Journal of Econometrics 193 (2016) 123–146

Contents lists available at ScienceDirect

Journal of Econometrics

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

Fixed-smoothing asymptotics in the generalized empirical likelihood estimation framework

Xianyang Zhang

Department of Statistics, Texas A&M University, College Station, TX 77843, United States

ARTICLE INFO

Article history:
Received 7 March 2014
Received in revised form
12 November 2015
Accepted 2 January 2016
Available online 27 February 2016

Keywords:
Blockwise empirical likelihood
Continuous updating estimator
Fixed-smoothing asymptotics
Generalized empirical likelihood
Heteroscedasticity and autocorrelation consistent
Over-identification test
Saddle point problem
Specification test

ABSTRACT

This paper concerns the fixed-smoothing asymptotics for two commonly used estimators in the generalized empirical likelihood estimation framework for time series data, namely the continuous updating estimator and the maximum blockwise empirical likelihood estimator. For continuously updating generalized method of moments (GMM) estimator, we show that the results for the two-step GMM estimator in Sun (2014a) continue to hold under suitable assumptions. For continuous updating estimator obtained through solving a saddle point problem (Newey and Smith, 2004) and the maximum blockwise empirical likelihood estimator (Kitamura, 1997), we show that their fixed-smoothing asymptotic distributions (up to an unknown linear transformation) are mixed normal. Based on these results, we derive the asymptotic distributions of the specification tests (including the over-identification testing and testing on parameters) under the fixed-smoothing asymptotics, where the corresponding limiting distributions are nonstandard yet pivotal. Simulation studies show that (i) the fixed-smoothing asymptotics provides better approximation to the sampling distributions of the continuous updating estimator and the maximum blockwise empirical likelihood estimator as compared to the standard normal approximation. The testing procedures based on the fixed-smoothing critical values are more accurate in size than the conventional chi-square based tests; (ii) the continuously updating GMM estimator is asymptotically more efficient and the corresponding specification tests are generally more powerful than the other two competitors. Finite sample results from an empirical data analysis are also reported.

1. Introduction

The generalized method of moments (GMM) has been a standard tool to estimate model parameters in the moment restriction models. However, it has been recognized that there are considerable problems with the GMM particularly in its finite sample performance, such as bias in point estimation, and size and power distortions in hypothesis testing (see e.g. Altonji and Segal (1996); Hansen et al. (1996)). In an effort to improve the finite sample properties of GMM, alternative estimation and inference methods have been developed such as the continuous updating GMM (Hansen et al., 1996), exponential tilting (Kitamura and Stutzer, 1997; Imbens et al., 1998), and empirical likelihood (EL) (Owen, 1988, 1990; Qin and Lawless, 1994), which were later integrated into a unified framework called the generalized empirical likelihood (GEL) by Newey and Smith (2004). The GEL estimator has been shown to enjoy better finite sample and large sample properties for independent data. In particular, the EL estimator and its associated test statistics have been shown to possess some desirable theoretical properties such as the Bartlett correctability (DiCiccio et al., 1991), the higher order efficiency (Newey and Smith, 2004) and the asymptotic efficiency in terms of large deviation (Kitamura, 2001; Kitamura et al., 2013).

Given these encouraging findings, the GEL methods have been extended to deal with moment restrictions defined based on a sequence of stationary time series (Smith, 2011), where kernel smoothing or data blocking schemes are employed to capture the underlying dependence structure. However, the performance of GEL can depend crucially on the choice of the bandwidth parameter involved in kernel smoothing or the block size of the data blocking scheme (Nordman, 2009; Nordman et al., 2013), for which no sound guidance seems available in the literature. To better reflect the finite sample situation and capture the choice of the bandwidth parameter or block size, one proposal is to embed the finite sample situation in a different limiting thought experiment, where the amount of smoothing (controlled by the bandwidth parameter or block size) is held fixed as the sample size grows. We shall call this type of asymptotics the fixed-smoothing asymptotics as compared to

E-mail address: zhangxiany@stat.tamu.edu.

http://dx.doi.org/10.1016/j.jeconom.2016.01.009
0304-4076/

the conventional increasing-smoothing asymptotics, where the amount of smoothing is assumed to be increasing with the sample size. Using the higher-order Edgeworth expansions, Jansson (2004), Sun et al. (2008), Sun (2014b) and Zhang and Shao (2013) rigorously proved that the fixed-smoothing asymptotics provides a high order refinement over the traditional increasing-smoothing asymptotics in the Gaussian location model. Because of its nice theoretical and finite sample properties, there is a growing body of literature on the development of inferential procedures under the fixed-smoothing asymptotics. See e.g., Kiefer and Vogelsang (2005), Sun (2011, 2013, 2014a), Sun and Kim (2012), Ibragimov and Müller (2010), Bester et al. (2011), Vogelsang (2012), Gonçaves (2011), Gonçalves and Vogelsang (2011), Shao (2010), and references therein. Most of the recent developments on the fixed-smoothing asymptotics have been devoted to the GMM framework and relatively few attention has been paid to the GEL estimation framework. Existing works (Nordman et al., 2013; Zhang and Shao, 2014, 2016) only concern with the asymptotic behavior of the EL ratio test statistic evaluated at the true parameter.

In this paper, we establish the fixed-smoothing asymptotics for two commonly used estimators in the GEL class namely the continuous updating estimator (CUE) and the maximum blockwise empirical likelihood estimator (MBELE), and their associated test statistics. As far as we are aware, these results are new. It is worth pointing out that our asymptotic results rely on some nonstandard arguments. On one hand, the standard arguments for GEL under the increasing-smoothing asymptotics (Smith, 2011) cannot be applied directly in the current setting. On the other hand, the criterion function and the convex hull constraint in the formulation of GEL bring technical difficulties when studying the asymptotic behavior of the GEL estimator under the fixed-smoothing asymptotics. It thus requires substantially different treatment as compared to that in Sun (2014a) for the two-step GMM estimator.

Following Hansen et al. (1996), the CUE can be formulated as a solution to a minimization problem, where the weighting matrix is continuously altered as the parameter is changed in the minimization. We shall call the CUE based on a minimization problem (with demeaned moment conditions) the continuously updating GMM (CU-GMM, hereafter) estimator. On the other hand, as shown in Newey and Smith (2004), in the independent and identically distributed (i.i.d) case, CUE is also a solution to a saddle point problem and thus is in the GEL class. It is worth noting that in the dependence case, these two formulations are not exactly the same. The CU-GMM involves the choice of a demeaned weighting matrix which is similar to the optimal weighting matrix involved in the two-step GMM approach. While the saddle point problem formulation requires the choice of smooth moment conditions or blocking schemes to preserve the underlying dependence among neighboring time observations. For CU-GMM estimator, we show that its fixed-smoothing limiting distribution is asymptotically mixed normal, which is consistent with the recent result in Sun (2014a) for the two-step GMM estimator. We study the fixed-smoothing asymptotics for the over-identification test and the Wald test in the CU-GMM framework. Under suitable assumptions, we are able to obtain similar results as those in Sun and Kim (2012) and Sun (2014a). For CUE based on the saddle point problem formulation, we show that its fixed-smoothing limit (up to an unknown linear transformation) is mixed normal. Based on this result, we derive the fixed-smoothing asymptotics for the over-identification test and the likelihood ratio (LR) type test. By rotating and transforming the limiting distributions, we show that both the over-identification test and the LR type test have nonstandard pivotal limiting distributions which can be approximated numerically.

Furthermore, we investigate the fixed-smoothing asymptotic behavior of the MBELE based on fully overlapping moment conditions (Kitamura, 1997). Surprisingly, under the fixed-smoothing asymptotics, the MBELE (based on fully overlapping moment conditions) is no longer consistent. The inconsistency result is closely related to the recent finding in Zhang and Shao (2016) which states that with a positive probability, the origin is outside the convex hull of the moment conditions evaluated at the true parameters even for large sample size. Nevertheless, we are able to derive the asymptotic distribution of the MBELE conditional on an event $\mathcal{D}_b$ with $P(\mathcal{D}_b) < 1$. Loosely speaking, conditional on $\mathcal{D}_b$, the origin is contained in the convex hull of the moment conditions when the sample size is large enough (in another word, the convex hull constraint is satisfied for large sample). The asymptotic distribution of the MBELE can be characterized by a mixed normal distribution. Based on these findings, we establish the fixed-smoothing asymptotics for the over-identification test and the LR type test in the blockwise empirical likelihood (BEL) framework, where the corresponding limiting distributions are nonstandard yet pivotal.

From a practical viewpoint, the contributions of the paper are two-folds: (i) *More accurate approximation to finite sample distribution*: we establish the fixed-smoothing asymptotics for CU-GMM, CUE based on the saddle point problem, and BEL. Through extensive simulations, it was demonstrated that the fixed-smoothing asymptotics provides a more accurate approximation than the conventional increasing-smoothing asymptotics in many situations. This is especially true if the degree of over-identification is large and the underlying dependence is strong; (ii) *Some guidance on the use of GEL estimators and GEL-based tests*: Under the fixed-smoothing asymptotics, we show that the CU-GMM estimator with the Bartlett kernel has smaller asymptotic variance compared to CUE and MBELE based on fully overlapping blocks. The corresponding CU-GMM based tests are asymptotically more powerful than the CUE and BEL based tests. A comparison between the CU-GMM based test and the CUE-based test reflects a size and power trade-off. The CUE-based test has more accurate size while it suffers from (asymptotic) power loss. The CU-GMM based test delivers higher power but its size is more distorted under strong dependence. Our numerical results also suggest that the BEL-based tests could be less preferable in practice due to relatively large upward size distortion and (asymptotic) power loss as compared to the CU-GMM based tests. To make the use of the fixed-smoothing asymptotics easy for practitioners, we tabulated the critical values for the specification tests for leading cases in Tables S.1–S.3 in the supplement, and we also provide the choice of bandwidth parameter and block size based on the MSE criterion, see Section 6.3.

The layout of the article is as follows. Section 2 establishes the fixed-smoothing asymptotics for the CU-GMM estimators and the CU-GMM based test statistics. In Section 3, we study the CUE based on the saddle point problem formulation and derive the fixed-smoothing asymptotics for the associated test statistics. Section 4 studies the fixed-smoothing asymptotics in the BEL framework. We compare the three methods in Section 5 in terms of asymptotic variance and local asymptotic power. Section 6 is devoted to the sampling distributions of CU-GMM estimator, CUE and MBELE, and the finite sample performance of the associated test statistics. An empirical data analysis is presented in Section 7. Section 8 concludes. The Appendix A contains the proofs for each of these sections. Additional technical details and numerical results are provided in a supplementary material (see Appendix B).

For notation, let $D[0, 1]$ be the space of functions on $[0, 1]$ which are right-continuous and have left limits, endowed with the Skorokhod topology (Billingsley, 1999). Weak convergence in $D[0, 1]$ or more generally in the $\mathbb{R}^{q_0}$-valued function space $D^{q_0}[0, 1]$ is denoted by "$\Rightarrow$", where $q_0 \in \mathbb{N}$. For any compact set

K, the space $l^\infty(K)$ is defined as the set of all uniformly bounded, real functions on K (van der Vaart and Wellner, 1996). Convergence in probability, almost sure convergence, and convergence in distribution are denoted by "$\to^p$", "$\to^{a.s.}$", and "$\to^d$" respectively. Denote by "$=^d$" equal in distribution. Let $\lfloor a \rfloor$ be the integer part of $a \in \mathbb{R}$. We use $|\cdot|$ to denote the Euclidean norm and use $\|\cdot\|$ to denote the matrix spectral norm.

2. Continuously updating GMM

2.1. Fixed-smoothing asymptotics for CU-GMM

Suppose we are interested in the estimation and inference of a p-dimensional parameter vector θ, which is identified by a set of moment conditions/restrictions. Denote by $\theta_0 = (\theta_{01}, \dots, \theta_{0p})' \in \mathbb{R}^p$ the true parameter of θ, an interior point of a compact parameter space $\Theta \subseteq \mathbb{R}^p$. Let $\{y_t\}_{t=1}^n$ be a sequence of $\mathbb{R}^v$-valued stationary time series and assume that the moment conditions

$$\mathbb{E}[f(y_t, \theta)] = 0, \quad t = 1, 2, \dots, n, \tag{1}$$

hold if and only if $\theta = \theta_0$, where $f(y_t, \theta) : \mathbb{R}^{v+p} \to \mathbb{R}^k$ is a map which is differentiable with respect to θ with $k \geq p$. Let $f_t(\theta) = f(y_t, \theta)$, $g_t(\theta) = \partial f_t(\theta)/\partial\theta' = (g_{t1}(\theta), \dots, g_{tp}(\theta)) \in \mathbb{R}^{k\times p}$, $\bar{f}_t(\theta) = \sum_{j=1}^t f_j(\theta)/t$ and $\bar{g}_t(\theta) = \sum_{j=1}^t g_j(\theta)/t$. Define $G(\theta) = \mathbb{E}[g_j(\theta)]$ and $G_0 = G(\theta_0)$. The commonly used estimator of θ is the two-step GMM estimator (Hansen, 1982) which is defined as

$$\hat{\theta}_{gmm} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \bar{f}_n(\theta)' V_n^{-1}(\check{\theta}_n) \bar{f}_n(\theta), \tag{2}$$

where $V_n(\check{\theta}_n)$ is a positive semi-definite weighting matrix which depends on some initial first step estimator $\check{\theta}_n$. Though the two-step GMM estimator is asymptotically optimal, Monte Carlo evidence indicated that it may be severely biased in small samples (see e.g. Altonji and Segal (1996)). As pointed out in Newey and Smith (2004), the poor bias properties of the two-step and iterated GMM estimators arise through the estimation of the Jacobian and efficient metric in the GMM criterion function. To reduce the finite sample bias, Hansen et al. (1996) proposed the CU-GMM, where the weighting matrix is continuously altered as the parameter is changed in the minimization. Below we study the fixed-smoothing asymptotics for the CU-GMM estimators based on suitable weighting matrix which takes the data dependence into account. To fix the idea, let $\mathcal{K}(\cdot,\cdot)$ be a positive semi-definite bivariate kernel. Consider

$$\Omega_n(\theta) = \frac{1}{n}\sum_{i,l=1}^n \mathcal{K}(i/n, l/n)(f_i(\theta) - \bar{f}_n(\theta))(f_l(\theta) - \bar{f}_n(\theta))', \tag{3}$$

that includes a large class of estimators as special cases such as the heteroskedasticity and autocorrelation consistent (HAC) covariance estimator (Kiefer and Vogelsang, 2005), and the orthonormal series variance estimator (Sun, 2011, 2013). In the econometrics and statistics literature, the bivariate kernel is usually defined through a positive semi-definite univariate kernel $\tilde{\mathcal{K}}(\cdot)$, that is, $\mathcal{K}(r_1, r_2) = \tilde{\mathcal{K}}((r_1 - r_2)/b_0)$ with b_0 being the bandwidth parameter. Here we suppose b_0 is fixed in the asymptotics and consider a general formulation based on the bivariate kernel $\mathcal{K}(\cdot,\cdot)$.

The CU-GMM estimator based on $\Omega_n(\theta)$ is defined as,

$$\hat{\theta}_{cue} = (\hat{\theta}_{cue,1}, \dots, \hat{\theta}_{cue,p})' = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \bar{f}_n'(\theta)\Omega_n^{-1}(\theta)\bar{f}_n(\theta). \tag{4}$$

The main contribution in this section is the establishment of the consistency and the asymptotic distribution of $\hat{\theta}_{cue}$, which requires considerable modifications of the arguments in Sun (2014a) and Sun and Kim (2012). To facilitate the theoretical derivation, we make the following assumptions.

Assumption 2.1. Assume that

$$\sup_{\theta\in\Theta}\sup_{1\le t\le n}\left|\frac{1}{n}\sum_{j=1}^t \{f_j(\theta) - \mathbb{E}f_j(\theta)\}\right| = o_p(1).$$

Assumption 2.2. Define $h(\theta) = \mathbb{E}f(y, \theta)$ with $y =^d y_i$. Assume that $h(\theta) = 0$ if and only if $\theta = \theta_0$, and $h(\theta)$ is continuous on Θ.

Assumption 2.3. Assume that $\sup_{1\le t\le n} \|\sum_{j=1}^t g_j(\tilde{\theta})/n - (t/n)G(\tilde{\theta})\| = \tau_n \to^p 0$ uniformly for all $\tilde{\theta}$ in $\mathcal{N}_{\epsilon^*}(\theta_0)$, where $\mathcal{N}_{\epsilon^*}(\theta_0) = \{\theta \in \mathbb{R}^p : |\theta - \theta_0| \le \epsilon^*\}$ with $\epsilon^* > 0$. Suppose that $G(\theta)$ is Lipschitz continuous on $\mathcal{N}_{\epsilon^*}(\theta_0)$ and $\text{rank}(G_0) = p$.

Assumption 2.4. Assume that $\sum_{t=1}^{\lfloor nr\rfloor} f_t(\theta_0)/\sqrt{n} \to^d \Lambda W_k(r)$, where $W_k(r)$ is a k-dimensional vector of independent standard Brownian motions, and $\Delta = \Lambda\Lambda' = \sum_{j=-\infty}^{+\infty} \mathbb{E}[f_t(\theta_0)f_{t+j}(\theta_0)']$ is the long run variance matrix of the moment conditions evaluated at θ_0.

Assumption 2.5. Let $\mathcal{K}_{k,l} = \mathcal{K}_{k,l}^{(n)} = \mathcal{K}(k/n, l/n)$ and $c_{k,l} = \mathcal{K}_{k,l} - \mathcal{K}_{k+1,l} - \mathcal{K}_{k,l+1} + \mathcal{K}_{k+1,l+1}$. Assume that $\sum_{k,l=1}^{n-1}|c_{k,l}| = O(1)$.

Assumption 2.1 requires that the uniform weak law of large number holds for the moment conditions. It is stronger than the usual requirement that $\sup_{\theta\in\Theta}\left|\sum_{j=1}^n \{f_j(\theta) - \mathbb{E}f_j(\theta)\}/n\right| = o_p(1)$ in the i.i.d case, because $\Omega_n(\theta_0)$ is allowed to converge to a stochastic limit under the fixed-smoothing asymptotics (Assumption 2.1 is imposed to bound $\Omega_n(\theta)$ from above). Assumption 2.1 can be verified under the more primitive conditions [see Lemma S.1 and Remark S.1 in the supplement]. Assumption 2.2 imposes an identification condition, which is required when proving the consistency. Assumptions 2.3–2.4 are standard in the literature (see e.g. Kiefer and Vogelsang (2005); Sun (2014a,b)). It can be verified under suitable mixing conditions (see e.g. Herrndorf (1984); Theorem 2.1 of Phillips and Durlauf (1986); Wu (2007)). Assumption 2.5 is a mild condition on the bounded variation of (the second derivative of) $\mathcal{K}(\cdot,\cdot)$. We point out that Assumption 2.5 does not rule out the Bartlett kernel, although it is non-smooth. It is also worth noting that by employing certain truncation arguments as in Sun (2014a), the functional central limit theorem in Assumption 2.4 might be weakened by just assuming the central limit theorem for the projection vectors of the moment conditions. However, to avoid complications, we do not pursue this point in the following derivations. The consistency of the CU-GMM estimator is established in the following lemma.

Lemma 2.1. *Under Assumptions 2.1, 2.2, 2.4 and 2.5,* $\hat{\theta}_{cue} \to^p \theta_0$.

Let $\mathcal{K}^*(r, s) = \mathcal{K}(r, s) - \int_0^1 \mathcal{K}(r_0, s)dr_0 - \int_0^1 \mathcal{K}(r, s_0)ds_0 + \int_0^1\int_0^1 \mathcal{K}(r_0, s_0)dr_0 ds_0$ be the demeaned kernel, and $B_k(r) = W_k(r) - rW_k(1)$ be a k-dimensional vector of independent Brownian bridges. For any constant $c > 0$, define the set $\mathcal{B}_c = \{\theta : |\theta - \theta_0| \le cn^{-1/3}\}$ and the event

$$\mathcal{A}_c = \{\text{there is a minimizer in the interior of } \mathcal{B}_c\}.$$

We present the main result in this section regarding the asymptotic distribution of the CU-GMM estimator.

Theorem 2.1. *Assume that $\tau_n = o_p(n^{-1/6})$ in Assumption 2.3. Under Assumptions 2.1–2.5, $P(\mathcal{A}_c) \to 1$ for any $c > 0$. Denote by $\hat{\theta}_{cue}$ the minimizer in the interior of $\mathcal{B}_c$, which satisfies the first order condition given in (37). Then we have*

$$\mathcal{S}_n := \sqrt{n}(\hat{\theta}_{cue} - \theta_0) \to^d \mathcal{S} := -(G_0'Q_0^{-1}G_0)^{-1}G_0'Q_0^{-1}\Lambda W_k(1), \tag{5}$$

where $Q_0 = \Lambda \int_0^1 \int_0^1 \mathcal{K}(r,s) dB_k(r) dB_k(s)' \Lambda' = \Lambda \int_0^1 \int_0^1 \mathcal{K}^*(r,s) dW_k(r) dW_k(s)' \Lambda'$, *and* $W_k(1)$ *and* Q_0 *are independent. Conditional on* Q_0, $\sqrt{n}(\hat{\theta}_{cue} - \theta_0)$ *converges to a mean-zero Gaussian distribution with the covariance matrix*

$$(G_0' Q_0^{-1} G_0)^{-1} G_0' Q_0^{-1} \Delta Q_0^{-1} G_0 (G_0' Q_0^{-1} G_0)^{-1}. \tag{6}$$

Theorem 2.1 shows that the CU-GMM estimator converges to a mixed normal distribution, which is consistent with the recent finding in Sun (2014a) for the two-step GMM estimator. It thus suggests that the CU-GMM estimator and the two-step GMM estimator are first order equivalent under the fixed-smoothing asymptotics.

2.2. CU-GMM based test statistics

In this subsection, we study the asymptotic behavior of the over-identification test and the Wald test in the framework of CU-GMM. Define

$$\mathcal{L}_{cue,n}(\theta) = n\bar{f}_n(\theta)' \Omega_n^{-1}(\theta) \bar{f}_n(\theta). \tag{7}$$

To test the validity of the moment restrictions, we consider the over-identification test statistic which is given by

$$J_{cue,n} = \mathcal{L}_{cue,n}(\hat{\theta}_{cue}) = \min_{\theta \in \Theta} n\bar{f}_n(\theta)' \Omega_n^{-1}(\theta) \bar{f}_n(\theta). \tag{8}$$

By Assumption 2.3, Theorem 2.1 and the Taylor expansion, we deduce that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} f_j(\hat{\theta}_{cue}) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} f_j(\theta_0) + \frac{1}{n} \sum_{j=1}^{\lfloor nr \rfloor} g_j(\tilde{\theta}) \sqrt{n}(\hat{\theta}_{cue} - \theta_0) \\ &\to^d \Lambda W_k(r) - r G_0 (G_0' Q_0^{-1} G_0)^{-1} G_0' Q_0^{-1} \Lambda W_k(1), \end{aligned}$$

where $\tilde{\theta}$ is on the line segment joining $\hat{\theta}_{cue}$ and θ_0, and it can vary across equations. It thus implies that $\sum_{j=1}^{\lfloor nr \rfloor} f_j(\hat{\theta}_{cue})/\sqrt{n} - r\sum_{j=1}^{n} f_j(\hat{\theta}_{cue})/\sqrt{n} \to^d \Lambda B_k(r)$. Using summation and integration by parts, and the continuous mapping theorem, we have

$$\Omega_n(\hat{\theta}_{cue}) \to^d Q_0 = \Lambda \int_0^1 \int_0^1 \mathcal{K}(r,s) dB_k(r) dB_k(s)' \Lambda'. \tag{9}$$

Again by the continuous mapping theorem, we derive that

$$\begin{aligned} J_{cue,n} \to^d J_\infty &= \{W_k(1) - \Lambda^{-1} G_0 (G_0' Q_0^{-1} G_0)^{-1} G_0' Q_0^{-1} \Lambda W_k(1)\}' \\ &\times \tilde{Q}_0^{-1} \{W_k(1) - \Lambda^{-1} G_0 (G_0' Q_0^{-1} G_0)^{-1} G_0' Q_0^{-1} \Lambda W_k(1)\}, \end{aligned} \tag{10}$$

where $\tilde{Q}_0 = \int_0^1 \int_0^1 \mathcal{K}(r,s) dB_k(r) dB_k(s)'$. By Sun and Kim (2012), we have for $k > p$,

$$\begin{aligned} J_\infty &=^d \mathcal{J}_{k-p} := W_{k-p}(1)' \\ &\times \left(\int_0^1 \int_0^1 \mathcal{K}^*(r,s) dW_{k-p}(r) dW_{k-p}(s)' \right)^{-1} W_{k-p}(1), \end{aligned} \tag{11}$$

which indicates that J_∞ is pivotal and its critical values can be obtained via simulation or i.i.d bootstrap (as the functional central limit theorem holds for the i.i.d bootstrap sample). For the series variance estimator that is $\mathcal{K}(r,s) = \sum_{j=1}^{M} \phi_j(r)\phi_j(s)/M$ with $\{\phi_j\}_{j=1}^{M}$ being a sequence of mean-zero orthonormal basis, it was shown in Sun and Kim (2012) that the nonstandard limiting distribution J_∞ reduces to a scaled F distribution.

Next, we investigate the asymptotic distribution of the Wald test. Consider the linear constraint $H_0 : \mathcal{R}\theta = 0$ with $\mathcal{R} \in \mathbb{R}^{m \times p}$ and $m \le p$. The Wald statistic for testing the null hypothesis H_0 is defined as

$$\begin{aligned} \mathcal{W}_n &:= n(\mathcal{R}\hat{\theta}_{cue})' \left\{ \mathcal{R} \left[\bar{g}_n(\hat{\theta}_{cue})' \Omega_n^{-1}(\hat{\theta}_{cue}) \bar{g}_n(\hat{\theta}_{cue}) \right]^{-1} \mathcal{R}' \right\}^{-1} \\ &\times (\mathcal{R}\hat{\theta}_{cue}). \end{aligned} \tag{12}$$

By Assumption 2.3, Theorem 2.1 and (9), it is straightforward to see that,

$$\begin{aligned} \mathcal{W}_n \to^d \mathcal{W}_\infty &= \left\{ \mathcal{R}(G_0' Q_0^{-1} G_0)^{-1} G_0' Q_0^{-1} \Lambda W_k(1) \right\}' \\ &\times \left\{ \mathcal{R}(G_0' Q_0^{-1} G_0)^{-1} \mathcal{R}' \right\}^{-1} \\ &\times \left\{ \mathcal{R}(G_0' Q_0^{-1} G_0)^{-1} G_0' Q_0^{-1} \Lambda W_k(1) \right\}. \end{aligned} \tag{13}$$

With some abuse of notation, write $W_k(r) = (W_m(r)', W_{p-m}(r)', W_q(r)')'$ with $p + q = k$, where $W_m(r)$, $W_{p-m}(r)$ and $W_q(r)$ are mutually independent standard Brownian motions of dimensions m, $p - m$ and q respectively. Define

$$\begin{aligned} C_{mm} &= \int_0^1 \int_0^1 \mathcal{K}^*(r,s) dW_m(r) dW_m(s)', \\ C_{mq} &= \int_0^1 \int_0^1 \mathcal{K}^*(r,s) dW_m(r) dW_q(s)', \\ C_{qq} &= \int_0^1 \int_0^1 \mathcal{K}^*(r,s) dW_q(r) dW_q(s)', \\ C_{mm,q} &= C_{mm} - C_{mq} C_{qq}^{-1} C_{mq}'. \end{aligned}$$

By rotation and transformation, it was shown in Sun (2014a) that

$$\begin{aligned} \mathcal{W}_\infty =^d \mathcal{W}_{m,q} &:= \left\{ W_m(1) - C_{mq} C_{qq}^{-1} W_q(1) \right\}' C_{mm,q}^{-1} \\ &\times \left\{ W_m(1) - C_{mq} C_{qq}^{-1} W_q(1) \right\}, \end{aligned} \tag{14}$$

which suggests that the limiting distribution does not depend on the unknown parameter Λ and hence can be approximated numerically. It is worth mentioning that the use of the fixed-smoothing asymptotic framework does not change the estimator, or the test statistics, but changes the asymptotic approximations we use.

Remark 2.1. Let $\bar{\theta}_{cue}$ be the minimizer of $\mathcal{L}_{cue,n}(\theta)$ subject to the constraint that $R\bar{\theta}_{cue} = 0$. We define the LR type statistic as

$$\mathcal{LR}_n := \mathcal{L}_{cue,n}(\bar{\theta}_{cue}) - \mathcal{L}_{cue,n}(\hat{\theta}_{cue}). \tag{15}$$

Let $D_n(\theta) = \bar{g}_n(\theta)' \Omega_n^{-1}(\theta) \bar{f}_n(\theta)$. The Lagrange multiplier (LM) statistic is defined as

$$\mathcal{LM}_n := n D_n(\bar{\theta}_{cue})' \left\{ \bar{g}_n(\bar{\theta}_{cue})' \Omega_n^{-1}(\bar{\theta}_{cue}) \bar{g}_n(\bar{\theta}_{cue}) \right\}^{-1} D_n(\bar{\theta}_{cue}). \tag{16}$$

In view of the proof of Theorem 2 in Sun (2014a), it is expected that the three tests are first order equivalent, that is $\mathcal{W}_n = \mathcal{LR}_n + o_p(1)$ and $\mathcal{W}_n = \mathcal{LM}_n + o_p(1)$. For brevity, we leave the analysis of the LR statistic and the LM statistic to future research.

Remark 2.2. For the series variance estimator, it has been shown in Sun (2014a) that $\mathcal{W}_\infty$ is a scaled mixed noncentral F random variable with the random noncentrality parameter $|C_{mq} C_{qq}^{-1} W_q(1)|^2$. To avoid simulations from the nonstandard limiting distribution $\mathcal{W}_\infty$, Hwang and Sun (2015) recently shows that after suitable studentization using the over-identification test, the test statistics (i.e., Wald test, LR test and LM test) in the two-step GMM setting have asymptotic central F distributions. Based on the above discussions, we expect that similar results hold in the CU-GMM framework when series variance estimator is used.

3. CUE based on the saddle point problem

As an alternative to the two-step GMM estimator, the GEL estimator has attracted considerable attention because of its competitive bias properties (Newey and Smith, 2004). It was shown in Newey and Smith (2004) that in the i.i.d case, the CU-GMM estimator is a solution to a saddle point problem and thus is in the GEL class. In the dependence case, the CU-GMM estimator and the CUE based on a saddle point problem are not exactly the same as the latter is internally studentized and does not require the choice of a weighting matrix.

In this section, we focus on the asymptotics of the CUE based on the saddle point problem formulation. Though in the case of CUE, the saddle point problem reduces to a minimization problem, it is generally not the case for other members in the GEL class. We believe that our arguments shed some light on the asymptotic behaviors of the general GEL estimators under the fixed-smoothing asymptotics.

3.1. Fixed-smoothing asymptotics for CUE

To deal with time series data, we consider the fully overlapping moment condition (Kitamura, 1997) which is given by $f_{tn}(\theta) = \frac{1}{B}\sum_{j=t}^{t+B-1} f_j(\theta)$ with $t = 1, 2, \ldots, n - B + 1 := N$ and $B = \lfloor nb \rfloor$ for $b \in (0, 1)$. The fully overlapping data blocking scheme aims to preserve the underlying dependence among neighboring time observations. Note that the amount of smoothing is controlled by b which is held fixed in our asymptotic analysis. It should be pointed out that the results presented below can be extended to the smoothed moment conditions (Smith, 2011),

$$f_{tn,\mathcal{g}}(\theta) = \frac{1}{nb}\sum_{j=t-n}^{t-1} \mathcal{g}(j/(nb)) f_{t-j}(\theta),$$

with $\mathcal{g}(\cdot)$ being a kernel function, or moment conditions based on alternative blocking/scanning schemes (McElroy and Politis, 2007; Nordman et al., 2013). Define the objective function $gel_{cue}(\theta) = \max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N}\{1 - (1 + \lambda' f_{tn}(\theta))^2\}/nb$. Consider the CUE based on the saddle point problem,[1]

$$\begin{aligned} \breve{\theta}_{cue} &= \operatorname*{argmin}_{\theta\in\Theta} gel_{cue}(\theta) \\ &= \operatorname*{argmin}_{\theta\in\Theta} \left(\sum_{t=1}^{N} f_{tn}(\theta)\right)' \left(\sum_{t=1}^{N} f_{tn}(\theta) f_{tn}(\theta)'\right)^{-1} \\ &\quad \times \left(\sum_{t=1}^{N} f_{tn}(\theta)\right). \end{aligned} \tag{17}$$

We first establish the consistency and the convergence rate of $\breve{\theta}_{cue}$ under the fixed-smoothing asymptotics.

Lemma 3.1. *Under Assumptions 2.1, 2.2 and 2.4, $\breve{\theta}_{cue} \to^p \theta_0$.*

Lemma 3.2. *Under Assumptions 2.1–2.4, $\mathcal{Z}_n := \sqrt{n}(\breve{\theta}_{cue} - \theta_0)$ is uniformly tight.*

[1] As pointed out by one referee, $\breve{\theta}_{cue}$ is also the solution to the following problem:

$$\operatorname*{argmin}_{\theta\in\Theta} \left(\sum_{t=1}^{N} f_{tn}(\theta)\right)' \left(\sum_{t=1}^{N} \tilde{f}_{tn}(\theta)\tilde{f}_{tn}(\theta)'\right)^{-1} \left(\sum_{t=1}^{N} f_{tn}(\theta)\right),$$

where $\tilde{f}_{tn}(\theta) = f_{tn}(\theta) - \sum_{t=1}^{N} f_{tn}(\theta)/N$. It can be proved by showing that the objective function in (17) is a monotonic function of the above objective function.

Based on the above $\sqrt{n}$-consistency result, we shall study the asymptotic distribution of $\breve{\theta}_{cue}$. Let $g_{tn}(\theta) = \frac{1}{B}\sum_{j=t}^{t+B-1} g_j(\theta)$. By the mean value theorem, we have $f_{tn}(\breve{\theta}_{cue}) = f_{tn}(\theta_0) + g_{tn}(\tilde{\theta})(\breve{\theta}_{cue} - \theta_0)$, where $\tilde{\theta} := \tilde{\theta}(\mathcal{Z}_n/\sqrt{n} + \theta_0, \theta_0)$ is on the line segment joining $\breve{\theta}_{cue}$ and θ_0, and it can vary across equations and t. Further define

$$Q_{cue,n}(Y) := \frac{1}{nb}\max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N}\{1 - (1 + \lambda'\sqrt{n} f_{tn}(\theta_0) + \lambda' g_{tn}(\tilde{\theta})Y)^2\},$$

where $\tilde{\theta} = \tilde{\theta}(Y/\sqrt{n} + \theta_0, \theta_0)$ is a function of Y ensuring that $f_{tn}(\theta_0) + g_{tn}(\tilde{\theta})Y/\sqrt{n} = f_{tn}(Y/\sqrt{n} + \theta_0)$ (by the mean value theorem). Notice that

$$\mathcal{Z}_n = \operatorname*{argmin}_{Y\in\mathbb{R}^p} Q_{cue,n}(Y). \tag{18}$$

Because θ_0 is an interior point of Θ, the minimization in (18) is taken over $\mathbb{R}^p$. For any $r \in [0, 1]$, note that $\sqrt{n} f_{tn}(\theta_0) \Rightarrow b^{-1}\Lambda D_k(r; b)$ and $g_{tn}(\tilde{\theta}) \to^p G_0$ with $D_k(r; b) := W_k(r+b) - W_k(r)$. Hence it is expected that $\mathcal{Z}_n$ converges to the following limit,

$$\begin{aligned} \mathcal{Z} &= \operatorname*{argmin}_{Y\in\mathbb{R}^p} Q_{cue}(Y), \\ Q_{cue}(Y) & \\ &= \frac{1}{b}\max_{\lambda\in\mathbb{R}^k} \int_0^{1-b} \{1 - (1 + \lambda' b^{-1}\Lambda D_k(r; b) + \lambda' G_0 Y)^2\} dr. \end{aligned} \tag{19}$$

Before rigorously justifying this claim, we provide some discussion on the nonstandard limit in (19). Assume that $\Lambda^{-1}G_0$ has the singular value decomposition $\Lambda^{-1}G_0 = U\Sigma V$ where $UU' = U'U = I_k$, $VV' = V'V = I_p$, and

$$\Sigma = \begin{pmatrix} P_{p\times p} \\ O_{(k-p)\times p} \end{pmatrix},$$

where P is a diagonal matrix with singular values on the main diagonal and O is the matrix of zeros. Let $\tilde{D}_k(r; b) = U' D_k(r; b) = (\tilde{D}_{k,1}(r; b), \ldots, \tilde{D}_{k,k}(r; b))'$ and $\tilde{D}_{k,a:a'}(r; b) = (\tilde{D}_{k,a}(r; b), \ldots, \tilde{D}_{k,a'}(r; b))'$ with $1 \le a \le a' \le k$. Note that $\tilde{D}_k(r; b) =^d D_k(r; b)$. By changing the variable, it is not hard to see that

$$Q_{cue}(Y) = \max_{\lambda\in\mathbb{R}^k} \frac{1}{b}\int_0^{1-b} \left\{1 - \left(1 + \lambda'\tilde{D}_k(r; b) + \lambda'_{1:p} bPVY\right)^2\right\} dr,$$

where $\lambda = (\lambda_1, \ldots, \lambda_k)'$ and $\lambda_{a:a'} = (\lambda_a, \ldots, \lambda_{a'})'$ with $1 \le a \le a' \le k$. following result shows that $\mathcal{Z}$ has a mixed normal distribution after a suitable linear transformation.

Lemma 3.3.

$$bPV\mathcal{Z} = -\frac{\int_0^{1-b}\left\{1 + \breve{\lambda}'_{p+1:k}\tilde{D}_{k,p+1:k}(r; b)\right\}\tilde{D}_{k,1:p}(r; b) dr}{\int_0^{1-b}\left\{1 + \breve{\lambda}'_{p+1:k}\tilde{D}_{k,p+1:k}(r; b)\right\} dr}, \tag{20}$$

where

$$\begin{aligned} \breve{\lambda}_{p+1:k} &= \max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p}} \int_0^{1-b}\left\{1 - \left(1 + \lambda'_{p+1:k}\tilde{D}_{k,p+1:k}(r; b)\right)^2\right\} dr \\ &= -\left(\int_0^{1-b} \tilde{D}_{k,p+1:k}(r; b)\tilde{D}_{k,p+1:k}(r; b)' dr\right)^{-1} \\ &\quad \times \int_0^{1-b} \tilde{D}_{k,p+1:k}(r; b) dr. \end{aligned} \tag{21}$$

Moreover, the minimizer $\mathcal{Z}$ is unique.

To understand the nonstandard limit $\mathcal{Z}$, we define the process

$$\varphi(r) = -\frac{1+\breve{\lambda}'_{p+1:k}\tilde{D}_{k,p+1:k}(r;b)}{\int_0^{1-b}\left\{1+\breve{\lambda}'_{p+1:k}\tilde{D}_{k,p+1:k}(r;b)\right\}dr}.$$

Then we have

$$\begin{aligned} bPV\mathcal{Z} &= \int_0^{1-b}\varphi(r)\tilde{D}_{k,1:p}(r;b)dr \\ &=^d \int_0^{1-b}\varphi(r)\int_0^1 \mathbf{I}\{r\le s\le r+b\}dW_p(s)dr \\ &= \int_0^1 \omega(s)dW_p(s), \end{aligned}$$

where $\omega(s) = \int_0^{1-b}\mathbf{I}\{s-b\le r\le s\}\varphi(r)dr$ with $\mathbf{I}$ denoting the indicator function, and $\omega(s)$ is independent of $W_p(s)$. Thus $bPV\mathcal{Z}$ is mixed normal, and it can be described via the stochastic integration of the form $\int_0^1\omega(s)dW_p(s)$.

Remark 3.1. Following the arguments in Remark 2 of Zhang and Shao (2014), we are able to show that when $b\to 0$,

$$\begin{aligned} &\frac{1}{b}\int_0^{1-b}\tilde{D}_{k,p+1:k}(r;b)dr \\ &\quad =^d \frac{1}{b}\int_0^{1-b}\int_r^{r+b} dW_{k-p}(s)dr \to^d W_{k-p}(1), \\ &\frac{1}{b}\int_0^{1-b}\tilde{D}_{k,p+1:k}(r;b)\tilde{D}_{k,1:p}(r;b)'dr \to^p O_{(k-p)\times p}, \\ &\frac{1}{b}\int_0^{1-b}\tilde{D}_{k,p+1:k}(r;b)\tilde{D}_{k,p+1:k}(r;b)'dr \to^p I_{k-p}. \end{aligned} \quad (22)$$

Simple calculation yields that $PV\mathcal{Z}\to^d W_p(1)$ as $b\to 0$, that is

$$\mathcal{Z}\to^d V'P^{-1}W_p(1) =^d N(0, V'P^{-1}P^{-1}V) = N(0,(G_0'\Delta G_0)^{-1}).$$

We are now ready to present the main result in this section, which establishes the asymptotic distribution for $\breve{\theta}_{cue}$.

Theorem 3.1. *Suppose that* $\sup_{1\le t\le n}\|\sum_{j=1}^t g_j(\theta)/n-(t/n)G(\theta)\| = o_{a.s.}(1)$ *uniformly for all* θ *in* $\mathcal{N}_{\epsilon^*}(\theta_0)$. *Assume that* $G(\theta)$ *is Lipschitz continuous on* $\mathcal{N}_{\epsilon^*}(\theta_0)$ *and* $rank(G_0)=p$. *Then under Assumptions 2.1, 2.2 and 2.4,* $\mathcal{Z}_n\to^d\mathcal{Z}$.

Remark 3.2. Our proof relies on the Skorokhod's embedding theorem and the Argmax continuous mapping theorem. For technical reasons, we need to impose the stronger assumption that $\sup_{\theta\in\mathcal{N}_{\epsilon^*}(\theta_0)}\sup_{1\le t\le n}\|\sum_{j=1}^t g_j(\tilde{\theta})/n-(t/n)G(\tilde{\theta})\| = o_{a.s.}(1)$ uniformly for all θ in $\mathcal{N}_{\epsilon^*}(\theta_0)$.

3.2. CUE-based test statistics

Building on the results in previous subsection, we derive the asymptotic distributions for the over-identification test and the LR type test under the fixed-smoothing asymptotics. In the current context, the over-identification test is defined as

$$\breve{J}_{cue,n} = \min_{\theta\in\Theta} gel_{cue}(\theta) = \frac{1}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N\{1-(1+\lambda' f_{tn}(\breve{\theta}_{cue}))^2\}.$$

By Theorem 3.1, it is not hard to show that

$$\begin{aligned} \breve{J}_{cue,n} &\to^d \breve{J}_{cue,\infty}(b) =^d \mathcal{J}_{cue,k-p}(b) \\ &:= \max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p}}\frac{1}{b}\int_0^{1-b}\left[1-\left\{1+\lambda'_{p+1:k}D_{k,p+1:k}(r;b)\right\}^2\right]dr \\ &= \frac{1}{b}\left(\int_0^{1-b}D_{k,p+1:k}(r;b)dr\right)' \\ &\quad\times\left(\int_0^{1-b}D_{k,p+1:k}(r;b)D_{k,p+1:k}(r;b)'dr\right)^{-1} \\ &\quad\times\left(\int_0^{1-b}D_{k,p+1:k}(r;b)dr\right). \end{aligned}$$

The nonstandard limit $\breve{J}_{cue,\infty}(b)$ has a quadratic form which clearly reflects the internal studentization. In practice, one can simulate critical values of $\breve{J}_{cue,\infty}(b)$ by approximating the process $D_{k,p+1:k}(r;b)$ numerically. By (22), we can deduce that $\breve{J}_{cue,\infty}(b)\to^d\chi^2_{k-p}$ as $b\to 0$. Therefore, the fixed-smoothing asymptotics is consistent with the increasing-smoothing asymptotics as the amount of smoothing increases (i.e. $b\to 0$).

We now turn our attention to the LR test. Consider the null hypothesis $\tilde{H}_0: r(\theta)=0$ versus the alternative that $\tilde{H}_a: r(\theta)\ne 0$, where $r(\theta):\mathbb{R}^p\to\mathbb{R}^m$ is a differentiable map with respect to θ and $m\le p$. Define $R(\theta)=\partial r(\theta)/\partial\theta'\in\mathbb{R}^{m\times p}$ and $R_0=R(\theta_0)$. Under the constraint that $r(\theta)=0$, the negative "log-likelihood" function is defined as

$$gel^c_{cue}(\theta) = \frac{1}{nb}\max_{\lambda\in\mathbb{R}^k,\tau\in\mathbb{R}^m}\sum_{t=1}^N\{1-(1+\lambda' f_{tn}(\theta)+\tau' r(\theta))^2\}.$$

The LR statistic is then given by

$$\mathcal{LR}^*_{cue,n} = \breve{J}^c_{cue,n}-\breve{J}_{cue,n},\quad \breve{J}^c_{cue,n}=\min_{\theta\in\Theta} gel^c_{cue}(\theta). \quad (23)$$

Using the arguments in the proof of Theorem 3.1 with some suitable modifications, we have

$$\begin{aligned} \breve{J}^c_{cue,n} &\to^d \breve{J}^c_{cue,\infty}(b) = \frac{1}{b}\min_{Y\in\mathbb{R}^p}\max_{\lambda\in\mathbb{R}^k,\tau\in\mathbb{R}^m} \\ &\quad\times\int_0^{1-b}\{1-(1+\lambda' D_k(r;b)+\lambda'_{1:p}bPVY+\tau' R_0 Y)^2\}dr. \end{aligned} \quad (24)$$

Let $\tilde{Y}$ be the minimizer and $\tilde{\lambda}$, $\tilde{\tau}$ be the corresponding maximizers in (24). Define $L=-R_0 b^{-1}V^{-1}P^{-1}$. Using the first order condition (FOC) with respect to λ and τ, we can show that

$$\begin{aligned} &\int_0^{1-b}(1+\tilde{\lambda}' D_k(r;b)+\tilde{\lambda}'_{1:p}bPV\tilde{Y}+\tilde{\tau}' R_0\tilde{Y})^2 dr \\ &\quad=\int_0^{1-b}(1+\tilde{\lambda}' D_k(r;b)+\tilde{\lambda}'_{1:p}bPV\tilde{Y}+\tilde{\tau}' R_0\tilde{Y})dr. \end{aligned}$$

Together with the FOC with respect to Y, we get

$$\begin{aligned} &\int_0^{1-b}(1+\tilde{\lambda}' D_k(r;b)+\tilde{\lambda}'_{1:p}bPV\tilde{Y}+\tau' R_0\tilde{Y})(\tilde{\lambda}'_{1:p}bPV+\tilde{\tau}' R_0)dr \\ &\quad=\int_0^{1-b}(1+\tilde{\lambda}' D_k(r;b)+\tilde{\lambda}'_{1:p}bPV\tilde{Y}+\tau' R_0\tilde{Y})^2 \\ &\quad\quad\times(\tilde{\lambda}'_{1:p}bPV+\tilde{\tau}' R_0)dr = 0, \end{aligned}$$

which implies that $\tilde{\lambda}'_{1:p}=-\tilde{\tau}' R_0 b^{-1}V^{-1}P^{-1}=\tilde{\tau}' L$. Thus we have

$$\begin{aligned} \breve{J}^c_{cue,\infty}(b) &= \frac{1}{b}\max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p},\tau\in\mathbb{R}^m} \\ &\quad\times\int_0^{1-b}\{1-(1+\tau' LD_{k,1:p}(r;b)+\lambda'_{p+1:k}D_{k,p+1:k}(r;b))^2\}dr. \end{aligned}$$

By the singular value decomposition for L and the change of variable, we derive that

$$\begin{aligned} \breve{J}^c_{cue,\infty}(b) &= \frac{1}{b}\max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p},\tau\in\mathbb{R}^m} \\ &\quad\times\int_0^{1-b}\{1-(1+\tau' D_{k,1:m}(r;b)+\lambda'_{p+1:k}D_{k,p+1:k}(r;b))^2\}dr. \end{aligned}$$

To simplify the expression for $\breve{J}^{c}_{cue,\infty}(b)$, define $\bar{D}_{m+k-p}(r) = (D_{k,1:m}(r;b)', D_{k,p+1:k}(r;b)')'$, where $D_{k,1:m}(r;b)$ and $D_{k,p+1:k}(r;b)$ are mutually independent. Simple algebra then yields

$$\breve{J}^{c}_{cue,\infty}(b) = \left(\frac{1}{b}\int_0^{1-b}\bar{D}_{m+k-p}(r)'dr\right) \times \left(\int_0^{1-b}\bar{D}_{m+k-p}(r)\bar{D}_{m+k-p}(r)'dr\right)^{-1}\left(\int_0^{1-b}\bar{D}_{m+k-p}(r)dr\right).$$

Notice that the processes $D_{k,p+1:k}(r;b)$ involved in $\breve{J}^{c}_{cue,\infty}(b)$ and $\breve{J}_{cue,\infty}(b)$ are identical. Therefore, by the continuous mapping theorem, we deduce that

$$\mathcal{LR}^{*}_{cue,n} \to^{d} \mathcal{LR}_{cue,m,q} := \breve{J}^{c}_{cue,\infty}(b) - \breve{J}_{cue,\infty}(b),$$

with $q = k - p$.

Remark 3.3. The limit $\mathcal{LR}_{cue,m,q}$ can be simplified using the inverse of Schur's complement. To see this, define

$$S_{mq} = \int_0^{1-b}D_{k,1:m}(r)dr - \left(\int_0^{1-b}D_{k,1:m}(r)D_{k,p+1:k}(r)'dr\right) \times \left(\int_0^{1-b}D_{k,p+1:k}(r)D_{k,p+1:k}(r)'dr\right)^{-1}\int_0^{1-b}D_{k,p+1:k}(r)'dr,$$

$$\begin{aligned} S_{mm,q} = &\int_0^{1-b}D_{k,1:m}(r)D_{k,1:m}(r)'dr \\ &- \left(\int_0^{1-b}D_{k,1:m}(r)D_{k,p+1:k}(r)'dr\right) \\ &\times \left(\int_0^{1-b}D_{k,p+1:k}(r)D_{k,p+1:k}(r)'dr\right)^{-1} \\ &\times \left(\int_0^{1-b}D_{k,p+1:k}(r)D_{k,1:m}(r)'dr\right), \end{aligned}$$

where $q = k - p$. The inverse of Schur's complement implies that $\mathcal{LR}_{cue,m,q} = S'_{mq}S^{-1}_{mm,q}S_{mq}/b$. Compared to the asymptotic distribution in (14), it is clear that the limiting distribution of $\mathcal{LR}^{*}_{cue,n}$ has a similar quadratic form which reflects the effect of the internal studentization.

Remark 3.4. We remark that $\mathcal{LR}_{cue,m,q}$ converges to χ^2_m as b approaches zero. To see this, note that as $b \to 0$,

$$\frac{1}{b}\int_0^{1-b}\bar{D}_{m+k-p}(r;b)dr \to^{d} \begin{pmatrix} W_m(1) \\ W_q(1)\end{pmatrix},$$

$$\frac{1}{b}\int_0^{1-b}\bar{D}_{m+k-p}(r)\bar{D}_{m+k-p}(r)'dr \to^{p} I_{m+k-p}.$$

It thus implies that $\mathcal{LR}_{cue,m,q} \to^{d} \chi^2_m$ as $b \to 0$.

4. Empirical likelihood

4.1. Fixed-smoothing asymptotics for MBELE

Another important estimator in the GEL class is the MBELE. Motivated by the connection between CUE and MBELE, we shall study the fixed-smoothing asymptotics for MBELE. Consider the profile empirical log-likelihood function based on the fully overlapping smoothed moment conditions (Kitamura, 1997),

$$\mathcal{L}_{el,n}(\theta) = \sup\left\{\sum_{t=1}^{N}\log(\pi_t) : \pi_t \geq 0, \sum_{t=1}^{N}\pi_t = 1, \sum_{t=1}^{N}\pi_t f_{tn}(\theta) = 0\right\}. \quad (25)$$

Standard Lagrange multiplier arguments imply that the maximum is attained when

$$\pi_t = \frac{1}{N\{1+\lambda' f_{tn}(\theta)\}}, \quad \text{with} \quad \sum_{t=1}^{N}\frac{f_{tn}(\theta)}{1+\lambda' f_{tn}(\theta)} = 0,$$

where λ is the Lagrange multiplier. Based on the dual problem (see e.g., Borwein and Lewis (1991)), the empirical log-likelihood ratio function (up to a multiplicative constant) is given by

$$elr(\theta) = \frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^{N}\log(1+\lambda' f_{tn}(\theta)), \quad \theta \in \Theta. \quad (26)$$

Under the traditional small-b asymptotics, i.e., $nb^2 + 1/(nb) \to 0$ as $n \to \infty$, and suitable weak dependence assumptions (Kitamura (1997); also see Theorem 1 of Nordman and Lahiri (2014)), it can be shown that

$$elr(\theta_0) \to^{d} \chi^2_k. \quad (27)$$

This result can be used as the basis for confidence region construction. As pointed out by Nordman et al. (2013), the coverage accuracy of BEL (based on the chi-square approximation) can depend crucially on the block length $B = \lfloor nb \rfloor$ and appropriate choices can vary with the underlying process. To capture the choice of block length in the asymptotics, Zhang and Shao (2014) showed that under Assumption 2.4, when $n \to +\infty$ and b is fixed,

$$elr(\theta_0) \to^{d} U_{el,k}(b) := \frac{2}{b}\max_{\lambda\in\mathbb{R}^k}\int_0^{1-b}\log(1+\lambda' D_k(r;b))dr, \quad (28)$$

where $\log(x) = -\infty$ for $x \leq 0$.

However, existing works (Nordman et al., 2013; Zhang and Shao, 2014, 2016) only concern with the asymptotic behavior of the EL ratio test statistic evaluated at the true parameter (i.e., $elr(\theta_0)$). And these results are not directly useful for studying the asymptotic behavior of the MBELE and the associated tests for over-identification testing and testing on subvector of parameters. Motivated by the results in the previous section, we shall establish the fixed-smoothing asymptotics for the MBELE and utilize this result to study the specification tests.

Formally, the MBELE is defined as

$$\hat{\theta}_{el} = \underset{\theta\in\Theta}{\operatorname{argmax}}\,\mathcal{L}_{el,n}(\theta) = \underset{\theta\in\Theta}{\operatorname{argmin}}\,elr(\theta),$$

where the second equality is due to the duality. Under Assumption 2.4, we can use the Skorokhod's embedding theorem (strong approximation) to embed $\{t\bar{f}_t/\sqrt{n}\}$ and $\{W_k(r)\}$ in a larger probability space such that

$$\sup_{0\leq r\leq 1}\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nr\rfloor}f_j(\theta_0) - \Lambda W_k(r)\right\| \to^{a.s.} 0.$$

The Skorokhod's embedding theorem is very useful in the subsequent derivation. Denote by $\mathcal{D}_b := \mathcal{D}_{b,k-p}$ the event that the origin of $\mathbb{R}^{k-p}$ is contained in the interior of the convex hull of $\{\tilde{D}_{k,p+1:k}(r;b)\}$, where $\tilde{D}_k(r;b) = U'D_k(r;b)$. It has been shown in Zhang and Shao (2016) that

$$P(\mathcal{D}_b) < 1,$$

for $b > 0$ and $k > p$ (see Proposition 3.1 therein). Due to the convex hull constraint in EL/BEL, the MBELE based on the fully overlapping moment conditions is no longer consistent under the fixed-smoothing asymptotics. However, we have the following result.

Lemma 4.1. *Suppose* $\sup_{1\leq t\leq n}\|\sum_{j=1}^{t}g_j(\tilde{\theta})/n - (t/n)G(\tilde{\theta})\| \to^{a.s.} 0$ *uniformly for all* $\tilde{\theta}$ *in* $\mathcal{N}_{\epsilon^*}(\theta_0)$, $G(\theta)$ *is Lipschitz continuous on* $\mathcal{N}_{\epsilon^*}(\theta_0)$ *and* $\operatorname{rank}(G_0) = p$. *Under Assumptions 2.1, 2.2 and 2.4, we have for any* $\epsilon > 0$, $P(|\hat{\theta}_{el} - \theta_0| > \epsilon | \mathcal{D}_b) \to 0$.

Remark 4.1. In view of the proof of Lemma 4.1, the inconsistency of $\hat{\theta}_{el}$ is closely related to the minimizer of $\tilde{Q}_{el}(z) := \max_{\lambda\in\mathbb{R}^k}\int_0^{1-b}\log(1+\lambda' D_k(r;b)+\lambda' b\Lambda^{-1}G_0 z)dr$. When the origin is not contained in the convex hull of $\{\tilde{D}_{k,p+1:k}(r;b)\}$, we see that $\tilde{Q}_{el}(z)=\infty$ for any z. Because the feasible set in the constraint optimization is empty (asymptotically), the consistency of $\hat{\theta}_{el}$ cannot be established in this case. It is worth pointing out that the inconsistency result is essentially due to the convex hull constraint in the formulation of BEL rather than the use of the fixed-smoothing asymptotics.

Conditional on $\mathcal{D}_b$, we establish the $\sqrt{n}$ convergence rate of $\hat{\theta}_{el}$.

Lemma 4.2. *Under the assumptions in Lemma 4.1 and conditional on $\mathcal{D}_b$, $\mathcal{Y}_n := \sqrt{n}(\hat{\theta}_{el}-\theta_0)$ is uniformly tight.*

Before presenting the main result regarding the limiting distribution for $\hat{\theta}_{el}$, we define the random variable,

$$\tilde{\mathcal{Y}} = \operatorname*{argmin}_{Y\in\mathbb{R}^p}\max_{\lambda\in\mathbb{R}^k}\int_0^{1-b}\log(1+\lambda'\tilde{D}_k(r;b)+\lambda_{1:p}' bPVY)dr,$$

where $\lambda=(\lambda_1,\ldots,\lambda_k)'$ and $\lambda_{1:p}=(\lambda_1,\ldots,\lambda_p)'$ with $k\ge p$. Conditional on $\mathcal{D}_b$, $\tilde{\mathcal{Y}}$ is well defined. Using the FOC, it is not hard to derive that on $\mathcal{D}_b$,

$$\begin{aligned} bPV\tilde{\mathcal{Y}} &= \operatorname*{argmin}_{z\in\mathbb{R}^p}\max_{\lambda\in\mathbb{R}^k}\int_0^{1-b}\log(1+\lambda'\tilde{D}_k(r;b)+\lambda_{1:p}' z)dr \\ &= -\left(\int_0^{1-b}\frac{1}{1+\check{\lambda}_{p+1:k}'\tilde{D}_{k,p+1:k}(r;b)}dr\right)^{-1} \\ &\quad\times\int_0^{1-b}\frac{\tilde{D}_{k,1:p}(r;b)}{1+\check{\lambda}_{p+1:k}'\tilde{D}_{k,p+1:k}(r;b)}dr, \end{aligned} \tag{29}$$

where $\check{\lambda}_{p+1:k}=\operatorname{argmax}_{\lambda\in\mathbb{R}^{k-p}}\int_0^{1-b}\log(1+\lambda'\tilde{D}_{k,p+1:k}(r;b))dr$. By letting

$$\psi(r) = -\left(\int_0^{1-b}\frac{1}{1+\check{\lambda}_{p+1:k}'\tilde{D}_{k,p+1:k}(r;b)}dr\right)^{-1}\times\frac{1}{1+\check{\lambda}_{p+1:k}'\tilde{D}_{k,p+1:k}(r;b)},$$

we deduce that

$$\begin{aligned} bPV\tilde{\mathcal{Y}} &= \int_0^{1-b}\psi(r)\tilde{D}_{k,1:p}(r;b)dr \\ &=^d \int_0^{1-b}\psi(r)\int_0^1 \mathbf{I}\{r\le s\le r+b\}dW_p(s)dr \\ &= \int_0^1\tilde{\omega}(s)dW_p(s), \end{aligned}$$

where $\tilde{\omega}(s)=\int_0^{1-b}\mathbf{I}\{s-b\le r\le s\}\psi(r)dr$. Notice that for any $\omega\notin\mathcal{D}_b$, $\check{\lambda}_{p+1:k}=\infty$ and $\psi(r)$ is not well defined. Fig. 3 (lower panels) plots the density of $PV\tilde{\mathcal{Y}}$ conditional on $\mathcal{D}_b$, where the conditional density of $PV\tilde{\mathcal{Y}}$ tends to have heavier tails as compared to those of the standard normal distribution. We are now in position to state the main theorem of this section.

Theorem 4.1. *Suppose the assumptions in Lemma 4.2 hold. Then conditional on $\mathcal{D}_b$, $\mathcal{Y}_n\to^d\tilde{\mathcal{Y}}$.*

Remark 4.2. In view of Proposition 3.1 Zhang and Shao (2016), $P(\mathcal{D}_b)$ decays at least exponentially to zero as the degrees of over-identification $q=k-p$ grows. Thus in the high-dimensional setting, where k is allowed to grow with the sample size, the convex hull constraint (i.e., the origin is contained in the convex hull of $\{f_{tn}(\theta)\}_{t=1}^N$ for some $\theta\in\Theta$) becomes more difficult to be satisfied even for large sample size. This phenomenon is also related to the mismatch between the domain of the EL and the parameter space (Tsao and Wu, 2013).

Remark 4.3. The asymptotic results above are established conditional on $\mathcal{D}_b$. This is essentially due to the fully overlapping smoothed moment conditions. The convex hull constraint violation problem depends crucially on how the smoothed moment conditions are constructed. There exists alternative blocking strategy (see e.g Nordman et al. (2013)) which avoids this issue asymptotically. However, Nordman et al.'s strategy suffers from the convex hull violation problem in finite sample (Zhang and Shao, 2016).

4.2. BEL-based test statistics

Utilizing the results in the previous subsection, we investigate the fixed-smoothing asymptotics for the BEL-based test statistics. Specifically, we shall focus on the over-identification test and the BEL ratio test for testing nonlinear constraint. The over-identification test is defined as

$$J_{el,n}=elr(\hat{\theta}_{el})=\frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N\log(1+\lambda' f_{tn}(\hat{\theta}_{el})),\quad\theta\in\Theta.$$

By Theorem 4.1, we have for $k>p$,

$$\begin{aligned} J_{el,n}=elr(\hat{\theta}_{el}) &= \frac{2}{nb}\min_{\theta\in\Theta}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N\log(1+\lambda' f_{tn}(\theta)) \\ \to^d J_{el,\infty}(b) &:= \frac{2}{b}\min_{z\in\mathbb{R}^p}\max_{\lambda\in\mathbb{R}^k} \\ &\quad\times\int_0^{1-b}\log(1+\lambda' D_k(r;b)+\lambda' b\Lambda^{-1}G_0 z)dr \\ &= \frac{2}{b}\max_{\tilde{\lambda}\in\mathbb{R}^k}\int_0^{1-b}\log(1+\tilde{\lambda}'\tilde{D}_k(r;b)+\tilde{\lambda}_{1:p}' bPV\tilde{\mathcal{Y}})dr \\ &= \frac{2}{b}\max_{\tilde{\lambda}_{p+1:k}\in\mathbb{R}^{k-p}}\int_0^{1-b}\log(1+\tilde{\lambda}_{p+1:k}'\tilde{D}_{k,p+1:k}(r;b))dr \\ &= U_{el,k-p}(b), \end{aligned} \tag{30}$$

where we have used the definition of $\tilde{\mathcal{Y}}$. It has been shown in Zhang and Shao (2016) that $P(J_{el,\infty}(b)=+\infty)>0$ (see Proposition 3.1 therein), that is with a positive probability, the limit of the over-identification test attains infinity. See Section 3.1 of Zhang and Shao (2016) for the discussion on the use of the (asymptotic) critical value based on the fixed-smoothing pivotal limit $U_{el,k-p}(b)$. As shown in Remark 2 of Zhang and Shao (2014), $J_{el,\infty}(b)\to^d\chi^2_{k-p}$ as $b\to 0$, which reveals the connection between the fixed-smoothing approximation and the traditional χ^2-based approximation.

Next, we consider the BEL ratio test for testing the hypothesis $\tilde{H}_0: r(\theta)=0$ versus the alternative $\tilde{H}_1: r(\theta)\ne 0$. Define $\check{\theta}_{el}$ the MBELE under the constraint that $r(\theta)=0$. The BEL ratio statistic is then given by

$$\mathcal{LR}_{el,n} := elr(\check{\theta}_{el})-elr(\hat{\theta}_{el}), \tag{31}$$

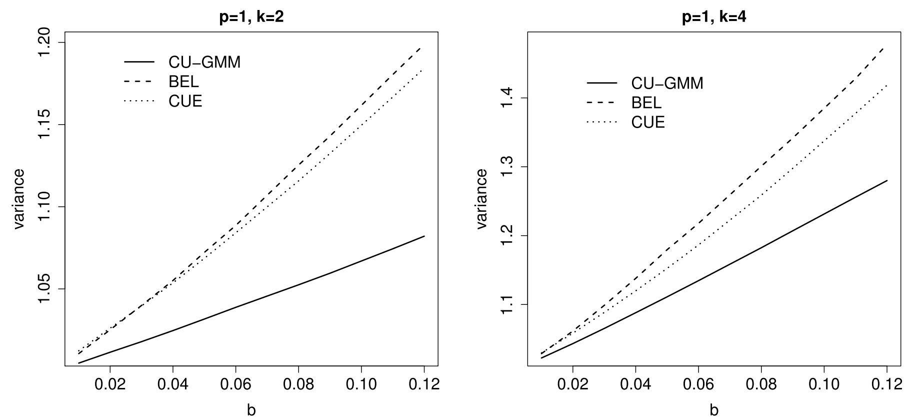

Fig. 1. Variances for $PV\mathcal{S}$, $PV\mathcal{Z}$ and $PV\tilde{\mathcal{Y}}$ conditional on $\mathcal{D}_b$, where $p=1$ and $k=2,4$.

where

$$elr(\breve{\theta}_{el}) = \frac{2}{nb}\min_{\theta\in\Theta}\max_{\lambda\in\mathbb{R}^k,\tau\in\mathbb{R}^m}\sum_{t=1}^{N}\log(1+\lambda' f_{tn}(\theta)+\tau' r(\theta)).$$

Using similar arguments as those in the proof of Theorem 4.1, we deduce that

$$elr(\breve{\theta}_{el}) \to^d U^c_{el,k}(b) := \frac{2}{b}\min_{Y\in\mathbb{R}^p}\max_{\lambda\in\mathbb{R}^k,\tau\in\mathbb{R}^m} \times \int_0^{1-b}\log(1+\lambda' D_k(r;b)+\lambda'_{1:p} bPVY + \tau' R_0 Y)dr.$$

By the FOC (with respect to Y, λ and τ), we obtain

$$\lambda_{1:p} = -\tau' R_0 b^{-1} V^{-1} P^{-1} = \tau' L,$$

which implies that

$$\begin{aligned} U^c_{el,k}(b) &= \frac{2}{b}\max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p},\tau\in\mathbb{R}^m} \\ &\quad\times\int_0^{1-b}\log(1+\tau' LD_{k,1:p}(r;b)+\lambda'_{p+1:k}D_{k,p+1:k}(r;b))dr \\ &= \frac{2}{b}\max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p},\tau\in\mathbb{R}^m} \\ &\quad\times\int_0^{1-b}\log(1+\tau' D_{k,1:m}(r;b)+\lambda'_{p+1:k}D_{k,p+1:k}(r;b))dr \\ &=^d U_{el,m+k-p}(b), \end{aligned}$$

where $D_{k,1:m}(r;b)$ and $D_{k,p+1:k}(r;b)$ are independent. Therefore, we get

$$\mathcal{LR}_{el,n} \to^d \mathcal{LR}_{el,m,q} := U_{el,m+q}(b) - U_{el,q}(b),$$

where the processes $D_{k,p+1:k}(r;b)$ involved in $U_{el,m+q}(b)$ and $U_{el,q}(b)$ are identical. In view of Remark 2 of Zhang and Shao (2014), it is also not hard to show that $U_{el,m+q}(b) - U_{el,q}(b) \to^d \chi^2_m$ as $b\to 0$.

5. Comparison

5.1. Asymptotic variance

The first order asymptotics for two types of CUEs and MBELE are nonequivalent under the fixed-smoothing asymptotics which is due to the choice of the criterion functions. In this subsection, we shall compare their asymptotic variances based on the fixed-smoothing limiting distributions. To this end, we need a simplified expression for the limiting distribution of $\mathcal{S}_n = \sqrt{n}(\hat{\theta}_{cue}-\theta_0)$. Suppose

$$\tilde{Q}_0 = \begin{pmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{pmatrix},$$

where $\tilde{Q}_{11}$ is a $p\times p$ matrix. The following result shows that $PV\mathcal{S}$ is pivotal, which allows us to compare the asymptotic variances of different estimators.

Lemma 5.1. $PV\mathcal{S} =^d -(I_p, -\tilde{Q}_{12}\tilde{Q}_{22}^{-1})W_k(1)$.

We compare the variances of $PV\mathcal{S}$, $PV\mathcal{Z}$ and $PV\tilde{\mathcal{Y}}$ conditional on $\mathcal{D}_b$ through simulations. The results for $p=1$ and $k=2,4$ are summarized in Fig. 1. It is clear that the CU-GMM estimator has the smallest asymptotic variance followed by the CUE based on the saddle point problem. The MBELE is less preferable due to its largest asymptotic variance as well as the inconsistency problem mentioned in Section 4.1.

5.2. Local power analysis

Our analysis has been so far focused on the null. In practice, it is of great interest to compare the local power properties of different inference procedures introduced above as they have different fixed-smoothing limits. For the sake of clarity, we consider the hypothesis $H_0: \mathcal{R}\theta_0 = 0$ versus the local alternative $H_{a,n}: \mathcal{R}\theta_0 = \delta/\sqrt{n}$ for $\mathcal{R}\in\mathbb{R}^{m\times p}$ with $m\le p$. We first consider the CU-GMM based Wald test $\mathcal{W}_n$ given in (12). Motivated by the connection between the Bartlett kernel and moving block approach, we shall focus on the Bartlett kernel, i.e., $\mathcal{K}(r_1,r_2) = \mathcal{K}_{\text{Bart}}((r_1-r_2)/b)$ in the CU-GMM based test, where $\mathcal{K}_{\text{Bart}}(r) = (1-|r|)\mathbf{I}\{|r|<1\}$ and $b\in(0,1]$. Based on the results in Section 2, it is not hard to show that under $H_{a,n}$,

$$\begin{aligned} \mathcal{W}_n &\to^d \mathcal{W}_{m,q}(b;\delta,G_0,\Lambda) \\ &= \left\{\mathcal{R}(G_0'Q_0^{-1}G_0)^{-1}G_0'Q_0^{-1}\Lambda W_k(1)+\delta\right\}'\left\{\mathcal{R}(G_0'Q_0^{-1}G_0)^{-1}\mathcal{R}'\right\}^{-1} \\ &\quad\times\left\{\mathcal{R}(G_0'Q_0^{-1}G_0)^{-1}G_0'Q_0^{-1}\Lambda W_k(1)+\delta\right\}. \end{aligned}$$

Here we suppress the dependency on m,p, and k for notational simplicity. Under alternative, the limiting distribution $\mathcal{W}_\infty(b;\delta,G_0,\Lambda)$ is non-pivotal but it can be simplified using rotation and transformation. Recall that $Q_0 = \Lambda\tilde{Q}_0\Lambda'$ and the singular value decomposition $\Lambda^{-1}G_0 = U\Sigma V$. Further define $L = -\mathcal{R}b^{-1}V'P^{-1}$ and let $L = U_L\Sigma_L V_L$ be the singular value decomposition for L with $\Sigma_L = (P_L, O)$, where P_L is a $m\times m$ diagonal matrix.

Lemma 5.2.

$$\mathcal{W}_\infty(b;\delta,G_0,\Lambda)=\mathcal{W}_{m,q}(b;\tilde{\delta})$$
$$=^d\left\{W_m(1)-C_{mq}C_{qq}^{-1}W_q(1)-b^{-1}\tilde{\delta}\right\}'C_{mm,q}^{-1}$$
$$\times\left\{W_m(1)-C_{mq}C_{qq}^{-1}W_q(1)-b^{-1}\tilde{\delta}\right\},\quad \tilde{\delta}=P_L^{-1}U_L^{-1}\delta.$$

Lemma 5.2 shows that $\mathcal{W}_\infty(b;\delta,G_0\Lambda)$ depends on δ, G_0 and Λ only through the quantity $\tilde{\delta}$. This observation allows us to make a fair comparison between different testing procedures.

Next we study the CUE and BEL based LR tests under the local alternative $H_{a,n}$. To this end, we define the following distributions,

$$\mathcal{J}^c_{cue,m,q}(b;\tilde{\delta})=\frac{1}{b}\max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p},\tau\in\mathbb{R}^m}\int_0^{1-b}\{1-(1+\tau'D_{k,1:m}(r;b)$$
$$+\lambda'_{p+1:k}D_{k,p+1:k}(r;b)+\tau'\tilde{\delta})^2\}dr,$$

and

$$U^c_{el,m,q}(b;\tilde{\delta})=\frac{2}{b}\max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p},\tau\in\mathbb{R}^m}$$
$$\times\int_0^{1-b}\log(1+\tau'D_{k,1:m}(r;b)+\lambda'_{p+1:k}D_{k,p+1:k}(r;b)+\tau'\tilde{\delta})dr,$$

where $\tilde{\delta}=P_L^{-1}U_L^{-1}\delta$. Using similar arguments in Sections 3.2 and 4.2, one can show that under $H_{a,n}$,

$$\mathcal{LR}^*_{cue,n}\to^d\mathcal{J}^c_{cue,m,q}(b;\tilde{\delta})-\mathcal{J}_{cue,q}(b),$$
$$\mathcal{LR}_{el,n}\to^d U^c_{el,m,q}(b;\tilde{\delta})-U_{el,q}(b).$$

To compare the local asymptotic powers, we define the following power functions

$$\mathcal{P}_{cugmm}(b;\tilde{\delta})=P\left(\mathcal{W}_{m,q}(b;\tilde{\delta})>C_\alpha(\mathcal{W}_{m,q}(b))\right),$$
$$\mathcal{P}_{cue}(b;\tilde{\delta})=P\left(\mathcal{J}^c_{cue,m,q}(b;\tilde{\delta})-\mathcal{J}_{cue,q}(b)>C_\alpha(\mathcal{LR}_{cue,m,q}(b))\right),$$
$$\mathcal{P}_{bel}(b;\tilde{\delta})=P\left(U^c_{el,m,q}(b;\tilde{\delta})-U_{el,q}(b)>C_\alpha(\mathcal{LR}_{el,m,q}(b))\right),$$

where $C_\alpha(\mathcal{F})$ denotes the $1-\alpha$ quantile of the distribution $\mathcal{F}$. Fig. 2 plots the power curves as functions of $\tilde{\delta}$ for $q=1$, $m=1,3$, and $b=0.05,0.1,0.15$. For $m>1$, we consider the direction $\tilde{\delta}=(\tilde{\delta}_0,\tilde{\delta}_0,\dots,\tilde{\delta}_0)'\in\mathbb{R}^m$. When $b=0.05$, the three power curves are very close to each other. As b increases, the asymptotic powers decrease for all methods. It is interesting to see that for the same b, the CU-GMM based test delivers the highest power among the three methods in all cases, and in contrast, the CUE-based test has the lowest asymptotic power.

6. Simulation studies

In this section, we compare the fixed-smoothing asymptotic approximation with the traditional normal approximation for two types of CUEs and MBELE. We also conduct simulation studies to evaluate the empirical size and size-adjusted power of the CU-GMM, CUE and BEL based specification tests. We only present the results for the Bartlett kernel in CU-GMM based estimation and inference. Similar results can be observed for other kernels. We follow the settings considered in Sun and Kim (2012) and Sun (2014a).

6.1. CUEs and MBELE

Consider the time series regression model,

$$y_t=x_t\theta+\delta z_{1,t}+\epsilon_{1,t},\quad x_t=\sum_{i=1}^{m_0}z_{i,t}+\epsilon_{2,t},\tag{32}$$

where the true parameter $\theta_0=0$ and $\delta=0$ under correct specification of the moment restrictions. We assume that $z_t=(z_{1,t},\dots,z_{m_0,t})'$ follows a VAR(1) process

$$z_t=\rho z_{t-1}+\sqrt{1-\rho^2}\epsilon_{3,t},$$
$$\epsilon_{3,t}=\left(\frac{\epsilon^{(1)}_{3,t}+\epsilon^{(0)}_{3,t}}{\sqrt{2}},\dots,\frac{\epsilon^{(m_0)}_{3,t}+\epsilon^{(0)}_{3,t}}{\sqrt{2}}\right),\tag{33}$$

where $(\epsilon^{(0)}_{3,t},\epsilon^{(1)}_{3,t},\dots,\epsilon^{(m_0)}_{3,t})\sim^{i.i.d}N(0,I_{m_0+1})$. Due to the presence of the common shock $\epsilon^{(0)}_{3,t}$, $\mathrm{cov}(z_{i,t},z_{j,t})=0.5$ for $i\neq j$. The data generating process (DGP) for $\epsilon_t=(\epsilon_{1,t},\epsilon_{2,t})'$ is the same as that for z_t except the dimensionality difference. The two vector processes ϵ_t and z_t are independent. We consider the moment conditions $f(v_t,\theta)=z_t(y_t-x_t\theta)$ with $v_t=(y_t,x_t,z_t')'$. Note that when $\delta=0$, $G_0=-(m_0+1)\mathbf{1}_{m_0}/2$ with $\mathbf{1}_{m_0}$ being the m_0-dimensional vector of all ones, and

$$\Delta=\frac{1+\rho^2}{2(1-\rho^2)}(I_{m_0}+\mathbf{1}_{m_0}\mathbf{1}'_{m_0}).$$

We consider the CU-GMM estimator, the CUE obtained through solving a saddle point problem and the MBELE. Set $m_0=2$ and 5, and the corresponding degrees of over-identification are $q=1$ and 4. For MBELE, we compute its value only when the origin of $\mathbb{R}^{k-p}$ is contained in the convex hull of $\{D_{tn,p+1:k}(\theta_0;b)\}_{t=1}^N$, where $D_{tn}(\theta_0;b)=\sqrt{n}bU'\Lambda^{-1}f_{tn}(\theta_0)=(D_{tn,1}(\theta_0;b),\dots,D_{tn,k}(\theta_0;b))'$. Fig. 3 presents the density curves for $PV\mathcal{S}_n$, $PV\mathcal{Z}_n$ and $PV\mathcal{Y}_n$ as well as the density curves for their fixed-smoothing limits (i.e. $PV\mathcal{S}$, $PV\mathcal{Z}$ and $PV\tilde{\mathcal{Y}}$), where $n=500$ and $\rho=0.5$. The fixed-smoothing asymptotics clearly outperforms the traditional normal approximation in terms of delivering better finite sample approximation, when b is away from zero. In particular, the fixed-smoothing asymptotics is able to capture the tail behavior of the finite sample distribution.

6.2. Empirical size

Following the setup in the previous subsection, we consider the over-identification tests ($J_{cue,n}$, $\breve{J}_{cue,n}$ and $J_{el,n}$) for testing the validity of the moment restrictions. Set $\rho=-0.5,0,0.5,0.8$, and $n=200,500$. The empirical sizes are summarized in Figs. 4–6. In the case of CUE-based test (i.e., $\breve{J}_{cue,n}$), over-rejection occurs for both the fixed-smoothing approximation and the chi-square based approximation when b is close to zero. For moderate or large b, the fixed-smoothing asymptotics provides consistently better approximation to the sampling distribution as compared to the increasing-smoothing asymptotics for which under-rejection occurs. In the case of CU-GMM based and BEL-based tests, the fixed-smoothing asymptotics outperforms the increasing-smoothing asymptotics in terms of delivering rejection probabilities closer to the nominal level. In contrast to the case of CUE (based on the saddle point problem), the rejection probabilities based on the chi-square approximation is severely upward biased for relatively large b.

To examine the performance of the Wald test and the LR type tests, we consider the DGP,

$$y_t=\beta_0+x_{1,t}\beta_1+x_{2,t}\beta_2+x_{3,t}\beta_3+\varepsilon_{0,t},\tag{34}$$

where the regressors $(x_{1,t},x_{2,t},x_{3,t})\in\mathbb{R}^3$ are correlated with $\varepsilon_{0,t}$. Consider m_1 instruments $z_{0,t},z_{1,t},\dots,z_{m_1-1,t}$ with $z_{0,t}\equiv1$. The reduced-form equations for the regressors are given by

$$x_{j,t}=z_{j,t}+\sum_{i=3}^{m_1-1}z_{i,t}+\varepsilon_{j,t},\quad j=1,2,3.\tag{35}$$

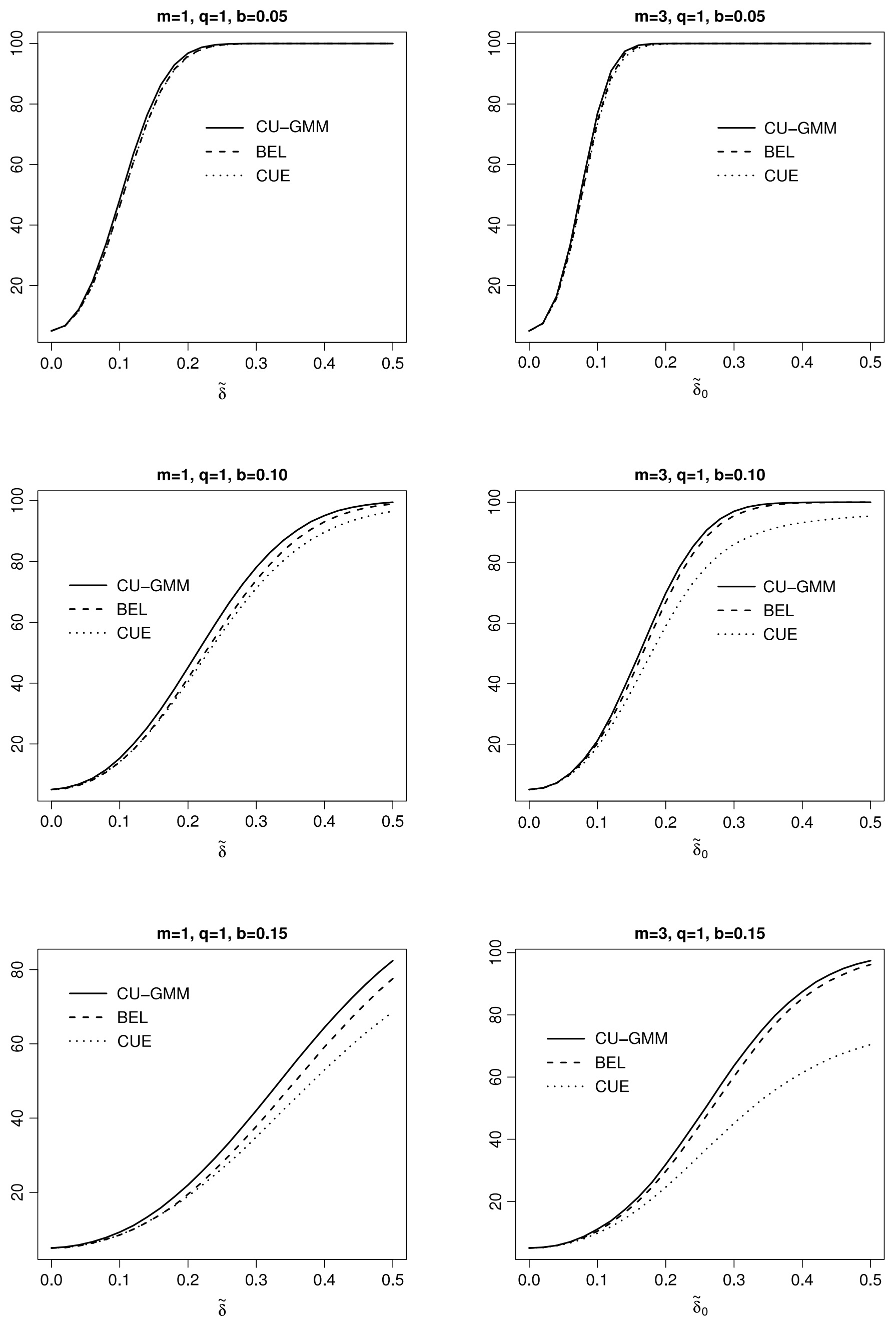

Fig. 2. Asymptotic power curves in % for the CU-GMM, CUE and BEL based tests.

We assume that $(z_{1,t}, z_{2,t}, \ldots, z_{m_1-1,t})'$ follows the VAR(1) model in (33). Suppose that $\varepsilon_t = (\varepsilon_{0,t}, \varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$ is independent with the instruments, and its DGP is the same as that for z_t except the dimensionality difference. The moment conditions are given by $f(v_t, \theta) = z_t(y_t - \beta_0 - x_{1,t}\beta_1 - x_{2,t}\beta_2 - x_{3,t}\beta_3)$, where $v_t = (y_t, x_t', z_t')'$ with $x_t = (1, x_{1,t}, x_{2,t}, x_{3,t})'$ and $z_t = (1, z_{1,t}, \ldots, z_{m_1-1,t})'$. We take $m_1 = 5$, $\rho = -0.5, 0, 0.5, 0.8$, and $n = 200, 500$. The null hypothesis of interest are

$$H_{01} : \beta_1 = 0,$$
$$H_{02} : \beta_1 = \beta_2 = \beta_3 = 0,$$

where $m = 1$ and 3 respectively. The empirical sizes are summarized in Figures S.1–S.3 in the supplement. Similar patterns can be observed for the Wald test and the LR type tests. Again it is clear that the fixed-smoothing asymptotics provides better approximation to the finite sample distributions.

6.3. Choice of bandwidth and block size

In this subsection, we compare the finite sample performance of different testing procedures based on data-dependent bandwidth and block size. We select the bandwidth parameter in CU-GMM based on the MSE criterion implemented using the VAR(1) plug-in procedure in Andrews (1991). For completeness, we reproduce the MSE optimal formula,

$$\hat{b} = 1.1447n^{-2/3}\left(\frac{2\mathrm{vec}(\Delta^{(1)})'W_k\mathrm{vec}(\Delta^{(1)})}{trW_k(I+K_{kk})\Delta\otimes\Delta}\right)^{1/3},$$

$$\Delta^{(1)} = \sum_{j=-\infty}^{+\infty} |j|\mathbb{E}f(y_t,\theta_0)f(y_{t-j},\theta_0)',$$

where W_k is a positive semi-definite weight matrix and K_{kk} is the $k^2 \times k^2$ commutation matrix. We fit a VAR(1) model to $\{f(y_t, \hat{\theta}^{(0)})\}$

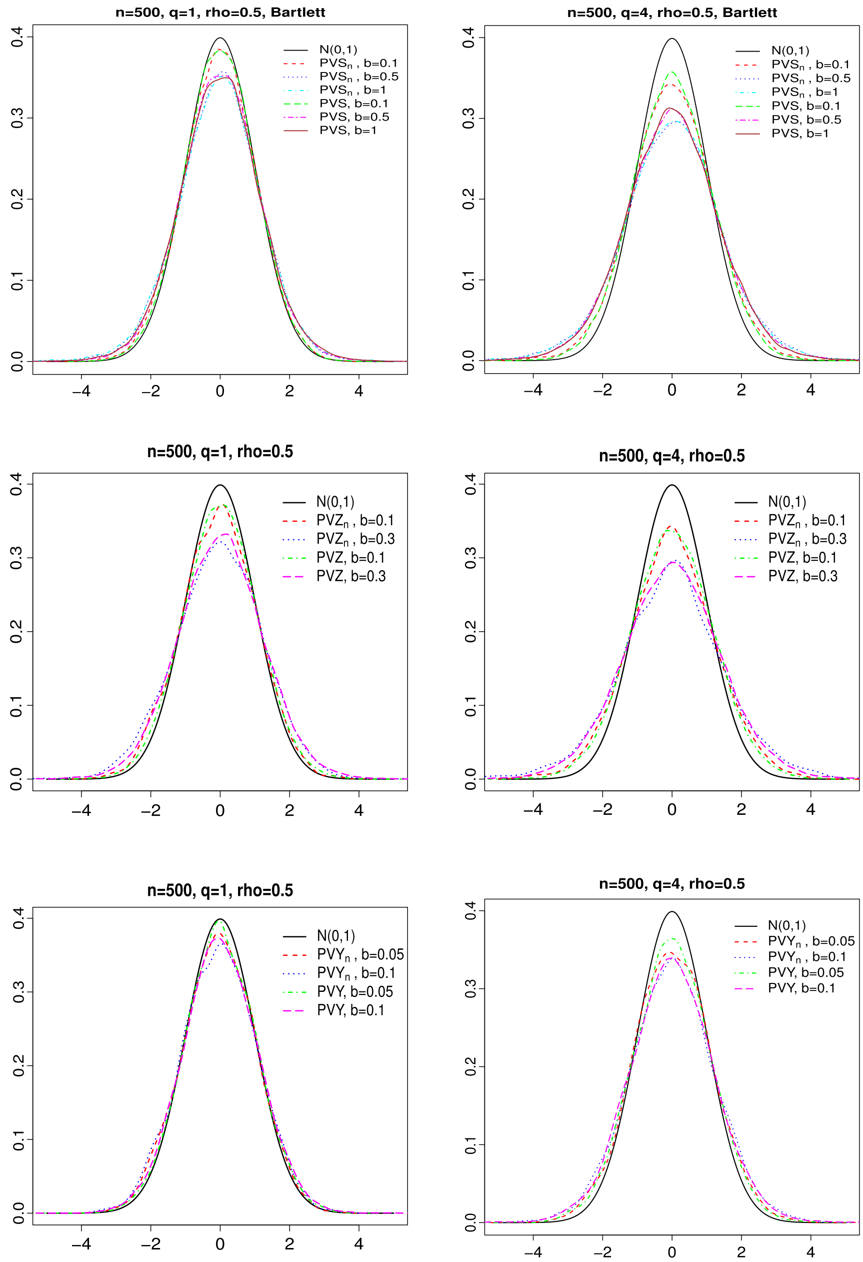

Fig. 3. Upper panels: the density curves for $PV\mathcal{S}_n$ and $PV\mathcal{S}$ (with the Bartlett kernel); middle panels: the density curves for $PV\mathcal{Z}_n$ and $PV\mathcal{Z}$; lower panels: the conditional density curves for $PV\mathcal{Y}_n$ and $PV\tilde{\mathcal{Y}}$.

with $\hat{\theta}^{(0)}$ being a preliminary estimator. Denote by $\hat{A}$ and $\hat{\Sigma}_0$ the estimators for the AR parameter matrix and the innovation covariance matrix respectively. Then the MSE optimal formula using the VAR(1) plug-in procedure is given by

$$\hat{b} = 1.1447n^{-2/3}\left(\frac{2\text{vec}(\hat{\Delta}^{(1)})'W_k\text{vec}(\hat{\Delta}^{(1)})}{trW_k(I+K_{kk})\hat{\Delta}\otimes\hat{\Delta}}\right)^{1/3},$$

where $\hat{\Delta} = (I_k - \hat{A})^{-1}\hat{\Sigma}_0(I_k - \hat{A}')^{-1}$ and $\hat{\Delta}^{(1)} = \hat{H} + \hat{H}'$ with $\hat{H} = (I-\hat{A})^{-2}\hat{A}\sum_{j=0}^{+\infty}\hat{A}^j\hat{\Sigma}_0(\hat{A}')^j$. It is worth noting that CUE and BEL use data blocks to form a block-based variance estimator for purposes of studentization which is asymptotically equivalent to a spectral density estimator at the origin based on Bartlett's lag-window kernel. Thus the above formula is also applicable to the choice of block size for the CUE-based and BEL-based tests. Alternatively one can use nonparametric estimators for the spectral quantities Δ and $\Delta^{(1)}$ by employing the flat-top kernel, see Politis and White (2004).

The simulation results are summarized in Tables 1–2. We have the following key observations: (1) the fixed-smoothing asymptotics generally provides more accurate approximation especially when the dependence is strong. Note that for the CUE-based tests, the chi-square approximation has a tendency to be undersized (see Fig. 5). Thus the empirical rejection probability based on chi-square approximation can be smaller; (2) the CUE-based tests deliver the most accurate size while the BEL-based tests have the largest size distortion under strong dependence.

6.4. Size-adjusted power

We further compare the size-adjusted powers of different testing procedures. For over-identification testing, we consider

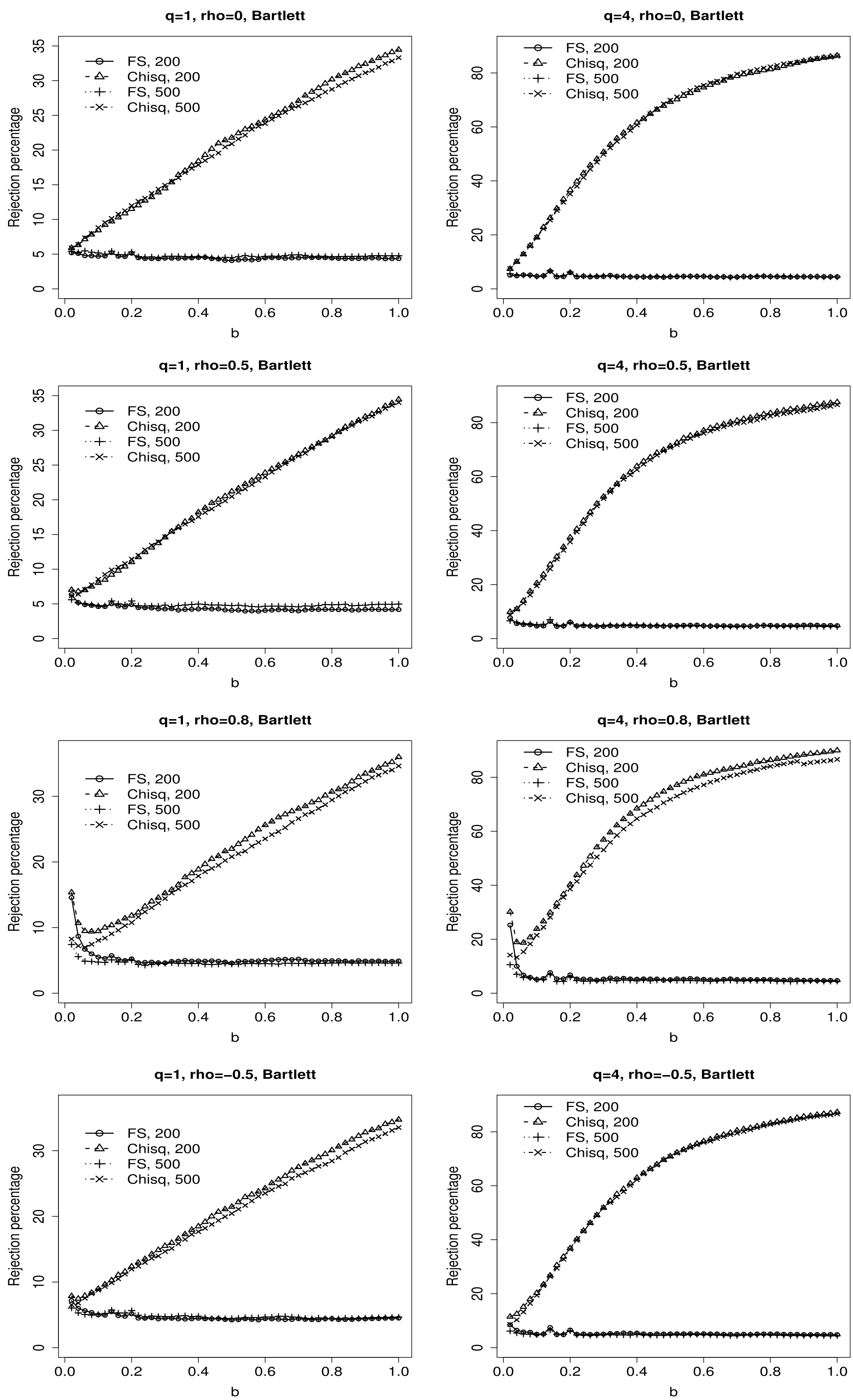

Fig. 4. Empirical rejection percentages for the CUGMM-based over-identification test (with the Bartlett kernel) based on the fixed-smoothing approximation and the chi-square approximation, where $q = 1$ for the left panels and $q = 4$ for the right panels. The nominal level is 5%, and the number of Monte Carlo replications is 5000.

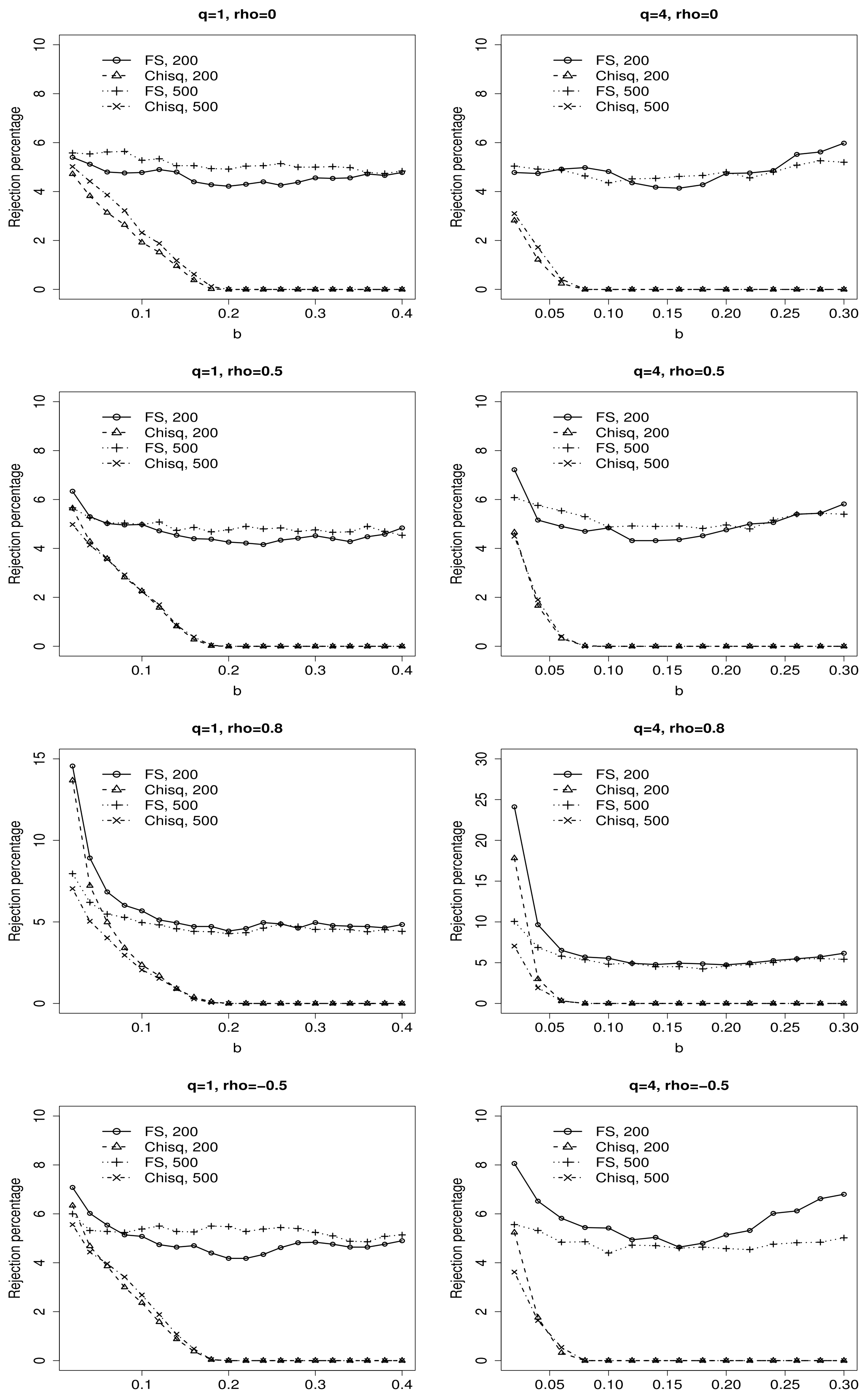

Fig. 5. Empirical rejection percentages for the CUE-based over-identification test based on the fixed-smoothing approximation and the chi-square approximation, where $q = 1$ for the left panels and $q = 4$ for the right panels. The nominal level is 5%, and the number of Monte Carlo replications is 5000.

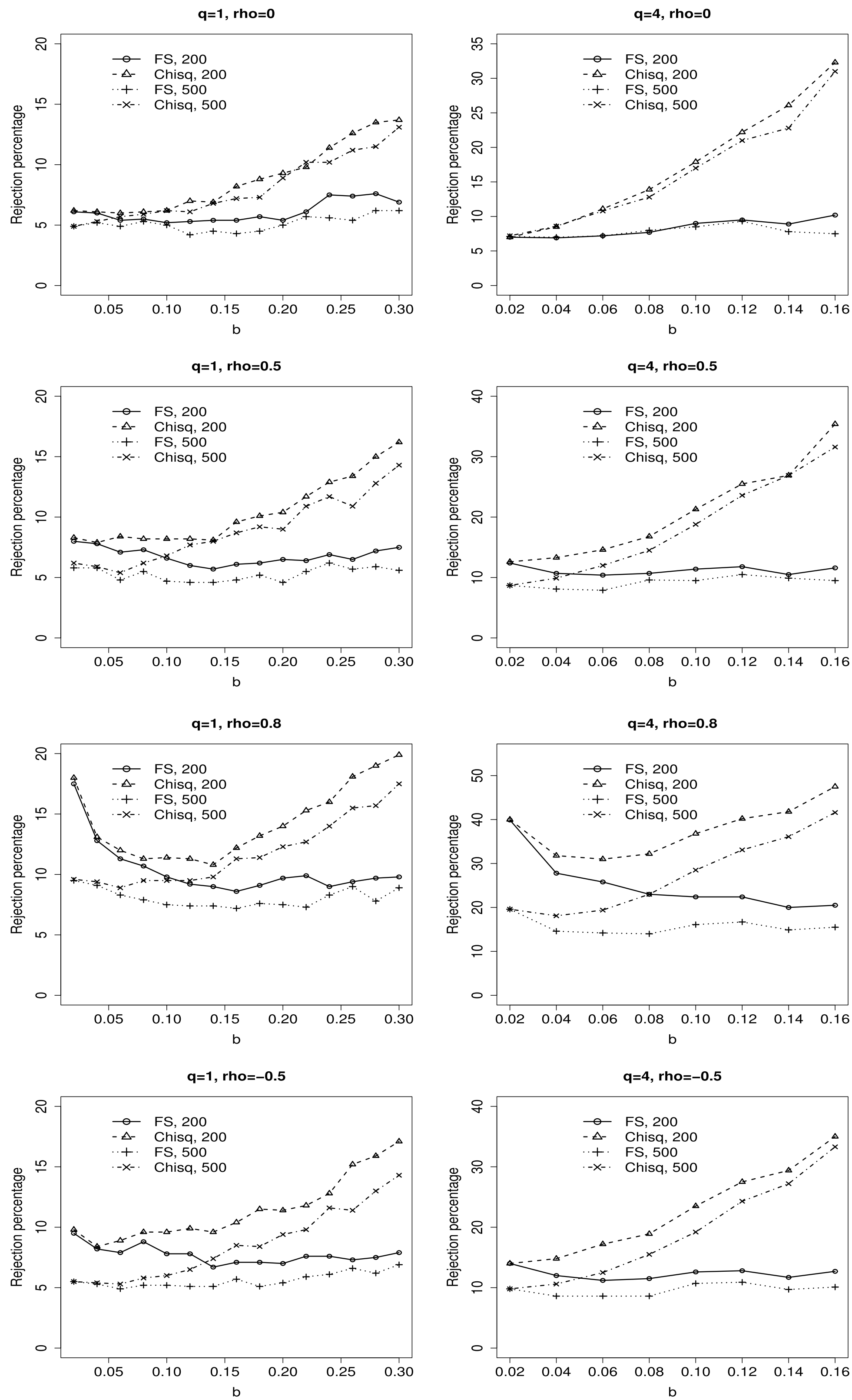

Fig. 6. Empirical rejection percentages for the BEL-based over-identification test based on the fixed-smoothing approximation and the chi-square approximation, where $q = 1$ for the left panels and $q = 4$ for the right panels. The nominal level is 5%, and the number of Monte Carlo replications is 1000.

model (32) with δ ranging from (0, 1]. For specification testing on parameters, we test the hypothesis $H_{01} : \beta_1 = 0$, where the true model is given by

$$y_t = x_{1,t}\beta_1 + \varepsilon_{0,t}, \quad \beta_1 = \delta \in (0, 0.5].$$

The size-adjusted powers for $n = 200$ are reported in Figs. 7–8. As seen from Fig. 7, the power curves are very close to each other when $\rho = 0.5$. As ρ increases, the CU-GMM based test becomes slightly more powerful than the other two competitors. Together with the results in Section 6.3, the CU-GMM based

Table 1
Empirical sizes in % for the CU-GMM, CUE and BEL based over-identification tests, where the bandwidth and block size are chosen using the VAR(1) plug-in procedure in Andrews (1991). The number of Monte Carlo replications is 5000.

n	ρ	CU-GMM				CUE				BEL			
		$q=1$		$q=4$		$q=1$		$q=4$		$q=1$		$q=4$	
		χ^2	FS	χ^2	FS	χ^2	FS	χ^2	FS	χ^2	FS	χ^2	FS
200	0.0	5.7	5.2	6.0	4.9	4.8	5.0	3.6	4.8	5.6	5.5	7.4	7.3
	0.5	6.8	6.0	10.2	7.0	5.5	5.8	4.0	7.0	7.6	7.3	13.4	12.4
	0.8	9.8	7.2	18.6	7.2	5.5	7.0	0.9	6.9	12.1	11.3	32.3	25.7
	−0.5	7.7	6.8	11.4	7.6	6.0	6.5	4.6	7.5	8.2	7.9	14.4	13.5
500	0.0	5.2	5.0	5.6	5.3	5.0	5.0	4.6	5.4	5.2	5.1	5.8	5.8
	0.5	6.2	5.8	8.4	7.1	5.5	5.7	5.8	7.1	6.4	6.3	9.2	9.1
	0.8	7.2	6.2	13.0	7.4	5.8	6.5	4.5	7.3	8.7	8.4	18.2	16.4
	−0.5	7.0	6.6	8.6	6.9	6.2	6.5	5.3	6.8	6.8	6.7	9.4	9.3

Table 2
Empirical sizes in % for the CU-GMM Wald test, the CUE and BEL based LR tests, where the bandwidth and block size are chosen using the VAR(1) plug-in procedure in Andrews (1991). The number of Monte Carlo replications is 5000.

n	ρ	CU-GMM				CUE				BEL			
		$m=1$		$m=3$		$m=1$		$m=3$		$m=1$		$m=3$	
		χ^2	FS	χ^2	FS	χ^2	FS	χ^2	FS	χ^2	FS	χ^2	FS
200	0.0	6.4	5.9	8.0	6.6	5.2	5.4	4.4	5.1	6.3	6.1	7.1	6.9
	0.5	9.9	8.3	15.1	9.7	5.9	6.9	3.5	5.7	9.9	9.4	14.1	12.3
	0.8	16.3	11.8	31.6	17.0	3.7	5.9	0.8	5.1	16.0	13.6	30.0	19.5
	−0.5	9.3	8.1	14.0	10.5	5.7	6.5	5.1	7.5	9.2	8.7	13.3	12.1
500	0.0	5.0	4.9	5.7	5.2	4.8	4.8	4.3	4.8	5.0	4.9	5.2	5.3
	0.5	7.6	6.6	10.2	8.1	5.5	5.9	4.8	6.3	7.4	7.1	9.4	8.9
	0.8	10.9	8.9	18.6	12.4	5.7	6.9	3.7	7.4	10.9	9.9	18.3	14.8
	−0.5	7.5	6.7	9.4	7.9	6.0	6.3	5.5	6.3	7.3	7.1	9.1	8.7

Table 3
Estimates of the coefficients, and p-values for testing the null $H_0 : \beta_1 = 0$ based on three different procedures with various block sizes (or bandwidths). For the CU-GMM method, $\mathcal{K}$ is chosen to be the Bartlett kernel, and $B = nb_0$. The last column corresponds to the results based on the MSE optimal B.

Method		B					
		1	3	6	20	30	MSE optimal
CUGMM	β_0	0.625	0.437	0.279	0.326	0.426	0.271
	β_1	−0.105	−0.072	−0.045	−0.056	−0.076	−0.044
	p-value (χ^2)	0.032	0.232	0.479	0.204	0.011	0.483
	p-value (FS)	0.036	0.257	0.520	0.371	0.147	0.532
CUE	β_0	0.625	0.490	0.376	0.358	0.403	0.378
	β_1	−0.105	−0.081	−0.062	−0.060	−0.070	−0.062
	p-value (χ^2)	0.096	0.280	0.402	0.343	0.433	0.388
	p-value (FS)	0.082	0.257	0.377	0.253	0.273	0.364
EL	β_0	0.700	0.609	0.463	0.501	0.534	0.472
	β_1	−0.118	−0.103	−0.079	−0.089	−0.098	−0.026
	p-value (χ^2)	0.031	0.105	0.140	0.068	0.638	0.111
	p-value (FS)	0.023	0.084	0.115	0.196	0.782	0.143

over-identification test seems preferable as it delivers reasonable size and the highest size-adjusted power among the three tests. For specification testing on parameters, a comparison between the CU-GMM based test and the CUE-based test reflects a size and power trade-off. The CUE-based test has more accurate size while it suffers from (asymptotic) power loss. The CU-GMM based test delivers more power in general but it may generate severe upward size distortions (see Table 2). This is similar to the size and power trade-off between different kernels in the GMM framework (see e.g. Kiefer and Vogelsang (2005)). The numerical results also suggest that the BEL-based tests could be less preferable in practice due to relatively large upward size distortion and (asymptotic) power loss compared to the CU-GMM based tests.

7. Data illustration

In this section, we estimate a quarterly time-series model relating the change in the US inflation rate to the unemployment rate for the time period from the third quarter of 1960 to the fourth quarter of 1999 (also see Baum et al. (2007) and Stock and Watson (2003)). As instruments, we use the second lag of quarterly GDP growth and the lagged values of the treasury bill rate, the trade-weighted exchange rate and the treasury medium-term bond rate. Specifically, the moment condition model is given by

$$f(w_t, \theta) = (y_t - \beta_0 - \beta_1 x_t) z_t, \tag{36}$$

where $w_t = (x_t, y_t, z_t')'$ with $z_t = (1, z_{t1}, \ldots, z_{t4})'$, and $\theta = (\beta_0, \beta_1)$. Here x_t, y_t and z_t denote the unemployment rate, the change in the US inflation, and the instruments respectively. We estimate θ using the CU-GMM with the Bartlett kernel, CU based on the saddle point problem, and BEL. We also implement the corresponding specification tests to test the null hypothesis $H_0 : \beta_1 = 0$. The results are summarized in Table 3. The estimates based on different methods and block sizes (or bandwidth) are generally close. In particular, the negative coefficient on the unemployment

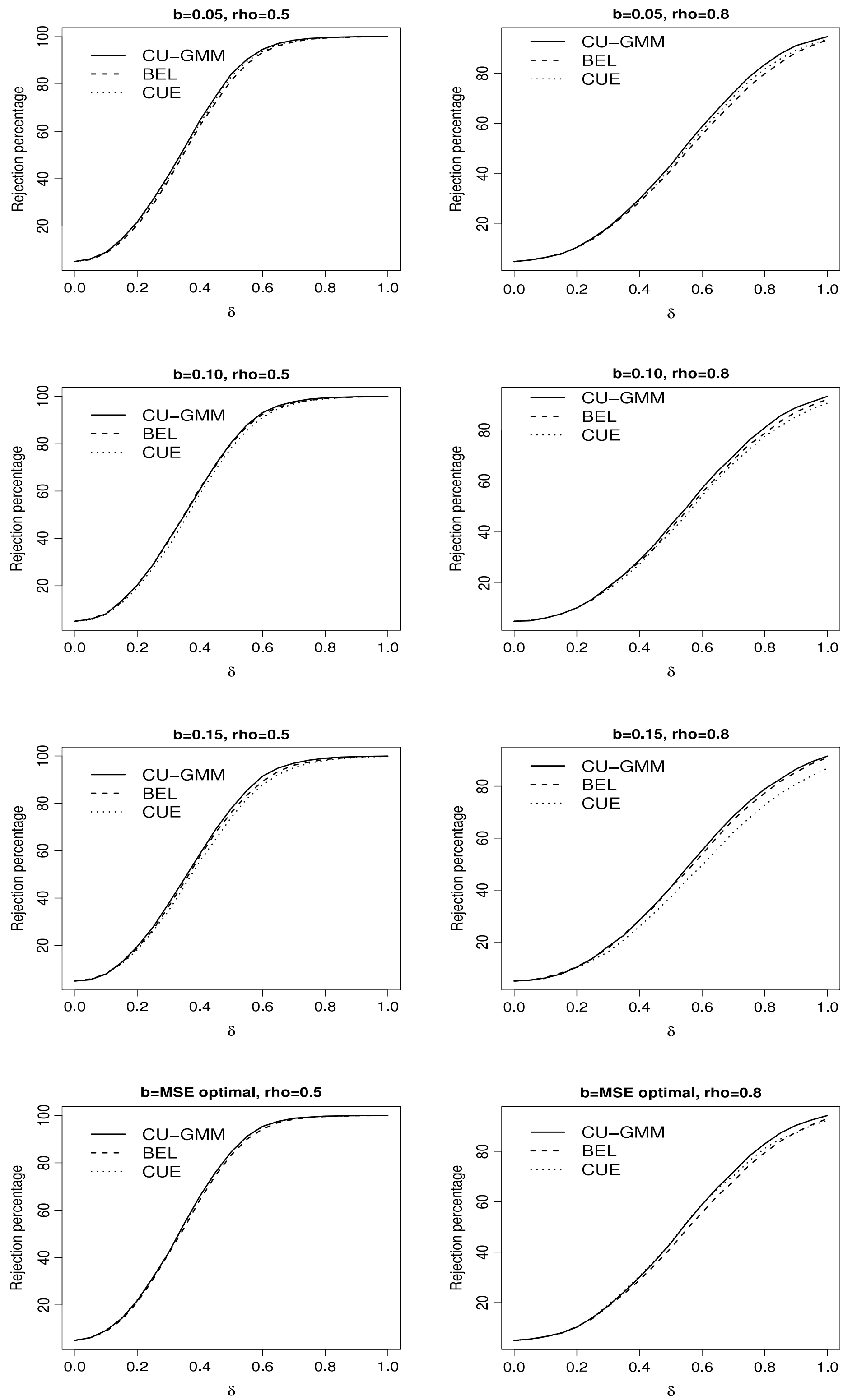

Fig. 7. Size-adjusted powers in % for the CU-GMM, CUE and BEL based over-identification tests. The nominal level is 5%, and the number of Monte Carlo replications is 5000.

rate is consistent with macroeconomic theories of the natural rate. In that context, lowering unemployment below the natural rate will cause an acceleration of price inflation.

The specification tests under fixed-smoothing and increasing-smoothing asymptotics both suggest the significance of β_1 when the dependence within the error process is neglected (i.e., $B = 1$). However, the analysis in Baum et al. (2007) indicated that the dependence is not negligible in this case (e.g., the Cumby–Huizinga test rejects the null hypothesis of the uncorrelatedness of the error process with a p-value 3.578×10^{-7}). Thus a relatively large block size or bandwidth is needed to account for the dependence within the error process. In this case, the statistical significance of the unemployment rate is questioned. It is worth noting that the fixed-smoothing asymptotics generally provides consistent

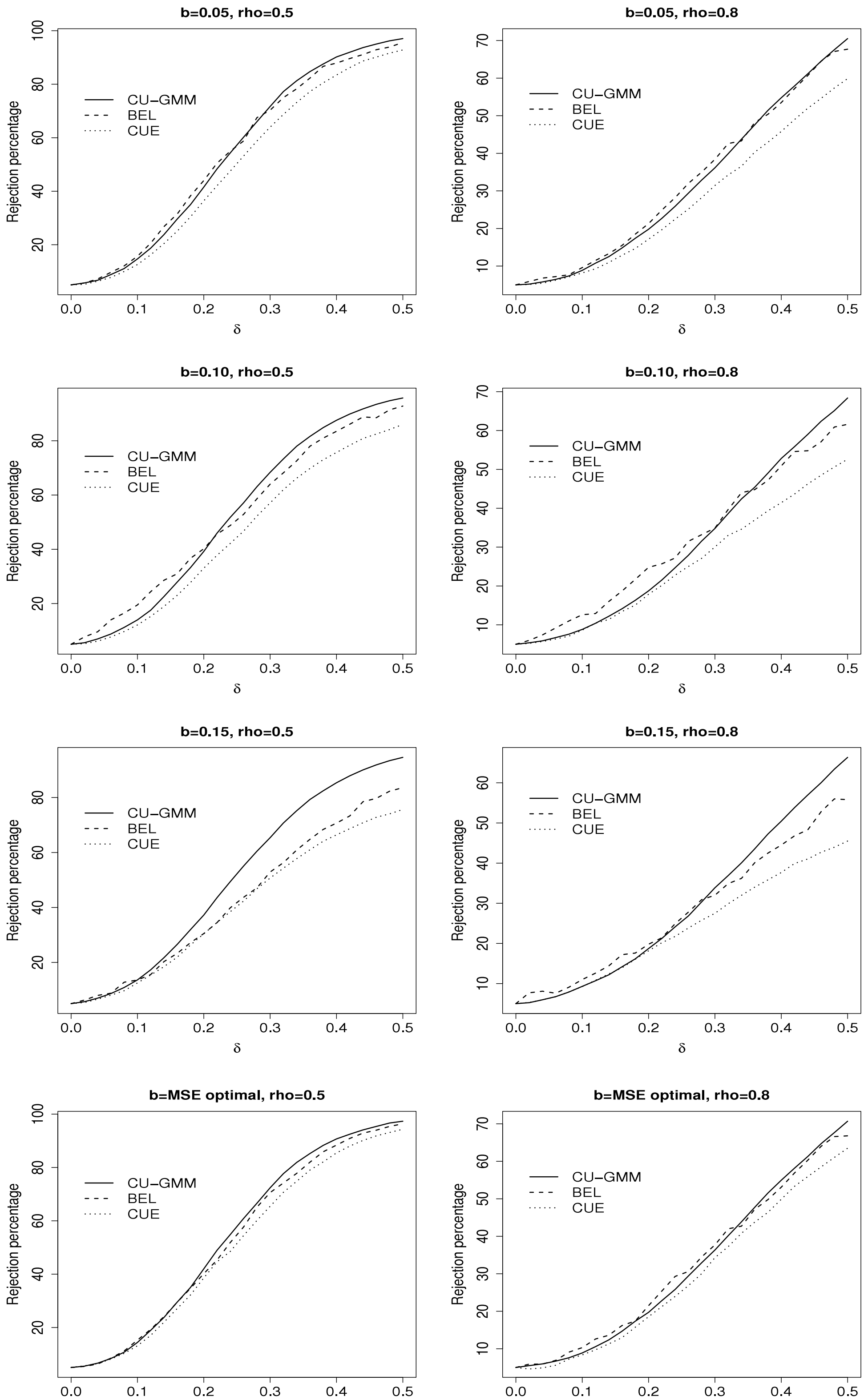

Fig. 8. Size-adjusted powers in % for the CU-GMM based Wald test, the CUE and BEL based LR tests. The nominal level is 5%, and the number of Monte Carlo replications is 5000.

results for relatively larger B, while the small-b asymptotics may deliver small p-values (below 5% or 10% nominal level) for large B, which is potentially due to its poor approximation to the sampling distributions when B/n is away from zero.

8. Concluding remarks

In this paper, we have established the fixed-smoothing asymptotics for CU-GMM, CUE based on the saddle point problem, and BEL. Through theoretical and numerical studies, we demonstrate that the fixed-smoothing asymptotics provides better approximation to the finite sample distributions in many situations. Our numerical results also reveal that the CU-GMM estimator is asymptotically more efficient and the corresponding specification tests are more powerful as compared to the other two competitors. To conclude, we point out a few directions for further research. First, building on the above techniques, we expect the results in this paper can be extended to other members in the GEL class such as exponential tilting (Kitamura and Stutzer, 1997) and the members of the Cressie–Read power divergence family of discrepancies (Imbens et al., 1998).[2] When the convex hull constraint is required in the formulation of GEL, special attention must be paid. We believe that the discussion for BEL sheds considerable light in this case. Second, it is of interest to study the high order properties of the GEL estimator and the GEL-based test statistics under the fixed-smoothing asymptotics. In view of the simulation results in Section 6, it is expected that the asymptotic limits derived under the fixed-smoothing asymptotics provide high order refinement over the traditional chi-square based inference. Finally, we have used the conventional MSE criterion to select the bandwidth parameter and block size in our simulation studies. It is also interesting to extend the methods proposed in Sun et al. (2008) and Sun (2014b) that are tailored to confidence interval construction and hypothesis testing to the current setting.

Acknowledgments

The author is grateful to Xiaofeng Shao for many helpful comments and suggestions on an earlier version. The author would like to thank the Editor, the Associate Editor and three reviewers for their insightful comments and constructive suggestions, which substantially improved the paper. All errors are my own.

Appendix A

Throughout the Appendix, let C be generic constant which is independent of n and it can be different from line to line. Denote by $\mathrm{tr}(\cdot)$ the trace of a matrix. For two square matrices $A, B \in \mathbb{R}^{k\times k}$, $A \le B$ means that $B - A$ is positive semi-definite. We use $|\cdot|_1$ and $|\cdot| := |\cdot|_2$ to denote the L_1 norm and the Euclidean norm respectively, and use $\|\cdot\|$ to denote the matrix spectral norm. With some abuse of notation, denote by $\tilde{\theta}$ the variable in the line segment joining the GEL estimator and θ_0, and $\tilde{\theta}$ can vary different from place to place.

A.1. Proofs of the main results in Section 2

Proof of Lemma 2.1. Using summation by parts, we have for any $\theta \in \Theta$,

$$\mathrm{tr}\{\Omega(\theta)\} = \mathrm{tr}\left[\frac{1}{n}\sum_{i,l=1}^{n}\mathcal{K}(i/n, l/n)(f_i(\theta) - \bar{f}_n(\theta))(f_l(\theta) - \bar{f}_n(\theta))'\right]$$

$$= \mathrm{tr}\left[\frac{1}{n}\sum_{k,l=1}^{n}k(\bar{f}_k(\theta) - \bar{f}_n(\theta))(\mathcal{K}_{k,l} - \mathcal{K}_{k+1,l})(f_l(\theta) - \bar{f}_n(\theta))'\right]$$

$$\le \sup_{1\le t\le n}\left|t\{\bar{f}_t(\theta) - \bar{f}_n(\theta)\}/n\right|\sum_{k,l=1}^{n-1}|l(\bar{f}_l(\theta) - \bar{f}_n(\theta))|\cdot|c_{k,l}|$$

$$\le n\sup_{1\le t\le n}\left|t\{\bar{f}_t(\theta) - \bar{f}_n(\theta)\}/n\right|^2\sum_{k,l=1}^{n-1}|c_{k,l}|,$$

where we have used the Cauchy–Schwarz inequality. Notice that

$$n\sup_{1\le t\le n}|t\{\bar{f}_t(\theta) - \bar{f}_n(\theta)\}/n|^2$$
$$\le 2n\sup_{1\le t\le n}\left|\frac{1}{n}\sum_{j=1}^{t}(f_j(\theta) - \mathbb{E}f_j(\theta))\right|^2 + 2n\sup_{1\le t\le n}\left|\frac{t}{n^2}\sum_{j=1}^{n}(f_j(\theta) - \mathbb{E}f_j(\theta))\right|^2 \le a_n,$$

where a_n does not depend on θ. Thus with probability approaching one, $\Omega_n(\theta) \le Ca_nI_k$ uniformly for all $\theta \in \Theta$ and $a_n/n \to 0$. Let $\mathcal{L}_{cue,n}(\theta) = n\bar{f}_n'(\theta)\Omega_n^{-1}(\theta)\bar{f}_n(\theta)$. Then we have $0 \le \mathcal{L}_{cue,n}(\hat{\theta}_{cue}) \le \mathcal{L}_{cue,n}(\theta_0)$, where $\mathcal{L}_{cue,n}(\theta_0) = O_p(1)$ by Assumption 2.4 and the continuous mapping theorem. Note with probability approaching one, $\Omega^{-1}(\hat{\theta}_{cue}) \ge Ca_n^{-1}I_k$ which implies that $a_n^{-1}|\bar{f}_n(\hat{\theta}_{cue})|^2 = O_p(1/n)$, that is $|\bar{f}_n(\hat{\theta}_{cue})| = o_p(1)$. Now by the assumption that $\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^{n}(f_j(\theta) - \mathbb{E}f_j(\theta))\right| = o_p(1)$, we get

$$|h(\hat{\theta}_{cue})| \le |\bar{f}_n(\hat{\theta}_{cue}) - h(\hat{\theta}_{cue})| + |\bar{f}_n(\hat{\theta}_{cue})|$$
$$\le \sup_{\theta\in\Theta}|\bar{f}_n(\theta) - h(\theta)| + |\bar{f}_n(\hat{\theta}_{cue})| = o_p(1).$$

By Assumption 2.2, we have $\hat{\theta}_{cue} \to^p \theta_0$, which completes the proof. $\diamond$

Proof of Theorem 2.1. It is shown in Lemma A.1 that $P(\mathcal{A}_c) \to 1$. Conditional on $\mathcal{A}_c$, denote by $\hat{\theta}_{cue}$ a minimizer in $\mathcal{B}_c$. For $\tau_n = o_p(n^{-1/6})$ and $\hat{\theta}_{cue} \in \mathcal{B}_c$, we have $\tau_n\sqrt{n}|\hat{\theta}_{cue} - \theta_0| = o_p(1)$. The FOC of the optimization problem (4) is given by

$$2n\bar{g}_{nj}'(\hat{\theta}_{cue})\Omega_n^{-1}(\hat{\theta}_{cue})\bar{f}_n(\hat{\theta}_{cue}) + n\bar{f}_n'(\hat{\theta}_{cue})\frac{\partial\Omega_n^{-1}}{\partial\theta_j}(\hat{\theta}_{cue})\bar{f}_n(\hat{\theta}_{cue}) = 0, \quad 1 \le j \le p, \tag{37}$$

where $\bar{g}_n(\theta) = (\bar{g}_{n1}(\theta), \ldots, \bar{g}_{np}(\theta))$. By matrix calculus, we have

$$\frac{\partial\Omega_n^{-1}}{\partial\theta_j}(\theta) = -\Omega_n^{-1}(\theta)\frac{\partial\Omega_n}{\partial\theta_j}(\theta)\Omega_n^{-1}(\theta), \quad 1 \le j \le p. \tag{38}$$

Let $A_{nj}(\theta) = \frac{1}{n}\sum_{i,l=1}^{n}\mathcal{K}(i/n, l/n)(g_{ij}(\theta) - \bar{g}_{nj}(\theta))(f_l(\theta) - \bar{f}_n(\theta))'$. It is not hard to see that $\partial\Omega_n(\theta)/\partial\theta_j = A_{nj}(\theta) + A_{nj}(\theta)'$. Using Taylor expansion $\bar{f}_t(\hat{\theta}_{cue}) = \bar{f}_t(\theta_0) + \bar{g}_t(\tilde{\theta}_t)(\hat{\theta}_{cue} - \theta_0)$ for $1 \le t \le n$, where $\tilde{\theta}_t$ is in the line segment joining $\hat{\theta}_{cue}$ and θ_0, and it could be different for each element of $\bar{g}_t$ with $1 \le t \le n$, we have

$$\sup_{1\le l\le n}\left|l\{\bar{f}_l(\hat{\theta}_{cue}) - \bar{f}_n(\hat{\theta}_{cue})\}/n\right|$$
$$\le \sup_{1\le l\le n}\left|l\{\bar{f}_l(\theta_0) - \bar{f}_n(\theta_0)\}/n\right| + |\hat{\theta}_{cue} - \theta_0|\sup_{1\le l\le n}\left\|l\{\bar{g}_l(\tilde{\theta}_l) - \bar{g}_n(\tilde{\theta}_n)\}/n\right\|$$
$$= O_p(\tau_n|\hat{\theta}_{cue} - \theta_0| + n^{-1/2}).$$

For $a_1, a_2 \in \mathbb{R}^k$ with $|a_1| = |a_2| = 1$, using summation by parts and the Cauchy–Schwarz inequality, we have

$$|a_1'A_{nj}(\hat{\theta}_{cue})a_2|$$
$$= \left|\frac{1}{n}\sum_{i,l=1}^{n}a_1'\mathcal{K}(i/n, l/n)(g_{ij}(\hat{\theta}_{cue}) - \bar{g}_{nj}(\hat{\theta}_{cue}))\right.$$

[2] The GEL estimator based on smoothed moment conditions (Smith, 2011) can be defined as $\hat{\theta}_{gel} = \arg\min_\theta\max_\lambda\sum_{t=1}^{n}\rho(\lambda' f_{tn,\mathcal{G}}(\theta))$, where $\rho(\cdot)$ is a concave function on its domain.

$$\times (f_l(\hat\theta_{cue}) - \bar f_n(\hat\theta_{cue}))' a_2 \Big|$$
$$= \Big| \frac{1}{n}\sum_{k,l=1}^{n} a_1' k(\bar g_{kj}(\hat\theta_{cue}) - \bar g_{nj}(\hat\theta_{cue}))(\mathcal{K}_{k,l} - \mathcal{K}_{k+1,l})$$
$$\times (f_l(\hat\theta_{cue}) - \bar f_n(\hat\theta_{cue}))' a_2 \Big|$$
$$\le \sup_{1\le t\le n} \left| t\{\bar g_{tj}(\hat\theta_{cue}) - \bar g_{nj}(\hat\theta_{cue})\}/n \right|$$
$$\times \sum_{k,l=1}^{n-1} |l(\bar f_l(\hat\theta_{cue}) - \bar f_n(\hat\theta_{cue}))| \cdot |c_{k,l}|$$
$$\le \sup_{1\le t\le n} \left| t\{\bar g_{tj}(\hat\theta_{cue}) - \bar g_{nj}(\hat\theta_{cue})\}/n \right|$$
$$\times \sup_{1\le l\le n} \left| l\{\bar f_l(\hat\theta_{cue}) - \bar f_n(\hat\theta_{cue})\}/n \right| \left(n \sum_{k,l=1}^{n-1} |c_{k,l}| \right)$$
$$= O_p(\tau_n^2 n |\hat\theta_{cue} - \theta_0| + \tau_n \sqrt{n}), \tag{39}$$

where we have used the fact that $\sup_{1\le t\le n} \|t\{\bar g_t(\hat\theta_{cue}) - \bar g_n(\hat\theta_{cue})\}/n\| \le \sup_{1\le t\le n} \|t\bar g_t(\hat\theta_{cue})/n - tG(\hat\theta_{cue})/n\| + \sup_{1\le t\le n} \|t\bar g_n(\hat\theta_{cue})/n - tG(\hat\theta_{cue})/n\| = O_p(\tau_n)$. Note that by Assumption 2.3, the above derivation also holds if $\hat\theta_{cue}$ is replaced by any $\tilde\theta$ between $\hat\theta_{cue}$ and θ_0. On the other hand, note that

$$\Omega_n(\hat\theta_{cue}) = \Omega_n(\theta_0) + \sum_{j=1}^{p} \frac{\partial \Omega_n}{\partial \theta_j}(\tilde\theta)(\hat\theta_{cue,j} - \theta_{0j})$$
$$= \Omega_n(\theta_0) + e_n, \tag{40}$$

where $\|e_n\| = O_p(\tau_n\sqrt{n}|\hat\theta_{cue} - \theta_0| + n\tau_n^2|\hat\theta_{cue} - \theta_0|^2) = o_p(1)$. Thus we must have $\Omega_n^{-1}(\hat\theta_{cue}) = \Omega_n^{-1}(\theta_0) + o_p(1)$. Note that $\bar f_n(\hat\theta_{cue}) = \bar f_n(\theta_0) + \bar g_n(\tilde\theta)(\hat\theta_{cue} - \theta_0)$. The FOC implies that

$$\sqrt{n}\bar g_{nj}'(\hat\theta_{cue})\Omega_n^{-1}(\hat\theta_{cue})\bar f_n(\hat\theta_{cue}) + \sqrt{n}\bar f_n'(\hat\theta_{cue})\frac{\partial \Omega_n^{-1}}{\partial \theta_j}(\hat\theta_{cue})\bar f_n(\hat\theta_{cue})/2$$
$$= \sqrt{n}\bar g_{nj}'(\hat\theta_{cue})(\Omega_n^{-1}(\theta_0) + o_p(1))(\bar f_n(\theta_0) + \bar g_n(\tilde\theta)(\hat\theta_{cue} - \theta_0))$$
$$- n\bar f_n'(\hat\theta_{cue})\Omega_n^{-1}(\hat\theta_{cue})O_p\left(\sqrt{n}\tau_n^2|\hat\theta_{cue} - \theta_0| + \tau_n\right)$$
$$\times \Omega_n^{-1}(\hat\theta_{cue})\bar f_n(\hat\theta_{cue})/2 = 0.$$

Therefore, we obtain

$$\sqrt{n}\{\bar f_n(\theta_0) + \bar g_n(\tilde\theta)(\hat\theta_{cue} - \theta_0)\}'\Omega_n^{-1}(\theta_0)\bar g_n(\hat\theta_{cue})(1 + o_p(1)) = 0.$$

Solving the above equation, we get

$$\sqrt{n}(\hat\theta_{cue} - \theta_0) = -\left(\bar g_n(\hat\theta_{cue})'\Omega_n^{-1}(\theta_0)\bar g_n(\tilde\theta)\right)^{-1}$$
$$\times \bar g_n(\hat\theta_{cue})'\Omega_n^{-1}(\theta_0)\sqrt{n}\bar f_n(\theta_0)(1 + o_p(1)).$$

The result thus follows from the functional central limit theorem and the continuous mapping theorem. ◇

Lemma A.1. *Under the assumptions in Theorem 2.1, we have* $P(\mathcal{A}_c) \to 1$.

Proof of Lemma A.1. Without loss of generality, we suppose $c = 1$. Note that for $\hat\theta_n = \theta_0 + c_0 n^{-1/3}$ with $|c_0| = 1$, we have

$$\bar f_n(\hat\theta_n) = \bar f_n(\theta_0) + \bar g_n(\breve\theta)c_0 n^{-1/3} = O_p(n^{-1/3}).$$

and

$$\Omega_n(\hat\theta_n) = \frac{1}{n}\sum_{i,l=1}^{n} \mathcal{K}(i/n, l/n)$$
$$\times \left\{ f_i(\theta_0) - \bar f_n(\theta_0) + (g_i(\breve\theta) - \bar g_n(\breve\theta))c_0 n^{-1/3} \right\}$$
$$\times \left\{ f_l(\theta_0) - \bar f_n(\theta_0) + (g_l(\breve\theta) - \bar g_n(\breve\theta))c_0 n^{-1/3} \right\}',$$

where $\breve\theta$ is between θ_0 and $\hat\theta_n$, and it could be different from place to place. Using similar arguments as in (39), it can be shown that $\Omega_n(\hat\theta_n) = \Omega_n(\theta_0) + o_p(1)$. Thus we have $P(\mathcal{L}_{cue,n}(\theta_0) < \sup_{c_0:|c_0|=1} \mathcal{L}_{cue,n}(\hat\theta_n)) \to 1$, which implies that $P(\mathcal{L}_{cue,n}(\theta_0) < \sup_{c_0:|c_0|=1} \mathcal{L}_{cue,n}(\hat\theta_n)) \le P(\mathcal{A}_c) \to 1$. ◇

A.2. Proofs of the main results in Section 3

Proof of Lemma 3.1. By definition, we have

$$\max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N} -(1 + \lambda' f_{tn}(\breve\theta_{cue}))^2 \le \max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N} -(1 + \lambda' f_{tn}(\theta_0))^2.$$

Define $V_{tn}(\theta) = f_{tn}(\theta) - h(\theta)$. Choosing $\lambda = -\text{sign}(h(\breve\theta_{cue}))/|h(\breve\theta_{cue})|_1$, where $\text{sign}(h(\breve\theta_{cue}))$ is a vector containing the signs for each element of $h(\breve\theta_{cue})$, we have

$$\max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N} -(1 + \lambda' f_{tn}(\breve\theta_{cue}))^2$$
$$= \max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N} -(1 + \lambda' V_{tn}(\breve\theta_{cue}) + \lambda' h(\breve\theta_{cue}))^2$$
$$\ge -NC\frac{\sup_{\theta\in\Theta}\max_{1\le t\le N} |V_{tn}(\theta)|^2}{|h(\breve\theta_{cue})|_1^2},$$

where C is a constant which does not depend on n. It implies that

$$\frac{\sup_{\theta\in\Theta}\max_{1\le t\le N} |V_{tn}(\theta)|^2}{|h(\breve\theta_{cue})|_1^2} \ge -\frac{1}{CN}\max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N} -(1 + \lambda' f_{tn}(\theta_0))^2.$$

Note that

$$-\frac{1}{N}\max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N} -(1 + \lambda' f_{tn}(\theta_0))^2 = \frac{1}{N}\sum_{t=1}^{N}(1 + \lambda_n(\theta_0)' f_{tn}(\theta_0))^2,$$

where

$$\lambda_n(\theta_0) = -\left\{\sum_{t=1}^{N} f_{tn}(\theta) f_{tn}(\theta)'\right\}^{-1} \left[\sum_{t=1}^{N} f_{tn}(\theta)\right].$$

Using the functional central limit theorem, we deduce that

$$\frac{1}{N}\sum_{t=1}^{N}(1 + \lambda_n(\theta_0)' f_{tn}(\theta_0))^2 \to^d \frac{1}{1-b}\int_0^{1-b}(1 + \mathcal{S}_k(r;b))^2 dr,$$

for

$$\mathcal{S}_k(r;b) = D_k(r;b)'\left[\int_0^{1-b} D_k(u;b)D_k(u;b)' du\right]^{-1}$$
$$\times \left[\int_0^{1-b} D_k(u;b) du\right].$$

Because $P(\int_0^{1-b}(1 + \mathcal{S}_k(r;b))^2 dr > 0) = 1$, we have $-\frac{1}{N}\max_{\lambda\in\mathbb{R}^k} \sum_{t=1}^{N} -(1 + \lambda' f_{tn}(\theta_0))^2$ bounded away from zero in probability. Thus we get $|h(\breve\theta_{cue})|_1^2 = o_p(1)$. By Assumption 2.2, we have $\breve\theta_{cue} \to^p \theta_0$. ◇

Proof of Lemma 3.3. Define $\breve{Y} = bPV\mathcal{Z}$. Note that for any Y and $k > p$,

$$\tilde{Q}_{cue}(Y) := \max_{\lambda\in\mathbb{R}^p}\int_0^{1-b}\left\{1-\left(1+\lambda'\tilde{D}_k(r;b)+\lambda'_{1:p}Y\right)^2\right\}dr \geq \max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p}}\int_0^{1-b}\left\{1-\left(1+\lambda'_{p+1:k}\tilde{D}_{k,p+1:k}(r;b)\right)^2\right\}dr,$$

where $\lambda = (\lambda_1,\dots,\lambda_k)'$ and $\lambda_{a:a'} = (\lambda_a,\dots,\lambda_{a'})'$ with $a \le a'$. We show that the lower bound can be achieved by setting $Y = \breve{Y}$. Because $h(\lambda) := -\int_0^{1-b}\left(1+\lambda'\tilde{D}_k(r;b)+\lambda'_{1:p}\breve{Y}\right)^2 dr$ is a concave function of λ, and its maximum is at least $b-1$, the maximizer of $h(\lambda)$ satisfies the FOC

$$\int_0^{1-b}\left(1+\lambda'\tilde{D}_k(r;b)+\lambda'_{1:p}\breve{Y}\right)(\tilde{D}_{k,1:p}(r;b)+\breve{Y})dr = 0,$$
$$\int_0^{1-b}\left(1+\lambda'\tilde{D}_k(r;b)+\lambda'_{1:p}\breve{Y}\right)\tilde{D}_{k,p+1:k}(r;b)dr = 0.$$

Notice that $\breve{Y}$ satisfies the equation

$$\int_0^{1-b}\left\{1+\breve{\lambda}'_{p+1:k}\tilde{D}_{k,p+1:k}(r;b)\right\}(\tilde{D}_{k,1:p}(r;b)+\breve{Y})dr = 0, \quad (41)$$

and $\breve{\lambda}_{p+1:k}$ is the maximizer of $\max_{\lambda_{p+1:k}\in\mathbb{R}^{k-p}} -\int_0^{1-b}\left\{1+\lambda'_{p+1:k}\tilde{D}_{k,p+1:k}(r;b)\right\}^2 dr$. It is not hard to check that $\breve{\lambda} = (0,\breve{\lambda}_{p+1:k})$ satisfies the FOC, and thus by the concavity,

$$\tilde{Q}_{cue,n}(\breve{Y}) = \int_0^{1-b}\left\{1-\left(1+\breve{\lambda}'_{p+1:k}\tilde{D}_{k,p+1:k}(r;b)\right)^2\right\}dr. \quad (42)$$

Hence we have $\tilde{Q}_{cue}(\breve{Y}) = \min_{Y\in\mathbb{R}^p}\tilde{Q}_{cue}(Y)$. If there is another minimizer $\hat{Y}$, then it satisfies the FOC

$$\int_0^{1-b}\left(1+\lambda'\tilde{D}_k(r;b)+\lambda'_{1:p}\hat{Y}\right)\lambda_{1:p}dr = 0,$$
$$\int_0^{1-b}\left(1+\lambda'\tilde{D}_k(r;b)+\lambda'_{1:p}\hat{Y}\right)(\tilde{D}_{k,1:p}(r;b)+\hat{Y})dr = 0,$$
$$\int_0^{1-b}\left(1+\lambda'\tilde{D}_k(r;b)+\lambda'_{1:p}\hat{Y}\right)\tilde{D}_{k,p+1:k}(r;b)dr = 0.$$

It is straightforward to check $\lambda_{1:p} = 0$, $\lambda_{p+1:k} = \breve{\lambda}_{p+1:k}$ and $\hat{Y} = \breve{Y}$. Therefore, the minimizer $\breve{Y}$ is unique. ⋄

Proof of Theorem 3.1. Suppose $\Lambda^{-1}g_{tn}(\tilde{\theta})$ with $\tilde{\theta} = \tilde{\theta}(Y/\sqrt{n}+\theta_0,\theta_0)$ has the singular value decomposition $\Lambda^{-1}g_{tn}(\tilde{\theta}) = U_{tn}\Sigma_{tn}V_{tn}$ where $U_{tn}U'_{tn} = U'_{tn}U_{tn} = I_k$, $V_{tn}V'_{tn} = V'_{tn}V_{tn} = I_p$, and

$$\Sigma_{tn} = \begin{pmatrix} P_{tn} \\ O_{(k-p)\times p}\end{pmatrix},$$

where P_{tn} is a diagonal matrix with singular values on the main diagonal, and O is the matrix of zeros. Then we have

$$Q_{cue,n}(Y) = \frac{1}{nb}\max_{\lambda\in\mathbb{R}^k} \times\sum_{t=1}^N\left\{1-(1+\lambda'U'\sqrt{n}b\Lambda^{-1}f_{tn}(\theta_0)+\lambda'U'U_{tn}\Sigma_{tn}bV_{tn}Y)^2\right\}$$
$$= \frac{1}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N\left\{1-\left(1+\lambda'D_{tn}(\theta_0;b)+\lambda'U'U_{tn}\begin{pmatrix}S_{tn}\\ O\end{pmatrix}\right)^2\right\}$$
$$= \frac{1}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N\left\{1-(1+\lambda'D_{tn}(\theta_0;b)+\lambda'\xi_{tn})^2\right\},$$

where $D_{tn}(\theta;b) = U'\sqrt{n}b\Lambda^{-1}f_{tn}(\theta)$, $S_{tn} = bP_{tn}V_{tn}Y$, and

$$\xi_{tn} = \begin{pmatrix}\xi_{tn,1}\\ \xi_{tn,2}\end{pmatrix} = U'U_{tn}\begin{pmatrix}S_{tn}\\ O\end{pmatrix}.$$

Under Assumption 2.4, we use the Skorokhod's embedding theorem (strong approximation) to embed $\{t\bar{f}_t/\sqrt{n}\}$ and $\{W_k(r)\}$ in a larger probability space such that

$$\sup_{0\le r\le 1}\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nr\rfloor}f_j(\theta_0)-\Lambda W_k(r)\right\| \to^{a.s.} 0.$$

Below the considered convergence will be pointwise along some fixed $\omega\in\mathcal{C}$ where $P(\mathcal{C}) = 1$. Let $\lambda_n(Y)$ and $\lambda(Y)$ be the maximizers in $Q_{cue,n}(Y)$ and $Q_{cue}(Y)$ respectively. The FOC implies that

$$\lambda_n(Y) = -\left(\frac{1}{n}\sum_{t=1}^N(D_{tn}(\theta_0;b)+\xi_{tn})(D_{tn}(\theta_0;b)+\xi_{tn})'\right)^{-1} \times\frac{1}{n}\sum_{t=1}^N(D_{tn}(\theta_0;b)+\xi_{tn})$$
$$= -\Psi_n^{-1}(Y)\frac{1}{n}\sum_{t=1}^N(D_{tn}(\theta_0;b)+\xi_{tn}),$$

and

$$\lambda(Y) = -\left(\int_0^{1-b}(\tilde{D}_k(r;b)+Y_{ex})(\tilde{D}_k(r;b)+Y_{ex})'dr\right)^{-1} \times\int_0^{1-b}(\tilde{D}_k(r;b)+Y_{ex})dr$$
$$= -\Psi^{-1}(Y)\int_0^{1-b}(\tilde{D}_k(r;b)+Y_{ex})dr,$$

where $Y_{ex} = (b(PVY)', 0'_{k-p})'$, and $\Psi_n(Y)$ and $\Psi(Y)$ are defined implicity. Note that $\Psi_n(Y) = \int_0^{1-b}(\tilde{D}_k(r;b)+Y_{ex})(\tilde{D}_k(r;b)+Y_{ex})'dr$ is positive semi-definite almost surely for any given Y. If there exists a vector β such that $\beta'\Psi_n(Y)\beta = 0$, then we have $\beta'\tilde{D}_k(r;b) = -\beta'Y_{ex}$ for all $r\in[0,1-b]$. By the independence among the components of $\tilde{D}_k(r;b)$, this is impossible almost surely.

Below, we show that $Q_{cue,n}$ is asymptotically tight in $l^\infty(K)$ for any compact set $K\subset\mathbb{R}^p$. Because $\sup_{Y\in K}|\tilde{\theta}(Y/\sqrt{n}+\theta_0,\theta_0)-\theta_0| \le \sup_{Y\in K}|Y|/\sqrt{n} \to 0$, $\sup_{\theta\in\mathcal{N}_{\epsilon^*}(\theta_0)}\sup_{1\le t\le N}\|g_{tn}(\theta)-G(\theta)\|\to^{a.s.}0$ and $G(\theta)$ is Lipschitz continuous on $\mathcal{N}_{\epsilon^*}(\theta_0)$, we have $\sup_{Y\in K}\sup_{1\le t\le N}|\xi_{tn,1}-bPVY|\to 0$ and $\sup_{Y\in K}\sup_{1\le t\le N}|\xi_{tn,2}|\to 0$ for any $\omega\in\mathcal{C}$. Notice that uniformly for $Y\in K$, $\|\Psi(Y)\|$ is bounded from below and hence

$$\left|\frac{1}{n}\sum_{t=1}^N(D_{tn}(\theta_0;b)+\xi_{tn})-\int_0^{1-b}(\tilde{D}_k(r;b)+Y_{ex})dr\right|\to 0,$$
$$\Psi_n^{-1}(Y)-\Psi^{-1}(Y) = \Psi_n^{-1}(Y)(\Psi(Y)-\Psi_n(Y))\Psi^{-1}(Y)\to 0.$$

Therefore, we have $|\lambda_n(Y)-\lambda(Y)|\to 0$ which implies that

$$b|Q_{cue}(Y)-Q_{cue,n}(Y)| = \left|\int_0^{1-b}\left(1+\lambda'(Y)\tilde{D}_k(r;b)+\lambda'_{1:p}(Y)bPVY\right)^2dr - \frac{1}{n}\sum_{t=1}^N(1+\lambda'_n(Y)D_{tn}(\theta_0;b)+\lambda'_n(Y)\xi_{tn})^2\right|+o(1)\to 0,$$

uniformly for $Y\in K$. Note for $Y_1, Y_2\in K$,

$$|Q_{cue,n}(Y_1) - Q_{cue,n}(Y_2)| \le 2 \sup_{Y \in K} |Q_{cue}(Y) - Q_{cue,n}(Y)| + |Q_{cue}(Y_1) - Q_{cue}(Y_2)|.$$

For any $\epsilon, \eta > 0$, there exists a $\delta > 0$ such that for large enough n,

$$\begin{aligned} &P\left(\sup_{|Y_1-Y_2|<\delta, Y_1, Y_2 \in K} |Q_{cue,n}(Y_1) - Q_{cue,n}(Y_2)| > \epsilon\right) \\ &\le P\left(\sup_{Y \in K} |Q_{cue}(Y) - Q_{cue,n}(Y)| > \epsilon/4\right) \\ &\quad + P\left(\sup_{|Y_1-Y_2|<\delta, Y_1, Y_2 \in K} |Q_{cue}(Y_1) - Q_{cue}(Y_2)| > \epsilon/2\right) \\ &< \eta, \end{aligned}$$

where we have used the continuity of Q_{cue}. For any given Y, it is not hard to verify that $Q_{cue,n}(Y)$ is asymptotically tight as it has a quadratic form. Thus we have shown that $Q_{cue,n}$ is asymptotically tight in $l^\infty(K)$. By Lemmas 3.2 and 3.3, we have $\mathcal{Z}_n = \mathrm{argmin} Q_{cue,n}(Y)$ is uniformly tight and the minimizer of Q_{cue} is unique. Therefore by the Argmax/Argmin continuous mapping theorem (see e.g. Theorem 3.22 of van der Vaart and Wellner (1996)), we have

$$\begin{aligned} \mathcal{Z}_n &= \mathrm{argmin} Q_{cue,n}(Y) \to^d \mathrm{argmin} Q_{cue}(Y) \\ &= b^{-1}V^{-1}P^{-1}\breve{Y} = \mathcal{Z}, \end{aligned} \tag{43}$$

which completes the proof. $\diamond$

Proof of Lemma 3.2. Let ϵ_n be a sequence of positive numbers such that $\epsilon_n = o_p(1)$, $1/(\sqrt{n}\epsilon_n) = o_p(1)$ and $\breve{\theta}_{cue} \in \mathcal{N}_{\epsilon_n}(\theta_0)$. Consider a shrinking neighborhood $\mathcal{N}_{\kappa/\sqrt{n}}(\theta_0)$ for some $\kappa > 0$. For large enough n, we have $\mathcal{N}_{\epsilon_n}(\theta_0) \setminus \mathcal{N}_{\kappa/\sqrt{n}}(\theta_0) \neq \emptyset$. For $\hat{\theta}_n \in \mathcal{N}_{\epsilon_n}(\theta_0) \setminus \mathcal{N}_{\kappa/\sqrt{n}}(\theta_0)$, by setting $\lambda = -\sqrt{n}G_0(\hat{\theta}_n - \theta_0)/|\sqrt{n}G_0(\hat{\theta}_n - \theta_0)|^2$, we deduce that

$$\begin{aligned} gel(\hat{\theta}_n) &= \frac{1}{nb} \max_{\lambda \in \mathbb{R}^k} \\ &\quad \times \sum_{t=1}^{N} \{1 - (1 + \lambda'\sqrt{n}f_{tn}(\theta_0) + \lambda'\sqrt{n}g_{tn}(\tilde{\theta})(\hat{\theta}_n - \theta_0))^2\} \\ &\ge \frac{1}{nb} \sum_{t=1}^{N} \{1 - (\sqrt{n}(\hat{\theta}_n - \theta_0)'G_0'b^{-1}\Lambda D_k(t/n; b)/|\sqrt{n}G_0 \\ &\quad \times (\hat{\theta}_n - \theta_0)|^2(1 + o_p(1)) + r_{tn})^2\}, \end{aligned}$$

where $\tilde{\theta}$ on the line segment joining $\hat{\theta}_n$ and θ_0, and it can vary across equations, and $r_{tn} = o_p(1)$ uniformly for t. Note that $|\sqrt{n}G_0(\hat{\theta}_n - \theta_0)|^2 = n(\hat{\theta}_n - \theta_0)'G_0'G_0(\hat{\theta}_n - \theta_0) \ge C\kappa^2$ since $\mathrm{rank}(G_0) = p$, and $\sqrt{n}(\hat{\theta}_n - \theta_0)'G_0'b^{-1}\Lambda D_k(t/n; b) = O_p(\kappa)$. For large enough κ, we must have $gel_{cue}(\hat{\theta}_n) > \min_{\theta \in \Theta} gel_{cue}(\theta)$ because $\min_{\lambda \in \mathbb{R}^k} \int_0^{1-b} (1 + \lambda' D_k(r; b))^2 dr > 0$ with probability one. Therefore we have $\breve{\theta}_{cue} \in \mathcal{N}_{\kappa/\sqrt{n}}(\theta_0)$ which implies the uniformly tightness of $\mathcal{Z}_n$. $\diamond$

A.3. Proofs of the main results in Section 4

Proof of Lemma 4.1. We use the Skorokhod's embedding theorem (strong approximation) to embed $\{t\bar{f}_t/\sqrt{n}\}$ and $\{W_k(r)\}$ in a larger probability space such that

$$\sup_{0 \le r \le 1} \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} f_j(\theta_0) - \Lambda W_k(r) \right\| \to^{a.s.} 0.$$

Let $g_{tn}(\theta) = \sum_{j=t}^{t+B-1} g_j(\theta)/B$ for $t = 1, 2, \ldots, N$, and $\hat{\theta}_n$ be a $\sqrt{n}$-consistent estimator of θ_0. We can always choose a subsequence $\hat{\theta}_{n_j}$ so that $\hat{\theta}_{n_j} \to^{a.s} \theta_0$. For the ease of notation, we assume that $\hat{\theta}_n$ is strongly consistent. Set $Z_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$. Using the mean value theorem, we have $\sqrt{n}f_{tn}(\hat{\theta}_n) = \sqrt{n}f_{tn}(\theta_0) + g_{tn}(\tilde{\theta})Z_n$, where $\tilde{\theta}$ is between θ_0 and $\hat{\theta}_n$, and it could be different for each element of g_{tn}. By the definition of $\hat{\theta}_{el}$, we have

$$\begin{aligned} elr(\hat{\theta}_{el}) &\le \frac{2}{nb} \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^{N} \log(1 + \lambda' f_{tn}(\hat{\theta}_n)) \\ &= \frac{2}{nb} \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^{N} \log(1 + \lambda'\sqrt{n}f_{tn}(\theta_0) + \lambda' g_{tn}(\tilde{\theta})Z_n) \\ &= \frac{2}{nb} \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^{N} \log(1 + \lambda'\sqrt{n}f_{tn}(\theta_0) + \lambda' G_0 Z_n \\ &\quad + \lambda'\{g_{tn}(\tilde{\theta}) - G_0\}Z_n). \end{aligned}$$

Below we consider the pointwise convergence for each $\omega \in \mathcal{C} \cap \mathcal{D}_b$ with $P(\mathcal{C}) = 1$. Define $D_{tn}(\theta_0; b) = \sqrt{n}bU'\Lambda^{-1}f_{tn}(\theta_0) = (D_{tn,1}(\theta_0; b), \ldots, D_{tn,k}(\theta_0; b))'$. For any $1 \le a \le a' \le k$, denote $D_{tn,a:a'}(\theta_0; b) = (D_{tn,a}(\theta_0; b), \ldots, D_{tn,a'}(\theta_0; b))'$. If the origin of $\mathbb{R}^{k-p}$ is contained in the interior of the convex hull of $\{\tilde{D}_{k,p+1:k}(r; b)\}$, then by strong approximation, the origin is contained in the interior of the convex hull of $\{D_{tn,p+1:k}(\theta_0; b)\}_{t=1}^N$ when n is large enough. By the independence, the convex hull of $\{D_{tn,p+1:k}(\theta_0; b)\}_{t=1}^N$ ($\{D_{tn,1:p}(\theta_0; b)\}_{t=1}^N$ or $\{D_{tn,1:k}(\theta_0; b)\}_{t=1}^N$) is of dimension $k-p$ (p or k) for large n. Set $\tilde{Z}_n = bPVZ_n$ and notice that

$$\begin{aligned} elr(\hat{\theta}_{el}) &\le \frac{2}{nb} \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^{N} \log(1 + \lambda' D_{tn}(\theta_0; b) + \lambda'_{1:p}\tilde{Z}_n \\ &\quad + \lambda' bU'\Lambda^{-1}\{g_{tn}(\tilde{\theta}) - G_0\}Z_n). \end{aligned}$$

By Lemma A.2, we can choose $\tilde{Z}_n$ such that the origin of $\mathbb{R}^k$ is contained in the interior of the convex hull of $\{D_{jn}(\theta_0; b) + \tilde{Z}_{n,ex}\}_{j=1}^N$ where $\tilde{Z}_{n,ex} = (\tilde{Z}_n', 0_{k-p}')'$. Also note that for large n, we have $\sup_{1 \le t \le N} |bU'\Lambda^{-1}\{g_{tn}(\tilde{\theta}) - G_0\}| \le \epsilon$. By choosing ϵ sufficiently small, the interior of the convex hull of $\{D_{jn}(\theta_0; b) + \tilde{Z}_{n,ex}\}_{j=1}^N$ can still contain the origin after a small perturbation caused by adding $bU'\Lambda^{-1}\{g_{tn}(\tilde{\theta}) - G_0\}$. Therefore $elr(\hat{\theta}_{el}) < \infty$. On the other hand, define $V_{tn}(\theta) = f_{tn}(\theta) - h(\theta)$. If there exists a $\delta > 0$ such that $|h(\hat{\theta}_{el})| > \sqrt{k}\max_{1 \le t \le n} \sup_{\theta \in \Theta} |V_{tn}(\theta)| + \delta$, then by choosing $\lambda = \tilde{a}\,\mathrm{sign}(h(\hat{\theta}_{el}))$ where $\mathrm{sign}(h(\hat{\theta}_{el}))$ is a vector of the signs of $h(\hat{\theta}_{el})$, we get

$$\begin{aligned} elr(\hat{\theta}_{el}) &\ge \frac{2}{nb} \sum_{t=1}^{N} \log(1 + \tilde{a}|h(\hat{\theta}_{el})| + \tilde{a}\,\mathrm{sign}(h(\hat{\theta}_{el}))' V_{tn}(\hat{\theta}_{el})) \\ &\ge \frac{2N}{bn} \log(1 + \tilde{a}|h(\hat{\theta}_{el})| - \tilde{a}\sqrt{k} \max_{1 \le t \le n} \sup_{\theta \in \Theta} |V_{tn}(\theta)|) \\ &\ge \frac{2N}{bn} \log(1 + \tilde{a}\delta), \end{aligned}$$

for any $\tilde{a} > 0$. By letting $\tilde{a} \to +\infty$, we get $elr(\hat{\theta}_{el}) = \infty$ which is a contradiction. Thus we should have $|h(\hat{\theta}_{el})| \le \sqrt{k}\max_{1 \le t \le N} \sup_{\theta \in \Theta} |V_{tn}(\theta)| + \delta$ for any $\delta > 0$. Because $\max_{1 \le t \le n} \sup_{\theta \in \Theta} |V_{tn}(\theta)| = o_p(1)$ and δ is arbitrary, we obtain $|h(\hat{\theta}_{el})| = o_p(1)$. By Assumption 2.2, we have $\hat{\theta}_{el} \to^p \theta_0$. $\diamond$

Lemma A.2. *Let $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{i(a+b)})' \in \mathbb{R}^{a+b}$ for $a, b \in \mathbb{N}$ and $i = 1, 2, \ldots, N$. Suppose that the convex hull of $\{\alpha_i\}_{i=1}^N$ is $a+b$ dimensional and the origin of $\mathbb{R}^b$ is an interior point of the convex* *hull of* $\{\alpha_{i,a+1:a+b}\}_{i=1}^N$. *Then there exists* $z = (z'_{1:a}, 0'_b)'$ *such that the origin of* $\mathbb{R}^{a+b}$ *is an interior point of the convex hull of* $\{\alpha_i + z\}_{i=1}^N$.

Proof. We prove the result by induction on a. Consider the case of $a = 1$. By the assumption, there exist a b dimensional simplex $\mathcal{S}$ which belongs to the convex hull of $\{\alpha_i\}_{i=1}^N$ and a point $(\beta, 0'_b)'$ such that $(\beta, 0'_b)'$ is in the relative interior of $\mathcal{S}$ (see Section 6 of Rockafellar (1970)). Since the convex hull of $\{\alpha_i\}_{i=1}^N$ is $b+1$ dimensional, we can find a point $\gamma \in \mathbb{R}^{b+1}$ which does not belong to the hyperplane containing $\mathcal{S}$. Thus by the convexity, there must exist a point $(\beta_1, 0'_b)' \in \{v\gamma + (1-v)\mathcal{S} : 0 < v < 1\}$ which is an interior point of the convex hull of $\{\alpha_i\}_{i=1}^N$. The conclusion thus follows by choosing $z = (-\beta_1, 0'_b)'$. Suppose the conclusion holds for $a \le k_0$. When $a = k_0 + 1$, we know that there exists $z = (z'_{k_0}, 0'_b)'$ such that the origin of $\mathbb{R}^{k_0+b}$ is an interior point of the convex hull of $\{\alpha_{i,2:a+b} + z\}_{i=1}^N$. Note that the convex hull of $\{\alpha_i + (0, z')'\}_{i=1}^N$ is still $a+b$ dimensional. By the above arguments, there exists a $\tilde{z} \in \mathbb{R}$ such that the origin of $\mathbb{R}^{k_0+1+b}$ is an interior point of the convex hull of $\{\alpha_i + (\tilde{z}, z')'\}_{i=1}^N$. Thus the conclusion holds for $a = k_0 + 1$. ◇

Proof of Lemma 4.2. Let ϵ_n be a sequence of positive numbers such that $\epsilon_n = o_p(1)$, $1/(\sqrt{n}\epsilon_n) = o_p(1)$ and $\hat{\theta}_{el} \in \mathcal{N}_{\epsilon_n}(\theta_0)$. Consider a shrinking neighborhood $\mathcal{N}_{\kappa/\sqrt{n}}(\theta_0)$. For large enough n, we have $\mathcal{N}_{\epsilon_n}(\theta_0) \setminus \mathcal{N}_{\kappa/\sqrt{n}}(\theta_0) \ne \emptyset$. For $\hat{\theta}_n \in \mathcal{N}_{\epsilon_n}(\theta_0) \setminus \mathcal{N}_{\kappa/\sqrt{n}}(\theta_0)$, we have

$$\begin{aligned}
&elr(\hat{\theta}_n)\\
&= \frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N \log(1 + \lambda'\sqrt{n}f_{tn}(\theta_0) + \lambda'\sqrt{n}g_{tn}(\tilde{\theta})(\hat{\theta}_n - \theta_0))\\
&\ge \frac{2}{nb}\sum_{t=1}^N \log(1 + \sqrt{n}(\hat{\theta}_n - \theta_0)'G_0'b^{-1}\Lambda D_k(t/n; b)\\
&\quad \times (1 + o_p(1)) + n(\hat{\theta}_n - \theta_0)'G_0'G_0(\hat{\theta}_n - \theta_0)(1 + o_p(1)) + r_{tn}),
\end{aligned}$$

where $\tilde{\theta}$ on the line segment joining $\hat{\theta}_n$ and θ_0, and it can vary across equations, and $r_{tn} = o_p(1)$ uniformly for t, Note that $n(\hat{\theta}_n - \theta_0)'G_0'G_0(\hat{\theta}_n - \theta_0) \ge C\kappa^2$ since $\text{rank}(G_0) = p$, and $\sqrt{n}(\hat{\theta}_n - \theta_0)'G_0'b^{-1}\Lambda D_k(t/n; b) = O_p(\kappa)$. For large enough κ, we have $elr(\hat{\theta}_n) > \min_{\theta\in\Theta} elr(\theta)$. Conditional on $\mathcal{D}_b$, $\min_{\theta\in\Theta} elr(\theta) = O_p(1)$. Therefore, we get $\hat{\theta}_{el} \in \mathcal{N}_{\kappa/\sqrt{n}}(\theta_0)$ which implies the uniformly tightness of $\mathcal{Y}_n$. ◇

Proof of Theorem 4.1. By the mean value theorem, we have $f_{tn}(\hat{\theta}_{el}) = f_{tn}(\theta_0) + g_{tn}(\tilde{\theta})(\hat{\theta}_{el} - \theta_0)$, where $\tilde{\theta}$ is between $\hat{\theta}_{el}$ and θ_0. It thus implies that

$$elr(\hat{\theta}_{el}) = \frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N \log(1 + \lambda'\sqrt{n}f_{tn}(\theta_0) + \lambda'\sqrt{n}g_{tn}(\tilde{\theta})(\hat{\theta}_{el} - \theta_0)).$$

Then we have $\mathcal{Y}_n = \text{argmin}_{Y\in\mathbb{R}^p} Q_{el,n}(Y)$, where

$$\begin{aligned}
Q_{el,n}(Y) &= \frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N \log(1 + \lambda'\sqrt{n}f_{tn}(\theta_0) + \lambda' g_{tn}(\tilde{\theta})Y)\\
&= \frac{2}{nb}\max_{\lambda\in\mathbb{R}^k}\sum_{t=1}^N \log(1 + \lambda' D_{tn}(\theta_0; b) + \lambda'\xi_{tn}),
\end{aligned}$$

where $\tilde{\theta} = \tilde{\theta}(Y/\sqrt{n} + \theta_0, \theta_0)$. For any $\omega \in \mathcal{C} \cap \mathcal{D}_b$ with $P(\mathcal{C}) = 1$, we define a sequence of sets $\mathcal{C}_{\omega,n} = \{Y \in \mathbb{R}^p :$ the origin of $\mathbb{R}^k$ is contained in the interior of the convex hull of $\{D_{tn}(\theta_0; b) + \xi_{tn}\}_{t=1}^N\}$. When $Y \notin C_{\omega,n}$, we have $Q_{cue,n}(Y) = \infty$. Thus we must have $\mathcal{Y}_n \in \mathcal{C}_{\omega,n}$. Based on the assumption, we also know that $\mathcal{C}_{\omega,n}$ is nonempty for large enough n and $\omega \in \mathcal{D}_b$ (see the proof of Lemma A.2). Let

$$Q_{el}(Y) = \frac{2}{b}\max_{\lambda\in\mathbb{R}^k}\int_0^{1-b} \log(1 + \lambda' D_k(r; b) + \lambda' Y_{ex})dr,$$

where $Y_{ex} = (b(PVY)', 0'_{k-p})'$, and define $\mathcal{C}_\omega$ the set of Y such that the origin of $\mathbb{R}^k$ is contained in the interior of the convex hull of $\{D_k(r; b) + Y_{ex}\}_{r\in[0,1-b]}$. Then by strong approximation, we have $\mu(\mathcal{C}_{\omega,n}\boldsymbol{\Delta}\mathcal{C}_\omega) \to 0$, where $\boldsymbol{\Delta}$ denotes the symmetric difference between two sets and μ denotes the Lebesgue measure. Note that the minimizer of Q_{el} is contained in $\mathcal{C}_\omega$, which is nonempty under the assumptions. For any compact set $K \subset \mathcal{C}_\omega$, because $\mathcal{C}_\omega$ and $\mathcal{C}_{\omega,n}$ are open, we have $K \subset \mathcal{C}_{\omega,n}$ for large enough n. We show in Lemma A.4 that,

$$|Q_{el,n}(Y) - Q_{el}(Y)| \to 0, \tag{44}$$

uniformly for $Y \in K$. By Lemma A.3, Q_{el} is continuous on $\mathcal{C}_\omega$. Then following the proofs in Theorem 3.1, we have $Q_{el,n}$ is asymptotically tight in $l^\infty(K)$ for any compact set $K \subset \mathcal{C}_\omega$. Using the FOC and the arguments in the proof of Lemma 3.3, it is not hard to verify that Q_{el} has a unique minimizer conditional on $\mathcal{D}_b$. By Lemma 4.2, we have $\{\mathcal{Y}_n\}$ is uniformly tight. Also note that when Y is close to the boundary of $\mathcal{C}_{\omega,n}$, $Q_{el,n}(Y)$ diverges to infinity. Thus we may restrict $\mathcal{Y}_n$ to a compact set contained in $\mathcal{C}_\omega$ for large enough n. We can then apply the Argmax/Argmin Theorem. Therefore we get $\mathcal{Y}_n \to^d \tilde{\mathcal{Y}}$, which completes the proof. ◇

Define $q_{el}(\lambda, Y) = 2\int_0^{1-b}\log(1 + \lambda' D_k(r; b) + \lambda' Y_{ex})dr/b$, and $G_k(r; b; Y) = D_k(r; b) + Y_{ex}$. Let $\lambda(Y) = \text{argmax}_{\lambda\in\mathbb{R}^k} q_{el}(\lambda, Y)$. Then we have $Q_{el}(Y) = q_{el}(\lambda(Y), Y)$. The constants C, C' and C'' below are understood to be independent of n. The proofs of the following two Lemmas are provided in the supplement (see Appendix B).

Lemma A.3. *For any* $\omega \in \mathcal{D}_b$, *the path of* $Q_{el}(Y)$ *and the maximizer* $\lambda(Y)$ *are continuous on* $\mathcal{C}_\omega$.

Lemma A.4. *Under the Assumptions in Theorem 4.1, we have* $|Q_{el,n}(Y) - Q_{el}(Y)| \to 0$ *uniformly for* $Y \in K \subset \mathcal{C}_\omega$, *where* K *is a compact set and* $\omega \in \mathcal{D}_b \cap \mathcal{C}$.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2016.01.009.

References

Altonji, J.G., Segal, L.M., 1996. Small-sample bias in GMM estimation of covariance structures. J. Bus. Econom. Statist. 14, 353–366.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Baum, C., Schaffer, M., Stillman, S., 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. Stata J. 7, 465–506.
Bester, C.A., Conley, T.G., Hansen, C.B., 2011. Inference with dependent data using cluster covariance estimators. J. Econometrics 165, 137–151.
Billingsley, P., 1999. Convergence of Probability Measures, second ed. Wiley, New York.
Borwein, J.M., Lewis, A.S., 1991. Duality relationships for entropy-type minimization problems. SIAM J. Control Optim. 29, 325–338.
DiCiccio, T., Hall, P., Romano, J., 1991. Empirical likelihood is Bartlett-correctable. Ann. Statist. 19, 1053–1061.
Gonçaves, S., 2011. The moving blocks bootstrap for panel linear regression models with individual fixed effects. Econometric Theory 27, 1048–1082.
Gonçalves, S., Vogelsang, T.J., 2011. Block bootstrap HAC robust tests: The sophistication of the naive bootstrap. Econometric Theory 27, 745–791.
Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. J. Bus. Econom. Statist. 14, 262–280.

Herrndorf, N., 1984. A functional central limit theorem for weakly dependent sequences of random variables. Ann. Probab. 12, 141–153.
Hwang, J., Sun, Y., 2015. Asymptotic F and t Tests in An Effcient GMM Setting. Working paper. Department of Economics, University of California, San Diego.
Ibragimov, R., Müller, U.K., 2010. t-statistic based correlation and heterogeneity robust inference. J. Bus. Econom. Statist. 28, 453–468.
Imbens, G.W., Spady, R.H., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.
Jansson, M., 2004. On the error of rejection probability in simple autocorrelation robust tests. Econometrica 72, 937–946.
Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21, 1130–1164.
Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. Ann. Statist. 25, 2084–2102.
Kitamura, Y., 2001. Asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica 69, 1661–1672.
Kitamura, Y., Santos, A., Shaikh, A., 2013. On the asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica 80, 413–423.
Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.
McElroy, T., Politis, D.N., 2007. Computer-intensive rate estimation, diverging statistics, and scanning. Ann. Statist. 35, 1827–1848.
Newey, W.K., Smith, R., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72, 219–255.
Nordman, D.J., 2009. Tapered empirical likelihood for time series data in time and frequency domains. Biometrika 96, 119–132.
Nordman, D.J., Bunzel, H., Lahiri, S.N., 2013. A non-standard empirical likelihood for time series. Ann. Statist. 41, 3050–3073.
Nordman, D.J., Lahiri, S.N., 2014. A review of empirical likelihood methods for time series. J. Statist. Plann. Inference 155, 1–18.
Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.
Owen, A.B., 1990. Empirical likelihood confidence regions. Ann. Statist. 18, 90–120.
Phillips, P.C.B., Durlauf, S.N., 1986. Multiple time series regression with integrated processes. Rev. Econ. Stud. 53, 473–495.
Politis, D.N., White, H., 2004. Automatic block-length selection for dependent bootstrap. Econometric Rev. 23, 53–70.
Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Ann. Statist. 22, 300–325.
Rockafellar, T.R., 1970. Convex Analysis. Princeton Univ. Press.
Shao, X., 2010. A self-normalized approach to confidence interval construction in time series. J. R. Stat. Soc. Ser. B Stat. Methodol. 72, 343–366.
Smith, R., 2011. GEL criteria for moment condition models. Econometric Theory 27, 1192–1235.
Stock, J., Watson, M., 2003. Introduction to Econometrics. Addison-Wesley, Reading, MA.
Sun, Y., 2011. Robust trend inference with series variance estimator and testing-optimal smoothing parameter. J. Econometrics 164, 345–366.
Sun, Y., 2013. A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. Econom. J. 16, 1–26.
Sun, Y., 2014a. Fixed-smoothing asymptotics in a two-step GMM framework. Econometrica 82, 2327–2370.
Sun, Y., 2014b. Let's fix it: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. J. Econometrics 178, 659–677.
Sun, Y., Kim, M.S., 2012. Simple and powerful GMM over-identification tests with accurate size. J. Econometrics 166, 267–281.
Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroscedasicity-autocorrlation robust testing. Econometrica 76, 175–194.
Tsao, M., Wu, F., 2013. Empirical likelihood on the full parameter space. Ann. Statist. 41, 2176–2196.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer Verlag, New York.
Vogelsang, T., 2012. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. J. Econometrics 166, 303–319.
Wu, W., 2007. Strong invariance principles for dependent random variables. Ann. Probab. 35, 2294–2320.
Zhang, X., Shao, X., 2013. Fixed-smoothing asymptotics for time series. Ann. Statist. 41, 1329–1349.
Zhang, X., Shao, X., 2014. Fixed-b asymptotics for blockwise empirical likelihood. Statist. Sinica 24, 1179–1194.
Zhang, X., Shao, X., 2016. On the coverage bound problem of empirical likelihood methods for time series. J. R. Stat. Soc. Ser. B 78, 395–421.

Supplement to "Fixed-smoothing Asymptotics in the Generalized Empirical Likelihood Estimation Framework"

Xianyang Zhang*

Texas A&M University

May 9, 2016

1 Additional technical details

Lemma **S.1**. *Assume that (i) Θ is compact and $f(y,\theta)$ is continuous in θ for all y; (ii) Suppose $|f(y,\theta)| \leq L(y)$, where $\mathbb{E}L(y) < \infty$; (iii) Define $\varphi(y,\theta,\rho) = \sup_{|\theta'-\theta|<\rho} S(y,\theta')$, where $S(y,\theta) = f(y,\theta) - h(\theta)$. For any $\theta \in \Theta$, there exists a $A_\theta > 0$, such that for all $0 \leq \rho < A_\theta$, $\varphi(y,\theta,\rho)$ is measurable in y and*

$$\max_{1\leq t\leq n} \left| \frac{1}{n} \sum_{j=1}^{t} \{\varphi(y_j,\theta,\rho) - \mathbb{E}\varphi(y_j,\theta,\rho)\} \right| = o_p(1).$$

Then Assumption 2.1 holds under the above conditions.

Proof of Lemma S.1. Since $f(y,\theta)$ is continuous in θ for all y, we know that $h(\theta)$ is continuous by the Dominated Convergence Theorem (DCT). Thus $S(y,\theta)$ is continuous in θ for all y, which implies that $\lim_{\rho\downarrow 0} \varphi(y,\theta,\rho) = S(y,\theta)$. By the DCT, we have $\lim_{\rho\downarrow 0} \mathbb{E}\varphi(y_j,\theta,\rho) = \mathbb{E}S(y_j,\theta) = 0$. Let $\epsilon > 0$. For each θ, we can find ρ_θ such that $\mathbb{E}\varphi(y_j,\theta,\rho_\theta) < \epsilon$. The spheres $Sp(\theta,\rho_\theta) = \{\theta' : |\theta'-\theta| < \rho_\theta\}$ cover Θ. Because Θ is compact, there exists a finite subcover say $\cup_{i=1}^m Sp(\theta_i,\rho_{\theta_i}) \supset \Theta$. Then we have

$$\begin{aligned}
\sup_{\theta} \sup_{1\leq t\leq n} \frac{1}{n} \sum_{j=1}^{t} S(y_j,\theta) &\leq \sup_{1\leq i\leq m} \sup_{1\leq t\leq n} \frac{1}{n} \sum_{j=1}^{t} \varphi(y_j,\theta_i,\rho_{\theta_i}) \\
&\leq \sup_{1\leq i\leq m} \sup_{1\leq t\leq n} \left| \frac{1}{n} \sum_{j=1}^{t} \{\varphi(y_j,\theta_i,\rho_{\theta_i}) - \mathbb{E}\varphi(y_j,\theta_i,\rho_{\theta_i})\} \right| + \epsilon.
\end{aligned}$$

*Department of Statistics, Texas A&M University, College Station, TX 77843. E-mail: zhangxiany@stat.tamu.edu. The author is grateful to Xiaofeng Shao for many helpful comments and suggestions on an earlier version. The author would like to thank the Editor, the Associate Editor and three reviewers for their insightful comments and constructive suggestions, which substantially improved the paper. All errors are my own.

By condition (iii),

$$P\left(\sup_{1\le t\le n}\left|\frac{1}{n}\sum_{j=1}^{t}\{\varphi(y_j,\theta_i,\rho_{\theta_i})-\mathbb{E}\varphi(y_j,\theta_i,\rho_{\theta_i})\}\right|<\epsilon,\text{for } i=1,2,\dots,m\right)\to 1.$$

Hence we have for any given $\epsilon>0$,

$$P\left(\sup_{\theta}\sup_{1\le t\le n}\frac{1}{n}\sum_{j=1}^{t}S(y_j,\theta)\le 2\epsilon\right)\to 1.$$

Using the same arguments, we can prove the result with S replaced by $-S$. Therefore, we get

$$P\left(\sup_{\theta}\sup_{1\le t\le n}\left|\frac{1}{n}\sum_{j=1}^{t}S(y_j,\theta)\right|\le 2\epsilon\right)\to 1,$$

which thus completes the proof. $\diamondsuit$

REMARK **S.1**. *We remark that condition (iii) in Lemma S.1 can be verified by assuming that for any $\theta\in\Theta$ and $0\le\rho<A_\theta$, the sequence $\{\varphi(y_j,\theta,\rho)\}$ is strongly mixing and satisfies suitable tail conditions. See e.g. Theorem 6.2 of Rio (2013).*

Proof of Lemma 9.3. Define the set $V_Y=\{\lambda\in\mathbb{R}^k:\min_{r\in[0,1-b]}(1+\lambda' G_k(r;b;Y))\ge 0\}$. For $Y\in\mathcal{C}_\omega$ with $\omega\in\mathcal{D}_b$, we know V_Y is a compact set [Lemma 3.1 of Zhang and Shao (2015)]. Below we show that Q_{el} is continuous at $Y_1\in\mathcal{C}_\omega$. Given any $\epsilon>0$, we have for any $|\lambda-\lambda(Y_1)|\ge\epsilon/2$, $q_{el}(\lambda(Y_1),Y_1)>q_{el}(\lambda,Y_1)+\delta$ for some $\delta>0$. Here we have used the strictly concavity of q_{el} on λ. For any $\epsilon_0>0$, define $\lambda_{\epsilon_0}=(1-\epsilon_0)\lambda$. Our goal below is to show that $q_{el}(\lambda,Y_2)<q_{el}(\lambda_{\epsilon_0}(Y_1),Y_2)$ for $|\lambda_{\epsilon_0}(Y_1)-\lambda(Y_1)|<\epsilon/2$ and any $|\lambda-\lambda(Y_1)|\ge\epsilon$. Thus we must have $|\lambda(Y_2)-\lambda(Y_1)|\le\epsilon$. Without loss of generality, we assume $\lambda\in V_{Y_2}$. Suppose $|Y_2-Y_1|<\varepsilon$. We pick a small ε so that $\min_{r\in[0,1-b]}(1+\lambda_{\epsilon_0}G(r;b;Y_1))\ge\min_{r\in[0,1-b]}\{1-\epsilon_0+\lambda_{\epsilon_0}G(r;b;Y_2)+\lambda_{\epsilon_0}(G(r;b;Y_1)-G(r;b;Y_2))+\epsilon_0\}>\epsilon_0/2$. By Taylor expansion, we deduce that

$$\begin{aligned}|q_{el}(\lambda_{\epsilon_0},Y_2)-q_{el}(\lambda_{\epsilon_0},Y_1)|\le&\frac{2}{b}\int_0^{1-b}\left(\frac{1}{1+\lambda'_{\epsilon_0}G(r;b;Y_1)}+\frac{1}{1+\lambda'_{\epsilon_0}G(r;b;Y_2)}\right)|\lambda'_{\epsilon_0}(Y_{ex,1}-Y_{ex,2})|dr\\ \le&C\epsilon_0^{-1}\varepsilon.\end{aligned}$$

Along with the concavity and the fact that $q_{el}(0,Y)=0$, we obtain that

$$(1-\epsilon_0)q_{el}(\lambda,Y_2)\le q_{el}(\lambda_{\epsilon_0},Y_2)\le q_{el}(\lambda_{\epsilon_0},Y_1)+C\varepsilon\epsilon_0^{-1}.$$

We can choose ϵ_0 small so that $|\lambda_{\epsilon_0}-\lambda(Y_1)|\ge\epsilon/2$. Hence we obtain

$$q_{el}(\lambda,Y_2)\le(q_{el}(\lambda(Y_1),Y_1)-\delta+C\varepsilon\epsilon_0^{-1})/(1-\epsilon_0).$$

Note that for small ε, we have $\min_{r\in[0,1-b]}(1+\lambda_{\epsilon_0}(Y_1)G(r;b;Y_2)) \geq \epsilon_0/2$. Using Taylor expansion again, we get

$$(1-\epsilon_0)q_{el}(\lambda(Y_1),Y_1) \leq q_{el}(\lambda_{\epsilon_0}(Y_1),Y_1) \leq q_{el}(\lambda_{\epsilon_0}(Y_1),Y_2) + C'\varepsilon\epsilon_0^{-1}, \tag{S.1}$$

which implies that

$$q_{el}(\lambda,Y_2) \leq (q_{el}(\lambda_{\epsilon_0}(Y_1),Y_2) + C'\varepsilon\epsilon_0^{-1} - (1-\epsilon_0)\delta + C(1-\epsilon_0)\varepsilon\epsilon_0^{-1})/(1-\epsilon_0)^2.$$

By picking a small enough ε and ϵ_0, we have $|\lambda_{\epsilon_0}(Y_1) - \lambda(Y_1)| < \epsilon/2$,

$$(2\epsilon_0 - \epsilon_0^2)q_{el}(\lambda(Y_2),Y_2) + C'\varepsilon\epsilon_0^{-1} + C(1-\epsilon_0)\varepsilon\epsilon_0^{-1} < (1-\epsilon_0)\delta,$$

which implies that $q_{el}(\lambda,Y_2) < q_{el}(\lambda_{\epsilon_0}(Y_1),Y_2)$. Thus we get $|\lambda(Y_1)-\lambda(Y_2)| \leq \epsilon$. From (S.1), we have

$$\begin{aligned} Q_{el}(Y_1) = q_{el}(\lambda(Y_1),Y_1) &\leq q_{el}(\lambda(Y_2),Y_2) + C'\varepsilon\epsilon_0^{-1} + \epsilon_0 q_{el}(\lambda(Y_1),Y_1) \\ &= Q_{el}(Y_2) + C'\varepsilon\epsilon_0^{-1} + \epsilon_0 q_{el}(\lambda(Y_1),Y_1). \end{aligned}$$

Switching the role of Y_1 and Y_2, we get

$$(1-\epsilon_0)q_{el}(\lambda(Y_2),Y_2) \leq q_{el}(\lambda_{\epsilon_0}(Y_2),Y_2) \leq q_{el}(\lambda(Y_1),Y_1) + C''\varepsilon\epsilon_0^{-1},$$

which implies that

$$Q_{el}(Y_2) \leq (Q_{el}(Y_1) + C''\varepsilon\epsilon_0^{-1})/(1-\epsilon_0).$$

Choosing sufficiently small ϵ_0 and ε, we can make $|Q_{el}(Y_1) - Q_{el}(Y_2)|$ small. The proof is thus completed. $\diamondsuit$

Proof of Lemma 9.4. Recall the definition of ξ_{tn} in the proof of Theorem 3.1. Let $q_{el,n}(\lambda;Y) = 2\sum_{i=1}^N \log(1+\lambda' G_{tn}(b;Y))/(nb)$ with $G_{tn}(b;Y) = D_{tn}(\theta_0;b) + \xi_{tn}$ and $\bar{q}_{el,n}(\lambda;Y) = 2\sum_{i=1}^N \log(1+\lambda' G_k(t/n;b;Y))/(nb)$ with $G_k(t/n;b;Y) = D_k(t/n;b) + Y_{ex}$. Define $\bar{Q}_{el,n}(Y) = \max_{\lambda\in\mathbb{R}^k} \bar{q}_{el,n}(\lambda;Y)$, and the maximizer of $\max_{\lambda\in\mathbb{R}^k} \bar{q}_{el,n}(\lambda;Y)$ is denoted by $\bar{\lambda}_n(Y)$. We claim that $\bar{\lambda}_n(Y)$ and $\lambda_n(Y)$ are uniformly bounded on K for large enough n. (For example in the univariate case that is $k=1$, we have $\min_{1\leq t\leq n}(1+\lambda_n(Y)'G_{tn}(b;Y)) \geq 0$ for large enough n, which implies that $-1/[\min_Y \max_t\{G_{tn}(b;Y)\}] \leq \lambda_n(Y) \leq -1/[\max_Y \min_t\{G_{tn}(b;Y)\}]$. Thus $\lambda_n(Y)$ is uniformly bounded.) Otherwise, there exists a sequence $\{Y_j\} \subset K$ such that there exists a $\lambda_j \in V_{n,Y_j} := \{\lambda \in \mathbb{R}^k : \min_{1\leq t\leq n}(1+\lambda' G_{tn}(b;Y_j)) \geq 0\}$ and the sequence $\{\lambda_j\}$ is unbounded. Since K is compact, we can assume that $Y_j \to Y_0$ without loss of generality. Thus for large enough j, $\lambda_j \in \{\lambda \in \mathbb{R}^k : \tilde{d}(\lambda, V_{n,Y_0}) \leq \epsilon\}$, where V_{n,Y_0} is bounded and $\tilde{d}(\cdot,\cdot)$ denotes the distance between two sets. This contradicts with the assumption that the sequence $\{\lambda_j\}$ is unbounded.

Because $\sup_{Y\in K}|\tilde{\theta}(Y/\sqrt{n}+\theta_0,\theta_0)-\theta_0|\leq \sup_{Y\in K}|Y|/\sqrt{n}\to 0$, $\sup_{\theta\in\mathcal{N}_{\epsilon^*}(\theta_0)}\sup_{1\leq t\leq N}||g_{tn}(\theta)-G(\theta)||\to^{a.s.} 0$ and $G(\theta)$ is continuous on $\mathcal{N}_{\epsilon^*}(\theta_0)$, we have $\sup_{1\leq t\leq n}|G_k(t/n;b;Y)-G_{tn}(b;Y)|\to^{a.s.} 0$ uniformly for $Y\in K$. Below we consider pointwise convergence for each $\omega\in\mathcal{C}\cap\mathcal{D}_b$ with $P(\mathcal{C})=1$. Let $\lambda_{n,\epsilon_0}=(1-\epsilon_0)\lambda_n$ for $0<\epsilon_0<1$. For large enough n, we have $\min_{r\in[0,1]}\{1+\lambda_{n,\epsilon_0}(Y)'G_k(t/n;b;Y)\}\geq \min_{r\in[0,1]}\{1+\lambda_{n,\epsilon_0}(Y)'G_{tn}(b;Y)+\lambda_{n,\epsilon_0}(Y)'(G_k(t/n;b;Y)-G_{tn}(b;Y))\}\geq\epsilon_0/2$. Then using the Taylor expansion and the concavity, we get

$$
\begin{aligned}
&|q_{el,n}(\lambda_{n,\epsilon_0}(Y),Y)-\bar{q}_{el,n}(\lambda_{n,\epsilon_0}(Y),Y)|\\
\leq&\frac{1}{n}\sum_{j=1}^{N}\left(\frac{1}{1+\lambda_{n,\epsilon_0}'G_k(t/n;b;Y)}+\frac{1}{1+\lambda_{n,\epsilon_0}(Y)'G_{tn}(b;Y)}\right)|\lambda_{n,\epsilon_0}(Y)'(G_k(t/n;b;Y)-G_{tn}(b;Y))|\\
\leq&C\epsilon_0^{-1}\varepsilon,
\end{aligned}
$$

which implies that

$$
(1-\epsilon_0)q_{el,n}(\lambda_n(Y);Y)\leq q_{el,n}(\lambda_{n,\epsilon_0}(Y),Y)\leq\bar{q}_{el,n}(\lambda_{n,\epsilon_0}(Y),Y)+C\epsilon_0^{-1}\varepsilon\leq\bar{Q}_{el,n}(Y)+C\epsilon_0^{-1}\varepsilon,
$$

where $\sup_{Y\in K}\max_{1\leq t\leq n}|G_k(t/n;b;Y)-G_{tn}(b;Y)|<\varepsilon$. It follows that $Q_{el,n}(Y)\leq(\bar{Q}_{el,n}(Y)+C\epsilon_0^{-1}\varepsilon)/(1-\epsilon_0)$. Similarly, we have

$$
(1-\epsilon_0)\bar{Q}_{el,n}(Y)\leq\bar{q}_{el,n}(\bar{\lambda}_{n,\epsilon_0}(Y),Y)\leq q_{el,n}(\bar{\lambda}_{n,\epsilon_0}(Y),Y)+C'\epsilon_0^{-1}\varepsilon\leq Q_{el,n}(Y)+C'\epsilon_0^{-1}\varepsilon.
$$

By picking small enough ϵ_0 and large enough n (and thus small enough ε), we have $|Q_{el,n}(Y)-\bar{Q}_{el,n}(Y)|$ converges to zero uniformly over K. Notice that

$$
q_{el}(\lambda;Y)=\sum_{i=1}^{N}\int_{(i-1)n}^{i/n}\log(1+\lambda'G_k(r;b;Y))dr+\int_{N/n}^{1-b}\log(1+\lambda'G_k(r;b;Y))dr.
$$

Note $Q_{el}(Y)$ is uniformly bounded over K. Following similar arguments above, we can show that $|\bar{Q}_{el,n}(Y)-Q_{el}(Y)|\to 0$ uniformly over K, which completes the proof. $\diamondsuit$

Proof of Lemma 5.1. Recall the singular value decomposition $G_\Lambda=\Lambda^{-1}G_0=U\Sigma V$ where $\Sigma'=(P,O)$ with P being a $p\times p$ diagonal matrix, and let

$$
\tilde{Q}_0^{-1}=\begin{pmatrix}\tilde{Q}^{11} & \tilde{Q}^{12}\\ \tilde{Q}^{21} & \tilde{Q}^{22}\end{pmatrix},
$$

where $\tilde{Q}^{11}$ is a $p\times p$ matrix. We have

$$
G_0'Q_0^{-1}G_0=G_\Lambda'\tilde{Q}_0^{-1}G_\Lambda=^d V'\Sigma'\tilde{Q}_0^{-1}\Sigma V=V'P\tilde{Q}^{11}PV,
$$

and

$$G_0' Q_0^{-1} \Lambda W_k(1) = G_\Lambda' \tilde{Q}_0^{-1} W_k(1) =^d V' \Sigma' \tilde{Q}_0^{-1} W_k(1) = V'(P\tilde{Q}^{11}, P\tilde{Q}^{12}) W_k(1)$$

where we have used the fact that $(U'\tilde{Q}_0^{-1}U, U'W_k(r)) =^d (\tilde{Q}_0^{-1}, W_k(r))$. Note that $\tilde{Q}^{12} = -\tilde{Q}^{11}\tilde{Q}_{12}\tilde{Q}_{22}^{-1}$. Therefore, we deduce that

$$\begin{aligned}\mathcal{S} =& -(G_0' Q_0^{-1} G_0)^{-1} G_0' Q_0^{-1} \Lambda W_k(1) = -V^{-1} P^{-1} (\tilde{Q}^{11})^{-1} (\tilde{Q}^{11}, \tilde{Q}^{12}) W_k(1) \\ =& -V^{-1} P^{-1} (I_p, -\tilde{Q}_{12}\tilde{Q}_{22}^{-1}) W_k(1).\end{aligned}$$

$\diamondsuit$

Proof of Lemma 5.2. The proof is very similar to that of Theorem 1 in Sun (2014a). For completeness, we present the details below. Note that

$$\begin{aligned}\mathcal{W}_\infty(b; \delta, G_0, \Lambda) = & \left\{\mathcal{R}(G_\Lambda' \tilde{Q}_0^{-1} G_\Lambda)^{-1} G_\Lambda' \tilde{Q}_0^{-1} W_k(1) + \delta\right\}' \left\{\mathcal{R}(G_\Lambda' \tilde{Q}_0^{-1} G_\Lambda)^{-1} \mathcal{R}'\right\}^{-1} \\ & \left\{\mathcal{R}(G_\Lambda' \tilde{Q}_0^{-1} G_\Lambda)^{-1} G_\Lambda' \tilde{Q}_0^{-1} W_k(1) + \delta\right\},\end{aligned}$$

with $G_\Lambda = \Lambda^{-1} G_0$. By the singular value decomposition $G_\Lambda = U\Sigma V$, we have

$$\begin{aligned}\mathcal{R}(G_\Lambda' \tilde{Q}_0^{-1} G_\Lambda)^{-1} G_\Lambda' \tilde{Q}_0^{-1} W_k(1) =^d & \mathcal{R}(V'\Sigma' \tilde{Q}_0^{-1} \Sigma V)^{-1} V' \Sigma' \tilde{Q}_0^{-1} W_k(1) \\ = & \mathcal{R} V' (\Sigma' \tilde{Q}_0^{-1} \Sigma)^{-1} \Sigma' \tilde{Q}_0^{-1} W_k(1),\end{aligned}$$

and

$$\mathcal{R}(G_\Lambda' \tilde{Q}_0^{-1} G_\Lambda)^{-1} \mathcal{R}' =^d \mathcal{R} V' (\Sigma' \tilde{Q}_0^{-1} \Sigma)^{-1} V \mathcal{R}',$$

where we have used the fact that $(U'\tilde{Q}_0^{-1}U, U'W_k(1)) =^d (\tilde{Q}_0^{-1}, W_k(1))$. Recall that

$$\tilde{Q}_0 = \begin{pmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{pmatrix}, \quad \tilde{Q}_0^{-1} = \begin{pmatrix} \tilde{Q}^{11} & \tilde{Q}^{12} \\ \tilde{Q}^{21} & \tilde{Q}^{22} \end{pmatrix},$$

where $\tilde{Q}_{11}$ and $\tilde{Q}^{11}$ are $p \times p$ matrices. Notice that $\tilde{Q}^{11} = (\tilde{Q}_{11} - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}\tilde{Q}_{21})^{-1}$ and $\tilde{Q}^{12} = -\tilde{Q}^{11}\tilde{Q}_{12}\tilde{Q}_{22}^{-1}$. Then we have

$$\begin{aligned}\mathcal{R} V' (\Sigma' \tilde{Q}_0^{-1} \Sigma)^{-1} V \mathcal{R}' = & \mathcal{R} V' (P' \tilde{Q}^{11} P)^{-1} V \mathcal{R}' = \mathcal{R} V' P^{-1} (\tilde{Q}_{11} - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}\tilde{Q}_{21}) P^{-1} V \mathcal{R}' \\ = & b^2 L (\tilde{Q}_{11} - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}\tilde{Q}_{21}) L',\end{aligned}$$

and

$$\begin{aligned}\mathcal{R} V' (\Sigma' \tilde{Q}_0^{-1} \Sigma)^{-1} \Sigma' \tilde{Q}_0^{-1} W_k(1) = & \mathcal{R} V' P^{-1} (I_{p\times p}, -\tilde{Q}_{12}\tilde{Q}_{22}^{-1}) W_k(1) \\ = & -bL (I_{p\times p}, -\tilde{Q}_{12}\tilde{Q}_{22}^{-1}) W_k(1).\end{aligned}$$

Let $W_k(1) = (W_p(1)', W_q(1)')'$ with $q = k - p$. It thus implies that

$$\mathcal{W}_\infty(b; \delta, G_0, \Lambda) =^d \left\{ L(W_p(1) - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}W_q(1)) - b^{-1}\delta \right\} \left\{ L(\tilde{Q}_{11} - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}\tilde{Q}_{21})L' \right\}^{-1} \left\{ L(W_p(1) - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}W_q(1)) - b^{-1}\delta \right\}.$$

Let $L = U_L\Sigma_L V_L$ be the singular value decomposition for L with $\Sigma_L = (P_L, O)$, where P_L is a $m \times m$ diagonal matrix. Again we have $(V_L W_p, V_L\tilde{Q}_{12}, \tilde{Q}_{22}) =^d (W_p, \tilde{Q}_{12}, \tilde{Q}_{22})$. Thus by the definitions of C_{mm}, C_{mq} and C_{qq}, we get

$$\begin{aligned}
\mathcal{W}_\infty(b; \delta, G_0, \Lambda) =^d & \left\{ U_L\Sigma_L(W_p(1) - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}W_q(1)) - b^{-1}\delta \right\}' \left\{ U_L\Sigma_L(\tilde{Q}_{11} - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}\tilde{Q}_{21})\Sigma_L' U_L' \right\}^{-1} \\
& \left\{ U_L'\Sigma_L(W_p(1) - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}W_q(1)) - b^{-1}\delta \right\} \\
=^d & \left\{ U_L P_L(W_m(1) - C_{mq}C_{qq}^{-1}W_q(1)) - b^{-1}\delta \right\}' \left\{ U_L P_L C_{mm,q} P_L' U_L' \right\}^{-1} \\
& \left\{ U_L' P_L(W_m(1) - C_{mq}C_{qq}^{-1}W_q(1)) - b^{-1}\delta \right\} \\
= & \left\{ W_m(1) - C_{mq}C_{qq}^{-1}W_q(1) - P_L^{-1}U_L^{-1}b^{-1}\delta \right\}' C_{mm,q}^{-1} \\
& \left\{ W_m(1) - C_{mq}C_{qq}^{-1}W_q(1) - P_L^{-1}U_L^{-1}b^{-1}\delta \right\},
\end{aligned}$$

which completes the proof. ♢

2 Additional numerical results

S.1 Simulated critical values

We tabulated the critical values for the specification tests for leading cases in Tables S.1-S.3. For the percentage points 90%, 95%, and 99%, we provide critical value functions for the specification tests using the cubic equation

$$cv(b) = a_0 + a_1 b + a_2 b^2 + a_3 b^3, \quad b \in (0, b_{up}],$$

with b being the proportion of the block size or the bandwidth parameter to the sample size n. Here $b_{up} = 1$ for the CUGMM-based tests, $b_{up} = 0.5$ for the CUE-based tests, and the choices of b_{up} for the BEL-based tests are given in Table S.3. Note that the limiting distributions of the specification tests can all be written in the form $\mathcal{C}_b(W_k(r), r \in [0, 1])$, where $W_k(r)$ is a k dimensional vector of independent standard Brownian motions, and $\mathcal{C}_b$ is a functional depending on the form of the limiting distribution. To simulate the critical values, we approximate the limiting distribution by $\hat{\mathcal{C}}_b(\sum_{i=1}^{\lfloor n_0 r \rfloor} e_i/\sqrt{n_0}, r \in [0, 1])$, where e_i's are i.i.d random vectors generated from $N(0, I_k)$ and $n_0 = 5000$, and $\hat{\mathcal{C}}_b$ is the sample counterpart for $\mathcal{C}_b$. Based on 10000 Monte Carlo replications, we simulate the critical values for $b \in [0.01, b_{up}]$ with the spacing 0.01. For each percentage point, the simulated critical values were used to fit the $cv(b)$ function by ordinary least squares (OLS) [also see Kiefer and Vogelsang (2005)]. The resulting multiple R^2

is very close to one in all cases. The excellent fit of OLS is indeed closely related with the high order expansions of the nonstandard limiting distributions [see e.g. Corollary 2 of Sun et al. (2008); Sun (2014b)].

S.2 Empirical size

Figures S.1-S.3 present the empirical sizes for the Wald test and the LR type tests. The simulation setting is described in Section 6.2 of the paper.

References

[1] Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* **21** 1130-1164.

[2] Rio, E. (2013). Inequalities and limit theorems for weakly dependent sequences. *Laboratoire de Mathmatiques de Versailles (LM-Versailles).*

[3] Sun, Y. (2014a). Fixed-smoothing asymptotics in a two-step GMM framework. *Econometrica* **82** 2327-2370.

[4] Sun, Y. (2014b). Let's fix it: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. *J. Econometrics* **178** 659-677.

[5] Sun, Y., Phillips, P. C. B. and Jin, S. (2008). Optimal bandwidth selection in heteroscedasicity-autocorrlation robust testing. *Econometrica* **76** 175-194.

[6] Zhang, X. and Shao, X. (2015). On the coverage bound problem of empirical likelihood methods for time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* To appear.

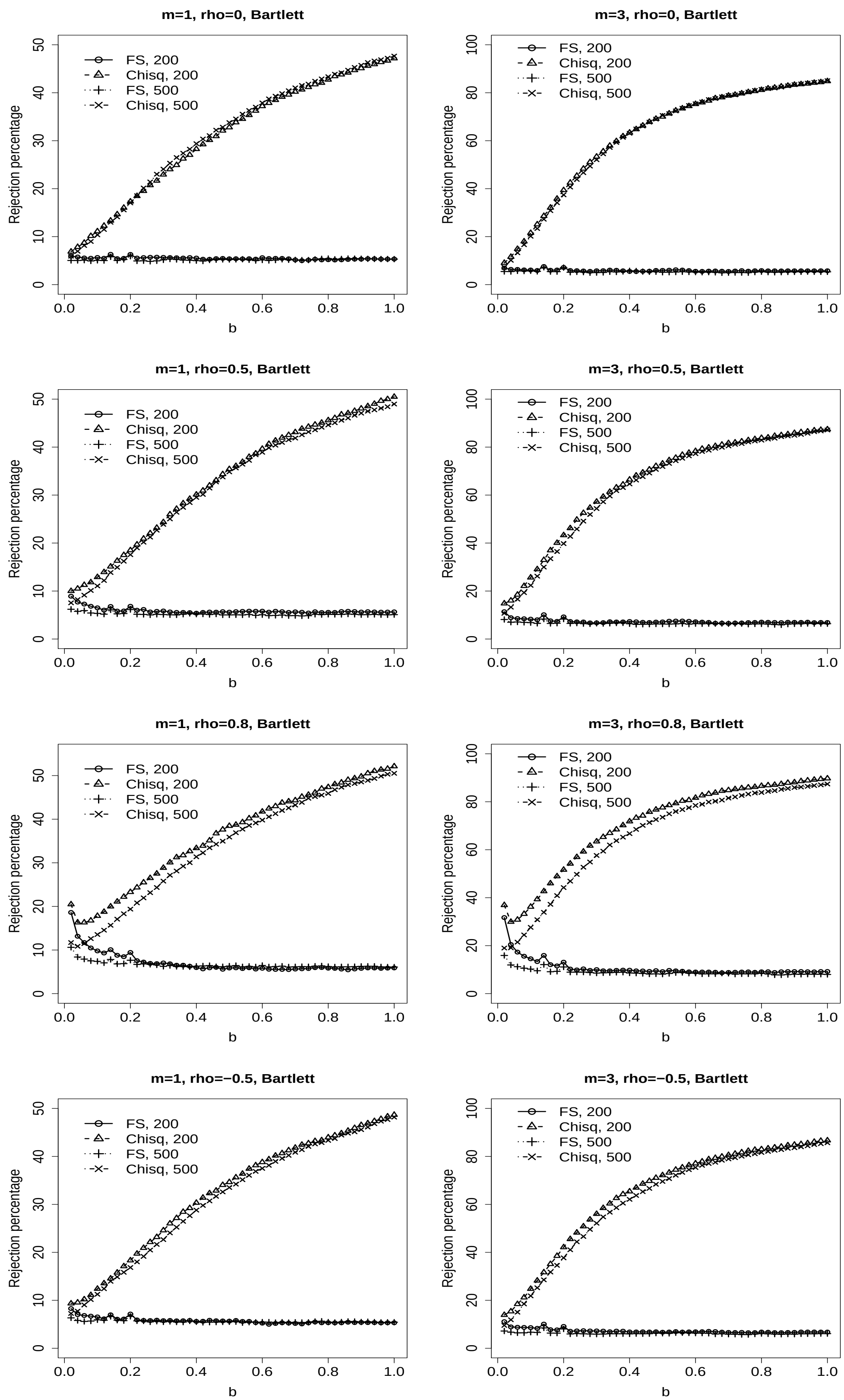

Figure S.1: Empirical rejection percentages for the CUGMM-based Wald test based on the fixed-smoothing approximation and the chi-square approximation, where $m = 1$ for the left panels and $m = 3$ for the right panels. The nominal level is 5%, and the number of Monte Carlo replications is 5000.

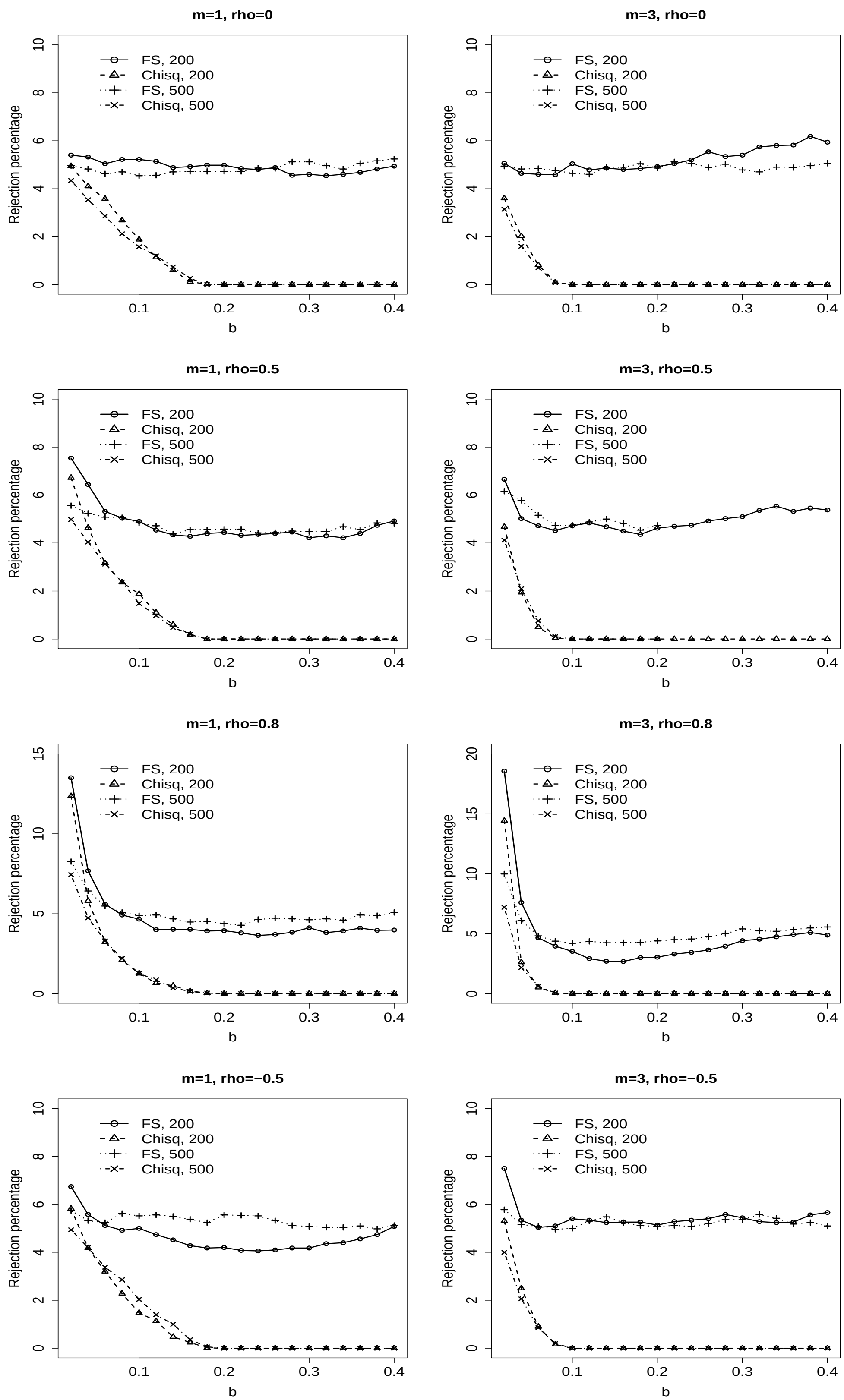

Figure S.2: Empirical rejection percentages for the CUE-based LR test based on the fixed-smoothing approximation and the chi-square approximation, where $m = 1$ for the left panels and $m = 3$ for the right panels. The nominal level is 5%, and the number of Monte Carlo replications is 5000.

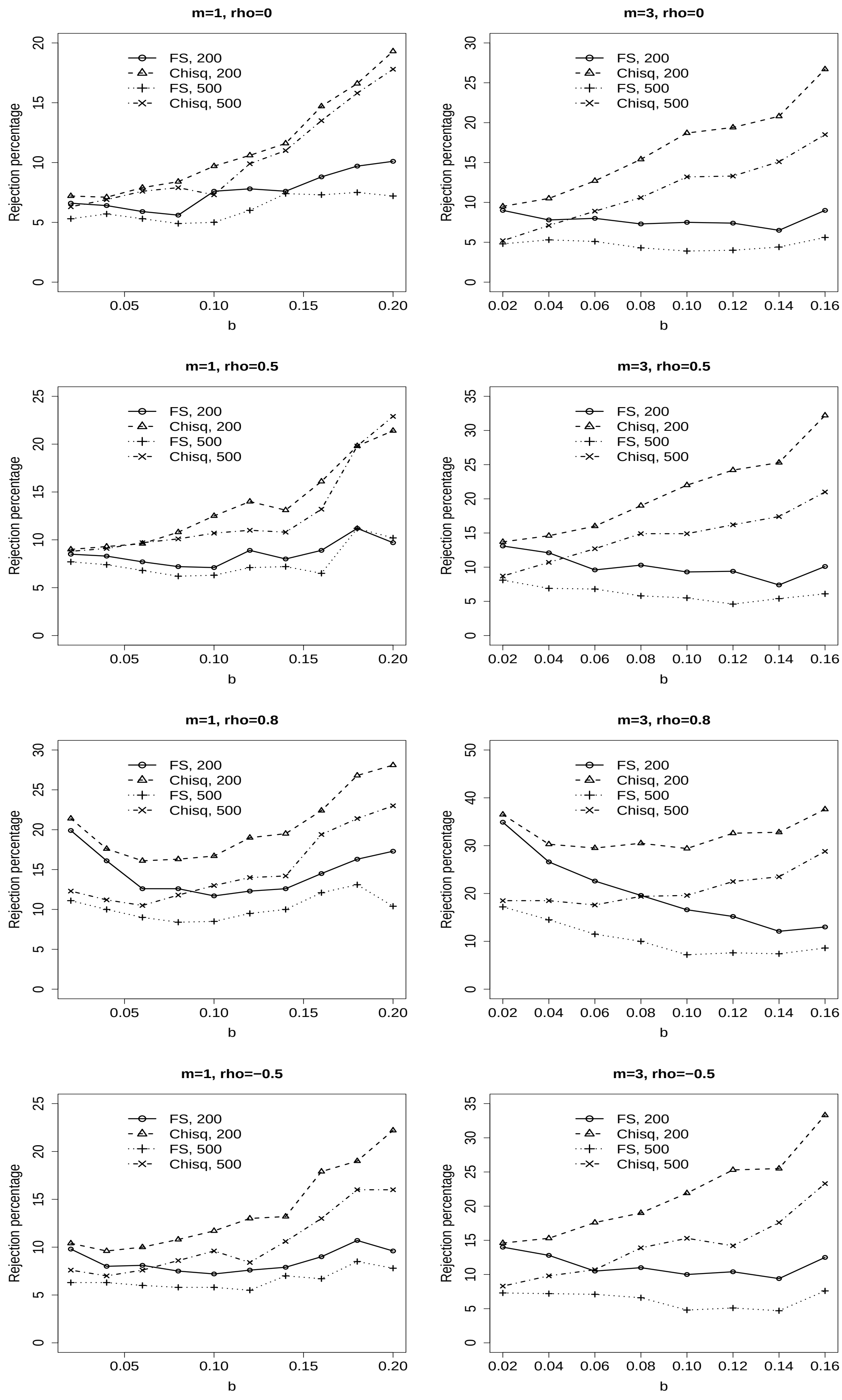

Figure S.3: Empirical rejection percentages for the BEL ratio test based on the fixed-smoothing approximation and the chi-square approximation, where $m = 1$ for the left panels and $m = 3$ for the right panels. The nominal level is 5%, and the number of Monte Carlo replications is 1000.

Table S.1: Critical value function coefficients for the limiting distributions of the CUGMM-based test statistics (with the Bartlett kernel).

	level	a_0	a_1	a_2	a_3	R^2
$\mathcal{J}_1$	90%	2.602	7.413	6.672	-2.596	0.9997
	95%	3.695	12.188	14.032	-6.474	0.9997
	99%	6.089	28.521	37.056	-21.385	0.9995
$\mathcal{J}_2$	90%	4.300	17.808	27.797	-14.024	0.9998
	95%	5.463	26.700	47.649	-28.790	0.9996
	99%	7.143	78.260	17.781	-3.653	0.9993
$\mathcal{J}_3$	90%	5.218	37.058	49.882	-31.132	0.9996
	95%	5.767	61.839	43.673	-27.484	0.9995
	99%	9.537	98.981	112.242	-77.201	0.9995
$\mathcal{J}_4$	90%	5.660	68.114	63.552	-42.807	0.9996
	95%	6.311	99.942	68.399	-47.213	0.9995
	99%	7.837	175.822	107.457	-83.295	0.9991
$\mathcal{W}_{1,1}$	90%	2.369	13.532	18.204	-9.634	0.9997
	95%	3.554	19.095	37.140	-21.783	0.9996
	99%	5.492	54.366	49.550	-29.963	0.9993
$\mathcal{W}_{2,1}$	90%	3.709	30.989	46.048	-29.165	0.9995
	95%	4.312	53.019	46.043	-29.352	0.9993
	99%	6.660	92.021	100.750	-70.834	0.9994
$\mathcal{W}_{3,1}$	90%	4.285	62.049	62.544	-43.175	0.9995
	95%	4.848	91.437	70.109	-50.819	0.9996
	99%	6.174	163.859	115.355	-90.885	0.9994
$\mathcal{W}_{1,2}$	90%	2.118	19.936	30.119	-19.418	0.9992
	95%	2.739	33.558	45.516	-31.914	0.9991
	99%	5.310	63.055	100.001	-70.130	0.9994
$\mathcal{W}_{2,2}$	90%	2.485	55.945	40.208	-29.234	0.9991
	95%	3.396	77.323	63.550	-46.911	0.9995
	99%	3.921	154.974	83.125	-70.862	0.9988
$\mathcal{W}_{3,2}$	90%	3.134	92.877	62.933	-50.271	0.9994
	95%	3.780	126.039	76.826	-60.206	0.9996
	99%	4.190	232.906	86.826	-80.971	0.9993
$\mathcal{W}_{1,3}$	90%	1.141	37.058	24.938	-20.737	0.9984
	95%	1.698	55.821	36.703	-28.288	0.9989
	99%	2.224	125.146	29.634	-24.904	0.9990
$\mathcal{W}_{2,3}$	90%	1.867	79.158	38.949	-32.154	0.9994
	95%	2.507	107.515	59.491	-50.412	0.9994
	99%	2.425	213.504	54.619	-50.578	0.9985
$\mathcal{W}_{3,3}$	90%	1.291	137.428	32.009	-33.539	0.9993
	95%	1.163	187.355	26.336	-39.138	0.9994
	99%	1.485	313.133	-0.554	-20.238	0.9992

Note: the critical values for $\mathcal{J}_{a_1}(b)$ and $\mathcal{W}_{a_2,a_3}(b)$ are approximated by cubic functions of the form $a_0 + a_1 b + a_2 b^2 + a_3 b^3$ for $b \in (0, 1]$. The estimated coefficients and multiple R^2 are reported. The Brownian motion is approximated by a normalized partial sum of 5000 i.i.d standard normal random variables and the number of Monte Carlo replications is 10000.

Table S.2: Critical value function coefficients for the limiting distributions of the CUE-based test statistics.

	level	a_0	a_1	a_2	a_3	R^2
$\mathcal{J}_{cue,1}$	90%	2.715	-2.089	-7.369	8.551	0.9997
	95%	3.832	-6.119	-4.846	11.133	0.9999
	99%	6.485	-21.798	26.704	-10.105	0.9997
$\mathcal{J}_{cue,2}$	90%	4.627	-9.525	-0.624	10.150	0.9999
	95%	5.963	-17.796	16.995	-2.569	0.9999
	99%	8.923	-40.204	74.718	-52.223	0.9998
$\mathcal{J}_{cue,3}$	90%	6.239	-18.513	15.808	0.634	0.9998
	95%	7.666	-29.417	44.235	-24.137	0.9999
	99%	11.119	-60.958	138.732	-115.646	0.9994
$\mathcal{J}_{cue,4}$	90%	7.785	-29.316	41.637	-20.160	0.9998
	95%	9.343	-42.616	79.325	-55.098	0.9999
	99%	12.719	-75.591	182.548	-158.035	0.9992
$\mathcal{LR}_{cue,1,1}$	90%	2.703	-3.956	-3.855	7.447	0.9997
	95%	3.878	-9.412	5.933	1.302	0.9995
	99%	6.556	-24.902	39.039	-23.578	0.9998
$\mathcal{LR}_{cue,2,1}$	90%	4.621	-12.963	11.982	-2.423	0.9999
	95%	5.926	-20.794	29.506	-16.094	1.0000
	99%	8.579	-40.106	79.468	-59.845	0.9998
$\mathcal{LR}_{cue,3,1}$	90%	6.199	-21.995	30.348	-15.210	0.9999
	95%	7.656	-32.739	58.642	-40.281	0.9999
	99%	10.754	-60.104	140.525	-120.087	0.9991
$\mathcal{LR}_{cue,1,2}$	90%	2.694	-6.680	2.446	3.081	0.9998
	95%	3.873	-13.098	18.250	-11.105	0.9997
	99%	6.141	-24.490	41.022	-27.170	0.9997
$\mathcal{LR}_{cue,2,2}$	90%	4.552	-15.671	19.790	-9.577	0.9999
	95%	5.915	-24.998	45.321	-33.163	0.9999
	99%	8.863	-48.372	113.278	-98.266	0.9982
$\mathcal{LR}_{cue,3,2}$	90%	6.230	-26.054	43.667	-28.803	1.0000
	95%	7.609	-35.898	70.870	-53.889	0.9999
	99%	10.819	-63.643	154.787	-135.947	0.9986
$\mathcal{LR}_{cue,1,3}$	90%	2.640	-8.146	3.710	5.432	0.9999
	95%	3.745	-14.142	18.297	-8.048	0.9999
	99%	6.363	-31.675	66.837	-54.577	0.9995
$\mathcal{LR}_{cue,2,3}$	90%	4.639	-19.402	28.364	-14.428	1.0000
	95%	5.923	-27.747	51.608	-37.222	0.9998
	99%	8.701	-48.235	111.143	-95.014	0.9987
$\mathcal{LR}_{cue,3,3}$	90%	6.196	-29.275	51.371	-33.200	0.9999
	95%	7.578	-39.330	80.944	-62.663	0.9997
	99%	10.729	-66.677	168.061	-152.191	0.9967

Note: the critical values for $\mathcal{J}_{cue,a_1}(b)$ and $\mathcal{LR}_{cue,a_2,a_3}(b)$ are approximated by cubic functions of the form $a_0 + a_1 b + a_2 b^2 + a_3 b^3$ for $b \in (0, 0.5]$. The estimated coefficients and multiple R^2 are reported. The Brownian motion is approximated by a normalized partial sum of 5000 i.i.d standard normal random variables and the number of Monte Carlo replications is 10000.

Table S.3: Critical value function coefficients for the limiting distributions of the BEL-based test statistics.

	level	b	a_0	a_1	a_2	a_3	R^2
$U_{el,1}$	90%	[0.01, 0.30]	2.722	2.063	11.061	10.825	0.9990
	95%	[0.01, 0.30]	3.636	9.405	-41.428	185.662	0.9933
	99%	[0.01, 0.24]	5.735	41.277	-408.287	1753.493	0.9952
$U_{el,2}$	90%	[0.01, 0.28]	3.687	46.967	-428.305	1408.895	0.9734
	95%	[0.01, 0.24]	4.937	57.530	-598.583	2366.409	0.9782
	99%	[0.01, 0.18]	7.566	87.454	-1153.026	6411.203	0.9805
$U_{el,3}$	90%	[0.01, 0.22]	5.069	70.426	-761.108	3205.284	0.9820
	99%	[0.01, 0.19]	6.600	70.110	-822.758	4301.616	0.9867
	99%	[0.01, 0.15]	8.320	248.957	-4456.415	26170.903	0.9718
$U_{el,4}$	90%	[0.01, 0.18]	6.505	88.910	-1114.448	5960.369	0.9910
	95%	[0.01, 0.16]	7.486	147.179	-2189.534	12392.553	0.9872
	99%	[0.01, 0.13]	9.473	326.726	-6372.303	42323.110	0.9779
$\mathcal{LR}_{el,1,1}$	90%	[0.01, 0.28]	1.893	42.649	-397.886	1305.620	0.9733
	95%	[0.01, 0.24]	2.808	61.112	-665.394	2523.659	0.9723
	99%	[0.01, 0.18]	5.528	76.797	-1047.148	5735.770	0.9916
$\mathcal{LR}_{el,2,1}$	90%	[0.01, 0.22]	3.539	67.616	-750.246	3148.968	0.9797
	95%	[0.01, 0.19]	4.929	69.195	-816.569	4167.509	0.9872
	99%	[0.01, 0.14]	7.515	109.648	-1753.763	11874.127	0.9886
$\mathcal{LR}_{el,3,1}$	90%	[0.01, 0.18]	4.987	91.407	-1168.053	6099.274	0.9895
	95%	[0.01, 0.16]	5.901	144.157	-2175.579	12292.074	0.9823
	99%	[0.01, 0.13]	7.714	312.234	-6177.733	41405.506	0.9776
$\mathcal{LR}_{el,1,2}$	90%	[0.01, 0.22]	1.868	53.727	-628.059	2647.589	0.9790
	95%	[0.01, 0.19]	2.900	68.442	-877.145	4204.341	0.9823
	99%	[0.01, 0.14]	5.418	78.612	-1239.375	8740.240	0.9882
$\mathcal{LR}_{el,2,2}$	90%	[0.01, 0.18]	3.304	94.149	-1268.588	6414.734	0.9850
	95%	[0.01, 0.15]	5.128	70.362	-949.549	6252.500	0.9941
	99%	[0.01, 0.12]	7.622	138.707	-2616.026	20554.413	0.9929
$\mathcal{LR}_{el,3,2}$	90%	[0.01, 0.15]	5.198	93.292	-1289.965	8290.256	0.9946
	95%	[0.01, 0.13]	6.624	112.562	-1753.285	12912.411	0.9941
	99%	[0.01, 0.10]	10.502	84.743	-1325.831	17707.469	0.9982
$\mathcal{LR}_{el,1,3}$	90%	[0.01, 0.18]	1.578	79.327	-1109.004	5472.033	0.9782
	95%	[0.01, 0.15]	3.020	67.878	-1027.325	6305.057	0.9857
	99%	[0.01, 0.12]	5.443	100.925	-1745.020	13718.961	0.9917
$\mathcal{LR}_{el,2,3}$	90%	[0.01, 0.15]	3.571	94.084	-1379.401	8495.686	0.9918
	95%	[0.01, 0.13]	4.942	103.104	-1637.590	11867.097	0.9898
	99%	[0.01, 0.10]	8.647	69.953	-1185.894	16329.897	0.9987
$\mathcal{LR}_{el,3,3}$	90%	[0.01, 0.13]	5.008	115.942	-1716.038	12257.481	0.9948
	95%	[0.01, 0.11]	6.815	101.232	-1444.346	13362.195	0.9947
	99%	[0.01, 0.09]	10.335	146.083	-2669.179	30187.183	0.9986

Note: the critical values for $U_{el,a_1}(b)$ and $\mathcal{LR}_{el,a_2,a_3}(b)$ are approximated by cubic functions of the form $a_0 + a_1 b + a_2 b^2 + a_3 b^3$ for b in a given range (see the third column). The estimated coefficients and multiple R^2 are reported. The Brownian motion is approximated by a normalized partial sum of 5000 i.i.d standard normal random variables and the number of Monte Carlo replications is 10000.