Doubly Robust GMM Inference and Differentiated Products Demand Models*

S. Auray[†] N. Lepage-Saucier[‡] P. Tuvaandorj[§]

August 25, 2018

Abstract

This paper develops robust inference methods for moment condition models implemented with a $n^{1/2}$-consistent auxiliary estimator of the nuisance parameters. When applied to models subject to weak identification and boundary parameter problems, they simultaneously overcome both irregularities and are asymptotically pivotal with minimal assumptions on the parameter space. If these problems are not present in the data, they are asymptotically equivalent to standard statistics for nonlinear models. They also have similar computational requirements.

We apply our tests to the differentiated products demand model, which may suffer from both problems: the variance of the random coefficients is often close to zero, causing the boundary parameter problem, and the strength of the available instruments is often put in doubt, which may cause weak identification. We evaluate the performance of the proposed tests by simulations.

Keywords: Boundary parameter, heterogeneity, pivotal statistic, random utility, robust inference, weak identification

*The authors thank participants at the 11th International Conference on Computational and Financial Econometrics (CFE 2017) conference, the 9th French Econometrics Conference, and CREST-Ensai seminar for several useful comments and suggestions. All errors are our own.

[†]CREST-Ensai and ULCO, `stephane.auray@ensai.fr`

[‡]CREST-Ensai, `nicolas.lepage-saucier@ensai.fr`

[§]CREST-Ensai, `purevdorj.tuvaandorj@ensai.fr`

1 Introduction

Models subject to multiple irregularities at once are frequently encountered in empirical economic work. For instance, weak identification is often combined with boundary parameters, persistence or instrument selection problems.[1] With statistical procedures addressing a single issue, practitioners often wrongly rule out any other irregularities by assumption, potentially leading to inference failure.

We develop identification- and boundary-robust generalized method of moments (GMM) test statistics for parameter subvector, implemented with a $n^{1/2}$-consistent auxiliary estimator of the nuisance parameters (n denotes the sample size). When applied to models with possible identification failure and boundary parameter problems, they enjoy the *double robustness* property: i.e., their null asymptotic distribution does not depend on unknown parameters, irrespective of identification strength or of the distance of any parameter to the boundary of the parameter space. The proposed statistics are an Anderson-Rubin (AR)-type statistic, a $C(\alpha)$-type statistic, and a Conditional Likelihood Ratio (CLR)-type statistic. They have computational costs similar to standard weak identification-robust versions of these statistics and have equivalent asymptotic properties in the absence of the boundary parameter problem.

The literature on weak identification and weak instruments is substantial and mainly aims at devising statistical procedures that are both "honest" and efficient.[2] The former property means that the procedures correctly indicate whether the data are

[1] Weak identification occurs when objective functions are relatively flat in parts of the parameter space or the data do not provide sufficient information. The boundary parameter problem may arise from any model with parameters subject to inequality constraints, such as those describing the curvature of utility or production functions, weights and probabilities constrained to the unit interval, or from the variance of random parameters which cannot be negative. The random coefficients demand model introduced by Berry, Levinsohn, and Pakes (1995) is a prominent example combining weak instruments and boundary parameters; other cases are discussed by Andrews (2002). Dynamic models such as GARCH or SVAR are subject to persistence and identification problems at the same time. Chernozhukov, Hansen, and Spindler (2015) develop inference methods about low-dimensional structural parameters when high-dimensional nuisance parameters are estimated using selection or regularization methods.

[2] See Stock, Wright, and Yogo (2002) and Dufour (2003) for literature surveys on weak IV.

informative by producing a wide (or infinite) confidence interval when identification is weak. The latter means that when identification is strong, the procedures should perform on par with usual test statistics based on t-ratios.

In the GMM context, identification-robust tests have been proposed by Stock and Wright (2000), Kleibergen (2005), Chaudhuri and Zivot (2011), Andrews and Cheng (2014), Andrews and Guggenberger (2015a), Andrews (2016) and Andrews and Mikusheva (2016). Of these papers, Stock and Wright (2000), Kleibergen (2005), Chaudhuri and Zivot (2011), and Andrews and Cheng (2014) consider tests for restrictions on a parameter subvector as we do in this paper. However, none of the identification-robust test statistics proposed in the literature is robust to the nuisance parameters lying on the boundary of the parameter space.

The boundary parameter problem has also received attention, notably by Andrews (1999, 2001) who studies the general properties of estimators and tests in extremum estimation setups. He shows that, in the presence of boundary effects, the usual estimators and tests such as Wald, LM, and LR statistics do not have the usual normal and chi-square distributions asymptotically. Additionally, Ketz (2018) proposes a boundary-robust LR-type statistic for extremum estimation.

While the literature addresses either weak identification or the boundary parameter problem separately, the extra challenge posed by their common occurrence has not been explored systematically. This gap in the literature is significant since both problems can occur simultaneously in many models with inequality restrictions on the parameters. In practice, they are sometimes ignored or addressed in unsatisfactory ways.[3]

[3] When boundary parameters or weak identification are suspected, a practical solution sometimes considered is pretesting to assess whether standard inference methods may be used. In general, however, such procedures suffer from the so-called pretest bias. Using the data twice in such a manner – first the pretesting and then using the standard t-statistic based confidence interval – creates a distortion in the null rejection probability of the test statistic used in the second stage because the two test statistics depend on one another in a complicated fashion.

An alternative strategy would be to use projection methods by including the parameters that are potentially at the boundary within the subvector under test. However, in addition to the standard drawbacks, these methods can be prohibitively slow and suffer from the curse of dimensionality for a model such as the BLP since the whole variance vector is potentially at the boundary.

We shall now describe our test statistics, which can be viewed as more general versions of their usual identification-robust counterparts. They merely require $n^{1/2}$-consistent estimator for the nuisance parameters. The usual restricted GMM estimator, which may have a non-normal asymptotic distribution, can serve as this auxiliary estimate in all of the proposed test statistics.

The first statistic is an Anderson-Rubin (AR)-type statistic that generalizes the GMM AR statistic of Stock and Wright (2000). It is a quadratic form of a moment function orthogonalized against the GMM score-type function associated with the nuisance parameters. It is asymptotically pivotal, even if the nuisance parameter is on the boundary of the parameter space. We also define a second doubly robust AR-type statistic which uses a one-step estimator of the nuisance parameters based on the initial $n^{1/2}$-consistent estimate.[4]

The next statistic is a Neyman's $C(\alpha)$-type statistic[5] implemented with a $n^{1/2}$-consistent auxiliary estimator of the nuisance parameters. It is comparable to the LM statistic of Kleibergen (2005) but employs a different Jacobian estimator constructed as follows. The Jacobian with respect to the parameters of interest and the sample moment functions are first orthogonalized against the nuisance parameters scores. This removes the effect of the estimation error asymptotically. The residual Jacobian is then orthogonalized against the residual sample moment functions, making the resulting Jacobian estimator asymptotically independent of the orthogonalized sample moment functions.

[4] The asymptotic properties of this second version of the AR statistic are equivalent to the first. However, it has a notable limitation in that the one-step estimator requires a non-restricted parameter space. Of course, restrictions on the parameter space are often the actual cause of the boundary parameter problem. The variance parameter of the BLP is a good example, since it is restricted to non-negative values.

[5] The $C(\alpha)$ statistic (non-identification-robust) is proposed by Neyman (1959). Dufour and Tuvaandorj (2018) develop identification-robust $C(\alpha)$-type statistics in likelihood and minimum distance models that can be reparameterized by identifiable parameters, and establish their bootstrap validity. For further references on the $C(\alpha)$ test and for the asymptotic distribution of a general $C(\alpha)$ statistic in the strongly identified GMM and estimating functions framework, see Dufour, Trognon, and Tuvaandorj (2016).

In a related paper, Chaudhuri and Zivot (2011) propose a projection-based $C(\alpha)$-type test in GMM with weakly identified nuisance parameters. The difference between their statistic and the one we propose are as follows. The procedure of Chaudhuri and Zivot (2011) employs Kleibergen (2005)'s Jacobian estimator. Their statistic is computationally expensive because it is projection-based, i.e., the auxiliary estimate of the nuisance parameters is determined by minimizing the test statistic over a confidence set obtained in the first stage. It is also conservative because of the use of the Bonferroni correction. However, it is sufficiently general to accommodate weakly identified nuisance parameters. In contrast, since our statistic assumes strongly identified nuisance parameters and uses the plug-in estimate, it is asymptotically chi-square distributed and is not conservative.

Furthermore, we do not need to specify the source from which the estimate is obtained for the general GMM result. As long as the auxiliary estimate can be computed easily, our statistic should be more straightforward to implement. Chaudhuri and Zivot (2011) present distributional results when the parameters are fixed as opposed to being estimated, which is the case considered in this paper. A final difference is that our $C(\alpha)$ statistic is based on a general GMM objective in contrast to the efficient GMM used by Chaudhuri and Zivot (2011). Andrews (2017) recently proposed a $C(\alpha)$-type procedure that has correct asymptotic size building on Chaudhuri and Zivot (2011). The aforementioned differences (except the conservativeness) between our test and Chaudhuri and Zivot (2011) apply there as well.

The last statistic is a boundary-robust CLR-type statistic which extends the identification-robust statistic of Andrews and Guggenberger (2015a). It is a function of the doubly robust AR and $C(\alpha)$ statistics and a rank statistic, and, like the others, it is implemented with a $n^{1/2}$-consistent auxiliary estimator.

We apply our statistics to the differentiated products demand model of Berry, Levinsohn, and Pakes (1995) (BLP). This model accounts for unobserved taste heterogeneity in a flexible way and only requires data on market shares, prices and product characteristics. The model and its variants have been used in numerous and various economic applications, including predicting the demand for a new product, evaluating its welfare effect, measuring the impact of mergers and exclusive dealing on prices, and measuring the impact of advertising and brand switching costs.[6]

There is also a substantial literature investigating the properties of estimators and tests in the BLP model. Berry, Linton, and Pakes (2004) derive the consistency and asymptotic normality of the GMM estimators when the number of products is large. Freyberger (2015) studies the properties of the GMM estimators when the number of markets is large. Reynaert and Verboven (2014) study the behavior of the optimal instruments in the BLP model and find that the efficiency gain from using them can be substantial.

Despite the model's popularity, inference in the BLP may be complicated by weak instruments and boundary parameters. To identify the impact of prices on product demand, commonly used instruments include cost shifters related to production inputs[7], and linear transformations of products characteristics which may affect a firm's market power, referred to as BLP instruments. Both types of instruments may be marginally relevant, causing weak identification. Often, variations in input prices are small and affect firms in similar ways, providing limited explanatory power for price differences between individual firms or individual products. As for BLP instruments, Armstrong (2016a) points out how instruments that depend on imperfect competition can be weak in large markets, where firms' market power tends to be small. The weak identification sequence of Armstrong (2016a) does not directly correspond to the drifting Jacobian assumption used in the literature. However, we demonstrate that it does upon reparameterization. Under the reparameterized model, we show the asymptotic pivotality of the doubly robust statistics by verifying the invariance properties of the test statistics.

Further complication arises from the need to estimate the heterogeneity of consumer tastes for characteristics. If tastes are in fact relatively homogeneous for some characteristics, their distribution can be clustered around a single parameter value.

[6] Knittel and Metaxoglou (2014) list 22 articles in leading journals using the main components of the BLP methodology. Recent examples include Dubois, Griffith, and O'Connell (2018), Miller and Weinberg (2017), Shcherbakov (2016), Nurski and Verboven (2016) and Eizenberg (2014).

[7] Common instruments include the cost of materials, components and transportation, or product prices in other regions, implying common cost shocks.

As a result, the variance parameter capturing this heterogeneity can be close to zero in empirical applications, causing a boundary parameter problem. Ketz (2018) shows how this distorts the size of standard t-ratio inference. He addresses the issue by developing a pivotal likelihood ratio-type test. In practice, since both weak instruments and the boundary parameter problem can be real features of the data in the BLP model, inference must be robust to both sources of irregularity.

The rest of the paper is organized as follows. The doubly robust statistics in a general GMM setup are described in detail in Section 2. The specific inference problems associated with the BLP are discussed in Section 3. Section 4 shows simulations exploring the finite sample properties of our test statistics applied to the BLP in a variety of parameter configurations. In particular, we consider cost shifters and simulations with BLP instruments in imperfect competition in the spirit of Armstrong (2016a). Finally, Section 5 concludes.

Notations

For variables with double indices e.g., y_{jt} with $j = 1, \ldots, J$ and $t = 1, \ldots, T$, define $y_{\cdot t} = [y_{1t}, \ldots, y_{Jt}]'$ and $y_{j\cdot} = [y_{j1}, \ldots, y_{jT}]'$. When it is clear from the context, we simplify the notations and write $y_t = [y_{1t}, \ldots, y_{Jt}]'$ and $y_j = [y_{j1}, \ldots, y_{jT}]'$. For a random or nonrandom function f that depends on a vector $\theta = (\theta_1', \theta_2')'$, we write $f(\theta) = f(\theta_1, \theta_2)$. The minimum eigenvalue of a matrix is denoted $\lambda_{\min}(.)$. $0_{r\times c}$ denotes $r \times c$ matrix of zeros, and I_k denotes the unit matrix of dimension k. For a $n \times k$ matrix X, let $P_X = X(X'X)^{-1}X'$ and $M_X = I_n - X(X'X)^{-1}X'$. $\|X\| = (\mathrm{tr}(X'X))^{1/2}$ denotes the matrix norm. $\xrightarrow{p}$ and $\xrightarrow{d}$ stand for the convergence in probability and in distribution, respectively. LLN abbreviates "Law of Large Numbers", CLT abbreviates "Central Limit Theorem", and CMT stands for "Continuous Mapping Theorem".

2 Doubly robust inference

In Subsection 2.1, we describe the testing problem and define the robust Jacobian estimator. In Subsection 2.2, we develop doubly robust test statistics and establish their asymptotic properties.

2.1 GMM setup

Let $m_i(\theta) \equiv m(y_i;\theta)$ be a $L \times 1$ moment function of a vector of observations y_i and a $d \times 1$ parameter vector $\theta \in \mathbb{R}^d$ satisfying the restriction:

$$E[m_i(\theta_0)] = 0, \quad i = 1, \dots, n,$$

where θ_0 denotes the true parameter value. The sample moment function and the GMM objective function are

$$\hat{m}_n(\theta) = n^{-1}\sum_{i=1}^{n} m_i(\theta), \quad Q_n(\theta) = \hat{m}_n(\theta)' W_n \hat{m}_n(\theta),$$

respectively, where W_n is a $L \times L$ weighting matrix. The efficient weight matrix is $W_n = \hat{\Sigma}_n^{-1}(\theta)$, where $\hat{\Sigma}_n(\theta)$ is a consistent estimate of the asymptotic variance $\Sigma(\theta)$ of $n^{1/2}\hat{m}_n(\theta)$. Under standard regularity conditions,

$$n^{1/2}\hat{m}_n(\theta_0) \xrightarrow{d} N[0, \Sigma],$$

with $\Sigma \equiv \Sigma(\theta_0)$. To describe the testing problem, partition $\theta = (\theta_1', \theta_2')'$, where $\theta_1 \in \Theta_1 \subseteq \mathbb{R}^{d_1}$ and $\theta_2 \in \Theta_2 \subseteq \mathbb{R}^{d_2}$ with $d = d_1 + d_2$. The corresponding partition of the true parameter value is $\theta_0 = [\theta_{01}', \theta_{02}']'$. Denote the Jacobian of the sample moment functions as

$$\hat{G}_n(\theta) \equiv \frac{\partial \hat{m}_n(\theta)}{\partial \theta'} = n^{-1}\sum_{i=1}^{n} G_i(\theta) = [\hat{G}_{1n}(\theta), \dots, \hat{G}_{dn}(\theta)], \tag{2.1}$$

where $G_i(\theta) = [G_{i1}(\theta), \dots, G_{id}(\theta)]$. It can be partitioned as $\hat{G}_n(\theta) = \left[\hat{G}_{n,1}(\theta), \hat{G}_{n,2}(\theta)\right]$, conformably to $\theta = (\theta_1', \theta_2')'$.

We develop tests for hypotheses on the subvector θ_1 (typically the coefficients on endogenous variables), treating θ_2 as nuisance parameters:

$$H_0 : \theta_1 = \theta_{01}, \quad \theta_2 \in \Theta_2.$$

We allow two nonstandard features for this testing problem: (i) The rank of $\operatorname*{plim}_{n\to\infty} \hat{G}_n(\theta_0)$ may be less than d, or close to being so, because θ_1 may not be identified or may be weakly identified.[8] In contrast, the nuisance parameter θ_2 is assumed to be strongly identified, with rank $\left[\operatorname*{plim}_{n\to\infty} \hat{G}_{n,2}(\theta_0)\right] = d_2$.[9] (ii) The nuisance parameter vector θ_2 (or some of its components) may be near the boundary of the parameter space Θ_2, causing the boundary parameter problem.

A confidence region for the parameters of interest θ_1 can then be built by collecting the parameter values that are not rejected by the tests of the null hypothesis.

It is well known in the weak identification literature that test statistics based on the Jacobian in (2.1) are not asymptotically pivotal. As shown by Kleibergen (2005), the asymptotic independence of the Jacobian estimator and the sample moment function is the key property allowing the construction of asymptotically pivotal test statistics; the robust Jacobian estimator of Kleibergen (2005) is defined as the residuals from the regression of the column vector $\hat{G}_{jn}(\theta)$ on $\hat{m}_n(\theta)$: for $j = 1, \ldots, d$,

$$\hat{G}^{\perp}_{jn}(\theta) = \hat{G}_{jn}(\theta) - \hat{C}_{jn}(\theta)\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta). \tag{2.2}$$

Our test statistics are based on the robust Jacobian estimator defined by

$$\hat{H}_n(\theta) = [\hat{H}_{1n}(\theta), \ldots, \hat{H}_{dn}(\theta)], \tag{2.3}$$

8 Suppose that $\hat{G}_n(\theta_0) \xrightarrow{p} G(\theta_0)$, where $G(\theta_0)$ is a $L \times d$ matrix. It is well known that the model parameters θ_0 are not identified when $G(\theta_0)$ is not full rank, i.e., $\operatorname{rank}(G(\theta_0)) < d$ (see Newey and McFadden (1994) for example). We use a drifting Jacobian assumption to model the weak identification, see Assumption A.4.

9 If some elements of θ_2 were also weakly identified, it could not be estimated consistently and statistics based on the plug-in estimator would not be pivotal.

where

$$\hat{H}_{jn}(\theta) = \hat{G}_{jn}(\theta) - \hat{C}_{jn}(\theta)\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta) + \Big(\hat{C}_{jn}(\theta)\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta) - \hat{D}_{jn}(\theta)\Big)$$
$$(\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta))^{-1}\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta), \tag{2.4}$$

$$\hat{C}_{jn}(\theta) = n^{-1}\sum_{i=1}^{n}\Big[G_{ij}(\theta) - \hat{G}_{jn}(\theta)\Big]\, m_i(\theta)', \tag{2.5}$$

$$\hat{D}_{jn}(\theta) = \frac{\partial \hat{G}_{jn}(\theta)}{\partial \theta_2'}, \quad j = 1, \ldots, d. \tag{2.6}$$

In contrast to (2.2), the Jacobian estimator in (2.3)-(2.6) is obtained by first orthogonalizing the Jacobian $\hat{G}_{jn}(\theta)$ and the sample moment function $\hat{m}_n(\theta)$ against the score-type function $\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta)$ with respect to the nuisance parameter vector θ_2 which yields

$$\bar{m}_n(\theta) = \hat{m}_n(\theta) - \hat{G}_{n,2}(\theta)(\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta))^{-1}\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta), \tag{2.7}$$

$$\bar{G}_{jn}(\theta) = \hat{G}_{jn}(\theta) - \hat{D}_{jn}(\theta)(\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta))^{-1}\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta), \tag{2.8}$$

and then orthogonalizing $\bar{G}_{jn}(\theta)$ against $\bar{m}_n(\theta)$. Thus, we may rewrite

$$\hat{H}_{jn}(\theta) = \bar{G}_{jn}(\theta) - \hat{C}_{jn}(\theta)\hat{\Sigma}_n(\theta)^{-1}\bar{m}_n(\theta). \tag{2.9}$$

The intuition behind (2.7)-(2.8) is simple. Consider the efficient CU-GMM objective function $Q_n(\theta) = \hat{m}_n(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta)$. In the absence of parameters at the boundary, the first order conditions of the constrained minimization of $Q_n(\theta)$ are met and $\hat{G}_{n,2}(\tilde{\theta})'\hat{\Sigma}_n(\tilde{\theta})^{-1}\hat{m}_n(\theta) = 0$. However, a parameter at the boundary imposes an additional constraint on θ_2 which distorts the distributions of the estimators and tests. The test statistics based on the transformations (2.7)-(2.8) are immune to such distortions. Obviously, if no parameters are at the boundary, they are asymptotically equivalent to their non-boundary-robust counterparts.

As before, we partition $\hat{H}_n(\theta) = \Big[\hat{H}_{n,1}(\theta), \hat{H}_{n,2}(\theta)\Big]$ conformably to $\theta = (\theta_1', \theta_2')'$. For an estimator $\tilde{\theta}^* = (\theta_{01}', \tilde{\theta}_2^{*\prime})'$ where $\tilde{\theta}_2^*$ is any $n^{1/2}$-consistent estimator of θ_{02}, we analogously set

$$\tilde{H}_n = \hat{H}_n(\tilde{\theta}^*) = [\tilde{H}_{n,1}, \tilde{H}_{n,2}] \quad \text{and} \quad \tilde{\Sigma}_n = \hat{\Sigma}_n(\tilde{\theta}^*). \tag{2.10}$$

Since the parameter θ_2 is strongly identified by assumption, we use the non-robust Jacobian $\hat{G}_{n,2}(\tilde{\theta}^*)$ instead of the robust Jacobian $\tilde{H}_{n,2}$, and set $\tilde{H}_{n,2} = \hat{G}_{n,2}(\tilde{\theta}^*)$ unless any confusion arises. We do not make any assumption regarding the location of the true parameters in the parameter space; we only assume that there exists a $n^{1/2}$-consistent estimate of the the subvector θ_2 under the null hypothesis $H_0 : \theta_1 = \theta_{01}$. The vector of strongly identified nuisance parameters is $n^{1/2}$-consistently estimable under mild regularity conditions (see for example Stock and Wright (2000)).

2.2 Doubly robust statistics

Our doubly robust statistics generalize the identification-robust statistics proposed in the framework of the efficient GMM and/or the continuous updating generalized method of moments (CU-GMM) of Hansen, Heaton, and Yaron (1996). We start by describing them to facilitate comparison with our statistics. Let $\tilde{\theta} = (\theta_{01}', \tilde{\theta}_2')'$ where $\tilde{\theta}_2$ is the restricted GMM estimator for the strongly identified nuisance parameter θ_2 under the constraint $\theta_1 = \theta_{01}$:

$$\tilde{\theta}_2 = \arg \min_{\theta_2 \in \Theta_2, \theta_1 = \theta_{01}} \hat{m}_n(\theta)' \hat{\Sigma}_n(\theta)^{-1} \hat{m}_n(\theta).$$

The GMM-AR statistic (also called the S statistic) of Stock and Wright (2000) and the GMM-LM statistic of Kleibergen (2005) for testing the hypothesis $H_0(\theta_{01}) : \theta_1 = \theta_{01}$ are given by

$$AR(\theta_{01}) = n\, \hat{m}_n(\tilde{\theta})' \hat{\Sigma}_n(\tilde{\theta})^{-1} \hat{m}_n(\tilde{\theta}), \tag{2.11}$$

$$LM(\theta_{01}) = n\, \hat{m}_n(\tilde{\theta})' \tilde{\Sigma}_n^{-1} \tilde{G}_{n,1}^{\perp} \big(\tilde{G}_{n,1}^{\perp\prime} \tilde{\Sigma}_n^{-1/2} M_{\tilde{\Sigma}_n^{-1/2} \tilde{G}_{n,2}^{\perp}} \tilde{\Sigma}_n^{-1/2} \tilde{G}_{n,1}^{\perp} \big)^{-1} \tilde{G}_{n,1}^{\perp\prime} \tilde{\Sigma}_n^{-1} \hat{m}_n(\tilde{\theta}). \tag{2.12}$$

These GMM-AR and GMM-LM statistics are robust to identification failure of θ_{01} having chi-square asymptotic distributions. However, they are not pivotal when the nuisance parameter θ_{02} is on the boundary of the parameter space Θ_2 because the restricted GMM estimator $\tilde{\theta}_2$ has non normal asymptotic distribution that depends on unknown parameters.

We shall now describe our doubly robust statistics for the hypothesis $H_0 : \theta = \theta_{01}$. The doubly robust $C(\alpha)$ statistic is defined by

$$C_\alpha(\theta_{01}) = n\,\hat{m}_n(\tilde{\theta}^*)'\tilde{B}_n' \left[\tilde{B}_n\tilde{\Sigma}_n\tilde{B}_n'\right]^{-1} \tilde{B}_n\hat{m}_n(\tilde{\theta}^*), \tag{2.13}$$

where $\tilde{B}_n = \tilde{H}_{n,1}'W_n\tilde{\Sigma}_nW_n - \tilde{H}_{n,1}'W_n\tilde{\Sigma}_nW_n\tilde{H}_{n,2}(\tilde{H}_{n,2}'W_n\tilde{\Sigma}_nW_n\tilde{H}_{n,2})^{-1}\tilde{H}_{n,2}'W_n\tilde{\Sigma}_nW_n$. For the efficient GMM, $W_n = \tilde{\Sigma}_n^{-1}$ and the $C(\alpha)$ statistic above simplifies to

$$\begin{aligned} C_\alpha(\theta_{01}) &= n\,\hat{m}_n(\tilde{\theta}^*)'\tilde{\Sigma}_n^{-1/2}P_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_n}\tilde{\Sigma}_n^{-1/2}\hat{m}_n(\tilde{\theta}^*) - n\,\hat{m}_n(\tilde{\theta}^*)'\tilde{\Sigma}_n^{-1/2}P_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}}\tilde{\Sigma}_n^{-1/2}\hat{m}_n(\tilde{\theta}^*), \\ &= n\,\hat{m}_n(\tilde{\theta}^*)'\tilde{\Sigma}_n^{-1/2}P_{M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}}\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,1}}\tilde{\Sigma}_n^{-1/2}\hat{m}_n(\tilde{\theta}^*). \end{aligned} \tag{2.14}$$

When $\tilde{\theta}^* = \tilde{\theta}$, that is, the auxiliary estimator is the restricted GMM estimator, $\tilde{G}_{n,2}'\tilde{\Sigma}_n^{-1}\hat{m}_n(\tilde{\theta}) = 0$ which holds if there is no boundary problem, the statistic (2.14) collapses to the subvector LM statistic of Kleibergen (2005) in (2.12). The statistic (2.14) differs from the projection-based $C(\alpha)$-type test procedure proposed by Chaudhuri and Zivot (2011) because we employ the Jacobian estimator (2.3) that is different from the Jacobian estimator of Kleibergen (2005) used by these authors. See Section 1 for further differences between our statistic and the Chaudhuri and Zivot (2011) procedure. The following proposition establishes the asymptotic distribution of the $C(\alpha)$ statistic in (2.13) under the assumptions stated in Appendix A.

Proposition 2.1. *Let Assumptions A.1-A.7 hold. For the $C(\alpha)$ statistic defined in (2.13), we have, under $H_0 : \theta_1 = \theta_{01}$, that*

$$C_\alpha(\theta_{01}) \xrightarrow{d} \chi^2_{d_1}.$$

The $C(\alpha)$ statistic is doubly robust under the sole requirement that the nuisance parameter estimate be $n^{1/2}$-consistent, a useful feature given the asymptotic non-normality of the restricted GMM estimator. In addition, it allows for a general weighting matrix W_n

as opposed to the existing identification-robust statistics based on the efficient GMM objective function.[10]

Next we construct two different doubly robust AR-type statistics. The first is defined as a $C(\alpha)$-type AR statistic based on an orthogonalized sample moment function (2.7):

$$AR_\alpha(\theta_{01}) = n\,\bar{m}_n(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\bar{m}_n(\tilde{\theta}^*). \tag{2.15}$$

The sample moment function in (2.7) is the residual vector from the projection of the sample moment function $\hat{m}_n(\theta)$ onto the space spanned by the nuisance parameter score vector $\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta)$ evaluated at the restricted GMM estimator $\tilde{\theta}$. In fact, the way the effect of the estimated nuisance parameters is removed from the moment in (2.7) underlies the mechanics behind all $C(\alpha)$-type statistics, hence the AR-type statistic in (2.15) may as well be called the $C(\alpha)$-type AR statistic.

The second is an AR-type statistic that uses the one-step estimator. Given an initial $n^{1/2}$-consistent estimator $\tilde{\theta}_2$ (possibly nonnormal asymptotically), a one-step Newton-Raphson iteration based on the GMM criterion function produces an asymptotically normal estimator (see Subsection 3.4 of Newey and McFadden (1994)). Ketz (2018) effectively uses the one-step estimator in boundary parameter testing problem with the goal of constructing a boundary-robust LR-type statistic (not necessarily identification-robust).

Let $\tilde{\theta} = (\theta_{01}', \tilde{\theta}_2')'$ where $\tilde{\theta}_2$ is a $n^{1/2}$-consistent estimator of θ_{02} under the null hypothesis $H_0 : \theta_1 = \theta_{01}$ e.g. the restricted GMM estimator of θ_2. Given $\tilde{\theta}$, the one-step estimator of θ_{02} may be defined as

$$\tilde{\theta}_2^* = \tilde{\theta}_2 - (\hat{G}_{n,2}(\tilde{\theta})'\hat{\Sigma}_n(\tilde{\theta})^{-1}\hat{G}_{n,2}(\tilde{\theta}))^{-1}\hat{G}_{n,2}(\tilde{\theta})'\hat{\Sigma}_n(\tilde{\theta})^{-1}\hat{m}_n(\tilde{\theta}). \tag{2.16}$$

[10] Moran (1973) and Chant (1974) examined the properties of the $C(\alpha)$ test the tested parameters are on the boundary of the maintained hypothesis. For results about the asymptotic distribution of the GMM $C(\alpha)$-type statistic in a weakly identified model when all parameters are fixed at their true values, see Chaudhuri and Zivot (2011). For general results in strongly identified models and further references about the $C(\alpha)$ statistic, see Dufour, Trognon, and Tuvaandorj (2016).

It is not difficult to see that the one-step estimator above is asymptotically normally distributed. In the following lemma, we present a mild extension of the asymptotic normality of the one-step estimator for a general nonlinear parameter transformation $\psi(\theta)$.

Lemma 2.2. *Let Assumptions A.1-A.5 hold. Let $\psi(\theta)$ be a $q \times 1$ ($q \leq d_2$) nonlinear transformation of the parameter that is continuously differentiable in a neighborhood of θ_0 with*

$$P\left[\text{rank}\left(\frac{\partial \psi(\theta)}{\partial \theta_2'}\right) = q\right] = 1.$$

For $\tilde{\theta} = (\theta_{01}', \tilde{\theta}_2')'$ such that $n^{1/2}(\tilde{\theta}_2 - \theta_{02}) = O_p(1)$, the one-step estimator defined as

$$\tilde{\psi}^* = \psi(\tilde{\theta}) - \frac{\partial \psi(\tilde{\theta})}{\partial \theta_2'}(\hat{G}_{n,2}(\tilde{\theta})'\hat{\Sigma}_n(\tilde{\theta})^{-1}\hat{G}_{n,2}(\tilde{\theta}))^{-1}\hat{G}_{n,2}(\tilde{\theta})'\hat{\Sigma}_n(\tilde{\theta})^{-1}\hat{m}_n(\tilde{\theta}) \tag{2.17}$$

satisfies

$$n^{1/2}(\tilde{\psi}^* - \psi(\theta_0)) \xrightarrow{d} N\left[0, \frac{\partial \psi(\theta_0)}{\partial \theta_2'}(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}\frac{\partial \psi(\theta_0)'}{\partial \theta_2}\right],$$

where $\hat{G}_{n,2}(\theta_0) \xrightarrow{p} H_2(\theta_0)$.[11]

Letting $\tilde{\theta}^* = (\theta_{01}', \tilde{\theta}_2^{*\prime})'$, where $\tilde{\theta}_2^*$ is as defined above in (2.16), define the doubly robust AR-type statistic as follows:

$$AR_\alpha^0(\theta_{01}) = n\,\hat{m}_n(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\hat{m}_n(\tilde{\theta}^*). \tag{2.18}$$

An advantage of the AR_α over the AR_α^0 statistic is that it does not require the estimator to be defined in the neighborhood of the true parameter. The AR_α^0 is applicable when an inequality constraint imposed on the nuisance parameter may be relaxed if the one-step estimator $\tilde{\theta}_2^*$ falls outside the constraint. This is not the case for the BLP model since it cannot be evaluated for negative values of the variance. Therefore, the AR_α^0 and other statistics based on it cannot be applied to the BLP model.

An alternative interpretation can be given to the orthogonalized sample moment

[11] We use the notation $H_2(\theta)$ to make it consistent with the robust Jacobian estimator $\hat{H}_{n,1}(\theta)$.

function (or the effective score-type sample moment function) in (2.7). Replacing $\psi(\theta)$ by $\hat{m}_n(\theta)$ in the formula (2.17), we can see that the orthogonalized sample moment function (2.7) can be viewed as the one-step estimator of the sample moment function given the initial $n^{1/2}$-consistent estimate $\tilde{\theta}$. The following proposition establishes the asymptotic distribution of the doubly robust AR statistics.

Proposition 2.3. *Let Assumptions A.1-A.4, A.6 and A.7 hold. Then, under $H_0: \theta_1 = \theta_{01}$, the AR-type statistics defined in (2.15) and (2.18) satisfy*

$$AR_\alpha(\theta_{01}), AR^0_\alpha(\theta_{01}) \xrightarrow{d} \chi^2_{L-d_2}.$$

By standard argument, the $AR_\alpha(\theta_{01})$ and $AR^0_\alpha(\theta_{01})$ statistics can be decomposed as the sums of the $C_\alpha(\theta_{01})$ statistic in (2.14) and statistics that test the overidentifying restriction $H_0 : E[m_i(\theta_0)] = 0$ for some $\theta_0 \in \Theta$:

$$\begin{aligned}
AR_\alpha(\theta_{01}) =& C_\alpha(\theta_{01}) + n\,\hat{m}_n(\tilde{\theta}^*)'\tilde{\Sigma}_n^{-1/2} M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}} M_{M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}}\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,1}} M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}} \tilde{\Sigma}_n^{-1/2}\hat{m}_n(\tilde{\theta}^*), \\
AR^0_\alpha(\theta_{01}) =& C_\alpha(\theta_{01}) + n\,\hat{m}_n(\tilde{\theta}^*)'\tilde{\Sigma}_n^{-1/2} M_{M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}}\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,1}} \tilde{\Sigma}_n^{-1/2}\hat{m}_n(\tilde{\theta}^*).
\end{aligned}$$

The last statistic that we consider is a conditional LR-type statistic for $H_0 : \theta_1 = \theta_{01}$ which builds on Andrews and Guggenberger (2015a) who in turn develop identification-robust statistic for the full parameter vector θ. Given the one-step estimator $\tilde{\theta}^* = (\theta_{01}', \tilde{\theta}_2^{*\prime})'$, we define the doubly robust subvector conditional LR-statistic as follows:

$$CLR_\alpha(\theta_{01}) = \tilde{S}_n' M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}} \tilde{S}_n - \lambda_{\min}\left[(\tilde{S}_n, \tilde{T}_n)' M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}} (\tilde{S}_n, \tilde{T}_n)\right], \tag{2.19}$$

where

$$\begin{aligned}
\tilde{S}_n &= \tilde{\Sigma}_n^{-1/2} n^{1/2} \hat{m}_n(\tilde{\theta}^*), \\
\tilde{T}_n &= \tilde{\Sigma}_n^{-1/2} n^{1/2} \hat{H}_{n,1}(\tilde{\theta}^*) \hat{U}_n(\tilde{\theta}^*)^{1/2}, \\
\hat{U}_n(\theta) &= [\theta_1, I_{d_1}](\hat{\Omega}^\varepsilon_n(\theta))^{-1}[\theta_1, I_{d_1}]'.
\end{aligned} \tag{2.20}$$

$\hat{\Omega}^\varepsilon_n(\theta)$ is an eigenvalue adjusted version of a $(d_1 + 1) \times (d_1 + 1)$ matrix $\hat{\Omega}_n(\theta)$ with

$(i, j), i, j = 1, \ldots, d_1 + 1$, element given by

$$\hat{\Omega}_{ij,n}(\theta) = \mathrm{tr}(\hat{K}_{ij,n}(\theta)' \hat{\Sigma}_n(\theta)^{-1})/L,$$

where the $L \times L$ matrix $\hat{K}_{ij,n}(\theta)$ is the (i, j) submatrix of $\hat{K}_n(\theta)$ defined as

$$\hat{K}_n(\theta) = (B(\theta)' \otimes I_L)\hat{V}_n(\theta)(B(\theta) \otimes I_L),$$

$\hat{V}_n(\theta)$ denotes a consistent estimator of the asymptotic variance $V(\theta_0)$ of

$$n^{1/2}\left[\hat{m}_n(\theta_0)', \mathrm{vec}\left(\hat{G}_{n,1}(\theta_0) - E[\hat{G}_{n,1}(\theta_0)]\right)'\right]',$$

and

$$B(\theta) = \begin{bmatrix} 1 & 0_{1\times d_1} \\ -\theta_1 & -I_{d_1} \end{bmatrix}.$$

For details on the eigenvalue adjustment, see Appendix D.

The doubly robust CLR test rejects at significance level $\zeta \in (0, 1)$ when

$$CLR_\alpha(\theta_{01}) \geq q_{1-\zeta}(\tilde{T}_n),$$

where $q_{1-\zeta}(\tilde{T}_n)$ is the $1 - \zeta$ quantile of the distribution of the random variable

$$S_\infty' M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}} S_\infty - \lambda_{\min}\left[(S_\infty, \tilde{T}_n)' M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}} (S_\infty, \tilde{T}_n)\right].$$

where $S_\infty \sim N[0, I_L]$ is asymptotically independent of $\tilde{T}_n$. The null asymptotic distribution of the subvector CLR statistic is given in the following proposition.

Proposition 2.4. *Let Assumptions A.1-A.4, A.6 and A.7 hold. Then, under $H_0 : \theta_1 = \theta_{01}$, the CLR statistic defined in (2.19) satisfies*

$$CLR_\alpha(\theta_{01}) \xrightarrow{d} S_\infty' M_{\Sigma^{-1/2}H_2} S_\infty - \lambda_{\min}\left[(S_\infty, T_\infty)' M_{\Sigma^{-1/2}H_2}(S_\infty, T_\infty)\right], \qquad (2.21)$$

where $S_\infty \sim N[0, I_L]$ is distributed independently of $T_\infty = \Sigma^{-1/2} H_{\infty,1}(\theta_0) U_\infty(\theta_0)$,

$$U_\infty(\theta_0) = [\theta_1, I_{d_1}](\Omega^\varepsilon(\theta_0))^{-1}[\theta_1, I_{d_1}]',$$

$\Omega^\varepsilon(\theta_0)$ is an eigenvalue adjusted version of a $(d_1+1)\times(d_1+1)$ matrix $\Omega(\theta_0)$ with $(i,j), i,j = 1,\ldots,d_1+1$, element given by

$$\Omega_{ij}(\theta_0) = \mathrm{tr}(K_{ij}(\theta_0)'\Sigma^{-1})/L,$$

where the $L\times L$ matrix $K_{ij}(\theta_0)$ is the (i,j) submatrix of $K(\theta_0)$ defined by

$$K(\theta_0) = (B(\theta_0)' \otimes I_L) V(\theta_0)(B(\theta_0) \otimes I_L),$$

and

$$\begin{aligned}
H_{\infty,1}(\theta_0) &= \left[H_{1\infty}(\theta_0), \ldots, H_{d_1\infty}(\theta_0)\right], \\
H_{i\infty}(\theta_0) &= \bar{G}_{i\infty}(\theta_0) - C_i(\theta_0)\Sigma^{-1}\bar{m}_\infty(\theta_0), \quad i = 1,\ldots,d_1, \\
\bar{m}_\infty(\theta_0) &= \left(I_L - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}\right) m_\infty(\theta_0), \\
\bar{G}_{i\infty}(\theta_0) &= G_{i\infty}(\theta_0) - D_i(\theta_0)(G_2(\theta_0)'\Sigma(\theta_0)^{-1}G_2(\theta_0))^{-1}H_2(\theta_0)'\Sigma^{-1}m_\infty(\theta_0).
\end{aligned}$$

When $d_1 = 1$, the doubly robust CLR statistic is simplified as

$$CLR_\alpha(\theta_{01}) = \frac{1}{2}\left(AR_\alpha(\theta_{01}) - R_\alpha(\theta_{01}) + \sqrt{(AR_\alpha(\theta_{01}) - R_\alpha(\theta_{01}))^2 + 4C_\alpha(\theta_{01})R_\alpha(\theta_{01})}\right), \tag{2.22}$$

where $C_\alpha(\theta_{01})$ and $AR_\alpha(\theta_{01})$ statistics are defined in (2.14) and (2.15), respectively, and

$$R_\alpha(\theta_{01}) = \tilde{T}_n' M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}} \tilde{T}_n. \tag{2.23}$$

The statistic (2.22) is a variant of the CLR-type statistic proposed in Kleibergen (2005) made robust to the boundary problem that can be implemented with $n^{1/2}$-consistent estimators of the nuisance parameters. The $AR_\alpha(\theta_{01})$ statistic may be replaced by the $AR^0_\alpha(\theta_{01})$ statistic in (2.18), and the $C_\alpha(\theta_{01})$ and $R_\alpha(\theta_{01})$ statistics with the initial

$n^{1/2}$-consistent estimator could be substituted by the $C_\alpha(\theta_{01})$ (or the LM statistic in (2.12)) and $R_\alpha(\theta_{01})$ statistics evaluated at the one-step estimator $\tilde{\theta}^* = (\theta_{01}', \tilde{\theta}_2^{*\prime})'$ with $\tilde{\theta}_2^*$ defined in (2.16).

The rank statistic (2.23) is a weighted version of the robust Jacobian, of which alternative forms can be considered. More details are provided in Appendix C, together with two alternative rank statistics used to construct alternative CLR-type statistics.

3 Inference in the BLP model

In this section, we describe the BLP model and the inferential challenges pertaining to it. We discuss the implementation of the proposed statistics in the BLP model including the appropriate form of the rank statistic used in the CLR-type test, and show their asymptotic validity under large and many market asymptotics.

3.1 The BLP model

In each market $t \in \{1, \ldots, T\}$ (which may represent periods), a consumer i chooses which product j to purchase. In total, J_t potential products are available, and an outside good, indexed by $j = 0$, which stands for a no purchase option. To simplify the exposition of this section, we assume that $J_t = J$ for all t; the argument would carry through with little modification for markets of various sizes. The utility of consumer i from choosing the alternative j in market t is

$$u_{ijt} = x_{jt}'\beta_i - p_{jt}\alpha + \xi_{jt} + \varepsilon_{ijt},$$

where x_{jt} is a $k \times 1$ vector of observed product characteristics, p_{jt} is the price of the jth product in period t and is endogenous, ξ_{jt} is a product characteristic unobserved by the researcher but observed by the consumer, and ε_{ijt} is the remaining unobserved component of the demand. The utility from the outside good is denoted $u_{i0t} = \varepsilon_{i0t}$. The consumer-specific $k \times 1$ vector of random coefficients $\beta_i = (\beta_{i1}, \ldots, \beta_{ik})'$ captures the variation of tastes between consumers.[12] It can be decomposed as the sum of its mean and deviation from its mean:

$$\beta_i = b + \Sigma_\beta^{1/2} v_i, \quad v_i \sim F_v,$$

where $b = (b_1, \ldots, b_k)'$, $\Sigma_\beta = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_k^2\}$, $\Sigma_\beta^{1/2}$ is the square root of the covariance matrix Σ_β such that $\Sigma_\beta = \Sigma_\beta^{1/2}\Sigma_\beta^{1/2}$, and F_v is the distribution function of the random vector v_i, and is known to the econometrician.

Letting $\delta_{jt} \equiv x_{jt}'b - p_{jt}\alpha + \xi_{jt}$ denote the mean utility common to all consumers, we may write

$$u_{ijt} = \delta_{jt} + x_{jt}'\Sigma_\beta^{1/2} v_i + \varepsilon_{ijt}.$$

Assuming that ε_{ijt} are i.i.d. Type-I extreme value distributed across i, j, and t, the conditional choice probability of the consumer i buying the jth product is given by

$$\pi_{ij} \equiv \frac{\exp(\delta_{jt} + x_{jt}'\Sigma_\beta^{1/2} v_i)}{1 + \sum_{j=1}^J \exp(\delta_{jt} + x_{jt}'\Sigma_\beta^{1/2} v_i)}.$$

The share of the product j in the market t is obtained by integrating the conditional choice probability π_{ij} with respect to the heterogeneity distribution:

$$S_{jt}(\delta_t, \sigma) = \int \frac{\exp(\delta_{jt} + x_{jt}'\Sigma_\beta^{1/2} v)}{1 + \sum_{j=1}^J \exp(\delta_{jt} + x_{jt}'\Sigma_\beta^{1/2} v)} dF_v(v), \quad j = 1, \ldots, J, \tag{3.1}$$

where $\delta_t = [\delta_{1t}, \ldots, \delta_{Jt}]'$ and $\sigma = [\sigma_1, \ldots, \sigma_k]'$. A common assumption is $F_v = \Phi_k$, where Φ_k is the cumulative distribution function of the multivariate standard normal distribution $N[0_{k\times 1}, I_k]$.

Equating the market share implied by the model $S_t(\delta_t, \sigma)$ to the observed market share s_t yields the market share system:

$$S_t(\delta_t, \sigma) = s_t,$$

[12] We assume that all characteristics have random coefficients while the price coefficient α is constant.

where $S_t(\delta_t, \sigma) = [S_{1t}(\delta_t, \sigma), \ldots, S_{Jt}(\delta_t, \sigma)]'$ and $s_t = [s_{1t}, \ldots, s_{Jt}]'$. Berry (1994) shows via a fixed point argument that there exists a unique vector $\delta(\sigma, p_t, s_t, x_t) \equiv \delta_t$ where $x_t = [x_{1t}, \ldots, x_{Jt}]$ $(k \times J)$ such that

$$S_t(\delta(\sigma, p_t, s_t, x_t), \sigma) = s_t$$

for any given s_t, x_t and σ.

The mean utility vector is obtained by inverting the implied market share function and can be written as

$$\delta_t = S_t^{-1}(\sigma, p_t, s_t, x_t).$$

Berry, Levinsohn, and Pakes (1995) compute δ_t by contraction mapping while Dubé, Fox, and Su (2012) obtain it indirectly through mathematical programming with equilibrium constraints (MPEC).

Let $\sigma^2 = [\sigma_1^2, \ldots, \sigma_k^2]'$ and collect the model parameters in $\theta = (\alpha, b', \sigma^{2\prime})'$.[13] From the last equation, the product-specific unobservable demand shock in market t is given by

$$\xi_{jt}(\theta, p_t, s_t, x_t) = \delta_{jt}(\sigma, p_t, s_t, x_t) - x_{jt}'b + p_{jt}\alpha, \quad j = 1, \ldots, J. \tag{3.2}$$

We focus on testing and building a confidence interval for the unknown scalar coefficient α on the price variable. The following mean independence assumption provides an exclusion restriction:

$$E[\xi_{jt}|z_t] = 0, \tag{3.3}$$

where z_t is a $L \times J$ matrix of instruments, and is used to identify the model parameters. The instruments z_t are variables or transformations thereof that shift the price without influencing tastes directly. One may, for example, consider the linear specification for

[13] We use the parameterization in terms of the variance rather than the dispersion because, as shown by Ketz (2017), the parameterization in terms of the latter in the BLP model results in a distorted inference.

the unobserved marginal cost

$$MC_{jt} = x'_{jt}\gamma_1 + z'_{jt,2}\gamma_2 + \omega_{jt}, \tag{3.4}$$

where $z_{jt,2}$ is a $l \times 1$ vector cost shifters (instruments) other than x_{jt}, and ω_{jt} is the unobserved component of the marginal cost. In vector notation,

$$MC_t = x'_t\gamma_1 + z'_{t,2}\gamma_2 + \omega_t, \tag{3.5}$$

where $MC_t = [MC_{1t}, \ldots, MC_{Jt}]'$, $z_{t,2} = [z_{1t,2}, \ldots, z_{Jt,2}]$, $p_t = [p_{1t}, \ldots, p_{Jt}]'$ and $\omega_t = [\omega_{1t}, \ldots, \omega_{Jt}]'$.

In perfectly competitive market, firms set a price equal to the marginal cost:

$$p_{jt} = MC_{jt}. \tag{3.6}$$

To gain a better understanding of the inferential problem, consider the linear case (3.4) which along with (3.2) and (3.6) yield the following system of equations:

$$\begin{aligned} \delta_{jt}(\theta, p_t, s_t, x_t) &= x'_{jt}b - p_{jt}\alpha + \xi_{jt}(\sigma, p_t, s_t, x_t), \\ p_{jt} &= x'_{jt}\gamma_1 + z'_{jt,2}\gamma_2 + \omega_{jt}. \end{aligned}$$

Let x_{-j} denote the characteristics of the products other than j, and let $h_j(x_{-j})$ be its transformation. As argued by Armstrong (2016a), for the model above, the characteristics of the rival product, $h_j(x_{-j})$, have no identifying power because they shift the price only through the markup which is zero in perfect competition. He considers the case where σ is known in which case, after $\delta_{jt}(\theta, p_t, s_t, x_t)$ is computed, the model is similar to the classical linear simultaneous equations (or the linear IV model). Even if cost shifters $z_{jt,2}$ uncorrelated with the demand shock ξ_{jt} are available, their identification strength may still be questionable.

Under imperfect competition, BLP instruments are made possible by exploiting firms' varying degree of market power which determines the markup they can extract from customers. Each period, there are F firms engaging in Bertrand competition.

Each firm $f \in F$ produces a subset of products denoted $\mathcal{J}(f) \subseteq \{1, \ldots, J\}$ and sets the prices so as to maximize the (per consumer) profit function

$$\sum_{k \in \mathcal{J}(f)} (p_{kt} - MC_{kt}) s_{kt}.$$

This yields the following first-order condition for Bertrand-Nash equilibrium at the interior solution:

$$s_{jt} + \sum_{k \in \mathcal{J}(f)} (p_{kt} - MC_{kt}) \frac{\partial s_{kt}}{\partial p_{jt}} = 0 \tag{3.7}$$

for any $j \in \mathcal{J}(f)$. Define a $J \times J$ matrix Δ_t with a (j, k) element given by

$$\Delta_{t,kj} = \begin{cases} -\frac{\partial s_{kt}}{\partial p_{jt}}, & \text{if} \quad j, k \in \mathcal{J}(f), \\ 0, & \text{otherwise.} \end{cases}$$

Then (3.7) in vectorized notation is

$$s_t - \Delta_t (p_t - MC_t) = 0.$$

Solving for c_t and using (3.5) gives

$$p_t - \Delta_t^{-1} s_t = x_t' \gamma_1 + z_{t,2}' \gamma_2 + \omega_t. \tag{3.8}$$

Unlike the perfect competition case, the markup $\Delta_t^{-1} s_t$ is subtracted from the equilibrium price in the equation (3.8). Since the characteristics of the competing products x_{-j} influence price through the markup term $\Delta_t^{-1} s_t$ and they are unrelated to taste shocks, they can be used as instruments. However, Armstrong (2016a) shows (assuming known σ) that in large market environment with J tending to infinity, the competing product characteristics instruments lose their identifying power when the markup converges to a constant quickly relative to the sampling error. Cost instruments may be included in $z_{jt,2}$ but, like for the perfect competition case, their explanatory power is not guaranteed.

As previously discussed, weak instruments are only one type of irregularity present in the BLP. If σ^2 or its components are on the boundary, i.e., they are close to zero, the restricted estimators of σ^2 will depend on unknown parameters which are not consistently estimable and the commonly used identification-robust statistics will not be pivotal.

Since both weak instruments and boundary parameters may be real features of the data, they cannot be ruled out a priori. Inference about α (as well as the other parameters) should thus be based on identification-robust and boundary-robust statistics that have asymptotically correct rejection rates, while retaining good power when these problems are absent from the data. The doubly robust GMM statistics introduced in the previous section have these properties.

We first derive the key quantities that are used to build the test statistics. We abstract from the simulation error in the evaluation of the integral (3.1), and the sampling error in the estimation of the market share s_{jt}.

Stack the observations into

$$
\begin{aligned}
&x_t = [x_{1t}, \ldots, x_{Jt}] \quad (k \times J), && z_{t,2} = [z_{1t,2}, \ldots, z_{Jt,2}] \quad (l \times J),\\
&X = [x_1, \ldots, x_T]' \quad (TJ \times k), && z_t = [x_t', z_{t,2}']' \quad ((k+l) \times J),\\
&\xi = [\xi_1', \ldots, \xi_T']' \quad (TJ \times 1), && Z_2 = [z_{1,2}, \ldots, z_{T,2}]' \quad (TJ \times l),\\
&\omega = [\omega_1', \ldots, \omega_T']' \quad (TJ \times 1), && Z = [X, Z_2] = [z_1, \ldots, z_T]' \quad (TJ \times (k+l)),\\
&p = [p_1', \ldots, p_T']' \quad (TJ \times 1).
\end{aligned}
$$

The model parameters are $\theta = (\alpha, b', \sigma^{2\prime})'$ with $\theta_1 = \alpha \in \mathbb{R}$ and $\theta_2 = (b', \sigma^{2\prime})' \in \Theta_2 \subset \mathbb{R}^{2k}$. Thus, $d_1 = 1$, $d_2 = 2k$, and $L = k + l$. As in Ketz (2017), we treat the variance vector σ^2, not the standard deviation, as part of the model parameters.

Two main asymptotics have been studied in the literature: many market asymptotics when T tends to infinity (Freyberger (2015) and Ketz (2017, 2018)) and large market asymptotics when J tends to infinity (Berry, Linton, and Pakes (2004), as well as Armstrong (2016a) who also considers the intermediate case of many large markets). If the number of markets is large relative to the number of products, the test statistics are derived by normalizing $Z'\xi$ by the number of markets $n = T$; the sample moment function is given by

$$\hat{m}_T(\theta) = T^{-1}Z'\xi = T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J} z_{jt}(\delta_{jt}(\sigma) + p_{jt}\alpha - x'_{jt}b), \tag{3.9}$$

where $\delta_t(\sigma) \equiv \delta_t(\sigma, p_t, s_t, x_t)$. Our results also carry over to the large market case (i.e., $n = J$) in which the moment restriction takes the following form:

$$\hat{m}_J(\theta) = J^{-1}Z'\xi = J^{-1}\sum_{j=1}^{J}\sum_{t=1}^{T} z_{jt}(\delta_{jt}(\sigma) + p_{jt}\alpha - x'_{jt}b). \tag{3.10}$$

The index i used to denote the observations in Section 2 corresponds either the product $i = j$ or the market $i = t$. In what follows, we consider two cases in which the asymptotics are taken with respect to either the number of products $n = J$ or the number of markets $n = T$.

3.2 Robust statistics under large market asymptotics

In this subsection, we show the asymptotic pivotality of our doubly robust statistics under weak identification sequences that arise endogenously from the market equilibrium. Armstrong (2016a) shows that the price coefficient α in the BLP model is weakly identified when the dependence of equilibrium markups on the characteristics of rival products is small relative to the number of products J. He also identifies conditions under which the BLP estimate remains consistent. We shall maintain assumptions similar to those in Armstrong (2016a).

Let $h_j(x_{-j})$ be a $l \times 1$ function of the characteristics of the products other than j referred to as the BLP instruments. For simplicity, we consider the single market case with $T = 1$, and write $\xi_{jt} = \delta_{jt}(\sigma_t, p_t, s_t, x_t) - x'_{jt}b + p_{jt}\alpha$ as

$$\xi_j = \delta_j(\sigma, p, s, x) - x'_j b + p_j\alpha$$

suppressing the subscript t.[14] We also write $\xi_j(\theta, p, s, x) = \xi_j(\alpha, \theta_2)$. Let $V[\xi_j] = V_{\xi\xi}$

[14] The results would be similar for a fixed number of large markets. The intermediate case of many

and $C[MC_j, \xi_j] = V_{MC\xi}$ denote the variance of ξ_j, and the covariance between MC_j and ξ_j respectively. The asymptotic distributions of the robust statistics are derived under the following conditions.

Assumption 3.1 (Parameter space)**.** *$\theta_2 = (b', \sigma^{2\prime})' \in \Theta_2 \subset [-b_l, b_u]^k \times [0, \sigma_u^2]^k$ for some $b_l, b_u, \sigma_u^2 > 0$.*

Assumption 3.2 (Conditions for large market asymptotics)**.** *(i) Let $\{[x_j', MC_j, \xi_j]'\}$ be i.i.d. with finite fourth moment where x_j contains a constant.*

(ii) The vector of instruments is denoted as $z_j = [x_j', h_j(x_{-j})']'$, the matrix $Z = [z_1, \ldots, z_J]'$ has full column rank with probability one: $P[\text{rank}(Z) = k+l] = 1$ with $l \geq k$, and as $J \to \infty$, the matrix $J^{-1}Z'\frac{\partial \delta(\sigma)}{\partial \sigma^{2\prime}}$ converges in probability uniformly over σ to a nonrandom matrix of full rank and continuous at σ_0. Moreover, $J^{-1}Z'Z - E[J^{-1}Z'Z] \xrightarrow{p} 0$ and $J^{-1}\sum_{j=1}^J \|z_j\|^4 - E[J^{-1}\sum_{j=1}^J \|z_j\|^4] \xrightarrow{p} 0$ as $J \to \infty$, where the stated expectations exist.

(iii) It holds that $J^{1/2}\text{max}_{1\leq j \leq J}|p_j - MC_j - \eta| \xrightarrow{p} 0$ for some constant η as $J \to \infty$. In addition, for all $j = 1, \ldots, J$ $E\left[(J^{1/2}|p_j - MC_j - \eta|)^{2+\varepsilon}\right] < \bar{U} < \infty$ for some $\bar{U}, \varepsilon > 0$.

(iv) It holds that

$$J^{-1/2}\sum_{j=1}^J \text{vec}\left(z_j\left(x_j', MC_j, \xi_j\right) - E\left[z_j\left(x_j', MC_j, \xi_j\right)\right]\right) \xrightarrow{d} N[0, V],$$

as $J \to \infty$, where the asymptotic covariance matrix

$$\begin{aligned} V = \lim_{J\to\infty} E&\left[J^{-1}\sum_{j=1}^J \text{vec}\left(z_j\left(x_j', MC_j, \xi_j\right)\right)\text{vec}\left(z_j\left(x_j', MC_j, \xi_j\right)\right)'\right] \\ &- E\left[J^{-1}\sum_{j=1}^J z_j\left(x_j', MC_j, \xi_j\right)\right] E\left[J^{-1}\sum_{j=1}^J z_j\left(x_j', MC_j, \xi_j\right)\right]' \end{aligned}$$

is invertible.

large markets, also discussed by Armstrong, is beyond the scope of this paper.

(v) It holds that

$J^{-1}z_jMC_j\xi_jz_j' - E[J^{-1}z_jMC_j\xi_jz_j'] \xrightarrow{p} 0$, $J^{-1}\sum_{j=1}^{J} z_jz_j'MC_j^2 - E[J^{-1}\sum_{j=1}^{J} z_jz_j'MC_j^2] \xrightarrow{p} 0$, *and* $J^{-1}\sum_{j=1}^{J} z_jz_j'MC_j - E[J^{-1}\sum_{j=1}^{J} z_jz_j'MC_j] \xrightarrow{p} 0$.

(vi) The conditional covariance of the marginal cost and the demand shock given z_j does not depend on z_j: $E[MC_j\xi_j|z_j] = V_{MC\,\xi}$.

(vii) $\sup_{\theta_2\in\Theta_2}\|J^{-1}Z'\xi(\alpha_0,\theta_2) - E[J^{-1}Z'\xi(\alpha_0,\theta_2)]\| \xrightarrow{p} 0$ *and* $\sup_{\theta_2\in\Theta_2}\|J^{-1}\sum_{j=1}^{J} z_jz_j'\xi_j^2(\alpha_0,\theta_2) - E\left[J^{-1}\sum_{j=1}^{J} z_jz_j'\xi_j^2(\alpha_0,\theta_2)\right]\| \xrightarrow{p} 0$ *where* $\lim_{J\to\infty} E[J^{-1}Z'\xi(\alpha_0,\theta_2)]$ *is zero uniquely at* $\theta_{02} = [b_0', \sigma_0^{2\prime}]'$ *and continuously differentiable with respect to* θ_2*, and* $\lim_{J\to\infty} E\left[J^{-1}\sum_{j=1}^{J} z_jz_j'\xi_j^2(\alpha_0,\theta_2)\right]$ *is positive definite, and continuous in* θ_2.

Remark 3.1. *Assumption 3.1 covers the realistic case where the variance vector or some of its components are zero, i.e., on the boundary of the parameter space.*

Assumption 3.2 is a modified version of the assumptions of Theorem 1 of Armstrong (2016a) and allows for unknown σ. Assumption 3.2 (ii) states that the Jacobian of the sample function with respect to the nuisance parameter vector is of full rank so that mean and variance parameters of the random coefficients are (locally) strongly identified. A similar high level rank condition for the Jacobian with respect to a full parameter vector (either $\theta = (\alpha, b', \sigma^{2\prime})'$ or $(\alpha, b', \sigma')'$) is used by Berry, Linton, and Pakes (2004), Freyberger (2015) and Ketz (2017). In contrast, Assumption 3.2 (ii) is with respect to the subvector $[b', \sigma^{2\prime}]'$, hence allows for weak identification of the price coefficient α.

Assumption 3.2 (iv) and (v) are stated in high level forms because, in general, the instrument vectors z_j's are not independently distributed and their dependence structure are left unspecified. In addition, they are correlated with the marginal costs MC_j's. However, Assumption 3.2 (iv) and (v) can be substituted by simpler conditions once the dependence structure of the instruments z_j's is characterized, or the marginal cost equation is specified explicitly.

Assumption 3.2 (vi) is an exogeneity-type condition that is needed for establishing the consistency of the covariance matrix estimates.

Assumption 3.2 (vii) guarantees the identification of θ_{02} and is used to establish the consistency of the restricted GMM estimator $\tilde{\theta}_2$.

As shown by Armstrong (2016a), the faster convergence of the markups to a vector of constants, η, than the $J^{1/2}$-rate, stated in Assumption 3.2 (iii), is essentially a weak identification condition which renders the BLP estimators inconsistent. This is intuitive because the characteristics of the rival products shift the price through the markup and have weak explanatory power if the markup is close to a constant.

On the other hand, the usual assumption of modeling the weak identification of the price coefficient α (see Stock and Wright (2000) and Kleibergen (2005) for analogous conditions) entails that

$$E[\hat{G}_{J,\alpha}(\theta_0)] = \bar{C}J^{-1/2} \tag{3.11}$$

for some vector of constants $\bar{C}$ as stated in Assumption A.4, where $\hat{G}_{J,\alpha}(\theta) = \frac{\partial \hat{m}_J(\theta)}{\partial \alpha}$.[15] The condition (3.11) is perhaps natural for modeling weak cost shifter instruments but its link with Assumption 3.2 (iii) may be somewhat elusive. It is therefore instructive to elaborate on the relationship between these two conditions. To this end, let $\bar{b} = [b', \alpha]'$, $x_j = [1, w_j']'$ and $\bar{X} = [X, -p]$ and define:

$$\mathit{\Gamma} = \begin{bmatrix} 1 & E[w_j'] & -E[MC_j] - \eta \\ E[w_j] & E[w_j w_j'] & -E[w_j MC_j] - \eta E[w_j] \\ 0 & 0 & 1 \end{bmatrix}^{-1} \mathcal{E}_k, \quad \mathcal{E}_k = \begin{bmatrix} 0 & 0_{1\times(k-1)} & 1 \\ 0_{(k-1)\times 1} & I_{k-1} & 0_{(k-1)\times 1} \\ 1 & 0_{1\times(k-1)} & 0 \end{bmatrix}.$$

Here $\mathcal{E}_k$ is the exchange matrix of dimension k where the counterdiagonal elements are 1's and all other elements are zero. Partition $\mathit{\Gamma} = [\mathit{\Gamma}_1, \mathit{\Gamma}_2]$, where

$$\mathit{\Gamma}_1 = \begin{bmatrix} (E[x_j x_j'])^{-1} E[x_j(MC_j + \eta)] \\ 1 \end{bmatrix}, \quad \mathit{\Gamma}_2 = \begin{bmatrix} (E[x_j x_j'])^{-1} \\ 0_{1\times k} \end{bmatrix} \mathcal{E}_k.$$

Following Armstrong (2016b), one can verify that

$$J^{-1} E[Z'(X, -MC - \eta \iota_J)] \mathit{\Gamma}_1 = 0. \tag{3.12}$$

[15] Andrews and Guggenberger (2015a) consider more general sequences of drifting parameters and establish the asymptotic similarity of GMM AR-type and GMM-CLR-type tests for full parameter vector. Although it is plausible that the subvector tests proposed here are also asymptotically similar under certain conditions, establishing such a property goes beyond the scope of this paper.

Rotate the parameter vector as $[\alpha, \phi']' = \mathit{\Gamma}^{-1}\bar{b}$ where

$$\phi = \begin{bmatrix} 0_{(k-1)\times 1} & I_{k-1} \\ 1 & 0_{1\times(k-1)} \end{bmatrix} \begin{bmatrix} E[x_j x_j'] & -E[x_j(MC_j + \eta)] \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix},$$

and rewrite the normalized sample moment function as

$$\begin{aligned} \hat{m}_J(\theta) &= J^{-1}Z'(\delta(\sigma) + p\alpha - Xb] = J^{-1}Z'(\delta(\sigma) - \bar{X}\mathit{\Gamma}\mathit{\Gamma}^{-1}\bar{b}] \\ &= J^{-1}Z'(\delta(\sigma) - \bar{X}\mathit{\Gamma}_1\alpha - \bar{X}\mathit{\Gamma}_2\phi] \equiv \hat{m}_J(\vartheta), \end{aligned} \tag{3.13}$$

where $\vartheta = (\alpha, \phi', \sigma^{2\prime})'$. It turns out that the Jacobian of the reparameterized moment function (3.15) with respect to α:

$$\hat{G}_{J,\alpha}(\vartheta) = \frac{\partial \hat{m}_J(\vartheta)}{\partial \alpha} = J^{-1}Z'\bar{X}\mathit{\Gamma}_1$$

is compatible with the drifting Jacobian condition in Assumption A.4. To see this, write

$$\begin{aligned} J^{1/2}E[\hat{G}_{J,\alpha}(\vartheta_0)] &= J^{-1/2}E[Z'\bar{X}\mathit{\Gamma}_1] \\ &= J^{-1/2}E[Z'\bar{X}\mathit{\Gamma}_1 - Z'(X, -MC - \eta\iota_J)\mathit{\Gamma}_1] + J^{-1/2}E[Z'(X, -MC - \eta\iota_J)\mathit{\Gamma}_1], \end{aligned}$$

where ι_J denotes a $J \times 1$ vector of ones. The second term is zero by (3.12). The weak identification condition in Assumption 3.2 (iii) along with moment bounds imply

$$J^{-1/2}E[Z'\bar{X}\mathit{\Gamma}_1 - Z'(X, -MC - \eta\iota_J)\mathit{\Gamma}_1] \to 0. \tag{3.14}$$

This shows that the reparameterized model is comptabile with the standard drifting Jacobian condition for weak identification: $E[\hat{G}_{J,\alpha}(\vartheta_0)] = \bar{C}J^{-1/2}$. In addition, it is easy to see that $J^{-1}E[Z'(X, -MC - \eta\iota_J)]\mathit{\Gamma}_2$ has full column rank k. These together with the fact that $J^{-1}Z'\frac{\partial\delta}{\partial\sigma^{2\prime}}$ converges in probability to a matrix of full rank, and ϑ and $\hat{m}_J(\vartheta)$ in place of θ and $\hat{m}_J(\theta)$ verify Assumption A.4.

The sample analog of the matrix Γ is given by $\hat{\Gamma} = [\hat{\Gamma}_1, \hat{\Gamma}_2]$ where

$$\hat{\Gamma}_1 = \begin{bmatrix} (X'X)^{-1}X'p \\ 1 \end{bmatrix}, \quad \hat{\Gamma}_2 = \begin{bmatrix} (J^{-1}X'X)^{-1}\mathcal{E}_k \\ 0_{1\times k} \end{bmatrix}.$$

The corresponding sample moment function is

$$\begin{aligned} \hat{m}_J(\theta) &= J^{-1}Z'(\delta(\sigma) + p\alpha - Xb] = J^{-1}Z'(\delta(\sigma) - \bar{X}\hat{\Gamma}_1\alpha - \bar{X}\hat{\Gamma}_2\phi^*] \\ &= J^{-1}Z'(\delta(\sigma) + M_X p\alpha - X(J^{-1}X'X)^{-1}\mathcal{E}_k\phi^*) \equiv \hat{m}_J(\vartheta^*), \end{aligned} \tag{3.15}$$

where $\vartheta^* = (\alpha, \phi^{*\prime}, \sigma^{2\prime})'$ and

$$\phi^* = \mathcal{E}_k \begin{bmatrix} J^{-1}X'X & -J^{-1}X'p \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix}. \tag{3.16}$$

Given an estimator $\tilde{b}^*$ of b, one can estimate the parameter subvector ϕ or ϕ^* under $H_0 : \alpha = \alpha_0$ by plugging $\tilde{b}^*$ in (3.16).

Ketz (2017) shows that, when $\sigma_\ell^2 = 0$ or $\sigma_\ell^2 \to 0$ (for any $\delta = \delta_t$, $t = 1, \dots, T$)

$$\frac{\partial \delta}{\partial \sigma_\ell^2} = \lim_{\sigma_\ell \to 0} \frac{1}{2} \frac{\partial^2 \delta}{\partial (\sigma_\ell)^2} \neq 0, \quad \ell = 1, \dots, k.$$

Thus, the Jacobian of the sample moment function with respect to an individual variance term is not degenerate. It does not, however, imply that the probability limit of the Jacobian with respect to the entire variance vector σ^2 is of full rank as maintained in Assumption 3.3.

The implementation of the robust test statistics requires a $J^{1/2}$-consistent estimator of the nuisance parameter vector θ_2. By definition, the $C(\alpha)$-type statistic uses a $J^{1/2}$-consistent estimator; the doubly robust C_α statistic in (2.13) is asymptotically pivotal with a plug-in $J^{1/2}$-consistent estimate. In addition, the AR_α statistic (2.15) and the corresponding CLR_α statistic (2.19) are also based on the orthogonalized sample moment function in (2.7), thus have $C(\alpha)$ statistic interpretation. Andrews (2002) and Ketz (2017) show that the GMM estimators of the boundary parameters are $J^{1/2}$-

consistent though not asymptotically normal in general. Accordingly, we estimate the nuisance parameters by the restricted CU-GMM under $H_0 : \alpha = \alpha_0$ and show their $J^{1/2}$-consistency so as to ensure Assumption A.6.

Since the model is not defined for negative values of σ^2, the doubly robust AR-type statistic cannot be the AR^0_α statistic computed using the one-step estimator defined by (2.16). Instead, we proceed with the AR_α statistic in (2.15). The orthogonalized moment function is given by

$$\begin{aligned}\bar{m}_J(\tilde{\theta}) &= \hat{m}_J(\tilde{\theta}) - \hat{G}_{J,2}(\tilde{\theta})'(\hat{G}_{J,2}(\tilde{\theta})'\hat{\Sigma}_J(\tilde{\theta})^{-1}\hat{G}_{J,2}(\tilde{\theta}))^{-1}\hat{G}_{J,2}(\tilde{\theta})'\hat{\Sigma}_J(\tilde{\theta})^{-1}\hat{m}_J(\tilde{\theta}),\\ &= J^{-1}\sum_{i=1}^{J} z_j'(\tilde{\delta}^*_j + p_j\alpha_0 - x_j'\tilde{b}^*),\end{aligned} \tag{3.17}$$

where

$$\tilde{\delta}^*_j = \delta_j(\tilde{\sigma}) - \left[0_{T\times k}, \frac{\partial \delta_j(\tilde{\sigma})}{\partial \sigma^{2\prime}}\right](\hat{G}_{J,2}(\tilde{\theta})'\hat{\Sigma}_J(\tilde{\theta})^{-1}\hat{G}_{J,2}(\tilde{\theta}))^{-1}\hat{G}_{J,2}(\tilde{\theta})'\hat{\Sigma}_J(\tilde{\theta})^{-1}\hat{m}_J(\tilde{\theta}) \tag{3.18}$$

and

$$\tilde{b}^* = \tilde{b} - [I_k, 0_{k\times k}]\,(\hat{G}_{J,2}(\tilde{\theta})'\hat{\Sigma}_J(\tilde{\theta})^{-1}\hat{G}_{J,2}(\tilde{\theta}))^{-1}\hat{G}_{J,2}(\tilde{\theta})'\hat{\Sigma}_J(\tilde{\theta})^{-1}\hat{m}_J(\tilde{\theta}). \tag{3.19}$$

From (3.18), it is clear that $\tilde{\delta}^*_j$ can be viewed as one-step estimator of $\delta_t(\sigma)$ which in turn is a nonlinear transformation of the model parameters and the data. In fact, one can proceed to show that $J^{1/2}[\tilde{\delta}^*_j - \delta_j(\sigma)]$ is asymptotically normally distributed.

Denote the AR-type statistic for $H_0 : \alpha = \alpha_0$ based on (3.17) corresponding to (2.15) by

$$AR_\alpha(\alpha_0) = J\,\bar{m}_J(\tilde{\theta})\tilde{\Sigma}_J^{-1}\bar{m}_J(\tilde{\theta}).$$

In the BLP context, contrary to the general GMM model of Subsection 2.1, we know that the instruments in Z_2 specifically are the ones that may lead to weak identification of α. Hence, in addition to the rank statistic defined in (2.23), we may also consider alternative CLR-type statistics based on simplified rank statistics obtained from the sample moment function given by

$$\hat{m}_J^M(\theta) = J^{-1}Z'M_X\xi, \tag{3.20}$$

where X is partialled out. These alternative statistics are detailed in Appendix C.

In the following proposition, we formally establish the results on the doubly robust statistics by first showing that the reparameterized sample moment function (3.15) satisfies the assumptions of the general GMM statistics of Section 2, and then appealing to the invariance properties of the test statistics to reparameterization. For general results about the invariance properties of various tests statistics in general estimating functions and GMM framework, see Dufour, Trognon, and Tuvaandorj (2017).

Proposition 3.1. *Let Assumption 3.1 and 3.2 hold. Under $H_0 : \alpha = \alpha_0$, the AR_α and C_α statistics based on (3.10) satisfy $AR_\alpha(\alpha_0) \xrightarrow{d} \chi^2_{l-k}$ and $C_\alpha(\alpha_0) \xrightarrow{d} \chi^2_1$ as $J \to \infty$. The asymptotic distribution of the CLR_α statistics based on the sample moment functions (3.10)and (3.20) statistic are as given in (2.21).*

The BLP model above resembles the linear IV model with a single endogenous right hand variable (the price term). In the latter model, the CLR statistic of Kleibergen (2005) reduces to the CLR statistic of Moreira (2003), which is known to have an optimal power property (see Andrews, Moreira, and Stock (2006) and Andrews and Guggenberger (2015a)). Thus it is plausible that the doubly robust CLR statistic in the BLP model may have similar optimal properties.

3.3 Robust statistics under many market asymptotics

The results of the previous subsection continue to hold in the many market scenario with T large. Here, weak identification may arise if the variation in the price induced by a change in the cost shifters is relatively small. We shall verify the conditions for the general GMM results of Subsection 2.1 on doubly robust tests under the following set of assumptions.

Assumption 3.3 (Jacobian and identifiability of model parameters)**.** *The matrix of IV's, Z, has full column rank with probability one: $P[\text{rank}(Z) = k + l] = 1$ with*

$l \geq k$. $E[T^{-1}Z'M_Xp] = \bar{C}T^{-1/2}$ for some $(k+l) \times 1$ vector $\bar{C}$. The matrices $E[z_t x_t']$ and $E\left[z_t \frac{\partial \delta_t(\sigma_0)}{\partial \sigma^{2\prime}}\right]$ have full rank. $E[z_t \xi_t(\alpha_0, \theta_2)] = 0$ uniquely at $\theta_2 = \theta_{02}$, and $E[z_t \xi_t(\alpha_0, \theta_2)\xi_t(\alpha_0, \theta_2)' z_t']$ is nonsingular for all $\theta_2 \in \Theta_2$.

Assumption 3.4 (Sampling regularity)**.** *The observed data $\{(p_t', s_t', z_t')'\}_{t=1}^T$ are independent with $E[\|z_t\|^{2+\varepsilon}] < M < \infty$ for some $M > 0$ and $\varepsilon > 0$, the random vector $(p_t', x_t')'$ lies in a compact set with probability one. The market shares satisfy $\varepsilon \leq s_{jt} \leq 1 - \varepsilon$ for some $\varepsilon > 0$ and for all $j = 1, \ldots, J$ and $t = 1, \ldots, T$ with $s_{0t} = 1 - \sum_{j=1}^J s_{jt}$.*

Remark 3.2. *Assumption 3.3 is the counterpart of Assumption 3.2 (ii) in many markets context, and guarantees the identification of the parameters other than α.*

Assumption 3.4 is used by Freyberger (2015) and Ketz (2017), and ensures the applications of LLN and CLT to the sample moment function and the estimators of covariance matrices.

In order establish the asymptotic properties, we use a reparameterization analogous to (3.15):

$$\begin{aligned} \hat{m}_T(\theta) &= T^{-1}Z'(\delta(\sigma) + p\alpha - Xb] = T^{-1}Z'(\delta(\sigma) - \bar{X}\hat{\Gamma}_1\alpha - \bar{X}\hat{\Gamma}_2\phi^*] \\ &= T^{-1}Z'(\delta(\sigma) + M_X p\alpha - X(T^{-1}X'X)^{-1}\mathcal{E}_k\phi^*) \equiv \hat{m}_T(\vartheta^*), \end{aligned} \tag{3.21}$$

where $\vartheta^* = (\alpha, \phi^{*\prime}, \sigma^{2\prime})'$ and $\phi^* = \mathcal{E}_k \begin{bmatrix} T^{-1}X'X & -T^{-1}X'p \end{bmatrix} [b, \alpha]'$. Let also

$$\hat{m}_T^M(\theta) = T^{-1}Z'M_X\xi. \tag{3.22}$$

In the following proposition, we establish the asymptotic distribution of the robust statistics:

Proposition 3.2. *Let Assumptions 3.1,3.3 and 3.4 hold. Under $H_0 : \alpha = \alpha_0$, the AR_α and C_α statistics based on (3.9) satisfy $AR_\alpha(\alpha_0) \xrightarrow{d} \chi^2_{l-k}$ and $C_\alpha(\alpha_0) \xrightarrow{d} \chi^2_1$ as $T \to \infty$. The asymptotic distribution of the CLR_α statistics based on the sample moment functions (3.9)and (3.22) statistic are as given in (2.21).*

4 Simulations

This Monte Carlo section explores the finite sample properties of our tests applied to the BLP model in a wide range of specifications with cost shifters and BLP instruments. The three boundary-robust statistics considered are: i) the AR_α statistic defined in (2.15), 2) the C_α statistic defined in (2.14), and 3) a CLR_α statistic based on the AR_α and C_α statistics, using the Robin and Smith (2000) rank statistic R_α defined in (2.23). The performance of two alternative robust CLR-type statistics are also presented in Table 3 Appendix C.

A complementary question to explore is how severely testing can be affected by assuming away boundary parameter problems. Within the same set of specifications, we do so by focusing on the following non-boundary-robust statistics: 1) an AR statistic defined in (2.11), 2) the subvector LM statistic of Kleibergen (2005) reproduced in (2.12), 3) a CLR statistics constructed with AR, LM and a non-boundary-robust version of the rank statistic R_α where $\hat{G}^{\perp}_{jn}(\theta)$ defined in (2.2) is used instead of $\hat{H}_n(\theta)$ defined in (2.3).

4.1 Simulations with cost shifters

Our main set of specifications with cost shifter instruments adapts the simple and flexible data generating process of Reynaert and Verboven (2014). A product j in market t is described by its price p_{jt} and exogenous characteristics x_{jt}. The characteristics include one constant term and either one or two additional random characteristics, according to the specification: $x_{jt} = (1, w^1_{jt})$ or $x_{jt} = (1, w^1_{jt}, w^2_{jt})$.

The endogenous price is a function of the product's exogenous characteristics and a vector of cost shifters $z_{jt,2}$ that do not influence consumer tastes: $p_{jt} = x'_{jt}\gamma_1 + z'_{jt,2}\gamma_2 + \omega_{jt}$. This price-setting equation reflects a perfect competition situation in which the price equals marginal cost. We include four cost shifter instruments: $z_{jt,2} = (z^1_{jt,2}, ..., z^4_{jt,2})$. Since AR-type statistics may entail a power loss when the model is overidentified compared to $C(\alpha)$ and CLR-type statistics, four cost shifters should make this difference more salient. The consumer's mean utility for product j at time t is $\delta_{jt} = x'_{jt}b - p_{jt}\alpha + \xi_{jt}$.

The characteristic w_{jt}^1, w_{jt}^2 and the cost shifters $z_{jt,2}'$ are all independently drawn from uniform distributions: $w_{jt}^1, w_{jt}^2 \sim U(1,2)$ and $z_{jt,2}^1, ..., z_{jt,2}^4 \sim U(0,1)$. The error terms of product demand and marginal cost are drawn from normal distributions and correlated:

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

We set $\rho = 0.3$ for specifications with weak endogeneity and $\rho = 0.8$ for strong endogeneity. Endogeneity levels of 0.8 may be high for differentiated product demand models, but the goal is to explore a wide parameter space.

The sample size is $T = 25$ markets and $J = 10$ products per markets, for a total of $N = T \times J = 250$ observations. Market size and sample size will vary in simulations using BLP instrument in the next subsection. The mean taste for product characteristics is $\alpha = 2$ and $b = (2,2)$ (or $b = (2,2,2)$). The value of α and the coefficient for the constant term have variance set to zero, which is assumed to be known in the estimation. The coefficients w_{jt}^1 (and w_{jt}^2) have i.i.d. normal distributions with variance $\Sigma_\beta = \sigma^2$ (or $\Sigma_\beta = \text{diag}\{\sigma^2, \sigma^2\}$). We use either $\sigma^2 = 0$ or $\sigma^2 = 1$, the former being more likely to generate boundary distortions. The price equation is set to $\gamma_1 = (0.7, 0.7)'$ (or $\gamma_1 = (0.7, 0.7, 0.7)'$) and $\gamma_2 = \kappa(1,1,1,1)'$, where κ determines the strength of the instruments. Three cases are considered: $\kappa = 3$ (as in Reynaert and Verboven (2014)) for strong instruments, $\kappa = 2\sqrt{C/(N(3l^2 + l))} = 0.055$ with a concentration parameter $C = 10$ for weak instruments (see Rothenberg (1984) or Stock and Yogo (2005)), and $\kappa = 0$ for non-informative instruments. Under the many-markets asymptotics, variances are computed at the market level, providing 25 i.i.d. observations. Appendix E provides extra details on the simulation design and the estimation algorithm.

Results. The rejection rates of boundary robust statistics at the true null H_0 presented in Table 1 conform well with the asymptotic results of Subsection 2.1. In all cases, the tests have rejection rates near 5%, irrespective of instrument strength or of σ^2 being at the boundary.

The motivation of the paper is reaffirmed by the results for non-boundary-robust

Table 1: Rejection rates at true value, boundary robust statistics (%)

	One random coefficient				Two random coefficients			
Boundary param.	Yes		No		Yes		No	
Endog. of price	Low	High	Low	High	Low	High	Low	High
Strong instruments								
AR_α	3.44	3.58	3.20	3.29	4.40	4.11	5.30	7.20
C_α	4.36	4.30	3.98	4.06	7.10	5.72	7.70	8.50
CLR_α	3.72	3.78	3.42	3.47	4.80	4.61	6.40	7.30
Weak instruments								
AR_α	3.24	2.52	2.94	2.40	2.80	2.40	2.00	2.10
C_α	3.96	4.62	4.10	4.58	4.10	3.10	2.90	3.60
CLR_α	3.18	2.80	3.08	2.64	2.30	2.40	1.60	1.80
Uninformative instruments								
AR_α	3.22	2.58	2.92	2.40	2.60	2.80	1.60	1.90
C_α	3.98	5.04	4.30	4.70	3.20	3.50	2.60	2.40
CLR_α	3.08	2.94	3.14	2.70	2.30	2.60	1.60	1.80

Note: Nominal sig. level: 0.05. Low endog.: $\rho = 0.3$; high endog.: $\rho = 0.8$. Boundary parameter: $\sigma^2 = 0$; no boundary parameter: $\sigma^2 = 1$. Strong instruments: $\kappa = 3$; weak instruments: $\kappa = 0.055$; uninformative instruments: $\kappa = 0$.

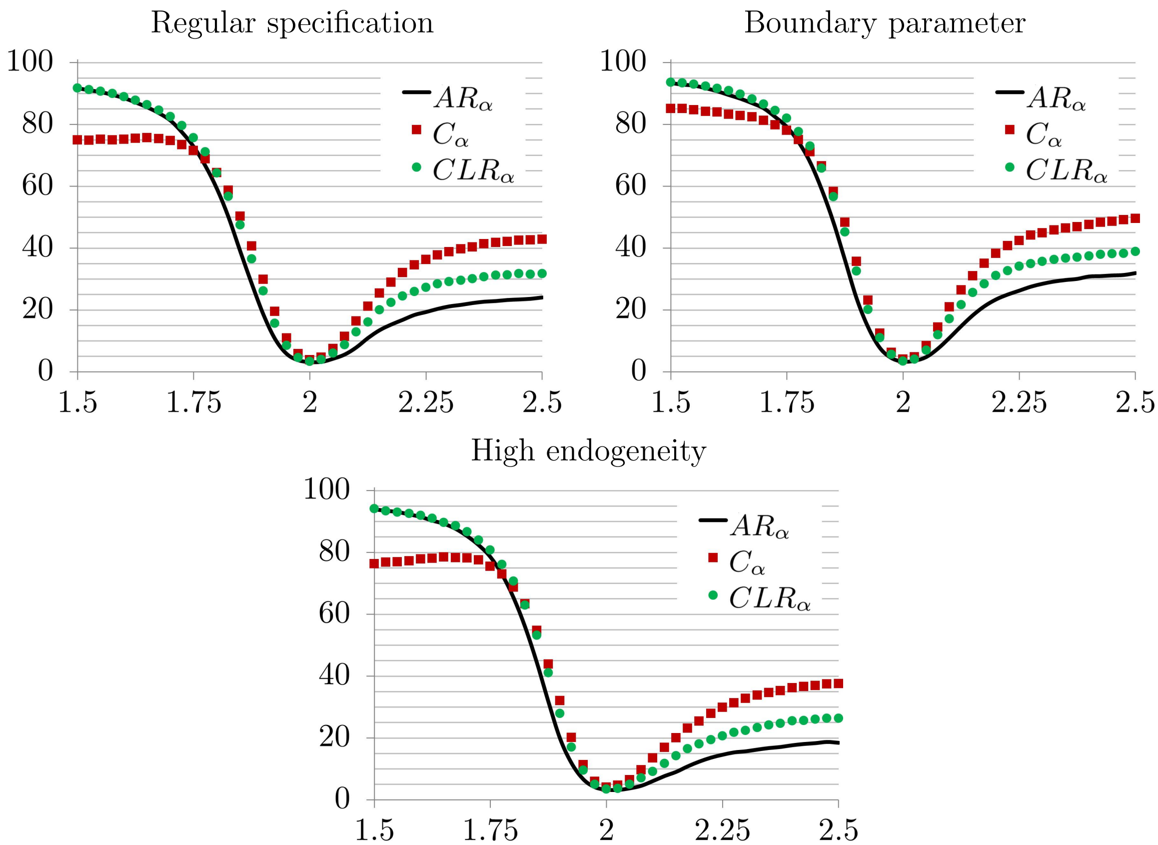

Figure 1: Power curves for strongly identified specifications

statistics shown in Table 4 of Appendix F. Over-rejection is especially serious for the LM and CLR statistics. With two random coefficients and strong instruments, even specifications with no true parameter at the boundary over-reject the true null hypothesis. This is understood once we realize that even with the true variance being $\Sigma_\beta = \mathrm{diag}\{\sigma^2, \sigma^2\} = \mathrm{diag}\{1, 1\}$, over 50% of the draws estimate at least one $\hat{\sigma}^2 = 0$.

Figure 1 shows the power of each test for cost shifters in various strongly-identified specifications ($\kappa = 3$). The regular specification has parameter values $\sigma^2 = 1$, $\rho = 0.3$, $\gamma_2 = 3$ and regular instruments. The specification with boundary parameter sets $\sigma^2 = 0$, and the specification with high endogeneity sets $\rho = 0.8$.

The asymmetric power of the tests with regular instruments reflects the nonlinearity of the BLP and its sensitivity to the exact parameter configurations. Different values of b would change the skewness of the confidence region. This is a clear advantage of confidence intervals obtained by inverting AR, C(α) or CLR-type statistics compared to standard two-sided tests which assume a symmetric sample distribution of the estimates. For regular instruments, no statistic clearly dominates all others in terms of power, but the CLR_α dominates the AR_α. Interestingly, the C_α has lower rejection rates than the other statistics for smaller values of α, but higher rejection rates for larger values.

4.2 Simulations with BLP instruments

We explore the finite sample properties of weak BLP instruments using a DGP similar to the one proposed by Armstrong (2016a). Products are still defined by their price p and two characteristics: a constant term and a random characteristic, $x_{jt} = (1, w_{jt})$, with $w_{jt} \sim U(0, 1)$. The true parameters are $\alpha = 1$ and $b = (3, 6)$. As before, the taste for the random characteristic is a random variable with variance $\sigma^2 = 9$, like in Armstrong (2016a)'s original DGP, or $\sigma^2 = 0$ to increase the risk of boundary distortions.

In contrast to the previous subsection, firms set their prices endogenously in a Nash-Bertrand market equilibrium. Prices are determined by a markup over marginal costs, described in (3.8), which provides identification power to BLP instruments. For a product j produced by firm f (which produces a set of goods $\mathcal{J}_f$) in market t, the two BLP instruments are the sum of the other products' characteristics in the same firm $\sum_{k\neq j, k\in\mathcal{J}_f} x_{k,t}$ and the sum of other products' characteristics in the same market $\sum_{k=1}^{J_t} x_{k,t} - x_{j,t}$.

Like in the previous subsection, we are interested in the properties of the tests for an overidentified model. However, in Armstrong (2016a)'s main specifications, all firms sell 10 products and markets have same sizes, either all 20, all 60 or all 100 products. Thus, no BLP instruments could be derived from the constant term, only 2 extra instruments from the random characteristics. Hence, we deviate from Armstrong (2016a) by using both the constant term and the random characteristic to generate a total of 4 BLP instruments. To do so, firms vary in size within markets, with approximately 1/3 selling 2 products, 1/3 selling 5 products and 1/3 selling 10 products.

Table 2: Simulations for BLP instruments, Boundary-robust statistics

Number. of markets	1		3				20			
Products per market	60	100	16 20 24	20 60 100	48 60 72	80 100 120	16 20 24	20 60 100	48 60 72	80 100 120
$\sigma^2 = 9$, null rejection rate										
AR_α	3.1	2.1	3.4	3.5	3.8	4.5	5.6	5.7	4.8	6.4
C_α	6.9	6.7	6.5	6.4	6.6	5.8	4.8	6.7	5.4	4.9
CLR_α	3.6	3.1	3.6	3.6	3.7	4.7	5.9	6.1	5.1	6.1
$\sigma^2 = 9$, power of test $\alpha = 0$										
AR_α	3.6	2.7	63.0	42.5	18.2	12.9	100.0	100.0	98.7	85.0
C_α	8.8	7.8	42.9	40.0	20.2	13.4	60.6	88.7	58.7	55.4
CLR_α	4.1	3.5	63.9	43.4	18.4	13.5	100.0	100.0	98.7	85.7
$\sigma^2 = 0$ (boundary parameter), null rejection rate										
AR_α	3.4	3.0	2.7	3.5	4.2	4.4	4.1	4.4	5.7	4.7
C_α	7.0	6.5	7.8	5.5	5.3	5.6	5.3	3.4	5.9	4.9
CLR_α	3.9	3.4	3.2	3.6	4.1	4.2	4.5	4.4	5.5	4.8
$\sigma^2 = 0$ (boundary parameter), power of test $\alpha = 0$										
AR_α	5.2	3.7	75.2	56.7	28.5	16.8	100.0	100.0	99.7	94.7
C_α	9.7	7.0	48.4	53.3	23.5	16.9	69.0	97.8	61.5	54.2
CLR_α	5.7	4.0	75.4	58.7	29.1	17.1	100.0	100.0	99.7	94.5

Note: Nominal sig. level: 0.05. Specifications with multiple markets contain approx. 1/3 of markets of each different size. All specifications have approx. 1/3 of firms of size 2, 1/3 of size 5 and 1/3 of size 10.

Simulated datasets contain either 1, 3 or 20 markets. Markets have on average 20, 60 or 100 products and, in specification with many markets, approximately 1/3 are 20% larger than average and 1/3 are 20% smaller. Following Armstrong (2016a), some cases allow for more variation, with 1/3 of markets having 20 products, 1/3 having 60 and 1/3 having 100 products. In specifications with a single market, since $\sum_{k=1}^{J_t} x_{k,t}$ is constant, it is replaced by $\sum_{k\in\mathcal{J}_f} x_{k,t}^2$ to provide the same number of instruments.[16] An additional cost shifter is drawn to further influence prices, but it is excluded from the estimation to maintain weak identification from BLP instruments. The structural errors are $\xi_i = \upsilon_{1,i}+\upsilon_{3,i}-1$ and $\omega_i = \upsilon_{1,i}+\upsilon_{2,i}-1$ where $\upsilon_1, \upsilon_2, \upsilon_3$ are independent $U(0,1)$ random variables. Appendix E provides extra details on the estimation procedure.

Results. Since the strength of BLP instruments depends on market power, they should be weak in larger markets. Table 2 confirms this intuition, especially in the single-market case where BLP instruments are completely uninformative for testing $H_0 : \alpha = 0$. But the weakness of the BLP instruments does not compromise validity of the doubly robust statistics, which show rejection rates close to 5% when testing the true null hypothesis $H_0 : \alpha = 1$. This contrasts with Armstrong (2016a) who found that when using BLP instruments, two-sided tests at nominal levels of 5% rejected the true null in corresponding DGP at rates over 15% when the number of products is large relative to the number of markets. Like for cost shifter instruments, the C_α statistic shows lower power for testing a small value of α, while the CLR_{alpha} shows good overall performances.

Table 5 of Appendix F confirms the wrong asymptotic size of the non-boundary-robust tests in the same set of simulations. As expected, the rejection rates are higher when the true σ^2 is at the boundary of the parameter space. Interestingly, the LM statistic over-rejects the true null hypothesis especially in smaller samples, while the other non-boundary-robust statistics slightly over-reject the true null in larger samples.

[16] Contrary to Armstrong (2016a), but following Berry, Levinsohn, and Pakes (1995), the product itself is excluded from the sums of characteristics that form the BLP instruments, which makes a difference in the estimation when taking higher power of the BLP instrument.

5 Conclusion

This paper proposes robust statistics that are implemented with $n^{1/2}$-consistent auxiliary estimator of the nuisance parameters in moment condition models. They are asymptotically pivotal under the null hypothesis, even if the parameters of interest are weakly identified and true parameters are close to the boundary of the parameter space. Our doubly robust statistics include a generalization of the GMM AR statistic of Stock and Wright (2000), a $C(\alpha)$-type statistic and a CLR-type statistic which extends the identification-robust GMM LM and CLR-type statistics of Kleibergen (2005) and Andrews and Guggenberger (2015a).

We apply our tests to the differentiated products demand models of Berry, Levinsohn, and Pakes (1995), which is subject to weak identification because of firms' limited market power or marginally relevant cost instruments, and to boundary parameter problems because of the heterogeneous tastes for products characteristics. In simulations, all tests are found to have correct level using as instruments standard cost shifters or BLP instruments in all parameter configurations. In contrast, similar non-boundary-robust statistics often show important size distortions.

Following Reynaert and Verboven (2014), seeing how optimal instruments may substantially improve estimation accuracy in the BLP context, making them robust to weak identification is a logical avenue for future work. Beyond the BLP, our test statistics can be adapted to conduct robust inference in other GMM applications where both boundary parameters and weak identification can be present. For example, using the methods developed in this paper, one may derive pivotal statistics for ARMA/GARCH-type models that are robust to weak identification, persistence and boundary parameter problems at the same time. Another important problem is the development of an identification-robust inference methods on structural coefficients after selecting the instruments by selection or regularization methods (see Chernozhukov, Hansen, and Spindler (2015)). We leave these topics for future research.

References

Andrews, D., and P. Guggenberger (2017): "Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models," *Econometric Theory*, 33(5), 1046–1080.

Andrews, D. W., and P. Guggenberger (2015a): "Identification-and Singularity-Robust Inference for Moment Condition Models," Discussion paper, Cowles Foundation Discussion Paper No. 1978, Yale University.

——— (2015b): "Supplemental material to "Identification-and Singularity-Robust Inference for Moment Condition Models"," Discussion paper, Cowles Foundation Discussion Paper No. 1978S, Yale University.

Andrews, D. W. K. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67(6), 1341–1383.

Andrews, D. W. K. (2001): "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–733.

Andrews, D. W. K. (2002): "Generalized Method of Moments Estimation when a Parameter is on a Boundary," *Journal of Business & Economic Statistics*, 20(4), 530–544.

——— (2017): "Identification-Robust Subvector Inference," Discussion paper, Cowles Foundation Discussion Paper No. 2015, Yale University.

Andrews, D. W. K., and X. Cheng (2014): "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure," *Econometric Theory*, 30(02), 287–333.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74(3), 715–752.

Andrews, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, 84(6), 2155–2182.

Andrews, I., and A. Mikusheva (2016): "Conditional Inference with a Functional Nuisance Parameter," *Econometrica*, 84(4), 1571–1612.

Armstrong, T. B. (2016a): "Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply," *Econometrica*, 84(5), 1961–1980.

——— (2016b): "Supplement to "Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply"," *Econometrica*.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S., O. B. Linton, and A. Pakes (2004): "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *The Review of Economic Studies*, 71(3), 613–654.

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

Chant, D. (1974): "On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions," *Biometrika*, 61(2), 291–298.

Chaudhuri, S., and E. Zivot (2011): "A New Method of Projection-based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics*, 164(2), 239–251.

Chernozhukov, V., C. Hansen, and M. Spindler (2015): "Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach," *Annual Review of Economics*, 7(1), 649–688.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012): "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, 80(5), 2231–2267.

Dubois, P., R. Griffith, and M. O'Connell (2018): "The Effects of Banning Advertising in Junk Food Markets," *The Review of Economic Studies*, 85(1), 396–436.

Dufour, J.-M. (2003): "Identification, Weak Instruments and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36(4), 767–808.

Dufour, J.-M., A. Trognon, and P. Tuvaandorj (2016): "Generalized $C(\alpha)$ Tests in Estimating Functions with Serial Dependence," in *Time Series Methods and Applications: the McLeod Festschrift*, ed. by W. K. Li, D. Stanford, and H. Yu, pp. 151–178. Springer, Berlin and New York.

Dufour, J.-M., A. Trognon, and P. Tuvaandorj (2017): "Invariant Tests based on M-estimators, Estimating Functions, and the Generalized Method of Moments," *Econometric Reviews*, 36(1-3), 182–204.

Dufour, J.-M., and P. Tuvaandorj (2018): "Identification-robust asymptotic and bootstrap $C(\alpha)$ tests for minimum distance and likelihood models," Discussion paper, McGill University, ENSAI.

Eizenberg, A. (2014): "Upstream Innovation and Product Variety in the U.S. Home PC Market," *Review of Economic Studies*, 81, 1003–1045.

Freyberger, J. (2015): "Asymptotic Theory for Differentiated Products Demand Models with Many Markets," *Journal of Econometrics*, 185(1), 162–181.

Guggenberger, P., J. J. S. Ramalho, and R. J. Smith (2012): "GEL Statistics under Weak Identification," *Journal of Econometrics*, 170(2), 331–349.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

Ketz, P. (2017): "A Simple Solution to Invalid Inference in the Random Coefficients Logit Model," Discussion paper, Paris School of Economics.

——— (2018): "Subvector Inference When the True Parameter Vector is near the Boundary," Discussion paper, Paris School of Economics.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming that They are Identified," *Econometrica*, 73, 1103–1124.

——— (2007): "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics," *Journal of Econometrics*, 139(1), 181–216.

Knittel, C. R., and K. Metaxoglou (2014): "Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective," *The Review of Economics and Statistics*, 96(1), 34–59.

Miller, N. H., and M. C. Weinberg (2017): "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica*, 85(6), 1763–1791.

Moran, P. A. P. (1973): "Asymptotic Properties of Homogeneity Tests," *Biometrika*, 60(1), 79–85.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 36, pp. 2111–2245. Elsevier Science.

Newey, W. K., and F. Windmeijer (2009): "Generalized Method of Moments with Many Weak Moment Conditions," *Econometrica*, 77(3), 687–719.

Neyman, J. (1959): "Optimal Asymptotic Tests of Composite Statistical hypotheses," in *Probability and Statistics, the Harald Cramér Volume*, ed. by U. Grenander, pp. 213–234, Uppsala, Sweden. Almqvist and Wiksell.

Nurski, L., and F. Verboven (2016): "Exclusive Dealing as a Barrier to Entry? Evidence from Automobiles," *Review of Economic Studies*, 83, 1156–1188.

Reynaert, M., and F. Verboven (2014): "Improving the Performance of Random Coefficients Demand Models: the Role of Optimal Instruments," *Journal of Econometrics*, 179(1), 83–98.

Robin, J.-M., and R. Smith (2000): "Tests of Rank," *Econometric Theory*, 16(2), 151–175.

Rothenberg, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, ed. by Z. G. M.D. Intriligator, vol. 2, chap. 15, pp. 881–935. Elsevier Science Publishers.

Shcherbakov, O. (2016): "Measuring Consumer Switching Costs," *RAND Journal of Economics*, 47(2), 366–393.

Skrainka, B. S., and K. L. Judd (2011): "High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation," Discussion paper, SSRN.

Smith, R. (2007): *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*vol. 3, chap. Weak Instruments and Empirical Likelihood: A Discussion of the Papers by D. W. K. Andrews, J. H. Stock, and Y. Kitamura, pp. 238–260. Cambridge University Press.

Stock, J., and M. Yogo (2005): *Identification and Inference for Econometric Models*chap. Testing for Weak Instruments in Linear IV Regression, pp. 80–108. Cambridge University Press, New York.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1097–1126.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20(4), 518–529.

A Regularity Conditions

Assumption A.1 (Jacobian of the sample moment function)**.** *The sample moment function $\hat{m}_n(\theta)$ has continuous left/right partial derivatives of order 1 with respect to θ, denoted*

$$\hat{G}_n(\theta) = \frac{\partial \hat{m}_n(\theta)}{\partial \theta'},$$

for all n on Θ (with probability one).[17] *The last d_2 columns of the $L \times d$ Jacobian matrix $\hat{G}_n(\theta) = \left[\hat{G}_{n,1}(\theta), \hat{G}_{n,2}(\theta)\right]$ partitioned conformably with $\theta = (\theta_1', \theta_2')'$, satisfies*

$$\sup_{\theta_2 \in \Theta_2 : \|\theta_2 - \theta_{02}\| \leq \varepsilon_n} \|\hat{G}_{n,2}(\theta_{01}, \theta_2) - H_2(\theta_{01}, \theta_2)\| \xrightarrow{p} 0,$$

for all $\varepsilon_n \to 0$, and a nonrandom matrix function $H_2(\theta)$.

Assumption A.2 (Derivative of the Jacobian)**.** *The $L \times 1$ column vectors, $\hat{G}_{in}(\theta), i = 1, \ldots, d_1$, of the Jacobian*

$$\hat{G}_n(\theta) = \frac{\partial \hat{m}_n(\theta)}{\partial \theta_1'} = [\hat{G}_{1n}(\theta), \ldots, \hat{G}_{d_1 n}(\theta)],$$

have continuous left/right partial derivatives of order 1 with respect to θ_2, denoted

$$\hat{D}_{in}(\theta) = \frac{\partial \hat{G}_{in}(\theta)}{\partial \theta_2'},$$

for all n, on Θ_2. Furthermore, for every $i = 1, \ldots, d_1$ and all $\varepsilon_n \to 0$

$$\sup_{\sup_{\theta_2 \in \Theta_2 : \|\theta_2 - \theta_{02}\| \leq \varepsilon_n}} \|\hat{D}_{in}(\theta_{01}, \theta_2) - D_i(\theta_{01}, \theta_2)\| \xrightarrow{p} 0,$$

for some $L \times d_2$ matrix function $D_i(\theta)$.

Assumption A.3 (Asymptotic normality of the sample moment function and the Jacobian)**.** *The random vector $n^{1/2}\left[\hat{m}_n(\theta_0)', \mathrm{vec}\left(\hat{G}_{n,1}(\theta_0) - E[\hat{G}_{n,1}(\theta_0)]\right)'\right]'$ obeys the*

[17] See Andrews (1999) for the definition of left/right partial derivative.

CLT:

$$n^{1/2}\begin{bmatrix}\hat{m}_n(\theta_0)\\ \mathrm{vec}\Big(\hat{G}_{n,1}(\theta_0)-E[\hat{G}_{n,1}(\theta_0)]\Big)\end{bmatrix}\xrightarrow{d}\begin{bmatrix}m_\infty(\theta_0)\\ G_{\infty,1}(\theta_0)\end{bmatrix},\quad\begin{bmatrix}m_\infty(\theta_0)\\ G_{\infty,1}(\theta_0)\end{bmatrix}\sim N[0,V(\theta_0)],$$

where the asymptotic covariance matrix partitioned conformably to $[m_\infty(\theta_0)', G_{\infty,1}(\theta_0)']'$,

$$V(\theta)=\begin{bmatrix}\Sigma(\theta) & \Sigma_{mG_1}(\theta)\\ \Sigma_{G_1m}(\theta) & \Sigma_{G_1G_1}(\theta)\end{bmatrix}, \tag{A.1}$$

is continuous at θ_0*,* $V(\theta_0)$ *is positive definite, and* $\Sigma_{mG_1}(\theta)=[C_1(\theta)',\ldots,C_{d_1}(\theta)']'$ *with* $C_i(\theta_0)\equiv C_i$*.*

Assumption A.4 (Identification status of parameter subvectors)**.** $E[\hat{G}_{n,1}(\theta_0)]=\bar{C}n^{-1/2}$ *where* $\bar{C}$ *is a* $L\times d_1$ *fixed matrix function of* θ_0*, and* $\mathrm{rank}[H_2(\theta_0)]=d_2$ *where* $\hat{G}_{n,2}(\theta_0)\xrightarrow{p} H_2(\theta_0)$*.*

Assumption A.5 (Consistent weighting matrix)**.** *The weighting matrix* W_n *is a consistent estimator of a positive definite matrix* W*:*

$$W_n\xrightarrow{p} W.$$

Assumption A.6 (Root-n consistent nuisance parameter estimate)**.** $\tilde{\theta}_2^*$ *is a consistent estimator of* θ_{02} *such that* $n^{1/2}(\tilde{\theta}_2^*-\theta_{02})$ *is asymptotically bounded in probability as* $n\to\infty$*, i.e.,*

$$n^{1/2}\big(\tilde{\theta}_2^*-\theta_{02}\big)=O_p(1).$$

Assumption A.7 (Consistent covariance matrix estimator)**.** *The estimator defined by*

$$\hat{V}_n(\theta)=\begin{bmatrix}\hat{\Sigma}_n(\theta) & \hat{\Sigma}_{n,mG_1}(\theta)\\ \hat{\Sigma}_{n,G_1m}(\theta) & \hat{\Sigma}_{n,G_1G_1}(\theta)\end{bmatrix}\quad \textit{with}\quad \hat{\Sigma}_{n,mG_1}(\theta)=[\hat{C}_{1n}(\theta)',\ldots,\hat{C}_{d_1n}(\theta)']' \tag{A.2}$$

is nonsingular for every $(\theta_{01}',\theta_2')'$*, where* $\theta_2\in\Theta_2, \|\theta_2-\theta_{02}\|\le\varepsilon_n$ *and* $\varepsilon_n\to 0$*, and satisfies* $\hat{V}_n(\tilde{\theta}^*)\xrightarrow{p} V(\theta_0)$*, where* $\tilde{\theta}^*=(\theta_{01}',\tilde{\theta}_2^*)'$ *with* $\tilde{\theta}_2^*$ *defined in Assumption A.6.*

By standard argument, one can show that the following assumption together with Assumption A.6 imply Assumption A.7.

Assumption A.8 (Local uniform convergence)**.** *$\hat{V}_n(\theta)$ is nonsingular for $(\theta_{01}', \theta_2')'$, for every $(\theta_{01}', \theta_2')'$, where $\theta_2 \in \Theta_2$, $\|\theta_2 - \theta_{02}\| \leq \varepsilon_n$ and $\varepsilon_n \to 0$, and satisfies* $\sup_{\theta_2 \in \Theta_2, \|\theta_2 - \theta_{02}\| \leq \varepsilon_n} \|\hat{V}_n(\theta_{01}, \theta_2) - V(\theta_{01}, \theta_2)\| \xrightarrow{p} 0$.

B Proofs

Proof of Proposition 2.1. We first show that

$$n^{1/2}\hat{H}_{jn}(\tilde{\theta}^*) = n^{1/2}\hat{H}_{jn}(\theta_0) + o_p(1), \quad j = 1, \ldots, d_1,$$
$$n^{1/2}\bar{m}_n(\tilde{\theta}^*) = n^{1/2}\bar{m}_n(\theta_0) + o_p(1).$$

By the triangle inequality, for any $i = 1, \ldots, d$

$$\|\hat{D}_{in}(\tilde{\theta}^*) - D_i(\theta_0)\| \leq \|\hat{D}_{in}(\tilde{\theta}^*) - D_i(\tilde{\theta}^*)\| + \|D_i(\tilde{\theta}^*) - D_i(\theta_0)\|.$$

Since $\tilde{\theta}^* \xrightarrow{p} \theta_0$, there exists a sequence $\varepsilon_n \to 0$ such that $P[\|\tilde{\theta}^* - \theta_0\| \leq \varepsilon_n] \to 1$. For any $\epsilon > 0$,

$$\begin{aligned}
&P[\|\hat{D}_{in}(\tilde{\theta}^*) - D_i(\tilde{\theta}^*)\| \leq \epsilon] \\
&\geq P[\|\tilde{\theta}^* - \theta_0\| \leq \varepsilon_n, \|\hat{D}_{in}(\tilde{\theta}^*) - D_i(\tilde{\theta}^*)\| \leq \epsilon] \\
&\geq P\left[\|\tilde{\theta}^* - \theta_0\| \leq \varepsilon_n, \sup_{\theta_2 \in \Theta_2, \|\theta_2 - \theta_{02}\| \leq \varepsilon_n} \|\hat{D}_{in}(\theta_{01}, \theta_2) - D_i(\theta_{01}, \theta_2)\| \leq \epsilon\right] \\
&\geq 1 - P[\|\tilde{\theta}^* - \theta_0\| > \varepsilon_n] - P\left[\sup_{\theta_2 \in \Theta_2, \|\theta_2 - \theta_{02}\| \leq \varepsilon_n} \|\hat{D}_{in}(\theta_{01}, \theta_2) - D_i(\theta_{01}, \theta_2)\| > \epsilon\right] \\
&\longrightarrow 1,
\end{aligned}$$

where the last convergence holds by Assumption A.2 and A.6. Thus, $\|\hat{D}_{in}(\tilde{\theta}^*) - D_i(\tilde{\theta}^*)\| \xrightarrow{p} 0$. By the continuity of $D_i(\theta)$ at θ_0 (Assumption A.2), $\|D_i(\tilde{\theta}) - D_i(\theta_0)\| \xrightarrow{p}$

0 hence

$$\hat{D}_{in}(\tilde{\theta}^*) \xrightarrow{p} D_i(\theta_0). \tag{B.1}$$

Using Assumptions A.1 and A.6,

$$\tilde{H}_{n,2} \xrightarrow{p} H_2(\theta_0). \tag{B.2}$$

By the mean value expansion, we have under Assumptions A.1, A.2 and A.6,

$$n^{1/2}\hat{m}_n(\tilde{\theta}^*) = n^{1/2}\hat{m}_n(\theta_0) + \hat{G}_{n,2}(\bar{\theta})n^{1/2}(\tilde{\theta}_2^* - \theta_{02}), \tag{B.3}$$

$$n^{1/2}\hat{G}_{in}(\tilde{\theta}^*) = n^{1/2}\hat{G}_{in}(\theta_0) + \hat{D}_{in}(\underline{\theta})n^{1/2}(\tilde{\theta}_2^* - \theta_{02}), \quad i = 1, \ldots, d_1, \tag{B.4}$$

where $\bar{\theta}$ and $\underline{\theta}$ are points on the segment joining θ_0 and $\tilde{\theta}^*$. Using Assumptions A.1 and A.6, we have $\hat{G}_{n,2}(\bar{\theta}) \xrightarrow{p} H_2(\theta_0)$. Substituting (B.3) in $\bar{m}_n(\tilde{\theta}^*)$, using (B.2), and Assumption A.6, we have

$$\begin{aligned}
n^{1/2}\bar{m}_n(\tilde{\theta}^*) &= \left(I_L - \hat{G}_{n,2}(\tilde{\theta}^*)(\hat{G}_{n,2}(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\hat{G}_{n,2}(\tilde{\theta}^*))^{-1}\hat{G}_{n,2}(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\right) n^{1/2}\hat{m}_n(\tilde{\theta}^*) \\
&= \left(I_L - \hat{G}_{n,2}(\tilde{\theta}^*)(\hat{G}_{n,2}(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\hat{G}_{n,2}(\tilde{\theta}^*))^{-1}\hat{G}_{n,2}(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\right) \\
&\quad \left(n^{1/2}\hat{m}_n(\theta_0) + \hat{G}_{n,2}(\bar{\theta})n^{1/2}(\tilde{\theta}_2^* - \theta_{02})\right) \\
&= \left(I_L - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}\right) n^{1/2}\hat{m}_n(\theta_0) + o_p(1) \\
&\xrightarrow{d} \left(I_L - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}\right) m_\infty(\theta_0) \\
&\equiv \bar{m}_\infty(\theta_0),
\end{aligned} \tag{B.5}$$

where the convergence holds by Assumption A.3 and Slutsky's lemma. Similarly, substituting (B.3) and (B.4) in $\bar{G}_n(\tilde{\theta}^*)$, using (B.1)and (B.2), and Assumption A.6, we have

$$
\begin{aligned}
& n^{1/2}\bar{G}_{in}(\tilde{\theta}^*) \qquad\qquad \text{(B.6)} \\
& = n^{1/2}\hat{G}_{jn}(\tilde{\theta}^*) - \hat{D}_{in}(\tilde{\theta}^*)(\hat{G}_{n,2}(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\hat{G}_{n,2}(\tilde{\theta}^*))^{-1}\hat{G}_{n,2}(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}n^{1/2}\hat{m}_n(\tilde{\theta}^*) \\
& = n^{1/2}\hat{G}_{in}(\theta_0) + D_i(\theta_0)n^{1/2}(\tilde{\theta}_2^* - \theta_{02}) \\
& \quad - D_i(\theta_0)(H_2(\theta_0)'\Sigma(\theta_0)^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)'\Sigma^{-1}\left(n^{1/2}\hat{m}_n(\theta_0) + H_2(\theta_0)n^{1/2}(\tilde{\theta}_2^* - \theta_{02}) + o_p(1)\right) \\
& = n^{1/2}\hat{G}_{in}(\theta_0) - D_i(\theta_0)(H_2(\theta_0)'\Sigma(\theta_0)^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)'\Sigma^{-1}n^{1/2}\hat{m}_n(\theta_0) + o_p(1) \\
& \xrightarrow{d} G_{i\infty}(\theta_0) - D_i(\theta_0)(H_2(\theta_0)'\Sigma(\theta_0)^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)'\Sigma^{-1}m_\infty(\theta_0) \\
& \equiv \bar{G}_{i\infty}(\theta_0), \qquad\qquad \text{(B.7)}
\end{aligned}
$$

where the convergence holds again by Assumption A.3 and Slutsky's lemma. Since $\hat{C}_{in}(\tilde{\theta}^*) \xrightarrow{p} C_i(\theta_0), i = 1, \ldots, d_1$, and $\hat{\Sigma}_n(\tilde{\theta}^*) \xrightarrow{p} \Sigma(\theta_0)$ by Assumption A.7, using (B.5) and (B.7) in the definition of $\tilde{H}_{n,1}$, we have for $i = 1, \ldots, d_1$

$$
\begin{aligned}
n^{1/2}\hat{H}_{in}(\tilde{\theta}^*) &= n^{1/2}\bar{G}_{in}(\tilde{\theta}^*) - \hat{C}_i(\tilde{\theta}^*)\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}n^{1/2}\bar{m}_n(\tilde{\theta}^*) \\
&= n^{1/2}\bar{G}_{in}(\theta_0) - C_i(\theta_0)\Sigma^{-1}n^{1/2}\bar{m}_n(\theta_0) + o_p(1) \\
&\xrightarrow{d} \bar{G}_{i\infty}(\theta_0) - C_i(\theta_0)\Sigma(\theta_0)^{-1}\bar{m}_\infty(\theta_0) \\
&\equiv H_{i\infty}(\theta_0), \qquad\qquad \text{(B.8)}
\end{aligned}
$$

where the convergence uses the CMT and Slutsky's lemma. Stack the vectors so obtained into

$$H_{\infty,1}(\theta_0) = [H_{1\infty}(\theta_0), \ldots, H_{d_1\infty}(\theta_0)].$$

Since $W_n \xrightarrow{p} W$ by Assumption A.5, $\tilde{\Sigma} \xrightarrow{p} \Sigma$ and $\tilde{H}_{n,2} \xrightarrow{p} H_2(\theta_0)$, invoking CMT gives

$$W_n\tilde{\Sigma}W_n\tilde{H}_{n,2}(\tilde{H}_{n,2}'W_n\tilde{\Sigma}W_n\tilde{H}_{n,2})^{-1}\tilde{H}_{n,2}'W_n \xrightarrow{p} W\Sigma W H_2(\theta_0)(H_2(\theta_0)'W\Sigma W H_2(\theta_0))^{-1}H_2(\theta_0)'W.$$

Using Slutsky's lemma, we then have

$$\begin{aligned} n^{1/2}\tilde{B}_n &= n^{1/2}\tilde{H}'_{n,1}W_n\tilde{\Sigma}_nW_n - \tilde{H}'_{n,1}W_n\tilde{\Sigma}_nW_n\tilde{H}_{n,2}(\tilde{H}'_{n,2}W_n\tilde{\Sigma}_nW_n\tilde{H}_{n,2})^{-1}\tilde{H}'_{n,2}W_n\tilde{\Sigma}_nW_n \\ &\xrightarrow{d} H_{\infty,1}(\theta_0)'W\Sigma W - H_{\infty,1}(\theta_0)'W\Sigma WH_2(\theta_0)(H_2(\theta_0)'W\Sigma WH_2(\theta_0))^{-1}H_2(\theta_0)'W\Sigma W \\ &\equiv B_\infty(\theta_0). \end{aligned} \tag{B.9}$$

Since $B_\infty(\theta_0)H_2(\theta_0) = 0$, we obtain

$$\begin{aligned} n^{1/2}\tilde{B}_n n^{1/2}\hat{m}_n(\tilde{\theta}^*) &= B_\infty(\theta_0)\left[n^{1/2}\hat{m}_n(\theta_0) + H_2(\theta_0)n^{1/2}(\tilde{\theta}_2^* - \theta_{02}) + o_p(1)\right] \\ &= B_\infty(\theta_0)m_\infty(\theta_0) + o_p(1) \\ &= B_\infty(\theta_0)\bar{m}_\infty(\theta_0) + o_p(1). \end{aligned} \tag{B.10}$$

Write

$$B_\infty(\theta_0)\bar{m}_\infty(\theta_0) = H_{\infty,1}(\theta_0)'(W\Sigma W)^{1/2}M_{(W\Sigma W)^{1/2}H_2(\theta_0)}(W\Sigma W)^{1/2}\bar{m}_\infty(\theta_0).$$

Since

$$\begin{aligned} E[\bar{G}_{i\infty}(\theta_0)\bar{m}_\infty(\theta_0)'] &= C_i(\theta_0)\Sigma^{-1}\left[\Sigma - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)'\right], \\ V[\bar{m}_\infty(\theta_0)] &= \Sigma - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)', \end{aligned}$$

it follows that

$$E[H_{i\infty}(\theta_0)\bar{m}_\infty(\theta_0)'] = 0.$$

Hence, $H_{i\infty}(\theta_0), i = 1, \ldots, d_1$, and $(W\Sigma W)^{1/2}M_{(W\Sigma W)^{1/2}H_2(\theta_0)}(W\Sigma W)^{1/2}\bar{m}_\infty(\theta_0)$ are independent Gaussian random vectors. Using $B_\infty(\theta_0)H_2(\theta_0) = 0$, we have conditional on $H_{\infty,1}(\theta_0)$

$$B_\infty(\theta_0)\bar{m}_\infty(\theta_0) \sim N[0_{d_1\times 1}, B_\infty(\theta_0)\Sigma B_\infty(\theta_0)']. \tag{B.11}$$

Therefore, conditional on $H_{\infty,1}(\theta_0)$

$$
\begin{aligned}
C_\alpha(\theta_{01}) &= n^{1/2}\hat{m}_n(\tilde{\theta}^*)' n^{1/2}\tilde{B}_n' \left[n^{1/2}\tilde{B}_n\tilde{\Sigma}_n n^{1/2}\tilde{B}_n' \right]^{-1} n^{1/2}\tilde{B}_n n^{1/2}\hat{m}_n(\tilde{\theta}^*) \\
&\xrightarrow{d} \bar{m}_\infty(\theta_0)' B_\infty(\theta_0)' \left[B_\infty(\theta_0)\Sigma B_\infty(\theta_0)' \right]^{-1} B_\infty(\theta_0)\bar{m}_\infty(\theta_0) \\
&\sim \chi^2_{d_1}.
\end{aligned}
$$

Since the latter is pivotal distribution, the result also holds unconditionally as claimed in the proposition. $\square$

Proof of Lemma 2.2. By the mean value expansion,

$$
\psi(\tilde{\theta}) = \psi(\theta_0) + \frac{\partial\psi(\bar{\theta})}{\partial\theta_2'}(\tilde{\theta}_2 - \theta_{02}),
$$

where $\bar{\theta}$ is the mean value lying on the segment joining $\tilde{\theta}$ and θ_0 and satisfies $\bar{\theta} \xrightarrow{p} \theta_0$. By the continuous differentiability of $\psi(\theta)$ we have

$$
\frac{\partial\psi(\bar{\theta})}{\partial\theta_2'} \xrightarrow{p} \frac{\partial\psi(\theta_0)}{\partial\theta_2'}.
$$

Since $n^{1/2}(\tilde{\theta}_2 - \theta_{02})$ is bounded in probability,

$$
n^{1/2}(\psi(\tilde{\theta}) - \psi(\theta_0)) = \frac{\partial\psi(\bar{\theta})}{\partial\theta_2'} n^{1/2}(\tilde{\theta}_2 - \theta_{02}) = \frac{\partial\psi(\theta_0)}{\partial\theta_2'} n^{1/2}(\tilde{\theta}_2 - \theta_{02}) + o_p(1). \tag{B.12}
$$

Proceeding similarly to (B.3), we have

$$
n^{1/2}\hat{m}_n(\tilde{\theta}) = n^{1/2}\hat{m}_n(\theta_0) + H_2(\theta_0) n^{1/2}(\tilde{\theta}_2 - \theta_{02}) + o_p(1) \tag{B.13}
$$

Thus, from (B.12) and (B.13)

$$\begin{aligned}
n^{1/2}(\tilde{\psi}^* - \psi(\theta_0)) &= \psi(\tilde{\theta}) - \frac{\partial\psi(\tilde{\theta})}{\partial\theta_2'}(\hat{G}_{n,2}(\tilde{\theta})'\hat{\Sigma}_n(\tilde{\theta})^{-1}\hat{G}_{n,2}(\tilde{\theta}))^{-1}\hat{G}_{n,2}(\tilde{\theta})'\hat{\Sigma}_n(\tilde{\theta})^{-1}\hat{m}_n(\tilde{\theta}) + o_p(1) \\
&= -\frac{\partial\psi(\theta_0)}{\partial\theta_2'}(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)'\Sigma^{-1}n^{1/2}\hat{m}_n(\theta_0) + o_p(1) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\text{B.14}) \\
&\xrightarrow{d} N\left[0, \frac{\partial\psi(\theta_0)}{\partial\theta_2'}(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}\frac{\partial\psi(\theta_0)'}{\partial\theta_2}\right].
\end{aligned}$$

□

Proof of Proposition 2.3. We first take up the $AR^0_\alpha(\theta_{01})$ statistic. From Theorem 3.5 of Newey and McFadden (1994),

$$n^{1/2}(\tilde{\theta}_2^* - \theta_{02}) = -(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}n^{1/2}\hat{m}_n(\theta_0) + o_p(1) \tag{B.15}$$

$$\xrightarrow{d} N[0, (H_2(\theta_0)'\Sigma H_2(\theta_0))^{-1}]. \tag{B.16}$$

From the equation (B.3), we have

$$n^{1/2}\hat{m}_n(\tilde{\theta}^*) = n^{1/2}\hat{m}_n(\theta_0) + H_2(\theta_0)n^{1/2}(\tilde{\theta}_2^* - \theta_{02}) + o_p(1), \tag{B.17}$$

hence using (B.15)

$$\begin{aligned}
n^{1/2}\hat{m}_n(\tilde{\theta}^*) &= n^{1/2}\hat{m}_n(\theta_0) - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}n^{1/2}\hat{m}_n(\theta_0) + o_p(1) \\
&= \left[I_L - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}\right] m_\infty(\theta_0) + o_p(1). \qquad (\text{B.18})
\end{aligned}$$

Thus,

$$\begin{aligned}
AR^0_\alpha(\theta_{01}) &= n\,\hat{m}_n(\tilde{\theta}^*)'\hat{\Sigma}_n(\tilde{\theta}^*)^{-1}\hat{m}_n(\tilde{\theta}^*) \\
&= m_\infty(\theta_0)'\left[I_L - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}\right]'\Sigma^{-1} \\
&\quad \left[I_L - H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}\right] m_\infty(\theta_0) + o_p(1) \\
&= m_\infty(\theta_0)'\Sigma^{-1/2}M_{\Sigma^{-1/2}H_2(\theta_0)}\Sigma^{-1/2}m_\infty(\theta_0) + o_p(1). \qquad (\text{B.19})
\end{aligned}$$

The result follows on noting that $M_{\Sigma^{-1/2}H_2(\theta_0)}$ is idempotent matrix of rank $L-d_2$ and $\Sigma^{-1/2}m_\infty(\theta_0)\sim N[0,I_L]$.

Next consider the $AR_\alpha(\theta_{01})$ statistic. Following the same argument that led to (B.14) (replacing $\tilde{\psi}^*$, $\psi(\theta_0)$ and $\frac{\partial\psi(\theta_0)}{\partial\theta_2'}$ by $\hat{m}_n^*(\tilde{\theta})$, $\hat{m}_n(\theta_0)$ and $H_2(\theta_0)$, respectively), we have

$$n^{1/2}\hat{m}_n^*(\tilde{\theta})=n^{1/2}\hat{m}_n(\theta_0)-H_2(\theta_0)(H_2(\theta_0)'\Sigma^{-1}H_2(\theta_0))^{-1}H_2(\theta_0)\Sigma^{-1}n^{1/2}\hat{m}_n(\theta_0)+o_p(1),$$

which corresponds to (B.17). Therefore, proceeding as in (B.18), we obtain the result $AR_\alpha(\theta_{01})\xrightarrow{d}\chi^2_{L-d_2}$. □

Proof of Proposition 2.4. Since $\hat{V}_n(\tilde{\theta}^*)\xrightarrow{p}V(\theta_0)$ by Assumption A.7, invoking the CMT gives

$$\hat{K}_n(\tilde{\theta}^*)\xrightarrow{p}K(\theta_0). \tag{B.20}$$

Then, using $\tilde{\Sigma}_n=\hat{\Sigma}_n(\tilde{\theta}^*)\xrightarrow{p}\Sigma$, (B.20), and the CMT yields $\hat{\Omega}_n(\tilde{\theta}^*)\xrightarrow{p}\Omega(\theta_0)$ from which it follows that $\hat{\Omega}_n^\varepsilon(\tilde{\theta}^*)\xrightarrow{p}\Omega^\varepsilon(\theta_0)$ by the continuity of $\hat{\Omega}_n^\varepsilon(\theta)$, see Lemma 17.1 (e) and Comments (iv) of Andrews and Guggenberger (2015b). Using the CMT once again gives

$$\hat{U}_n(\tilde{\theta}^*)\xrightarrow{p}U_\infty(\theta_0)=[\theta_0,I_{d_1}]\Omega^\varepsilon(\theta_0)^{-1}[\theta_0,I_{d_1}]'.$$

In the proof of Proposition 2.3, we have shown that

$$\begin{aligned}M_{\tilde{\Sigma}_n^{-1/2}\tilde{H}_{n,2}}\tilde{\Sigma}_n^{-1/2}\hat{m}_n(\tilde{\theta}^*)&=M_{\Sigma^{-1/2}H_2(\theta_0)}\Sigma^{-1/2}m_\infty(\theta_0)+o_p(1)\\&\sim M_{\Sigma^{-1/2}H_2(\theta_0)}S_\infty+o_p(1),\end{aligned}$$

where $S_\infty\sim N[0,I_L]$ and the proof of Proposition 2.1 shows that

$$n^{1/2}\hat{H}_{in}(\tilde{\theta}^*)\xrightarrow{d}H_{i\infty}(\theta_0),\quad i=1,\ldots,d_1,$$

where $H_{i\infty}(\theta_0)$ is distributed independently of S_∞. Next we shall determine the limit of $\tilde{T}_n$. By Slutsky's lemma,

$$\begin{aligned} n^{1/2}\tilde{T}_n &\xrightarrow{d} \Sigma^{-1/2} H_{\infty,1}(\theta_0) U_\infty(\theta_0) \\ &\equiv T_\infty \end{aligned}$$

and by the CMT, we have

$$\begin{aligned} CLR &\xrightarrow{d} S_\infty' M_{\Sigma^{-1/2}H_2(\theta_0)} S_\infty - \lambda_{\min}\left[(S_\infty, T_\infty)' M_{\Sigma^{-1/2}H_2(\theta_0)}(S_\infty, T_\infty)\right] \\ &\equiv CLR_\infty. \end{aligned}$$

□

Proof of Proposition 3.1. We only prove the asymptotic results for the robust statistics based on the sample moment function (3.10). The proof of the CLR statistic based on (3.20) is similar thus omitted. The reparameterized sample moment function

$$\hat{m}_J(\vartheta) = J^{-1}Z'\xi = J^{-1}Z'(\delta(\sigma) - \bar{X}\mathit{\Gamma}_1\alpha - \bar{X}\mathit{\Gamma}_2\phi] \tag{B.21}$$

is continuously differentiable with respect to θ because $\delta(\sigma)$ is so as shown by Berry (1994). This together with the uniform convergence in Assumption 3.2 (ii) implies Assumption A.1. Assumption A.2 is clearly satisfied because

$$\hat{D}_{1n}(\vartheta) = \left[Z'\frac{\partial^2\xi}{\partial\alpha\partial\phi'}, Z'\frac{\partial^2\xi}{\partial\alpha\partial\sigma^{2\prime}}\right] = 0. \tag{B.22}$$

To verify Assumption A.3, partition $\mathit{\Gamma}_1 = [\mathit{\Gamma}_{11}', 1]'$ where $\mathit{\Gamma}_{11} \in \mathbb{R}^k$, and write

$$\begin{aligned} J^{1/2}\hat{G}_{J,\alpha}(\vartheta_0) &= -J^{-1/2}Z'\bar{X}\hat{\mathit{\Gamma}}_1 \\ &= J^{-1/2}Z'p - J^{-1/2}Z'X\mathit{\Gamma}_{11} \\ &= J^{-1/2}Z'(p - MC - \iota_J\eta) + J^{-1/2}Z'(MC + \iota_J\eta) - J^{-1/2}Z'X\mathit{\Gamma}_{11}. \end{aligned} \tag{B.23}$$

Note that $J^{-1/2}\text{vec}(Z'X)$, $J^{1/2}Z'MC$ and $J^{1/2}Z'\xi$ are jointly asymptotically normal.

By Cauchy-Schwarz inequality

$$
\begin{aligned}
J^{-1/2}\|Z'(p - MC - \iota_J\eta)\| &\leq J^{-1}\|Z\|J^{1/2}\|p - MC - \iota_J\eta\| \\
&\xrightarrow{p} 0, \qquad \text{(B.24)}
\end{aligned}
$$

where the convergence follows from Assumption 3.2 (ii) and (iii). Furthermore, by Assumption 3.2 (iv)

$$
J^{-1}Z'(MC + \iota_J\eta) - J^{-1}E[Z'MC] - J^{-1}E[Z'\iota_J]\eta \xrightarrow{p} 0. \qquad \text{(B.25)}
$$

Since $J^{-1}Z'X$ converges in probability to matrices of full rank by Assumption 3.2 (ii), (B.23)-(B.25) together with Slutsky's lemma yield Assumption A.3. Note that

$$
\begin{aligned}
J^{1/2}E[\hat{G}_{J,\alpha}(\vartheta_0)] &= J^{-1/2}E[Z'\bar{X}\hat{\Gamma}_1] \\
&= J^{-1/2}E[Z'\bar{X}\hat{\Gamma}_1 - Z'(X, -MC - \eta\iota_J)\hat{\Gamma}_1] + J^{-1/2}E[Z'(X, -MC - \eta\iota_J)\hat{\Gamma}_1],
\end{aligned} \qquad \text{(B.26)}
$$

where ι_J denotes a $J \times 1$ vector of ones. By direct calculation,

$$
J^{-1/2}E[Z'(X, -MC - \eta\iota_J)]\Gamma_1 = 0. \qquad \text{(B.27)}
$$

Furthermore,

$$
\begin{aligned}
J^{-1/2}Z'\bar{X}\Gamma_1 - Z'(X, -MC - \eta\iota_J)\Gamma_1 &= -J^{-1/2}Z'\left[0_{J\times k}, p - MC - \eta\iota_J\right]\Gamma_1 \\
&\xrightarrow{p} 0,
\end{aligned}
$$

because using Cauchy-Schwarz inequality, Assumption 3.2 (ii) and (iii),

$$
\begin{aligned}
J^{-1/2}\|Z'(p - MC - \iota_J\eta)\| &\leq J^{-1}\|Z\|J^{1/2}\|p - MC - \iota_J\eta\| \\
&\xrightarrow{p} 0. \qquad \text{(B.28)}
\end{aligned}
$$

Let $\psi_j = |J^{1/2}(p_j - MC_j - \eta)|$ and $\psi = [\psi_1, \ldots, \psi_J]'$. Then, by Cauchy-Schwarz inequality

$$
\begin{aligned}
E\left[\left(J^{-1/2}\|Z'(p-MC-\iota_J\eta)\|\right)^{1+\varepsilon/2}\right] &= E\left[\left(J^{-1/2}\|Z\|\right)^{1+\varepsilon/2}\left(J^{-1/2}\|\psi\|\right)^{1+\varepsilon/2}\right] \\
&\leq \left(E\left[\left(J^{-1/2}\|Z\|\right)^{2+\varepsilon}\right]\right)^{1/2}\left(E\left[\left(J^{-1/2}\|\psi\|\right)^{2+\varepsilon}\right]\right)^{1/2} \\
&= \left(E\left[\left(J^{-1}\|Z\|^2\right)^{1+\varepsilon/2}\right]\right)^{1/2}\left(E\left[\left(J^{-1}\|\psi\|^2\right)^{1+\varepsilon/2}\right]\right)^{1/2}.
\end{aligned} \tag{B.29}
$$

By Jensen's inequality and Assumption 3.2 (ii) and (iii),

$$
E\left[\left(J^{-1}\|\psi\|^2\right)^{1+\varepsilon/2}\right] \leq E\left[J^{-1}\sum_{j=1}^{J}\psi_j^{2+\varepsilon}\right] < \infty, \tag{B.30}
$$

and

$$
E\left[\left(J^{-1}\|Z\|^2\right)^{1+\varepsilon/2}\right] \leq \left(E\left[\left(J^{-1}\|Z\|^2\right)^2\right]\right)^{(1+\varepsilon/2)/2} \leq \left(E\left[J^{-1}\sum_{j=1}^{J}\|z_j\|^4\right]\right)^{(1+\varepsilon/2)/2} < \infty. \tag{B.31}
$$

From (B.29)-(B.31),

$$
E\left[\left(J^{-1/2}\|Z'(p-MC-\iota_J\eta)\|\right)^{1+\varepsilon/2}\right] < \infty. \tag{B.32}
$$

It follows from (B.28) and (B.32) that

$$
J^{-1/2}E[Z'\bar{X}\mathit{\Gamma}_1 - Z'(X,-MC-\eta\iota_J)\mathit{\Gamma}_1] \to 0. \tag{B.33}
$$

Finally, (B.26), (B.27) and (B.33) together yield

$$
E[\hat{G}_{J,\alpha}(\vartheta_0)] = \bar{C}J^{-1/2}.
$$

Assumption A.4 is satisfied because the probability limit of the Jacobian $J^{-1}Z'\frac{\partial\delta(\sigma)}{\partial\sigma^{2\prime}}$ is of full rank as maintained in Assumption 3.2 (ii), and the Jacobian with respect to ϕ

$$-J^{-1}Z'\bar{X}\Gamma_2 = -J^{-1}Z'X(E[x_j x_j'])^{-1}\begin{bmatrix} 0_{1\times(k-1)} & 1 \\ I_{k-1} & 0 \end{bmatrix}$$

converges in probability to a matrix of full rank.

In order to obtain the $J^{1/2}$-consistent estimator of $\vartheta_2 = (\phi', \sigma^{2\prime})'$, we estimate the nuisance parameter vector θ_2 by the restricted GMM under $H_0 : \alpha = \alpha_0$. We verify the assumptions of Andrews (2002) (see also Proposition 1 of Ketz (2017)). By standard argument (e.g., Theorem 2.1 of Newey and McFadden (1994)), the compactness of the parameter space in Assumption 3.1 and Assumption 3.2 (vii) together imply the consistency of the restricted GMM estimator $\tilde{\theta}_2$. Thus, Assumption GMM1 of Andrews (2002) holds. Assumption 3.2 (vii) implies Assumption GMM2 (a), (c) and (e) of Andrews (2002). Thus, the part (a) of Assumption GMM2^{2*} of Andrews (2002) holds. The part (b) therein is satisfied by Assumption 3.1 and 3.2 (vii). The part (c) follows because $\hat{m}_J(\theta)$ is continuously differentiable with respect to θ_2. The part (d) is satisfied because $J^{-1}Z\frac{\partial\xi(\alpha_0,\theta_2)}{\partial\theta_2'}$ converges in probability to a nonrandom matrix uniformly over Θ_2 that is continuous at σ_0. The part (e) follows from Assumption 3.2 (ii). Thus, Assumption GMM2^{2*} of Andrews (2002) is verified. Assumption GMM3 of Andrews (2002) holds by Assumption 3.2 (iv). Then by Theorem 1 of Andrews (1999), the restricted GMM estimator $\tilde{\theta}_2$ is $J^{1/2}$-consistent though not asymptotically normal in general. Given the $J^{1/2}$-consistent estimator $\tilde{b}$, the estimator $\tilde{\phi}$ is obtained as in (3.16):

$$\tilde{\phi} = \begin{bmatrix} 0_{(k-1)\times 1} & I_{k-1} \\ 1 & 0_{1\times(k-1)} \end{bmatrix}\Big(E[x_j x_j'] \quad -E[x_j(MC_j+\eta)]\Big)\begin{bmatrix} \tilde{b} \\ \alpha_0 \end{bmatrix}.$$

Because $\tilde{b}$ is $J^{1/2}$-consistent, so is $\tilde{\phi}$. This verifies Assumption A.6.

Next we verify Assumption A.7. Note that

$$
\begin{aligned}
\xi(\tilde{\vartheta}) &\equiv \delta(\tilde{\sigma}) - \bar{X}\hat{\Gamma}_1\alpha_0 - \bar{X}\hat{\Gamma}_2\tilde{\phi} \\
&= \delta(\tilde{\sigma}) + p\alpha_0 - X\tilde{b} \\
&= \xi(\tilde{\theta}). \qquad \text{(B.34)}
\end{aligned}
$$

Let $\tilde{\xi} \equiv \xi(\tilde{\vartheta}) = \xi(\tilde{\theta}) = [\tilde{\xi}_1, \ldots, \tilde{\xi}_J]'$, $\tilde{\delta}_j = \delta_j(\tilde{\sigma})$ and $\tilde{\xi}_j = \tilde{\delta}_j + p_j\alpha_0 - x_j'\tilde{b}$. The covariance matrix estimates are given by

$$
\begin{aligned}
\tilde{C}_J &\equiv \hat{C}_J(\tilde{\vartheta}) = J^{-1}\sum_{j=1}^{J} z_j p_j \tilde{\xi}_j z_j', \\
\tilde{\Sigma}_J &\equiv \hat{\Sigma}_J(\tilde{\vartheta}) = J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j^2.
\end{aligned}
$$

We will show that

$$
\tilde{C}_J - V_{MC\,\xi}E[J^{-1}Z'Z] \xrightarrow{p} 0, \qquad \text{(B.35)}
$$

$$
\tilde{\Sigma}_J - V_{\xi\xi}E[J^{-1}Z'Z] \xrightarrow{p} 0. \qquad \text{(B.36)}
$$

Write

$$
\begin{aligned}
\tilde{C}_J &= J^{-1}\sum_{j=1}^{J} z_j p_j \tilde{\xi}_j z_j' \\
&= J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j (p_j - MC_j - \eta) + J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j (MC_j + \eta). \qquad \text{(B.37)}
\end{aligned}
$$

Note that by Assumption 3.2 (ii)

$$
\begin{aligned}
J^{-1}\sum_{j=1}^{J} \|z_j\|^2 &= J^{-1}\sum_{j=1}^{J} \mathrm{tr}(z_j z_j') \\
&= \mathrm{tr}\left(J^{-1}Z'Z\right) \\
&= O_p(1), \qquad \text{(B.38)}
\end{aligned}
$$

The first term in (B.37) can be written as

$$J^{-1}\sum_{j=1}^{J} z_j z_j'\tilde{\xi}_j(p_j - MC_j - \eta) = J^{-1}\sum_{j=1}^{J} z_j z_j'(\xi_j + (\tilde{\delta}_j - \delta_j) - x_j'(\tilde{b} - b))(p_j - MC_j - \eta).$$

Note that

$$\begin{aligned}\|J^{-1}\sum_{j=1}^{J} z_j z_j'\xi_j(p_j - MC_j - \eta)\| &\leq J^{-1}\sum_{j=1}^{J}\|z_j z_j'\| \max_{1\leq j\leq J}|\xi_j| \max_{1\leq j\leq J}|p_j - MC_j - \eta| \\ &\to 0,\end{aligned}$$

where the convergence is due to $\max_{1\leq j\leq J}|\xi_j| = o_p(J^{1/2})$ which holds by the independence of ξ_j's and its finite second moment, $J^{1/2}\max_{1\leq j\leq J}|p_j - MC_j - \eta| = o_p(1)$, and (B.38). Moreover,

$$\begin{aligned}\sum_{j=1}^{J} z_j z_j'(\tilde{\delta}_j - \delta_j)(p_j - MC_j - \eta) &\leq \sum_{j=1}^{J}\|z_j z_j'\| \max_{1\leq j\leq J}|\tilde{\delta}_j - \delta_j| \max_{1\leq j\leq J}|p_j - MC_j - \eta| \\ &\to 0,\end{aligned} \tag{B.39}$$

where the convergence is due to $\max_{1\leq j\leq J}|\tilde{\delta}_j - \delta_j| = o_p(1)$ and $J^{1/2}\max_{1\leq j\leq J}|p_j - MC_j - \eta| = o_p(1)$, and (B.38). In addition,

$$\begin{aligned}\sum_{j=1}^{J} z_j z_j' x_j'(\tilde{b} - b)(p_j - MC_j - \eta) &\leq \sum_{j=1}^{J}\|z_j z_j'\| \max_{1\leq j\leq J}\|x_j\|\|\tilde{b} - b\| \max_{1\leq j\leq J}|p_j - MC_j - \eta| \\ &\to 0,\end{aligned} \tag{B.40}$$

where the convergence holds because $\max_{1\leq j\leq J}\|x_j\| = o_p(J^{1/2})$, $J^{1/2}(\tilde{b} - b) = O_p(1)$, $J^{1/2}\max_{1\leq j\leq J}|p_j - MC_j - \eta| = o_p(1)$, and (B.38). Thus the first term on the right-hand side (RHS) of (B.37) is asymptotically negligible:

$$J^{-1}\sum_{j=1}^{J} z_j z_j'\tilde{\xi}_j(p_j - MC_j - \eta) \xrightarrow{p} 0. \tag{B.41}$$

Next consider the second term $J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j (MC_j + \eta)$ in (B.37). Write

$$J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j (MC_j+\eta) - J^{-1}\sum_{j=1}^{J} z_j z_j' \xi_j MC_j = J^{-1}\sum_{j=1}^{J} z_j z_j' (\tilde{\xi}_j - \xi_j) MC_j - J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j \eta. \tag{B.42}$$

Using the triangle inequality and the matrix norm product inequality, the first term on the RHS of (B.42) can be bounded as

$$\begin{aligned}\|J^{-1}\sum_{j=1}^{J} z_j z_j' (\tilde{\xi}_j - \xi_j) MC_j\| &= \|J^{-1}\sum_{j=1}^{J} z_j z_j' (\tilde{\delta}_j - \delta_j) MC_j - x_j'(\tilde{b} - b) MC_j\| \\ &\leq J^{-1}\sum_{j=1}^{J} \|z_j z_j' MC_j\| \max_{1\leq j\leq J} |\tilde{\delta}_j - \delta_j| + J^{-1}\sum_{j=1}^{J} \|z_j z_j'\| \|\tilde{b} - b\| \max_{1\leq j\leq J} \|x_j\|\end{aligned} \tag{B.43}$$

Using Cauchy-Schwarz inequality,

$$J^{-1}\sum_{j=1}^{J} \|z_j z_j' MC_j\| \leq \left(J^{-1}\sum_{j=1}^{J} \|z_j\|^4\right)^{1/2} \left(J^{-1}\sum_{j=1}^{J} MC_j^2\right)^{1/2}. \tag{B.44}$$

The last term on the RHS of (B.44) is bounded in probability because MC_j's are i.i.d. with finite second moments. The first term on the RHS of (B.44) is bounded in probability by Assumption 3.2 (ii). Using $\max_{1\leq j\leq J} |\tilde{\delta}_j - \delta_j| = o_p(1)$, the first term of (B.43) is bounded. The second term of (B.43) is bounded using (B.38) and the fact that $\|\tilde{b} - b\| \max_{1\leq j\leq J} \|x_j\| = o_p(1)$. Finally, we show that

$$J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j \xrightarrow{p} 0. \tag{B.45}$$

Proceeding similarly to previous calculations, we obtain

$$J^{-1}\sum_{j=1}^{J} z_j z_j' (\tilde{\xi}_j - \xi_j) \xrightarrow{p} 0.$$

Let z_{js} denote the s-th element of z_j. Since

$$E\left[J^{-1}\sum_{j=1}^{J} z_j z_j' \xi_j\right] = 0,$$

and by Assumption 3.2 (ii)

$$\begin{aligned} V\left[J^{-1}\sum_{j=1}^{J} z_j z_{js} \xi_j\right] &= E\left[J^{-2}\sum_{j=1}^{J} z_j z_j' z_{js}^2\right] \\ &= J^{-1}E\left[J^{-1}\sum_{j=1}^{J} \|z_j\|^2 z_{js}^2\right] \\ &\to 0, \end{aligned}$$

(B.45) follows. Combining (B.42) and (B.45), we have

$$J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j (MC_j + \eta) - J^{-1}\sum_{j=1}^{J} z_j z_j' \xi_j MC_j \xrightarrow{p} 0. \tag{B.46}$$

By Assumption 3.2 (v) and (vi),

$$\begin{aligned} J^{-1}\sum_{j=1}^{J} z_j z_j' \xi_j MC_j &= E\left[J^{-1}\sum_{j=1}^{J} z_j z_j' \xi_j MC_j\right] + o_p(1) & \text{(B.47)} \\ &= V_{MC\,\xi} E[J^{-1}Z'Z] + o_p(1). & \text{(B.48)} \end{aligned}$$

We next show that

$$J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j^2 - J^{-1}\sum_{j=1}^{J} E[z_j z_j' \xi_j^2] \xrightarrow{p} 0. \tag{B.49}$$

Write

$$J^{-1}\sum_{j=1}^{J} z_j z_j' \tilde{\xi}_j^2 = J^{-1}\sum_{j=1}^{J} z_j z_j' \xi_j^2 + J^{-1}\sum_{j=1}^{J} z_j z_j' (\tilde{\xi}_j^2 - \xi_j^2).$$

It suffices to show that the second term converges to 0. By the triangle inequality,

$$\|J^{-1}\sum_{j=1}^{J} z_j z_j'(\tilde{\xi}_j^2 - \xi_j^2)\| \leq J^{-1}\sum_{j=1}^{J} \|z_j z_j'\| |\tilde{\xi}_j^2 - \xi_j^2|. \tag{B.50}$$

On noting that $\tilde{\xi}_j = \xi_j + \delta_j - \delta_j - x_j'(\tilde{b} - b)$,

$$\hat{\xi}_j^2 = \xi_j^2 + (\tilde{\delta}_j - \delta_j)^2 + (\tilde{b} - b)' x_j x_j'(\tilde{b} - b) + 2\xi_j(\tilde{\delta}_j - \delta_j) - 2\xi_j x_j'(\tilde{b} - b) - 2(\tilde{\delta}_j - \delta_j) x_j'(\tilde{b} - b). \tag{B.51}$$

Hence

$$\begin{aligned} J^{-1}\sum_{j=1}^{J} \|z_j z_j'\| |\tilde{\delta}_j^2 - \delta_j^2| &\leq \left(J^{-1}\sum_{j=1}^{J} \|z_j z_j'\|\right) \max_{1\leq j\leq J} |\tilde{\delta}_j^2 - \delta_j^2| \\ &\xrightarrow{p} 0. \end{aligned} \tag{B.52}$$

Using $\max_{1\leq j\leq J} \|x_j\| = o_p(J^{1/2})$,

$$\begin{aligned} J^{-1}\sum_{j=1}^{J} \|z_j z_j'\| (\tilde{b} - b)' x_j x_j'(\tilde{b} - b) &\leq \left(J^{-1}\sum_{j=1}^{J} \|z_j z_j'\|\right) \max_{1\leq j\leq J} |x_j'(\tilde{b} - b)|^2 \\ &\to 0. \end{aligned} \tag{B.53}$$

Using Cauchy-Schwarz inequality gives the following convergence results:

$$J^{-1}\sum_{j=1}^{J}\|z_j\|^2|\xi_j||\tilde{\delta}_j-\delta_j| \leq \sqrt{J^{-1}\sum_{j=1}^{J}\|z_j\|^4}\sqrt{J^{-1}\sum_{j=1}^{J}\xi_j^2}\max_{1\leq j\leq J}|\tilde{\delta}_j-\delta_j| \xrightarrow{p} 0, \tag{B.54}$$

$$J^{-1}\sum_{j=1}^{J}\|z_j\|^2|\xi_j x_j'(\tilde{b}-b)| \leq \sqrt{J^{-1}\sum_{j=1}^{J}\|z_j\|^4}\sqrt{J^{-1}\sum_{j=1}^{J}\|x_j\xi_j\|^2}\|\tilde{b}-b\| \xrightarrow{p} 0, \tag{B.55}$$

$$J^{-1}\sum_{j=1}^{J}\|z_j\|^2|(\tilde{\delta}_j-\delta_j)x_j'(\tilde{b}-b)| \leq \left(J^{-1}\sum_{j=1}^{J}\|z_j\|^2\right)\max_{1\leq j\leq J}|\tilde{\delta}_j-\delta_j|\max_{1\leq j\leq J}|x_j'(\tilde{b}-b)| \xrightarrow{p} 0. \tag{B.56}$$

From (B.50)-(B.56),

$$J^{-1}\sum_{j=1}^{J}z_jz_j'\tilde{\xi}_j^2 - J^{-1}\sum_{j=1}^{J}z_jz_j'\xi_j^2 \xrightarrow{p} 0.$$

Note that

$$E\left[J^{-1}\sum_{j=1}^{J}z_jz_j'\xi_j^2\right] = E[J^{-1}Z'Z]V_{\xi\xi}$$

and by the independence of ξ_j's

$$V[J^{-1}\sum_{j=1}^{J}z_jz_{js}\xi_j^2] = E\left[J^{-2}\sum_{j=1}^{J}z_jz_j'z_{js}^2(\xi_j^2-V_{\xi\xi})^2\right],$$

where z_{js} denotes the s-th element of z_j. Since

$$E\left[J^{-1}\sum_{j=1}^{J}z_jz_j'z_{js}^2\right] = E\left[J^{-1}\sum_{j=1}^{J}\|z_j\|^2z_{js}^2\right] = O(1),$$

we have

$$V[J^{-1}\sum_{j=1}^{J}z_jz_{js}\xi_j^2] = E\left[J^{-2}\sum_{j=1}^{J}z_jz_j'z_{js}^2(\xi_j^2-V_{\xi\xi})^2\right] \to 0.$$

Then (B.49) follows by Chebyshev's inequality.

To verify Assumption A.7, it remains to show that $\hat{\Sigma}_{G_1G_1} - \Sigma_{G_1G_1} \xrightarrow{p} 0$ where

$$\hat{\Sigma}_{G_1G_1} = J^{-1}\sum_{j=1}^{J} \bar{z}_j\bar{z}_j' p_j^2 - J^{-1}\bar{Z}'p(J^{-1}\bar{Z}'p)'. \tag{B.57}$$

Since $J^{-1/2}\bar{Z}'p = J^{-1/2}\bar{Z}'(MC+\eta\iota_J)+o_p(1) = J^{-1/2}\sum_{j=1}^{J}(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)(MC_j+\eta)$, the covariance matrix $\Sigma_{G_1G_1}$ of the Jacobian $J^{-1/2}\bar{Z}'p$ is given by the limit of

$$\begin{aligned} &E[J^{-1}\sum_{j=1}^{J}(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)'(MC_j+\eta)^2] \\ &- E[J^{-1}\bar{Z}'(MC+\eta\iota_J)](E[J^{-1}\bar{Z}'(MC+\eta\iota_J)])'. \end{aligned} \tag{B.58}$$

The first term of (B.58) can be rewritten as

$$\begin{aligned} &E[J^{-1}\sum_{j=1}^{J}(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)'(MC_j+\eta)^2] \\ &= J^{-1}E\left[\sum_{j=1}^{J} z_jz_j'(MC_j+\eta)^2\right] - E[J^{-1}Z'X](E[x_jx_j'])^{-1}E\left[J^{-1}\sum_{j=1}^{J} x_jz_j'(MC_j+\eta)^2\right] \qquad \text{(B.59)} \\ &\quad - E\left[J^{-1}\sum_{j=1}^{J} z_jx_j'(MC_j+\eta)^2\right](E[x_jx_j'])^{-1}E[J^{-1}X'Z] \qquad \text{(B.60)} \\ &\quad + E[J^{-1}Z'X](E[x_jx_j'])^{-1}E\left[J^{-1}\sum_{j=1}^{J} x_jx_j'(MC_j+\eta)^2\right](E[x_jx_j'])^{-1}E[J^{-1}X'Z]. \qquad \text{(B.61)} \end{aligned}$$

Rewrite the first term on the RHS of (B.57) as

$$\begin{aligned} J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_jp_j^2 &= J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(p_j - MC_j - \eta)^2 + 2J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j+\eta)(p_j - MC_j - \eta) \\ &\quad + J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j+\eta)^2. \end{aligned} \tag{B.62}$$

The first and the second on the RHS of (B.62) are $o_p(1)$ because, using $J^{-1}\bar{Z}'\bar{Z} = O_p(1)$ and $J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j' MC_j = O_p(1)$ (Assumption 3.2 (ii) and (v)), we have

$$J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(p_j - MC_j - \eta)^2 \leq J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j' \max_{1\leq j\leq J}(p_j - MC_j - \eta)^2 \xrightarrow{p} 0,$$

$$J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j + \eta)(p_j - MC_j - \eta) \leq J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j + \eta)\max_{1\leq j\leq J}|p_j - MC_j - \eta| \xrightarrow{p} 0.$$

Thus,

$$J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j p_j^2 = J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j+\eta)^2 + o_p(1) = J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j^2+2MC_j\eta+\eta^2)+o_p(1), \tag{B.63}$$

Using $\bar{z}_j = z_j - Z'X(X'X)^{-1}x_j$, rewrite

$$\begin{aligned}
&J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j + \eta)^2 \\
&= J^{-1}\sum_{j=1}^{J}(z_j - Z'X(X'X)^{-1}x_j)(z_j - Z'X(X'X)^{-1}x_j)'(MC_j + \eta)^2 \\
&= J^{-1}\sum_{j=1}^{J}z_j z_j'(MC_j + \eta)^2 - Z'X(X'X)^{-1}J^{-1}\sum_{j=1}^{J}x_j z_j'(MC_j + \eta)^2 \\
&\quad - J^{-1}\sum_{j=1}^{J}z_j x_j'(MC_j + \eta)^2(X'X)^{-1}X'Z + Z'X(X'X)^{-1}J^{-1}\sum_{j=1}^{J}x_j x_j'(MC_j + \eta)^2(X'X)^{-1}X'Z.
\end{aligned}$$

Invoking Assumption 3.2 (ii), (v) and CMT, and using the expressions (B.59)-(B.61), we obtain

$$\begin{aligned}
&J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j'(MC_j^2 + 2MC_j\eta + \eta)^2 \\
&- E[J^{-1}\sum_{j=1}^{J}(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)'(MC_j + \eta)^2] \xrightarrow{p} 0,
\end{aligned}$$

From (B.63),

$$J^{-1}\sum_{j=1}^{J}\bar{z}_j\bar{z}_j p_j^2 - E[J^{-1}\sum_{j=1}^{J}(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)(z_j - E[J^{-1}Z'X](E[x_jx_j'])^{-1}x_j)'(MC_j+\eta)^2]$$
$$\xrightarrow{p} 0. \tag{B.64}$$

Furthermore, using $J^{-1}\bar{Z}'p = J^{-1}\bar{Z}'(MC+\eta\iota_J)+o_p(1)$ and Assumption 3.2 (i) and (ii), we have

$$\begin{aligned}
&J^{-1}\bar{Z}'(MC+\eta\iota_J) - E[J^{-1}\bar{Z}'(MC+\eta\iota_J)]\\
&= J^{-1}Z'(MC+\eta\iota_J) - E[J^{-1}Z'(MC+\eta\iota_J)]\\
&\quad - \left(J^{-1}Z'X(J^{-1}X'X)^{-1}J^{-1}X'(MC+\eta\iota_J) - E[J^{-1}Z'X](E[x_jx_j'])^{-1}E[x_j(MC_j+\eta)]\right) + o_p(1)\\
&\xrightarrow{p} 0.
\end{aligned}$$

Since $E[J^{-1}\bar{Z}'(MC+\eta\iota_J)] - E[J^{-1}Z'X](E[x_jx_j'])^{-1}E[x_j(MC_j+\eta)] = 0$ by (B.27), it follows that $J^{-1}\bar{Z}'p \xrightarrow{p} 0$. The latter combined with (B.64) yields $\hat{\Sigma}_{G_1G_1} - \Sigma_{G_1G_1} \xrightarrow{p} 0$. Therefore, Assumption A.7 is satisfied. It follows that the $C(\alpha)$-type statistic based on the reparameterized moment function is asymptotically χ_1^2 distributed under $H_0: \alpha = \alpha_0$. Finally, we verify the invariance properties of the test statistics. From (3.15), the sample moment function is invariant to reparameterization:

$$\hat{m}_J(\theta) = \hat{m}_J(\vartheta). \tag{B.65}$$

For $\hat{G}_{J,2}(\theta) = \left[\hat{G}_{J,b}(\theta), \hat{G}_{J,\sigma^2}(\theta)\right] = \left[-J^{-1}Z'X, J^{-1}Z'\frac{\partial\delta(\sigma)}{\partial\sigma^{2\prime}}\right]$,

$$\begin{aligned}
\hat{G}_{J,2}(\vartheta) &= [\hat{G}_{J,\phi}(\vartheta), \hat{G}_{J,\sigma^2}(\vartheta)]\\
&= \left[-J^{-1}Z'X(E[x_jx_j'])^{-1}\mathcal{E}_k, J^{-1}Z'\frac{\partial\delta(\sigma)}{\partial\sigma^{2\prime}}\right]\\
&= \left[-J^{-1}Z'X, J^{-1}Z'\frac{\partial\delta(\sigma)}{\partial\sigma^{2\prime}}\right]\begin{bmatrix}(E[x_jx_j'])^{-1}\mathcal{E}_k & 0_{k\times k}\\ 0_{k\times k} & J^{-1}Z'\frac{\partial\delta(\sigma)}{\partial\sigma^{2\prime}}\end{bmatrix}\\
&= \hat{G}_{J,2}(\theta)\bar{B}.
\end{aligned} \tag{B.66}$$

Therefore, using (B.66), $\hat{G}_{n,2}(\theta) = \hat{H}_{n,2}(\theta)$ and $\hat{\Sigma}_n(\vartheta) = \hat{\Sigma}_n(\theta)$

$$
\begin{aligned}
&M_{\hat{\Sigma}_n(\vartheta)^{-1/2}\hat{H}_{n,2}(\vartheta)} \\
&= I_L - \hat{\Sigma}_n(\vartheta)^{-1/2}\hat{H}_{n,2}(\vartheta)(\hat{H}_{n,2}(\vartheta)'\hat{\Sigma}_n(\vartheta)^{-1}\hat{H}_{n,2}(\vartheta))^{-1}\hat{H}_{n,2}(\vartheta)'\hat{\Sigma}_n(\vartheta)^{-1/2}, \\
&= I_L - \hat{\Sigma}_n(\theta)^{-1/2}\hat{H}_{n,2}(\theta)\bar{B}(\bar{B}\hat{H}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{H}_{n,2}(\theta)\bar{B})^{-1}\bar{B}\hat{H}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1/2}, \\
&= M_{\hat{\Sigma}_n(\theta)^{-1/2}\hat{H}_{n,2}(\theta)}. \qquad (B.67)
\end{aligned}
$$

This combined with (B.65) yields the invariance of the AR_α statistic to reparameterization:

$$
\begin{aligned}
AR_\alpha(\alpha_0) &= J\,\bar{m}_J(\tilde{\theta})'\hat{\Sigma}_J(\tilde{\theta})^{-1}\bar{m}_J(\tilde{\theta}), \\
&= J\,\bar{m}_J(\tilde{\vartheta})'\hat{\Sigma}_J(\tilde{\vartheta})^{-1}\bar{m}_J(\tilde{\vartheta}) \\
&\xrightarrow{d} \chi^2_{l-k}. \qquad (B.68)
\end{aligned}
$$

Let $\bar{X} = [X, -p] = [\bar{x}_1, \ldots, \bar{x}_J]'$. Next we show the invariance of the $C(\alpha)$-type statistic. The robust Jacobian estimator defined in (2.9) takes the following form in the reparameterized model:

$$
\hat{H}_{J,1}(\vartheta) = \hat{H}_{J,\alpha}(\vartheta) = -J^{-1}Z'\bar{X}\mathit{\Gamma}_1 + \left(J^{-1}\sum_{j=1}^{J} z_j\bar{x}_j'\mathit{\Gamma}_1\xi_j(\vartheta)z_j'\right)\hat{\Sigma}_J(\vartheta)^{-1}\bar{m}_J(\vartheta),
$$

where we used (B.22). Since $\xi_j(\vartheta)z_j'\hat{\Sigma}_J(\vartheta)^{-1}\bar{m}_J(\vartheta)$ is scalar, we may rewrite the Jacobian estimator above as

$$
\begin{aligned}
\hat{H}_{J,\alpha}(\vartheta) &= -J^{-1}Z'\bar{X}\mathit{\Gamma}_1 + \left(J^{-1}\sum_{j=1}^{J} z_j\bar{x}_j'\mathit{\Gamma}_1\xi_j(\vartheta)z_j'\right)\hat{\Sigma}_J(\vartheta)^{-1}\bar{m}_J(\vartheta), \\
&= \left[-J^{-1}Z'\bar{X} + J^{-1}\sum_{j=1}^{J} z_j\bar{x}_j'\xi_j(\vartheta)z_j'\hat{\Sigma}_J(\vartheta)^{-1}\bar{m}_J(\vartheta)\right]\mathit{\Gamma}_1 \\
&= [\hat{H}_{J,b}(\theta), \hat{H}_{J,\alpha}(\theta)]\mathit{\Gamma}_1.
\end{aligned}
$$

Similarly, we have

$$\hat{H}_{J,\phi}(\vartheta) = \left[-J^{-1} Z'\bar{X} + J^{-1} \sum_{j=1}^{J} z_j \bar{x}_j' \xi_j(\vartheta) z_j' \hat{\Sigma}_J(\vartheta)^{-1} \bar{m}_J(\vartheta) \right] \Gamma_2 = [\hat{H}_{J,b}(\theta), \hat{H}_{J,\alpha}(\theta)]\Gamma_2.$$

In sum, $[\hat{H}_{J,\alpha}(\vartheta), \hat{H}_{J,\phi}(\vartheta)] = [\hat{H}_{J,b}(\theta), \hat{H}_{J,\alpha}(\theta)]\Gamma$ and

$$[\hat{H}_{J,\alpha}(\vartheta), \hat{H}_{J,\phi}(\vartheta), \hat{H}_{J,\sigma^2}(\vartheta)] = [\hat{H}_{J,b}(\theta), \hat{H}_{J,\alpha}(\theta), \hat{H}_{J,\sigma^2}(\theta)] \begin{bmatrix} \Gamma & 0 \\ 0 & I_k \end{bmatrix}.$$

Now applying the Proposition 4.2 of Dufour, Trognon, and Tuvaandorj (2017), we obtain the invariance of the robust $C(\alpha)$ statistic from which it follows that

$$C_\alpha(\alpha_0) \xrightarrow{d} \chi_1^2. \tag{B.69}$$

Next we consider the rank statistic (2.23). Using (B.67),

$$\begin{aligned} M_{\hat{\Sigma}_J(\vartheta)^{-1/2}\hat{H}_{J,2}(\vartheta)} \hat{\Sigma}_J(\vartheta)^{-1/2} \hat{H}_{J,1}(\vartheta) &= M_{\hat{\Sigma}_J(\vartheta)^{-1/2}\hat{H}_{J,2}(\vartheta)} \hat{\Sigma}_J(\vartheta)^{-1/2} \left[\hat{H}_{J,b}(\theta), \hat{H}_{J,\alpha}(\theta) \right] \Gamma_1 \\ &= M_{\hat{\Sigma}_J(\vartheta)^{-1/2}\hat{H}_{J,2}(\vartheta)} \hat{\Sigma}_J(\vartheta)^{-1/2} \left[\hat{H}_{J,b}(\theta)\Gamma_{11} + \hat{H}_{J,\alpha}(\theta) \right] \\ &= M_{\hat{\Sigma}_J(\theta)^{-1/2}\hat{H}_{J,2}(\theta)} \hat{\Sigma}_J(\theta)^{-1/2} \hat{H}_{J,\alpha}(\theta). \end{aligned} \tag{B.70}$$

Since $\hat{U}_J(\vartheta) = [\alpha, 1]\hat{\Omega}_J^{\varepsilon}(\vartheta)^{-1}[\alpha, 1]'$, where $\hat{\Omega}_J^{\varepsilon}(\theta)$ is as defined in (2.20), is scalar, we thus have

$$\begin{aligned} \hat{T}_J(\vartheta)' M_{\hat{\Sigma}_J(\vartheta)^{-1/2}\hat{H}_{J,2}(\vartheta)} \hat{T}_J(\vartheta) &= J\, \hat{H}_{J,\alpha}(\vartheta)' \hat{\Sigma}_J(\vartheta)^{-1/2} M_{\hat{\Sigma}_J(\vartheta)^{-1/2}\hat{H}_{J,2}(\vartheta)} \hat{\Sigma}_J(\vartheta)^{-1/2} \hat{H}_{J,\alpha}(\vartheta) \hat{U}_J(\vartheta) \\ &= J\, \hat{H}_{J,\alpha}(\theta)' \hat{\Sigma}_J(\theta)^{-1/2} M_{\hat{\Sigma}_J(\theta)^{-1/2}\hat{H}_{J,2}(\theta)} \hat{\Sigma}_J(\theta)^{-1/2} \hat{H}_{J,\alpha}(\theta) \hat{U}_J(\vartheta) \\ &= \hat{T}_J(\theta)' M_{\hat{\Sigma}_J(\theta)^{-1/2}\hat{H}_{J,2}(\theta)} \hat{T}_J(\theta) \hat{U}_J(\vartheta)/\hat{U}_J(\theta) \\ &\xrightarrow{d} T_\infty' M_{\Sigma^{-1/2}H_2} T_\infty U_\infty(\vartheta_0)/U_\infty(\theta_0). \end{aligned} \tag{B.71}$$

Therefore, the rank statistic (2.23) is invariant under the reparameterization up to the scale factor $\hat{U}_J(\vartheta)/\hat{U}_J(\theta)$. Using (B.68), (B.69) and (B.71) and invoking the CMT, we obtain the asymptotic distribution of the CLR statistic.

□

Proof of Proposition 3.2. Again, we only provide the proof for the test statistics based on the sample moment function (3.9). We write $\xi(\alpha_0, \theta_2) = \xi(\alpha_0, \theta_2, p_t, s_t, x_t)$. The differentiability of the sample moment function holds as in the proof of Proposition 3.1. The compact support condition for the observed data $(p_t', s_t', x_t')'$, the parameter space assumption and the continuity of $\xi_t(\alpha_0, \theta_2)$ and $\partial\xi_t(\alpha_0, \theta_2)/\partial\sigma^{2\prime}$ in their arguments imply that $\xi_t(\alpha_0, \theta_2)$ and $\partial\xi_t(\alpha_0, \theta_2)/\partial\sigma^{2\prime}$ are bounded random vectors with finite moments. Therefore, $E[\|z_t\partial\xi_t(\alpha_0, \theta_2)/\partial\sigma^{2\prime}\|] < E[\|z_t\|]\bar{M} < \infty$ for some constant $\bar{M} > 0$. Lemma 2.4 of Newey and McFadden (1994) then gives

$$\sup_{\theta_2\in\Theta_2} \|T^{-1}Z'\partial\xi(\alpha_0, \theta_2)/\partial\theta_2' - E[z_t\partial\xi_t(\alpha_0, \theta_2)/\partial\theta_2']\| \xrightarrow{p} 0, \tag{B.72}$$

where $E[z_t\partial\xi_t(\alpha_0, \theta_2)/\partial\theta_2']$ is continuous in θ_2. This verifies Assumption A.1. Similarly to (B.72), we obtain

$$\sup_{\theta_2\in\Theta_2} \|T^{-1}Z'\xi(\alpha_0, \theta_2) - E[z_t\xi_t(\alpha_0, \theta_2)]\| \xrightarrow{p} 0. \tag{B.73}$$

Assumption A.2 is verified as in Proposition 3.1. Assumption A.4 follows from Assumption 3.3 on noting that $\hat{G}_T(\vartheta) = T^{-1}Z'M_Xp$ and the probability limits of $T^{-1}Z'X$ and $T^{-1}Z'\partial\delta(\sigma_0)/\partial\sigma^{2\prime}$ have full rank. Assumptions A.3 is implied by Assumptions 3.1 and 3.4. To see this, using the fact that p_t lies in a compact set, note that $E[\|z_t\xi_t\|^{2+\varepsilon}] < E[\|z_t\|^{2+\varepsilon}]\bar{M} < \infty$ and $E[\|z_tp_t\|^{2+\varepsilon}] < E[\|z_t\|^{2+\varepsilon}]\bar{M} < \infty$ for some constant $\bar{M} > 0$ Therefore, $T^{-1/2}Z'M_Xp$ and $T^{-1/2}Z'\xi$ obey the CLT (see also Freyberger (2015)).

Next, we verify Assumption A.6. Again, this is established by verifying the assumptions of Andrews (2002). On noting that $E[\|z_t\xi_t(\alpha_0, \theta_2)\xi_t(\alpha_0, \theta_2)'z_t'\|] \leq E[\|z_t\|^2\|\xi_t(\alpha_0, \theta_2)\|^2] < E[\|z_t\|^2]\bar{M} < \infty$ for some $\bar{M} > 0$, by Lemma 2.4 of Newey and McFadden (1994) and i.i.d. assumption

$$\sup_{\theta_2\in\Theta_2} \|T^{-1}\sum_{t=1}^{T} z_t\xi_t(\alpha_0, \theta_2)\xi_t(\alpha_0, \theta_2)'z_t' - E[z_t\xi_t(\alpha_0, \theta_2)\xi_t(\alpha_0, \theta_2)'z_t']\| \xrightarrow{p} 0, \tag{B.74}$$

where $E[z_t\xi_t(\alpha_0,\theta_2)\xi_t(\alpha_0,\theta_2)'z_t']$ is continuous in θ_2. Combining this with the compactness of the parameter space in Assumption 3.1, the identifiability of θ_{02} and the nonsingularity of $E[\|z_t\xi(\alpha_0,\theta_2)\xi_t(\alpha_0,\theta_2)'z_t'\|]$ in Assumption 3.3, and using Theorem 2.1 of Newey and McFadden (1994), we obtain the consistency of the restricted GMM estimator $\tilde{\theta}_2$. Thus, Assumption GMM1 of Andrews (2002) holds. Assumption GMM2 (a), (c) and (e) of Andrews (2002) follows from the uniform LLN for $T^{-1}Z'\xi(\alpha,\theta_2)$ in (B.73), $E[z_t\xi_t(\theta_0)] = 0$, and (B.74) combined with the continuity of $E[z_t\xi(\alpha_0,\theta_2)\xi_t(\alpha_0,\theta_2)'z_t']$ at θ_{02}, respectively. Thus, the part (a) of Assumption GMM2^{2*} of Andrews (2002) holds. The part (b) therein is satisfied by Assumption 3.1 and the continuous differentiability of $E[z_t\xi_t(\alpha_0,\theta_2)]$ with respect to θ_2 which follows from the fact that $\xi_t(\alpha_0,\theta_2)$ is continuously differentiable with respect to θ_2, $E[\sup_{\theta_2\in\Theta_2}\|z_t\xi_t(\alpha_0,\theta_2)\|]\leq E[\|z_t\|\sup_{\theta_2\in\Theta_2}\|\xi_t(\alpha_0,\theta_2)\|]<\infty$ and Lemma 3.6 of Newey and McFadden (1994). The part (c) follows because $z_t\xi_t(\alpha_0,\theta)$ is continuously differentiable with respect to θ_2. The part (d) holds because $T^{-1}Z\frac{\partial\xi(\alpha_0,\theta_2)}{\partial\theta_2'}$ converges in probability uniformly over Θ_2 to a nonrandom matrix continuous at θ_{02} as shown in (B.72). Assumption GMM3 of Andrews (2002) holds because $T^{-1/2}Z'\xi \xrightarrow{d} N[0,E[z_t\xi_t\xi_t'z_t']]$ as verified above. Then by Theorem 1 of Andrews (1999), the restricted GMM estimator $\tilde{\theta}_2$ is $T^{1/2}$-consistent.

Assumption A.8 follows from the uniform LLN for the following sums established similarly to (B.74):

$$T^{-1}\sum_{t=1}^{T} z_t\xi_t(\alpha_0,\theta_2)\frac{\partial\xi_t(\alpha_0,\theta_2)'}{\partial\theta_{2l}}z_t', \quad T^{-1}\sum_{t=1}^{T} z_t\frac{\partial\xi_t(\alpha_0,\theta_2)}{\partial\theta_{2l}}\frac{\partial\xi_t(\alpha_0,\theta_2)'}{\partial\theta_{2l}}z_t',$$

where θ_{2l} denotes the l-th element of θ_2. Finally, noting that Assumptions A.6 and A.8 imply Assumption A.7, and using the invariance properties of the test statistics we obtain the result. $\square$

C Alternative rank and CLR statistics

As in Kleibergen (2005, 2007) and Smith (2007), one may consider CLR-type statistic based on the Robin and Smith (2000) rank statistic in which the robust Jacobian is weighted by $\hat{V}_{Hn}(\theta)^{-1/2}$, the inverse square root of the Jacobian variance matrix (Jacobian-variance weighting):

$$\begin{aligned}\hat{V}_{Hn}(\theta) = &\sum_{i=1}^{n}\left[\operatorname{vec}(G_i(\theta) - \hat{G}_n(\theta))\right]\left[\operatorname{vec}(G_i(\theta) - \hat{G}_n(\theta))\right]' \\ &+ \left[\hat{C}_n(\theta)\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta) - \hat{D}_n(\theta)\right]\left[\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta)\right]^{-1} \\ &\quad \left[\hat{C}_n(\theta)\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta) - \hat{D}_n(\theta)\right]' - \hat{C}_n(\theta)\hat{\Sigma}_n(\theta)^{-1}\hat{C}_n(\theta)', \qquad \text{(C.1)} \\ \hat{C}_n(\theta) = &\left[\hat{C}_{1n}(\theta)', \ldots, \hat{C}_{dn}(\theta)'\right]' \in \mathbb{R}^{dL\times L}, \\ \hat{D}_n(\theta) = &\left[\hat{D}_{1n}(\theta)', \ldots, \hat{D}_{dn}(\theta)'\right]' \in \mathbb{R}^{dL\times L},\end{aligned}$$

where $\hat{C}_{jn}(\theta), i = 1, \ldots, d$, denotes the estimator of the covariance between the sample moment functions and the Jacobian vector $\hat{G}_{jn}(\theta)$, and $\hat{D}_n(\theta)$ is the left/right derivative of $\hat{G}_{jn}(\theta)$ with respect to θ_2. The variance matrix (C.1) accounts for the additional variation induced by the term

$$\left(\hat{C}_{jn}(\theta)\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta) - \hat{D}_{jn}(\theta)\right)(\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{G}_{n,2}(\theta))^{-1}\hat{G}_{n,2}(\theta)'\hat{\Sigma}_n(\theta)^{-1}\hat{m}_n(\theta)$$

in (2.4). For testing the full parameter vector with $d \geq 2$, Andrews and Guggenberger (2017) show that the CLR tests based on the Jacobian-variance weighting may not lead to tests with correct asymptotic size in general. Andrews and Guggenberger (2017) also show that the rank statistic of Robin and Smith (2000) weighted by $\hat{\Sigma}_n(\theta)^{-1/2}$, the inverse square root of the sample moment variance matrix (moment-variance weighting), as considered by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012), leads to tests with asymptotically correct size but may entail a power loss.

For the BLP application – a nonlinear IV model with a single endogenous variable – in addition to the CLR statistic (2.19) (or (2.22)), we will use a simpler rank statistic in which exogenous variables that do not affect the identification of the price coefficient are partialled out. Let $\bar{Z} = M_X Z = [\bar{z}_1, \ldots, \bar{z}_J]'$ with $\bar{z}_j = z_j - Z'X(X'X)^{-1}x_j, j = 1, \ldots, J$. The robust Jacobian corresponding to the sample moment function (3.20) is

$$\hat{H}^M_{J,\alpha}(\theta) = \hat{G}^M_{J,\alpha}(\theta) - \hat{C}^M_{J,\alpha}(\theta)\hat{\Sigma}^M_J(\theta)^{-1}\bar{m}^M_J(\theta), \tag{C.2}$$

where

$$\begin{aligned}
&\hat{G}^M_{J,\alpha}(\theta) = J^{-1}\bar{Z}'p, \quad \hat{C}^M_{J,\alpha}(\theta) = J^{-1}\sum_{j=1}^J \bar{z}_j p_j \xi_j \bar{z}_j', \quad \hat{\Sigma}^M_J(\theta) = J^{-1}\sum_{j=1}^J \bar{z}_j \bar{z}_j' \xi_j^2, \\
&\bar{m}^M_J(\theta) = \hat{m}^M_J(\theta) - \hat{G}^M_{J,\sigma^2}(\theta)(\hat{G}^M_{J,\sigma^2}(\theta)'\hat{\Sigma}^M_J(\theta)^{-1}\hat{G}^M_{J,\sigma^2}(\theta))^{-1}\hat{G}^M_{J,\sigma^2}(\theta)'\hat{\Sigma}^M_J(\theta)^{-1}\hat{m}^M_J(\theta), \\
&\hat{G}_{J,\sigma^2}(\theta) = J^{-1}\bar{Z}'\frac{\partial\delta(\sigma)}{\partial\sigma^{2\prime}}.
\end{aligned}$$

From the simplified moment condition (3.20) and the corresponding Jacobian (C.2), a rank statistic $R_{\alpha M}(\theta_{01})$, an analog of $R_\alpha(\theta_{01})$, can be constructed. We may also define a simpler rank statistic

$$R_{\alpha M1}(\theta_{01}) = J\,\hat{H}^M_{J,\alpha}(\tilde{\theta})'\hat{V}^M_{HJ}(\tilde{\theta})^{-1}\hat{H}^M_{J,\alpha}(\tilde{\theta}), \tag{C.3}$$

where

$$\begin{aligned}
\hat{V}^M_{HJ}(\theta) &= \hat{\Sigma}^M_{J,G_\alpha G_\alpha}(\theta) - \hat{C}^M_{J,\alpha}(\theta)\hat{\Sigma}^M_J(\theta)^{-1/2}M_{\hat{\Sigma}^M_J(\theta)^{-1/2}\hat{H}_{J,\sigma^2}(\theta)}\hat{\Sigma}^M_J(\theta)^{-1/2}\hat{C}^M_{J,\alpha}(\theta)', \\
\hat{\Sigma}^M_{J,G_\alpha G_\alpha}(\theta) &= J^{-1}\sum_{j=1}^J \bar{z}_j\bar{z}_j' p_j^2 - J^{-1}\bar{Z}'pp'\bar{Z}.
\end{aligned}$$

Since CLR-type statistics based on the three rank statistics $R_\alpha(\theta_{01})$, $R_{\alpha M}(\theta_{01})$, and $R_{\alpha M1}(\theta_{01})$ could have different properties, we investigate all versions in simulations.

Table 3 shows the performance of all versions of our tests for cost shifters with a single random coefficient. It includes two alternative doubly-robust CLR-type statistics: a $CLR_{\alpha M}$ statistic computed from the simplified moment function and a $CLR_{\alpha M1}$ statistic also computed from (3.20) where the rank statistic R_α is replaced by the simplified rank statistic $R_{\alpha M1}$ defined in (C.3). We show the rejection rates for the true value of $\alpha = 2$, as well as testing power for $\alpha = 1.75$ and $\alpha = 2.25$.

The results confirm the good performance of the CLR-type statistics and the weaker power of the AR_α statistic. Even with the potential for greater accuracy, the properties of the $CLR_{\alpha M}$ and $CLR_{\alpha M1}$ tests are nearly identical to those of the CLR_α. Interestingly, the $CLR_{\alpha M1}$ places more weight on the C_α statistic, providing better power when testing larger values of α, but less when testing smaller values.

D Eigenvalue adjustment

The eigenvalue adjustment procedure of Andrews and Guggenberger (2015a) is described as follows. Let A be a nonzero positive semidefinite matrix of dimension $p \times p$, and $\varepsilon > 0$ be some constant. The spectral decomposition of A is given by $A = \Gamma\Delta\Gamma'$, where $\Delta = \text{diag}(\lambda_1, \ldots, \lambda_p)$, $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, is the diagonal matrix that consists of the eigenvalues of A, and Γ is an orthogonal matrix of the corresponding eigenvectors. The eigenvalue adjusted matrix is defined as $A^\varepsilon = \Gamma\Delta^\varepsilon\Gamma'$ where $\Delta^\varepsilon = \text{diag}(\max\{\lambda_1, \lambda_1\varepsilon\}, \ldots, \max\{\lambda_p, \lambda_1\varepsilon\})$. The matrix A^ε enjoys a number of important properties (see Andrews and Guggenberger (2015a)).

E Details on estimation and simulations

The model is estimated by MPEC and sparse grid integration. Dubé, Fox, and Su (2012) showed that MPEC displays speed and convergence improvement over the traditional Nested Fixed Point algorithm using contraction mapping which is often used with looser tolerance level. The normal distribution of β is estimated using sparse-grid integration, which Skrainka and Judd (2011) showed to be faster, more precise and less prone to convergence to local minima than Monte Carlo integration. For the cost shifters, we use a 7-node approximation for the creation of the data and for the estimation to avoid sampling error. We parameterize the model in terms of σ^2 to avoid the problem of reduced-rank Jacobian and improve convergence properties. The code was written by Dubé, Fox, and Su (2012) and modified by Ketz (2017) and ourselves. The need to estimate the model with α constrained far away from its true value makes

Table 3: Rejection rates for all doubly-robust statistics (%)

Bound. param.	Yes						No					
Endog. of price	Low			High			Low			High		
α_0	1.75	2.00	2.25	1.75	2.00	2.25	1.75	2.00	2.25	1.75	2.00	2.25
Strong, regular instruments												
AR_α	80.5	3.4	18.3	84.9	3.6	14.1	73.2	3.2	15.9	78.5	3.3	14.3
C_α	78.8	4.4	35.9	81.2	4.3	30.6	71.8	4.0	33.2	75.5	4.1	29.4
CLR_α	83.0	3.7	26.9	86.7	3.8	20.7	75.9	3.4	23.8	80.8	3.5	20.4
$CLR_{\alpha M}$	82.7	3.7	26.4	86.5	3.7	20.6	75.8	3.4	23.7	80.4	3.2	19.8
$CLR_{\alpha M1}$	82.2	3.9	32.6	85.2	3.9	26.8	75.9	3.7	29.6	79.6	3.7	26.4
Weak, regular instruments												
AR_α	3.3	3.2	3.0	2.8	2.5	2.4	3.1	2.9	2.9	2.6	2.4	2.3
C_α	3.8	4.0	4.3	4.5	4.6	4.7	4.1	4.1	4.5	4.6	4.6	4.8
CLR_α	3.2	3.2	3.2	3.0	2.8	2.6	3.3	3.1	3.1	3.0	2.6	2.5
$CLR_{\alpha M}$	3.1	2.9	2.9	2.7	2.5	2.4	3.2	2.9	2.8	2.6	2.4	2.4
$CLR_{\alpha M1}$	3.3	3.0	3.1	3.1	2.8	2.8	3.2	3.2	3.1	3.0	2.8	2.7
Uninformative, regular instruments												
AR_α	3.0	3.2	2.9	2.7	2.6	2.5	3.0	2.9	3.0	2.6	2.4	2.4
C_α	3.7	4.0	4.4	4.6	5.0	5.2	4.2	4.3	4.4	4.8	4.7	4.8
CLR_α	3.1	3.1	3.0	3.0	2.9	2.8	3.2	3.1	3.0	2.8	2.7	2.7
$CLR_{\alpha M}$	2.9	2.8	2.7	2.6	2.5	2.5	3.1	3.0	2.8	2.6	2.4	2.5
$CLR_{\alpha M1}$	3.2	3.0	3.0	3.1	3.1	3.1	3.1	3.1	2.9	2.8	2.8	3.0

Note: Nominal sig. level: 0.05. Low endogeneity: $\rho = 0.3$; high endogeneity: $\rho = 0.8$. Boundary parameter: $\sigma^2 = 0$; no boundary parameter: $\sigma^2 = 1$. Strong instruments: $\kappa = 3$; weak instruments: $\kappa = 0.055$.

convergence of the MPEC algorithm significantly more challenging than for $\alpha = \alpha_0$. To improve convergence of the algorithm, we bound market shares away from zero to avoid rank degeneracy of the Jacobian and Hessian matrices in intermediate steps. Also, logit probability market shares cannot be computed if $\delta_{jt} + x'_{jt}\Sigma_\beta^{1/2} v$ takes exceedingly large values for some j, t because of exponentiation. Since these quantities enter the logit fractions at the numerator and denominator, the problem can be avoided by appropriate scaling.

The estimation is done by CU-GMM. As previously mentioned, the CU-GMM auxiliary estimate is not necessary for the validity of our robust statistics, but it is necessary for the non-boundary-robust statistics to which we compare them. The starting values are based on a two-stage least squares estimation of a homogeneous model and a starting guess for σ^2 based on the magnitude of the OLS estimate of b. Two other random starting points are used to avoid the convergence to local minima which arises when testing values of α that are far from its true value. For the power curves shown in Figure 1, instead of the first stage estimates, we use the estimates of a neighboring value of α as starting values to increase convergence speed.

For simulations with BLP instruments, we generate the data and compute equilibrium prices by adapting the program written by Armstrong (2016a). Estimation is done with the same algorithm as for the cost shifters simulations. To minimize the risk of simulation errors, the Monte Carlo integration is done with a 100 support point distribution instead of the original 10.

Rejection rates are computed using 5,000 replications for all specifications, except those with two random coefficients and with BLP instruments, which use 1,000 replications to avoid prohibitive computing time.

F Simulation results, non-boundary-robust statistics

This section presents simulation results for non-boundary-robust statistics using cost shifters in Table 4 and BLP instruments in Table 5.

Table 4: Rejection rates at true value, non-boundary-robust statistics (%)

	One random coefficient				Two random coefficients			
Boundary param.	Yes		No		Yes		No	
Endog. of price	Low	High	Low	High	Low	High	Low	High
Strong instruments								
AR	5.78	5.80	3.56	3.77	13.40	12.44	9.00	10.40
LM	13.12	13.42	6.96	8.33	26.10	27.48	17.40	18.10
CLR	8.26	8.30	4.64	5.08	15.90	15.65	10.20	10.70
Weak instruments								
AR	5.68	4.68	4.10	3.58	8.60	7.10	4.71	4.10
LM	8.36	9.62	6.28	7.90	14.80	15.10	11.01	12.20
CLR	5.44	5.00	4.34	4.00	8.30	7.00	4.71	4.20
Uninformative instruments								
AR	5.64	4.80	4.14	3.60	8.10	7.10	3.80	3.90
LM	8.26	10.02	6.60	8.38	14.10	14.70	12.00	10.50
CLR	5.46	5.20	4.44	3.92	7.80	6.60	4.00	4.40

Note: Nominal sig. level: 0.05. Low endog.: $\rho = 0.3$; high endog.: $\rho = 0.8$. Boundary parameter: $\sigma^2 = 0$; no boundary parameter: $\sigma^2 = 1$. Strong instruments: $\kappa = 3$; weak instruments: $\kappa = 0.055$; uninformative instruments: $\kappa = 0$.

Table 5: Simulations for BLP instruments, Non-boundary-robust statistics

Number. of markets	1		3				20			
Products per market	60	100	16 20 24	20 60 100	48 60 72	80 100 120	16 20 24	20 60 100	48 60 72	80 100 120
$\sigma^2 = 9$, null rejection rate										
AR	3.3	2.7	3.4	3.5	3.9	4.7	5.6	5.7	4.8	6.4
LM	7.4	7.6	6.6	6.5	6.9	6.4	4.8	6.7	5.4	4.9
CLR	3.5	3.1	3.6	3.6	4.0	4.5	5.4	5.7	4.8	6.7
$\sigma^2 = 9$, power of test $\alpha = 0$										
AR	4.4	3.4	68.2	47.2	19.2	13.3	100.0	100.0	98.7	85.0
LM	11.8	9.2	45.3	45.5	21.9	14.4	61.2	89.9	58.8	55.3
CLR	5.6	4.2	68.1	48.1	19.6	13.0	100.0	100.0	98.6	85.2
$\sigma^2 = 0$ (bound. param.), null rejection rate										
AR	5.4	5.4	5.4	5.7	6.3	6.6	7.5	7.4	8.0	6.4
LM	12.4	10.2	11.0	9.4	9.7	8.8	5.8	5.8	6.8	6.5
CLR	5.7	6.4	6.2	6.2	6.5	6.8	7.7	7.2	7.6	6.2
$\sigma^2 = 0$ (bound. param.), power of test $\alpha = 0$										
AR	10.4	8.3	83.6	74.4	35.9	22.7	100.0	100.0	99.9	96.2
LM	16.0	13.7	50.7	72.6	30.5	22.2	72.8	99.8	62.6	56.1
CLR	10.7	8.9	83.5	76.0	35.5	23.4	100.0	100.0	99.9	96.0

Note: Nominal sig. level: 0.05. Specifications with multiple markets contain 1/3 of markets of each different size. All specifications have 1/3 of firms of size 2, 1/3 of size 5 and 1/3 of size 10.