}$$

where

$$\hat{G}_{\theta^*} = \frac{1}{T} \sum_{t=1}^T z_t \frac{\partial \xi(\hat{\theta}_T^*, s_t, x_t)}{\partial \theta^{*'}}. \tag{17}$$

Given the conditions in (9), under which Assumptions 1–3 hold, we have $\sqrt{T}(\tilde{\theta}_T^* - \theta^*) \xrightarrow{d} \mathcal{Z}$ under $\{\gamma_T\} \in \Gamma(\gamma_0)$; see Appendix C for details. We note that Assumption 3(i) is essential, implying that a *quasi* unconstrained estimator is not available for θ . Let

$$\tilde{t}_{\sigma^2, T} \equiv \tilde{t}_{\sigma^2, T}(\sigma_{\text{null}}^2) \equiv \sqrt{T} \frac{\tilde{\sigma}_{T,1}^2 - \sigma_{\text{null}}^2}{\sqrt{\hat{V}_{K_1+1}}} \left(\tilde{t}_{\mu, T} \equiv \tilde{t}_{\mu, T}(\mu_{\text{null}}) \equiv \sqrt{T} \frac{\tilde{\mu}_{T,1} - \mu_{\text{null}}}{\sqrt{\hat{V}_1}} \right)$$

denote the t-statistic for testing $H_0 : \sigma_1^2 = \sigma_{\text{null}}^2$ ($H_0 : \mu_1 = \mu_{\text{null}}$) that uses $\tilde{\theta}_T^* = (\tilde{\mu}_T, \tilde{\sigma}_T^2)$ rather than $\hat{\theta}_T = (\hat{\mu}_T, \hat{\sigma}_T)$. A natural candidate for a CI for σ_1^2 that by construction is asymptotically similar is given by the CI based on $|\tilde{t}_{\sigma^2, T}|$, i.e.,

$$\left[\tilde{\sigma}_{T,1}^2 - z_{1-\alpha/2} \sqrt{\hat{V}_{K_1+1}/T}, \tilde{\sigma}_{T,1}^2 + z_{1-\alpha/2} \sqrt{\hat{V}_{K_1+1}/T} \right] \cap \mathbb{R}_+.$$

However, since $\tilde{\sigma}_{T,1}^2$ may take on negative values,²³ this CI is (asymptotically) empty or arbitrarily short with positive probability under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ with $h_1 < \infty$ and, thus, “unreasonable” using the terminology in Müller and Norets (2016). We note that under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$ with $h_1 < \infty$, the problem of making inference about σ_1^2 using $\tilde{\sigma}_{T,1}^2$ only, as opposed to $\tilde{\theta}_T^*$, asymptotically reduces to the problem of making inference about the unknown mean in a scalar Gaussian shift problem, where the mean is *a priori* known to be nonnegative. This inference problem is well-known and one solution, suggested by Feldman and Cousins (1998), is to invert the test based on the (generalized) Likelihood Ratio statistic. The finite-sample analogue of the latter is given by

$$\text{LR}_T \equiv \text{LR}_T(\sigma_{\text{null}}^2) \equiv \tilde{t}_{\sigma^2, T}^2(\sigma_{\text{null}}^2) - \min_{s \in [0, \infty)} \tilde{t}_{\sigma^2, T}^2(s).$$

The corresponding test rejects $H_0 : \sigma_1^2 = \sigma_{\text{null}}^2$ if $\text{LR}_T(\sigma_{\text{null}}^2)$ exceeds $\text{cv}_{1-\alpha}^{\text{LR}}(\sigma_{\text{null}}^2 / \sqrt{\hat{V}_{K_1+1}/T})$, where

$$\text{cv}_{1-\alpha}^{\text{LR}}(v) \equiv \inf\{q \in \mathbb{R} : P(Z^2 - \min_{s \in [-v, \infty)} (Z - s)^2 \leq q) \geq 1 - \alpha\}$$

and where $Z \sim N(0, 1)$. This test is appealing as it reduces to the one-sided t-test when testing $H_0 : \sigma_1^2 = \sigma_{\text{null}}^2 = 0$, as long as $\alpha \leq 0.5$, while it behaves like a two-sided t-test for large values of σ_{null}^2 or, rather, $\sigma_{\text{null}}^2 / \sqrt{\hat{V}_{K_1+1}/T}$, cf. Figure 10 in Feldman and Cousins (1998). Furthermore, the resulting CI is, by construction, never empty or arbitrarily short.

The test that rejects $H_0 : \mu_1 = \mu_{\text{null}}$ when $|\tilde{t}_{\mu, T}|$ exceeds $z_{1-\alpha/2}$ and the above test based on LR_T for testing $H_0 : \sigma_1^2 = \sigma_{\text{null}}^2$ can also be obtained as special cases of the test proposed in Ketz (2018a) and it follows from Lemma 7 in Ketz (2018a) that they satisfy certain asymptotic optimality properties, as long as $\alpha \leq 0.5$ for the test based on LR_T . Similarly, Corollary 2 in Ketz (2018a) implies that the corresponding CIs are asymptotically similar under the conditions in (9).

6. Monte Carlo

In this section, we perform a simulation study to investigate the quality of the approximation provided by the asymptotic theory derived above. 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 $x_{jt,3}$. $x_{jt,3}$ can be thought to represent price and is modelled as being endogenous. In particular, $x_{jt,3}$ is generated as follows

$$x_{jt,3} = w'_{jt} \pi_1 + z'_{jt} \pi_2 + \zeta_{jt},$$

²³ We, therefore, suggest to continue to report $\hat{\theta}_T^*$ as a point estimate for θ^* . Alternatively, as $\hat{\theta}_T^*$ may fall outside the CIs based on $|\tilde{t}_{\mu, T}|$ and LR_T defined below, one may report $\hat{\mu}_{T,k}$ ($k \in \{1, \dots, K_1\}$) and $\max\{0, \tilde{\sigma}_{T,k}^2\}$ ($k \in \{1, \dots, K_2\}$) that, by construction, lie inside the CIs based on $|\tilde{t}_{\mu, T}|$ and LR_T .

Table 3
Monte Carlo results.

Estimator	Average	MCCP	Quantiles				
			0.05	0.25	0.5	0.75	0.95
$\hat{\sigma}_T$	0.185	0.833	0.000	0.000	0.101	0.340	0.573
$\hat{\sigma}_T^2$	0.079	0.977	0.000	0.000	0.010	0.115	0.328
$\hat{\mu}_{T,1}$	1.918	0.968	1.474	1.756	1.918	2.092	2.311
$\hat{\sigma}_T^2$	0.001	0.951	−0.353	−0.115	0.010	0.115	0.328
$\hat{\mu}_{T,1}$	2.007	0.951	1.503	1.801	2.003	2.198	2.542
SE($\hat{\sigma}_T$)	10,807.094	–	0.161	0.269	0.874	11,485.022	49,797.439
SE($\hat{\sigma}_T^2$)	0.202	–	0.115	0.148	0.181	0.229	0.353

where $w_{jt} = (x_{jt,1}, x_{jt,2})'$ is the set of exogenous product characteristics and z_{jt} is 3-dimensional vector of cost shifters. This reflects the case of perfect competition, where there is no markup on price. The endogeneity of $x_{jt,3}$ 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, $x_{jt,1}$ and $x_{jt,2}$, are given by $U[1, 2]$ and 1, 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 with $\sigma_1 = 1$ and $\sigma_2 = \sigma_3 = 0$, but only estimate σ_1 , i.e., $K_2 = 1$. Here, we choose $\sigma = \sigma_1 = 0$. The rest of the parameter values is chosen as follows: $\mu = (\mu_1, \mu_2, \mu_3)' = (2, 2, -2)'$, $\pi_1 = (0.7, 0.7)'$, $\pi_2 = (3, 3, 3)'$. We set $T = 25$ and $J = 10$.

We also implement the same estimator as RV, which uses (an approximation to) optimal instruments and, thus, an optimal weighting matrix. The optimal instruments for the exogenous product characteristics, $x_{jt,1}$ and $x_{jt,2}$, are the product characteristics themselves. The optimal instrument for price, $x_{jt,3}$, is given by $w'_{jt}\hat{\pi}_1 + z'_{jt}\hat{\pi}_2$, where $\hat{\pi}_1$ and $\hat{\pi}_2$ denote the ordinary least squares estimators from a regression of $x_{jt,3}$ 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_1 , suggested by RV, at a random guess of σ_1 , drawn from $|N(0, 1)|$, rather than at a first-step estimate. 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 Eq. (4) 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 u 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 et al. (2012a).

Table 3 reports the Monte Carlo results, obtained using 1000 simulations. Column 1 reports the average of $\hat{\sigma}_T$, $\hat{\sigma}_T^2$, $\hat{\mu}_{T,1}$, $\tilde{\sigma}_T^2$, and $\tilde{\mu}_{T,1}$ over the simulations. In addition, it reports the average of – what we refer to as – the standard errors of $\hat{\sigma}_T$ and $\hat{\sigma}_T^2$, namely $\frac{\sqrt{\hat{V}_{\hat{\sigma}_T/T}}}{2\hat{\sigma}_T}$ and $\sqrt{\hat{V}_{\hat{\sigma}_T^2/T}}$. Column 2 reports the “Monte Carlo Coverage Probability” (MCCP), i.e., the proportion of simulated samples for which the corresponding CIs based on $|t_{\sigma,T}|$, $|t_{\sigma^2,T}|$, $|t_{\mu,T}|$, LR_T , and $|\tilde{t}_{\mu,T}|$ cover the true parameter, at the 95% nominal level. Columns 3–7 report the 0.05, 0.25, 0.5, 0.75, and 0.95 quantiles of the 5 estimators and the standard errors of $\hat{\sigma}_T$ and $\hat{\sigma}_T^2$ over the simulations. The motivation for looking at the quantiles of the standard errors stems from the fact that $\hat{\sigma}_T$ enters the denominator of $\frac{\sqrt{\hat{V}_{\hat{\sigma}_T/T}}}{2\hat{\sigma}_T}$ 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 constrained optimization algorithms typically 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 3 shows that the asymptotic theory derived in this paper provides good approximations to the finite-sample behavior of estimators and CIs. In particular, Table 3 shows that the distribution of $\hat{\theta}_T^*$ is subject to boundary effects, cf. Proposition 1, while $\hat{\theta}_T^*$ is approximately normally distributed, cf. Section 5. Furthermore, the MCCPs of 83.3% and 97.7% of the CIs based on $|t_{\sigma,T}|$ and $|t_{\sigma^2,T}|$ for $\sigma = \sigma^2 = 0$ are very close to the corresponding asymptotic coverage probabilities of 83.65% and 97.5%, cf. Fig. 1 for $h = 0$. Similarly, the MCCPs of the CIs based on LR_T and $|\tilde{t}_{\mu,T}|$ are very close to the nominal level of 95%. Lastly, the MCCP of the CI based on $|t_{\mu,T}|$ is above the nominal level of 95%, which is also in line with the results in Section 4 given the use of an optimal weighting matrix and given that $K_2 = 1$. As expected, the standard error for $\hat{\sigma}_T$ can be huge with the 95th quantile at about 50,000, while the standard error for $\hat{\sigma}_T^2$ does not display this irregularity.

To further illustrate the nonstandard behavior of $\hat{\sigma}_T$ apparent in Table 3, Fig. 2 plots the histograms of the finite-sample and the asymptotic distribution of $\hat{\sigma}_T$, see also footnote 13.

Table 4
Numerical performance.

K_2	θ			θ^*		
	# IAC Average	Median	# CF	# IAC Average	Median	# CF
1	6.64	6	5	4.86	4	2
2	20.35	37	7	6.77	5	2

6.1. Computational efficiency

So far, we have been quiet 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 $\hat{\sigma}_{T,k}$ for all $k \in \{1, \dots, K_2\}$, which might be tempting given that the publicly available code for estimating the random coefficients logit model uses θ . However, since typical optimization algorithms, which use variants of the Newton–Raphson method, rely on either closed form expressions or numerical approximations of the Jacobian and the Hessian of the objective function, which are functions of the Jacobian of the sample moments, 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 one random coefficient in estimation, we also perform estimation with respect to an additional variance parameter, namely on the constant. These two cases are denoted as $K_2 = 1$ and $K_2 = 2$ in Table 4, which reports the average and the median number of iterations that the optimization algorithm needs in order to achieve convergence (IAC). For each cell, the total number of optimizations is 2000, using 2 different starting values for each of the 1000 simulated samples. In addition, Table 4 reports the number of convergence failures (CF) for the algorithm, defined as having failed to find a local minimum after 100 iterations.

Table 4 shows that on average the algorithm needs more iterations in order to achieve convergence when optimization is performed with respect to θ , 6.64 vs. 4.86. The difference in the average number of iterations is more pronounced when $K_2 = 2$, 20.35 vs. 6.77. This is partly due to the high number of convergence failures that are encountered when $K_2 = 2$. 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, $K_2 \in \{1, 2\}$.

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 et al. (2012b).

7. Applications

In Appendix B, we show that the standard error of $\hat{\sigma}_{T,k}^2$ is obtained by multiplying the standard error of $\hat{\sigma}_{T,k}$ by $2 \cdot \hat{\sigma}_{T,k}$ for all $k \in \{1, \dots, K_2\}$; see also equations (10) and (12). Therefore, a reparameterization can be performed without direct access to the data, as long as $\hat{\sigma}_T$ and its standard error(s) 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 size distortions of the CI based on $|t_{\sigma,T}|$ may have influenced conclusions with respect to the presence of heterogeneity in consumer preferences in previous work. In what follows, we reproduce the estimation results of a few published articles that use the random coefficients logit model and apply the reparameterization in terms of variances. In particular, we reproduce estimates of standard deviations, σ , and corresponding standard errors as reported in Berry et al. (1995), Berry et al. (1999), and Reynaert and Verboven (2014). In addition, we report estimates of σ^2 along with corresponding standard errors. We also compute the corresponding nominal 95% CIs based on $|t_{\sigma,T}|$ and LR_T , which is possible since, in all three papers, the estimates of σ and, thus, σ^2 are in the interior of the parameter space such that $\tilde{\theta}_T^* = \hat{\theta}_T^*$, cf. equation (16). The reason for reporting CIs based on LR_T is that the underlying test reduces to the powerful one-sided t-test when testing $H_0 : \sigma_k^2 = \sigma_{\text{null}}^2 = 0$ ($k \in \{1, \dots, K_2\}$) such that our conclusions with respect to whether previous work may have found spurious evidence of heterogeneity in consumer preferences are conservative. We note that the analysis merely serves to illustrate the potential relevance of our findings and is not meant to question the validity of the findings in these papers.

All three papers analyze the demand for cars. Berry et al. (1995, 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

Table 5
Berry et al. (1995) – Table 4 (left panel).

Variable	σ	SE	CI - $ t_{\sigma,T} $		σ^2	SE	CI - LR _T	
Constant	3.612	1.485	0.701	6.523	13.047	10.728	0	34.063
HP/Weight	4.628	1.885	0.933	8.323	21.418	17.448	0	55.600
Air	1.818	1.695	0	5.140	3.305	6.163	0	15.379
MP\$	1.050	0.272	0.517	1.583	1.103	0.571	0.163	2.222
Size	2.056	0.585	0.909	3.203	4.227	2.406	0.270	8.940

Table 6
Berry et al. (1999) – Table 5.

Variable	σ	SE	CI - $ t_{\sigma,T} $		σ^2	SE	CI - LR _T	
Constant	1.112	1.171	0	3.407	1.237	2.604	0	6.339
HP/Weight	0.167	4.652	0	9.285	0.028	1.554	0	3.072
Size	1.392	0.707	0.006	2.778	1.938	1.968	0	5.794
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	0.388

Table 7
Reynaert and Verboven (2014) - Table 6 Optimal instruments (ii).

Variable	σ	SE	CI - $ t_{\sigma,T} $		σ^2	SE	CI - LR _T	
Price/Inc.	0.524	0.168	0.195	0.853	0.274	0.176	0	0.619
Hp/We.	3.202	0.679	1.872	4.532	10.252	4.346	2.856	18.767
Foreign	0.718	0.513	0	1.723	0.515	0.736	0	1.957
Size	0.239	0.394	0	1.011	0.057	0.188	0	0.427
Height	0.104	0.030	0.044	0.163	0.011	0.006	0	0.023
€/km	2.103	4.715	0	11.345	4.424	19.835	0	43.282

characteristics in Berry et al. (1995, 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).

Table 5 reproduces part of the left panel of Table 4 in Berry et al. (1995). The CIs based on $|t_{\sigma,T}|$ suggest the presence of heterogeneity (in consumer preferences) with respect to four out of five product characteristics, while the CIs based on LR_T only provide evidence for heterogeneity with respect to MP\$ and Size.

Table 6 reproduces part of Table 5 in Berry et al. (1999). The CIs based on $|t_{\sigma,T}|$ are indicative of heterogeneity with respect to Size and MP\$. The CIs based on LR_T, on the other hand, all include 0 and, thus, provide no evidence of heterogeneity.

Table 7 reproduces part of Table 6 in Reynaert and Verboven (2014). Here, the CIs based on $|t_{\sigma,T}|$ provide evidence for heterogeneity with respect to three out of six product characteristics, namely Price/Inc., Hp/We., and Height. The CIs based on LR_T, however, include 0 for all but one product characteristic, namely Hp/We.

Appendix A. Derivatives of sample moments

First, we provide the derivative of $G_T(\theta)$ 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 et al. (2012a) in order to estimate the model with respect to θ^* .

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

Define

$$s_{jt}(\sigma, u) \equiv \frac{e^{\delta_{jt} + \sum_{k=1}^{K_2} x_{jt,k} \sigma_k u_k}}{1 + \sum_{l=1}^J e^{\delta_{lt} + \sum_{k=1}^{K_2} x_{lt,k} \sigma_k u_k}}$$

With slight abuse of notation, let δ_{jt} , ξ_{jt} , and s_{jt} denote the j th entry of $\delta_t \equiv \delta(\sigma, s_t, x_t)$, $\xi_t \equiv \xi(\theta, s_t, x_t)$, and $s_t \equiv s_j(\sigma^2, \delta_t, x_t)$, respectively. Then, we have for $k \in \{1, \dots, K_1\}$

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

and for $i, i' \geq 1$ and $k, k' \in \{1, \dots, K_1\}$

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

Also for $i, i' \geq 1, k \in \{1, \dots, K_1\}$, and $k' \in \{1, \dots, K_2\}$

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

Furthermore, we have for $k \in \{1, \dots, K_2\}$ (by the implicit function theorem)

$$\frac{\partial \xi_t}{\partial \sigma_k} = \frac{\partial \delta_t}{\partial \sigma_k} = - \left(\frac{\partial s_t}{\partial \delta'_t} \right)^{-1} \frac{\partial s_t}{\partial \sigma_k},$$

where $\frac{\partial s_t}{\partial \delta'_t}$ has typical elements ($j' \neq j$)

$$\frac{\partial s_{jt}}{\partial \delta_{jt}} = \int S_{jt}(\sigma, u) (1 - S_{jt}(\sigma, u)) dF_u(u) \text{ and } \frac{\partial s_{jt}}{\partial \delta_{j't}} = - \int S_{jt}(\sigma, u) S_{j't}(\sigma, u) dF_u(u)$$

and $\frac{\partial s_t}{\partial \sigma_k}$ has typical element

$$\frac{\partial s_{jt}}{\partial \sigma_k} = \int S_{jt}(\sigma, u) \left(x_{jt,k} - \sum_m S_{mt}(\sigma, u) x_{mt,k} \right) u_k dF_u(u).$$

Note that $S_{jt}(\sigma, u)$ does not depend on u_k when $\sigma_k = 0$ and, therefore, can be written as $S_{jt}(\sigma, u_{-k})$ when $\sigma_k = 0$, where u_{-k} denotes u without u_k . Thus, evaluated at $\sigma_k = 0$, the typical element of $\frac{\partial s_t}{\partial \sigma_k}$ equals

$$\frac{\partial s_{jt}}{\partial \sigma_k} \Big|_{\sigma_k=0} = \int S_{jt}(\sigma, u_{-k}) \left(x_{jt,k} - \sum_m S_{mt}(\sigma, u_{-k}) x_{mt,k} \right) dF_{u_{-k}}(u_{-k}) \int u_k dF_{u_k}(u_k) = 0, \tag{18}$$

since u_k has mean zero. Evaluated at $\sigma_k = 0$, we therefore have

$$\frac{\partial \xi_t}{\partial \sigma_k} \Big|_{\sigma_k=0} = 0_j.$$

Thus, $\frac{\partial}{\partial \sigma_k} G_T(\theta) = 0_L$ for every $T \in \mathbb{N}$ and, consequently, $\frac{\partial}{\partial \sigma_k} G(\theta, \gamma_0) |_{\theta=\theta_0} = 0_L$ for all $\gamma_0 = (\theta_0^*, \phi_0) \in \Gamma$ with $\sigma_{0,k} = \sqrt{\sigma_{0,k}^2} = 0$ such that Assumption 3(i) cannot hold when the model is parameterized with respect to $\theta = (\mu, \sigma)$. Furthermore, for $k, k' \in \{1, \dots, K_2\}$

$$\frac{\partial^2 \xi_t}{\partial \sigma_{k'} \partial \sigma_k} = \frac{\partial}{\partial \sigma_{k'}} \frac{\partial \xi_t}{\partial \sigma_k} = \left(\frac{\partial s_t}{\partial \delta'_t} \right)^{-1} \frac{\partial^2 s_t}{\partial \sigma_{k'} \partial \delta'_t} \left(\frac{\partial s_t}{\partial \delta'_t} \right)^{-1} \frac{\partial s_t}{\partial \sigma_k} - \left(\frac{\partial s_t}{\partial \delta'_t} \right)^{-1} \frac{\partial^2 s_t}{\partial \sigma_{k'} \partial \sigma_k}.$$

Next to $\frac{\partial s_t}{\partial \delta'_t}$ and $\frac{\partial s_t}{\partial \sigma_k}$, we have $\frac{\partial^2 s_t}{\partial \sigma_{k'} \partial \delta'_t}$ and $\frac{\partial^2 s_t}{\partial \sigma_{k'} \partial \sigma_k} \cdot \frac{\partial^2 s_t}{\partial \sigma_{k'} \partial \delta'_t}$ has typical elements ($j' \neq j$)

$$\begin{aligned} \frac{\partial^2 s_{jt}}{\partial \sigma_{k'} \partial \delta_{jt}} &= \int S_{jt}(\sigma, u) (1 - 2S_{jt}(\sigma, u)) \left(x_{jt,k'} - \sum_m S_{mt}(\sigma, u) x_{mt,k'} \right) u_{k'} dF_u(u) \text{ and} \\ \frac{\partial^2 s_{jt}}{\partial \sigma_{k'} \partial \delta_{j't}} &= - \int S_{jt}(\sigma, u) S_{j't}(\sigma, u) \left(x_{jt,k'} + x_{j't,k'} - 2 \sum_m S_{mt}(\sigma, u) x_{mt,k'} \right) u_{k'} dF_u(u). \end{aligned}$$

For $k' = k$, $\frac{\partial^2 s_{jt}}{\partial \sigma_{k'} \partial \sigma_k}$ is given by

$$\begin{aligned} \frac{\partial^2 s_{jt}}{\partial^2 \sigma_k} &= \int S_{jt}(\sigma, u) \left(x_{jt,k} - \sum_m S_{mt,k}(\sigma, u) x_{mt,k} \right)^2 u_k^2 dF_u(u) \\ &+ \int S_{jt}(\sigma, u) \left(- \sum_m \{ S_{mt}(\sigma, u) \left[x_{mt,k} - \sum_n S_{nt}(\sigma, u) x_{nt,k} \right] u_k \} x_{mt,k} \right) u_k dF_u(u) \\ &= \int S_{jt}(\sigma, u) \left(x_{jt,k} - \sum_m S_{mt}(\sigma, u) x_{mt,k} \right)^2 u_k^2 dF_u(u) \\ &- \int S_{jt}(\sigma, u) \left(\sum_m S_{mt}(\sigma, u) \left[x_{mt,k} - \sum_n S_{nt}(\sigma, u) x_{nt,k} \right] x_{mt,k} \right) u_k^2 dF_u(u), \end{aligned}$$

which evaluated at $\sigma_k = 0$ does not equal zero. The expression factorizes as in (18) above, but $\int u_k^2 dF_{u_k}(u_k) = 1 \neq 0$. For $k' \neq k$, $\frac{\partial^2 s_{jt}}{\partial \sigma_{k'} \partial \sigma_k}$ is given by

$$\begin{aligned} \frac{\partial^2 s_{jt}}{\partial \sigma_{k'} \partial \sigma_k} &= \int S_{jt}(\sigma, u) \left(x_{jt,k} - \sum_m S_{mt,k}(\sigma, u) x_{mt,k} \right) u_k \left(x_{jt,k'} - \sum_m S_{mt,k'}(\sigma, u) x_{mt,k'} \right) u_{k'} dF_u(u) \\ &\quad - \int S_{jt}(\sigma, u) \left(\sum_m \{ S_{mt}(\sigma, u) \left[x_{mt,k'} - \sum_n S_{nt}(\sigma, u) x_{nt,k'} \right] u_{k'} \} x_{mt,k} \right) u_k dF_u(u), \end{aligned}$$

which evaluated at $\sigma_k = 0$ equals zero, again by the same argument as in (18).

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 et al. (2012b). Let

$$S_{jt}(\sigma^2, u) \equiv \frac{e^{\delta_{jt} + \sum_{k=1}^{K_2} x_{jt,k} \sqrt{\sigma_k^2} u_k}}{1 + \sum_{l=1}^J e^{\delta_{lt} + \sum_{k=1}^{K_2} x_{lt,k} \sqrt{\sigma_k^2} u_k}},$$

such that with a slight abuse of notation $S_{jt}(\sigma^2, u) = S_{jt}(\sigma, u)$. $\frac{\partial s_{jt}}{\partial \sigma_k^2}$ ($k \in \{1, \dots, K_2\}$) has typical element

$$\frac{\partial s_{jt}}{\partial \sigma_k^2} = \frac{1}{2\sqrt{\sigma_k^2}} \int S_{jt}(\sigma^2, u) \left(x_{jt,k} - \sum_m S_{mt}(\sigma^2, u) x_{mt,k} \right) u_k dF_u(u).$$

Therefore, $\frac{\partial^2 s_{jt}}{\partial \sigma_k^2} = \frac{1}{2\sigma_k} \frac{\partial s_{jt}}{\partial \sigma_k}$, where $\sigma_k = \sqrt{\sigma_k^2}$. The elements of $\frac{\partial^2 s_{jt}}{\partial \sigma_{k'}^2 \partial \sigma_k^2}$ are given by ($j' \neq j$ and $k \in \{1, \dots, K_2\}$)

$$\begin{aligned} \frac{\partial^2 s_{jt}}{\partial \sigma_k^2 \partial \delta_{jt}} &= \frac{1}{2\sqrt{\sigma_k^2}} \int S_{jt}(\sigma^2, u) (1 - 2S_{jt}(\sigma^2, u)) \left(x_{jt,k} - \sum_m S_{mt}(\sigma^2, u) x_{mt,k} \right) u_k dF_u(u) \text{ and} \\ \frac{\partial^2 s_{jt}}{\partial \sigma_k^2 \partial \delta_{j't}} &= -\frac{1}{2\sqrt{\sigma_k^2}} \int S_{jt}(\sigma^2, u) S_{j't}(\sigma^2, u) \left(x_{jt,k} + x_{j't,k} - 2 \sum_m S_{mt}(\sigma^2, u) x_{mt,k} \right) u_k dF_u(u). \end{aligned}$$

Thus, $\frac{\partial^2 s_{jt}}{\partial \sigma_k^2 \partial \delta_{jt}} = \frac{1}{2\sigma_k} \frac{\partial^2 s_{jt}}{\partial \sigma_k \partial \delta_{jt}}$ and $\frac{\partial^2 s_{jt}}{\partial \sigma_k^2 \partial \delta_{j't}} = \frac{1}{2\sigma_k} \frac{\partial^2 s_{jt}}{\partial \sigma_k \partial \delta_{j't}}$. For $k' = k$ ($k, k' \in \{1, \dots, K_2\}$), $\frac{\partial^2 s_{jt}}{\partial \sigma_k^2 \partial \sigma_k^2}$ is given by

$$\begin{aligned} \frac{\partial^2 s_{jt}}{\partial \sigma_k^2} &= \frac{1}{4\sigma_k^2} \int S_{jt}(\sigma^2, u) \left(x_{jt,k} - \sum_m S_{mt,k}(\sigma^2, u) x_{mt,k} \right)^2 u_k^2 dF_u(u) \\ &\quad - \frac{1}{4\sigma_k^2} \int S_{jt}(\sigma^2, u) \left(\sum_m S_{mt}(\sigma^2, u) \left[x_{mt,k} - \sum_n S_{nt}(\sigma^2, u) x_{nt,k} \right] x_{mt,k} \right) u_k^2 dF_u(u) \\ &\quad - \frac{1}{4(\sigma_k^2)^{\frac{3}{2}}} \int S_{jt}(\sigma^2, u) \left(x_{jt,k} - \sum_m S_{mt}(\sigma^2, u) x_{mt,k} \right) u_k dF_u(u). \end{aligned}$$

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

$$\begin{aligned} \frac{\partial^2 s_{jt}}{\partial \sigma_{k'}^2 \partial \sigma_k^2} &= \frac{1}{4\sqrt{\sigma_k^2} \sqrt{\sigma_{k'}^2}} \int S_{jt}(\sigma^2, u) \left(x_{jt,k} - \sum_m S_{mt,k}(\sigma^2, u) x_{mt,k} \right) u_k \left(x_{jt,k'} - \sum_m S_{mt,k'}(\sigma^2, u) x_{mt,k'} \right) u_{k'} dF_u(u) \\ &\quad - \frac{1}{4\sqrt{\sigma_k^2} \sqrt{\sigma_{k'}^2}} \int S_{jt}(\sigma^2, u) \left(\sum_m \{ S_{mt}(\sigma^2, u) \left[x_{mt,k'} - \sum_n S_{nt}(\sigma^2, u) x_{nt,k'} \right] u_{k'} \} x_{mt,k} \right) u_k dF_u(u). \end{aligned}$$

It follows that $\frac{\partial^2 s_{jt}}{\partial \sigma_{k'}^2 \partial \sigma_k^2} = \frac{1}{4\sigma_k \sigma_{k'}} \frac{\partial^2 s_{jt}}{\partial \sigma_{k'} \partial \sigma_k}$.

Appendix B. Asymptotic variance estimator

In this section, we define the standard “plug-in” estimator of $V(\gamma_0)$, the asymptotic variance of the normal random vector that underlies the asymptotic distribution result for $\sqrt{T}(\hat{\theta}_T^* - \theta_T^*)$, cf. Proposition 1. Furthermore, we establish the relationship with the standard “plug-in” estimator that is used in the construction of test statistics based on $\hat{\theta}_T$, cf. equation (10). For the sake of brevity, we only consider the standard one-step GMM estimator that uses $\mathcal{W}_T \equiv \frac{1}{T} \sum_{t=1}^T z_t' z_t$ and assume that $\mathcal{W} \neq c \cdot \Omega(\gamma_0)^{-1}$ for any $0 < c < \infty$. Then, we have

$$V(\gamma_0) = (G_{\theta^*}' \mathcal{W} G_{\theta^*})^{-1} G_{\theta^*}' \mathcal{W} \Omega(\gamma_0) \mathcal{W} G_{\theta^*} (G_{\theta^*}' \mathcal{W} G_{\theta^*})^{-1}.$$

Let

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

Then, under $\{\gamma_T\} \in \Gamma(\gamma_0)$, we have $\mathcal{W}_T \xrightarrow{P} \mathcal{W} = E_{\phi_0} z_t' z_t$, $\hat{\Omega} \xrightarrow{P} \Omega(\gamma_0)$, $\hat{G}_{\theta^*} \xrightarrow{P} G_{\theta^*}$ (with \hat{G}_{θ^*} defined in Eq. (17)), and

$$\hat{V} \equiv (\hat{G}_{\theta^*}' \mathcal{W}_T \hat{G}_{\theta^*})^{-1} \hat{G}_{\theta^*}' \mathcal{W}_T \hat{\Omega} \mathcal{W}_T \hat{G}_{\theta^*} (\hat{G}_{\theta^*}' \mathcal{W}_T \hat{G}_{\theta^*})^{-1} \xrightarrow{P} V(\gamma_0).$$

This follows from (repeated application of) Lemma 12.2 in Andrews and Cheng (2014b) given the conditions in (9), together with Slutsky’s Theorem and the Continuous Mapping Theorem. Let – recall the abuse of notation in defining $\xi(\hat{\theta}_T^*, s_t, x_t)$ and $\xi(\theta_T, s_t, x_t)$ –

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

Since constrained optimizers typically stay strictly within their bounds, we have (up to a small, predefined numerical error)

$$\hat{G}_{\theta^*} = \hat{G}_{\theta} \begin{bmatrix} I_{K_1} & 0 \\ 0 & S(\hat{\sigma}_T^2) \end{bmatrix}, \text{ where } S(\sigma^2) \equiv \begin{bmatrix} \frac{1}{2\sqrt{\sigma_1^2}} & 0 & \dots & 0 \\ 0 & \frac{1}{2\sqrt{\sigma_2^2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{2\sqrt{\sigma_{k_2}^2}} \end{bmatrix}.$$

It follows that

$$\hat{V} = \begin{bmatrix} I_{K_1} & 0 \\ 0 & S(\hat{\sigma}_T^2)^{-1} \end{bmatrix} (\hat{G}_{\theta}' \mathcal{W}_T \hat{G}_{\theta})^{-1} \hat{G}_{\theta}' \mathcal{W}_T \hat{\Omega} \mathcal{W}_T \hat{G}_{\theta} (\hat{G}_{\theta}' \mathcal{W}_T \hat{G}_{\theta})^{-1} \begin{bmatrix} I_{K_1} & 0 \\ 0 & S(\hat{\sigma}_T^2)^{-1} \end{bmatrix},$$

where the term in the middle corresponds to the standard “asymptotic variance” estimator used in practice, cf. equation (10).

Appendix C. Verification of assumptions in Ketz (2018a,b)

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

Proof of Proposition 1. Proposition 1 in Ketz (2018b) obtains the same asymptotic distribution result as Proposition 1, but under general high-level assumptions as in Andrews (1999) suitably adapted to accommodate drifting sequences of true parameters, namely Assumptions 2 and 3 in Ketz (2018a) and Assumptions 6 and 7 in Ketz (2018b). Therefore, the proof proceeds by verifying these Assumptions. As mentioned above, Assumptions 1–3 correspond to slightly modified versions of Assumptions GMM1, GMM2, and GMM5 in Andrews and Cheng (2014a). Similarly, Assumptions 2 and 3 in Ketz (2018a) and Assumption 7 in Ketz (2018b) correspond to slightly modified versions of 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(b) in Andrews and Cheng (2014b) and Lemma 3.1(b) in Andrews and Cheng (2012), Assumption GMM1 in Andrews and Cheng (2014a) implies Assumption 6 in Ketz (2018b). Therefore, Assumptions 1–3 imply Assumptions 2 and 3 in Ketz (2018a) and Assumptions 6 and 7 in Ketz (2018b) with $DQ_n(\theta) = G_{\theta^*}' \mathcal{W}_T(\theta^*)$ and $D^2Q_n(\theta) = \mathcal{J}(\gamma_0) = G_{\theta^*}' \mathcal{W} G_{\theta^*}$. □

²⁴ Note that the proof of Lemma 10.3 in Andrews and Cheng (2014b) goes through when continuous differentiability is replaced by 1/r continuous differentiability.

The aforementioned asymptotic normality result, $\sqrt{T}(\tilde{\theta}_T^* - \theta_T^*) \xrightarrow{d} \mathcal{Z}$ under $\{\gamma_T\} \in \Gamma(\gamma_0)$, follows from Theorem 1 in Ketz (2018a), which applies under Assumptions 1–4 in Ketz (2018a). The proof of Proposition 1 shows that Assumptions 1–3, which hold given the conditions in (9), imply Assumptions 2 and 3 in Ketz (2018a). Furthermore, Proposition 1 implies Assumption 1 in Ketz (2018a). It remains to show that Assumption 4 in Ketz (2018a) holds. By Assumptions 1 and 2 and Slutsky’s Theorem, we have $\hat{G}'_{\theta^*} \mathcal{W}_T \hat{G}_{\theta^*} = G'_{\theta^*} \mathcal{W} G_{\theta^*} + o_p(1)$ under $\{\gamma_T\} \in \Gamma(\gamma_0)$ such that Assumption 4(ii) in Ketz (2018a) is satisfied. Similarly, we have $\hat{G}'_{\theta^*} \mathcal{W}_T = G'_{\theta^*} \mathcal{W} + o_p(1)$ under $\{\gamma_T\} \in \Gamma(\gamma_0)$. Given Assumption 1 in Ketz (2018a), it, therefore, suffices to show that under $\{\gamma_T\} \in \Gamma(\gamma_0)$

$$\sup_{\theta^* \in \Theta: \|\sqrt{T}(\theta^* - \theta_T^*)\| \leq \epsilon} \|G_T(\theta^*) - G_T(\theta_T^*) - G_{\theta^*}(\theta^* - \theta_T^*)\| = o_p(1/\sqrt{T})$$

for all $0 < \epsilon < \infty$, 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).

Appendix D. Details for Section 4

Before verifying Assumptions 1–3 for the random coefficients logit model under the conditions in (9), we define $M_1(s_t, x_t, z_t) - M_4(s_t, x_t, z_t)$:

$$M_1(s_t, x_t, z_t) = \sup_{\theta^* \in \Theta} \|z_t' \xi(\theta^*, s_t, x_t)\|^{2+\epsilon}, M_2(s_t, x_t, z_t) = \sup_{\theta^* \in \Theta} \left\| z_t \frac{\partial \xi(\theta^*, s_t, x_t)}{\partial \theta^{*'}} \right\|^{1+\epsilon},$$

$$M_3(s_t, x_t, z_t) = \sup_{\theta^* \in \Theta} \left\| \frac{\partial}{\partial \theta^{*'}} \text{vec} \left(z_t \frac{\partial \xi(\theta^*, s_t, x_t)}{\partial \theta^{*'}} \right) \right\|, \text{ and } M_4(s_t, x_t, z_t) = \sup_{\theta^* \in \Theta} \left\| \frac{\partial}{\partial \theta^{*'}} \text{vec} \left(z_t' \xi(\theta^*, s_t, x_t) \xi'(\theta^*, s_t, x_t) z_t \right) \right\|.$$

Instead of verifying Assumption 2, we verify Assumption 2*. Assumption 2*(i) follows from Appendix A. Assumption 2*(ii) and Assumption 1(i), (iii), and (iv) hold by Lemma 12.2 in Andrews and Cheng (2014b) given the conditions in (9). Assumption 1(ii) follows from a mean value expansion together with the conditions in (9). Assumption 3(i) follows immediately from the conditions in (9). Lastly, Assumption 3(ii) follows by Lemma 12.3 in Andrews and Cheng (2014b), which applies under the conditions in (9).

Next, we derive equation (11). Under $\gamma_T = (\theta_T^*, \phi_T)$, the test based on $|t_{\sigma, T}|$ rejects $H_0 : \sigma_1 = \sigma_{T,1}$ if

$$\sqrt{T} \frac{\hat{\sigma}_{T,1} - \sigma_{T,1}}{\frac{\sqrt{\hat{V}_{\sigma_1^2}}}{2\hat{\sigma}_{T,1}}} < -z_{1-\alpha/2} \text{ or } \sqrt{T} \frac{\hat{\sigma}_{T,1} - \sigma_{T,1}}{\frac{\sqrt{\hat{V}_{\sigma_1^2}}}{2\hat{\sigma}_{T,1}}} > z_{1-\alpha/2}.$$

It never rejects if $\hat{\sigma}_{T,1} = 0$. For $\hat{\sigma}_{T,1} > 0$, we can solve the resulting quadratic equations, $\hat{\sigma}_{T,1}^2 - \sigma_{T,1} \hat{\sigma}_{T,1} \pm \frac{1}{2} z_{1-\alpha/2} \sqrt{\hat{V}_{\sigma_1^2}/T} \leq 0$, for the “unknown” $\hat{\sigma}_{T,1}$. The first rejection region is “active” only if $\sigma_{T,1}^2 - 2 \cdot z_{1-\alpha/2} \sqrt{\hat{V}_{\sigma_1^2}/T} > 0$ and is given by

$$\frac{\sqrt{\sigma_{T,1}^2} - \sqrt{\sigma_{T,1}^2 - 2 \cdot z_{1-\alpha/2} \sqrt{\hat{V}_{\sigma_1^2}/T}}}{2} < \hat{\sigma}_{T,1} < \frac{\sqrt{\sigma_{T,1}^2} + \sqrt{\sigma_{T,1}^2 - 2 \cdot z_{1-\alpha/2} \sqrt{\hat{V}_{\sigma_1^2}/T}}}{2}.$$

The second rejection region is given by

$$\hat{\sigma}_{T,1} > \frac{\sqrt{\sigma_{T,1}^2} + \sqrt{\sigma_{T,1}^2 + 2 \cdot z_{1-\alpha/2} \sqrt{\hat{V}_{\sigma_1^2}/T}}}{2},$$

as we never reject for $\hat{\sigma}_{T,1} < 0$. Multiplying through by $T^{1/4}$, squaring and subtracting $\sqrt{T} \sigma_{T,1}^2$ gives equation (11) under $\{\gamma_T\} \in \Gamma(\gamma_0, h)$.

Proof of Proposition 2. As mentioned above, the proof of Proposition 2 relies on Corollary 2.1(b) in Andrews et al. (2011), which uses Assumptions B1, B2*, C1, and C2 in Andrews et al. (2011). Assumptions B1, C1, and C2 are satisfied by $CP_\sigma(\hat{h})$, $CP_{\sigma^2}(\hat{h})$, and $CP_\mu(\hat{h})$ (where $CP^-(\hat{h}) = CP^+(\hat{h}) = CP(\hat{h})$ using their notation), while Assumption B2* is satisfied given the definition of Γ (with $h_n(\gamma) = (\sqrt{n}\sigma^2, \gamma)$ using their notation). Therefore, Corollary 2.1(b) in Andrews et al. (2011) implies $AsySz. = \inf_{\hat{h} \in \hat{H}} CP(\hat{h})$ for all three subscripts, σ , σ^2 , and μ . \square

References

- Andrews, D.W.K., 1992. Generic uniform convergence. *Econom. Theory* 8 (2), 241–257.
- Andrews, D.W.K., 1999. Estimation when a parameter is on a boundary. *Econometrica* 67 (6), 1341–1383.
- Andrews, D.W.K., 2002. Generalized method of moments estimation when a parameter is on a boundary. *J. Bus. Econom. Statist.* 20 (4), 530–544.
- Andrews, D.W.K., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80 (5), 2153–2211.
- Andrews, D.W.K., Cheng, X., 2014a. Gmm estimation and uniform subvector inference with possible identification failure. *Econom. Theory* 30 (2), 287–333.
- Andrews, D.W.K., Cheng, X., 2014b. Supplement to 'GMM estimation and uniform subvector inference with possible identification failure'.
- Andrews, D.W.K., Cheng, X., Guggenberger, P., 2011. Generic results for establishing the asymptotic size of confidence sets and tests. Cowles Foundation Discussion Paper No. 1813.
- Andrews, D.W.K., Guggenberger, P., 2010a. Applications of subsampling, hybrid, and size-correction methods. *J. Econometrics* 158 (2), 285–305.
- Andrews, D.W.K., Guggenberger, P., 2010b. Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econom. Theory* 26 (2), 426–468.
- Armstrong, T.B., 2016. Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica* 84 (5), 1961–1980.
- Berry, S., 1994. Estimating discrete choice models of product differentiation. *RAND J. Econ.* 25 (2), 242–262.
- Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. *Econometrica* 63 (4), 841–890.
- Berry, S., Levinsohn, J., Pakes, A., 1999. Voluntary export restraints on automobiles: evaluating a trade policy. *Am. Econ. Rev.* 89 (3), 400–430.
- Berry, S., Linton, O., Pakes, A., 2004. Limit theorems for estimating the parameters of differentiated product demand systems. *Rev. Econom. Stud.* 71 (3), 613–654.
- Cox, D., Hinkley, D., 1974. *Theoretical Statistics*. Chapman and Hall.
- Dovonon, P., Renault, E., 2013. Testing for common conditionally heteroskedastic factors. *Econometrica* 81 (6), 2561–2586.
- Dubé, J.-P., Fox, J., Su, C.-L., 2012a. Improving the numerical performance of blp static and dynamic discrete choice random coefficients demand estimation. *Econometrica* 80 (5), 2231–2267.
- Dubé, J.-P., Fox, J., Su, C.-L., 2012b. Supplement to 'improving the numerical performance of blp static and dynamic discrete choice random coefficients demand estimation'. *Econometrica Supplemental Material* 80 (5).
- Feldman, G.J., Cousins, R.D., 1998. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D* 57 (7), 3873–3889.
- Freyberger, J., 2015. Asymptotic theory for differentiated products demand models with many markets. *J. Econometrics* 185 (1), 162–181.
- Goeree, M.S., 2008. Limited information and advertising in the us personal computer industry. *Econometrica* 76 (5), 1017–1074.
- Ketz, P., 2017. Testing overidentifying restrictions when the true parameter vector is near or at the boundary of the parameter space. Working Paper.
- Ketz, P., 2018a. Subvector inference when the true parameter vector may be near or at the boundary. *J. Econom.* 207 (2), 285–306.
- Ketz, P., 2018b. Supplementary material for 'Subvector inference when the true parameter vector may be near or at the boundary'.
- Knittel, C.R., Metaxoglou, K., 2014. Estimation of random-coefficient demand models: two empiricists' perspective. *Rev. Econ. Stat.* 96 (1), 34–59.
- Lee, J.H., Liao, Z., 2018. On standard inference for gmm with local identification failure of known forms. *Econom. Theory* 34 (4), 790–814.
- Liu, X., Shao, Y., 2003. Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.* 31, 807–832.
- Müller, U.K., Norets, A., 2016. Credibility of confidence sets in nonstandard econometric problems. *Econometrica* 84 (6), 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 (2), 307–342.
- Petrin, A., 2002. Quantifying the benefits of new products: the case of the minivan. *J. Polit. Econ.* 110 (4), 705–729.
- Reynaert, M., Verboven, F., 2014. Improving the performance of random coefficients demand models: the role of optimal instruments. *J. Econometrics* 179 (1), 83–98.
- Rotnitzky, A., Cox, D.R., Bottai, M., Robins, J., 2000. Likelihood-based inference with singular information matrix. *Bernoulli* 6 (2), 243–284.
- Sargan, J., 1983. Identification and lack of identification. *Econometrica* 51 (6), 1605–1633.
- Skrainka, B., Judd, K., 2011. High performance quadrature rules: how numerical integration affects a popular model of product differentiation. Working Paper.
- Staiger, D., Stock, J., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65 (3), 557–586.