The University of Manchester

Economics Discussion Paper Series EDP-1205

Testing for Structural Instability in Moment Restriction Models: an Info-metric Approach

Alastair R. Hall

Yuyi Li

Chris D. Orme

January 2012

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

Testing for Structural Instability in Moment Restriction Models: an Info-metric Approach[1]

Alastair R. Hall

University of Manchester[2]

Yuyi Li

University of Manchester

Chris D. Orme

University of Manchester

April 4, 2011

[1] This paper is based on results developed as part of the thesis for which Li was awarded a PhD in Economics at the University of Manchester, UK in 2011. The first author acknowledges the support of the ESRC grant RES-062-23-1351. We are grateful to Amos Golan for useful discussions on aspects of this work.

[2] Corresponding author. Economics, SoSS, University of Manchester, Manchester M13 9PL, UK. Email: alastair.hall@manchester.ac.uk

Abstract

In this paper, we develop an info-metric framework for testing hypotheses about structural instability in nonlinear, dynamic models estimated from the information in population moment conditions. Our methods are designed to distinguish between three states of the world: (i) the model is structurally stable in the sense that the population moment condition holds at the same parameter value throughout the sample; (ii) the model parameters change at some point in the sample but otherwise the model is correctly specified; (iii) the model exhibits more general forms of instability than a single shift in the parameters. An advantage of the info-metric approach is that the null hypotheses concerned are formulated in terms of distances between various choices of probability measures constrained to satisfy (i) and (ii), and the empirical measure of the sample. Under the alternative hypotheses considered, the model is assumed to exhibit structural instability at a single point in the sample, referred to as the break point; our analysis allows for the break point to be either fixed *a priori* or treated as occuring at some unknown point within a certain fraction of the sample. We propose various test statistics that can be thought of as sample analogs of the distances described above, and derive their limiting distributions under the appropriate null hypothesis. In principle, there are a number of possible measures of distance that can be used in this context but we focus on the measure associated with Empirical Likelihood estimation. The limiting distributions of our statistics are non-standard but coincide with various distributions that arise in the literature on structural instability testing within the Generalized Method of Moments framework. A small simulation study illustrates the finite sample performance of our test statistics.

Keywords: Moment condition models, structural instability, parameter variation, Empirical Likelihood.

1 Introduction

There has been considerable interest in the development of tests for structural instability in moment condition models. In the majority of this literature, the null hypothesis is structural stability in the sense that the population moment condition holds at the same parameter value throughout the sample, and the alternative involves instability at single point in the sample, known as the break-point. Depending on the setting, this break-point can be treated as known, in which case the potential point of instability is specified *a priori*, or unknown, in which case the point of potential instability is left unspecified. The earliest contributions to this literature considered inference procedures within the Generalized Method of Moments (GMM) framework (Hansen, 1982). For the fixed-break point case, Andrews and Fair (1988) introduced tests for parameter variation, and Ghysels and Hall (1990) introduced so-called predictive tests that Ghysels, Guay, and Hall (1997) show test jointly parameter constancy and the overidentifying restrictions in one sub-sample. For the unknown break-point case, Andrews (1993) proposes so-called sup-tests for parameter variation, Sowell (1996) considers a general framework for the construction of tests for parameter variation, and Ghysels, Guay, and Hall (1997) propose extensions of the predictive test to this setting. Building from these earlier works, Hall and Sen (1999) show that the hypothesis of structural stability can be decomposed into one of parameter constancy and another concerning the validity of the overidentifying restrictions in each sub-sample, and propose tests for each component. They further show this approach has the potential to discriminate between states of the world in which violation of the null is caused by neglected parameter variation and those in which violation of the null is caused by more general forms of misspecification of the moment condition.

While all these tests are valid in their own terms, they are developed within the GMM framework and the latter has received some criticism in recent years because it has been found that GMM inference procedures can be unreliable in certain settings of interest.[1] This criticism has led to the development of alternative methods for estimation in moment condition models, leading examples of which are empirical likelihood (EL) (Qin and Lawless, 1994) and exponential tilting (ET) (Kitamura and Stutzer, 1997). Both EL and ET have a common structure, and this insight has led to the development of two generic frameworks for the estimation of moment

[1] For a review of this literature see *inter alia* Hall (2005)[Ch. 6].

condition models that include EL and ET (and other estimators of interest) as special cases. The first such framework is the Generalized Empirical Likelihood (GEL) introduced by Smith (1997). The second framework is the information-theoretic framework of Kitamura and Stutzer (1997) and its extensions in Golan (2002,2006). It is therefore of interest to develop tests for structural instability within these more general frameworks.

In a recent paper, Guay and Lamarche (2010) propose analogous tests to those of Hall and Sen (1999) for the GEL framework, and present a limiting distribution theory for these statistics under both null and local alternatives. They observe that the GEL statistics have the same first order asymptotic properties as their GMM counterparts under null and local alternatives. They report simulation evidence on their tests based on ET, and find the tests to perform comparably to their GMM counterparts for the most part but one particular GEL test based on the LM principle is superior.

In this paper, we consider the derivation of the same tests as Guay and Lamarche (2010) but from an information-theoretic - or equivalently - info-metric perspective. While the same tests result, we argue that the info-metric approach has considerable advantage in terms of the specification of the hypotheses and thus interpretation of the outcome of the tests.[2] This advantage stems from the info-metric approach being based on the concept of minimizing the distance between the class of probability distributions restricted to satisfy the moment condition and the true probability distribution. This allows us to relate the various hypotheses of interest in structural instability testing to the distance between certain classes of probability distributions and the true distribution. We believe this is a more fundamental - and also more instructive - representation of these hypotheses than their expression in terms of identifying restrictions (parameter variation) and overidentifying restrictions as is done in both the GMM and GEL frameworks.

In principle, there are a number of possible measures for the distance between probability distributions that can be used in developing our info-metric tests for structural instability. Here, we focus on the distance measure associated with Empirical Likelihood estimation and develop test statistics within the EL framework. We also explore certain other issues relating to structural instability testing in this context. Like Guay and Lamarche (2010), we assume the data to be

[2] Our results are based on Li's (2011) PhD thesis. This work was performed independently of and contemporaneously to Guay and Lamarche (2010).

weakly dependent and account for this dependence in estimation using the kernel-smoothing methods advocated by Smith (2004). However, given the nature of the null and alternative hypotheses here, there are two possible ways to proceed. One way involves kernel-smoothing the moment functions first and then splitting the resulting smoothed values into sub-samples of before-the-break and after-the-break (smoothed) observations. An alternative is to split the data into two sub-samples of before-the-break and after-the-break, and then kernel-smooth the moment functions using only observations from within that sub-sample. Guay and Lamarche (2010) employ the first approach. In this paper, we consider the second approach as well and demonstrate that both methods yield test statistics that are first order asymptotic equivalent under both null and local alternatives.

An outline of the paper is as follows. Section 2 presents the info-metric approach to the specification of the null and alternative hypotheses of our structural instability. Section 3 demonstrates that the order of kernel-smoothing and sample splitting does not affect the first order asymptotic properties of the partial EL - estimators under null and local alternatives. Section 4 presents the test statistics and discusses the connection between our info-metric methods and various structural instability tests derived within the GMM framework. Section 5 presents results from a small simulation study that indicates the finite sample performance of our methods. Section 6 concludes. All proofs are relegated to a mathematical appendix.

2 An info-metric approach to structural stability testing

In this section we propose an information-theoretic (IT) approach to testing for evidence of structural instability in population moment condition models. However, to motivate our approach, it is useful to begin by briefly reviewing IT estimation of moment condition models absent of any concerns regarding structural stability.

Suppose a researcher is interested in estimating the $k \times 1$ vector of parameters β_0 based on the information in the $\ell \times 1$ moment condition $E[g(Z, \beta_0)] = 0$ where Z is a $d \times 1$ random vector. It is assumed that $\ell > k$. This model is said to be structurally stable because the moment condition holds at the same parameter value throughout the sample. Following Kitamura (2006), we can characterize IT estimation of this model at the population level using the following framework.

Let $\mathbf{M}$ denote the set of all probability measures on $\Re^d$,

$$\mathbf{P}(\beta) = \left\{ P \in \mathbf{M} : \int g(z, \beta) dP = 0 \right\},$$

and

$$\mathbf{P} = \cup_{\beta \in \mathcal{B}} \mathbf{P}(\beta),$$

where $\mathcal{B}$ is the parameter space. Note that $\mathbf{P}$ is the set of all probability measures that are compatible with the moment condition, and is referred to as a statistical model in this context. This model is correctly specified if and only if $\mathbf{P}$ contains the true measure μ; that is, the data satisfies the population moment condition at $\beta = \beta_0$. A class of IT estimators of β can be defined as

$$\arg\inf_{\beta \in \mathcal{B}} \rho(\beta, \mu), \quad \text{where } \rho(\beta, \mu) = \inf_{P \in \mathbf{P}(\beta)} D(P \,\|\, \mu)$$

where $D(\cdot \,\|\, \cdot)$ is a divergence measure between two probability measures[3] and $\rho(\cdot)$ is referred to as the contrast function. Kitamura (2006) shows that if the model is correctly specified then the minimum of the contrast function is attained at $\beta = \beta_0$, the true parameter value.

Now consider the problem of testing structural stability. Define $Z(r)$ to be a stochastic process on $r \in [0, 1]$. We focus exclusively on the case where the alternative hypothesis involves instability at a single point and so we define

$$\begin{aligned} Z(r) &= Z^{(1)}, \text{ for } r \leq \pi \\ &= Z^{(2)}, \text{ for } r > \pi \end{aligned}$$

where $\pi \in (0, 1)$ and is referred to as the break-fraction. Notice that the break-fraction is embedded in the definition of $Z^{(i)}$ but we suppress this for notational brevity. In structural stability testing, π may be fixed *a priori*, the so-called "known break-point case", or it may be left unrestricted beyond $\pi \in \Pi \subset (0, 1)$, the so-called "unknown break-point case". Our methods can handle both cases, but for purposes of exposition here, it is most convenient to treat π as fixed and then to discuss the extension to the unknown break-point case at the end of the section.

To formalize the null and alternative hypotheses, we need to introduce two sets of probability measures. First, we define

$$\mathbf{P}_0 = \cup_{\beta \in \mathcal{B}} \mathbf{P}_0(\beta)$$

[3] This divergence measure must be non-negative and satisfy $D(P \,\|\, Q) = 0$ if and only if $P = Q$.

where

$$\mathbf{P}_0(\beta) \;=\; \left\{ (P_1, P_2) \in \mathbf{M} \times \mathbf{M} : \int g(z_i, \beta) dP_i \;=\; 0, \text{ for } i = 1, 2 \right\},$$

so that $\mathbf{P}_0$ is the set of all pairings of probability measures that are compatible with moment condition holding at the same parameter value in both sub-samples. Notice that this model specification differs from $\mathbf{P}$ by allowing for the measures for $Z^{(1)}$ and $Z^{(2)}$ to be potentially different. Second, we define the set

$$\mathbf{P}_1 \;=\; \cup_{(\beta_1, \beta_2) \in \mathcal{B} \times \mathcal{B}} \mathbf{P}_1(\beta_1, \beta_2),$$

where

$$\mathbf{P}_1(\beta_1, \beta_2) \;=\; \left\{ (P_1, P_2) \in \mathbf{M} \times \mathbf{M} : \int g(z_i, \beta_i) dP_i \;=\; 0, \;\; \text{for } i = 1, 2 \right\},$$

so that $\mathbf{P}_1(\beta_1, \beta_2)$ is the set of all pairings of probability measures that are compatible with moment condition holding in both sub-samples but at potentially different parameter values.

Using these definitions, the hypotheses of interest can be expressed in terms of (μ_1, μ_2), the true measures for $(Z^{(1)}, Z^{(2)})$. The null hypothesis is:

$$H_0(\pi) : \; (\mu_1, \mu_2) \in \mathbf{P}_0. \tag{1}$$

Thus under H_0 the model is structurally stable in the sense that the population moment condition holds at the same value in both sub-samples. One potential alternative of interest is:

$$H_A(\pi) : \; (\mu_1, \mu_2) \in \mathbf{P}_0^c, \tag{2}$$

which equates to "not $H_0(\pi)$". While this alternative is of interest in its own right, we show below that the states of the world under this alternative can be split into two groups, and such a decomposition can provide useful model building information. The first such group is captured by the hypothesis:

$$H_{PV}(\pi) : \; (\mu_1, \mu_2) \in \mathbf{P}_1 \setminus \mathbf{P}_0. \tag{3}$$

Under $H_{PV}(\pi)$, the moment condition is satisfied in both sub-samples but at different parameter values. This situation is commonly referred to as "parameter variation" which is reflected in the "PV" subscript. The second group is the hypothesis:

$$H_{MS}(\pi) : \; (\mu_1, \mu_2) \in \mathbf{P}_1^c. \tag{4}$$

Under $H_{MS}(\pi)$, the population moment condition is not satisfied in one or both sub-samples - even allowing for the possibility of a parameter shift - indicating the model is misspecified in that the moment condition fails to hold over the entire sample, which is reflected in the "MS" subscript.

While both $H_{PV}(\pi)$ and $H_{MS}(\pi)$ imply $H_0(\pi)$ is false, they have very different model building implications. $H_{PV}(\pi)$ implies that the model is correctly specified once allowance is made for the change in parameters, whilst $H_{MS}(\pi)$ implies the moment condition does not hold and hence the model is more fundamentally misspecified. As argued by Hall and Sen (1999), it therefore seems valuable to develop inference procedures that can distinguish these two cases. Hall and Sen (1999) achieve this goal within a GMM framework by developing separate tests based on the stability of the identifying restrictions and the stability of the overidentifying restrictions. Here we develop IT methods that provide similar model-building information. We believe that the IT approach is more attractive than the GMM framework of Hall and Sen (1999) and also the GEL framework of Guay and Lamarche (2010) because it is fundamentally anchored in distances between the underlying probability measures satisfying the various hypotheses considered.

To motivate the form of our inferential procedures, it is useful to consider population measures for discriminating between $H_0(\pi)$, $H_{PV}(\pi)$, and $H_{MS}(\pi)$. To this end, let $\rho_\pi([\beta_1, \beta_2], [\mu_1, \mu_2])$ denote the contrast function for estimation that allows for a break at the point indexed by π, and let $D_\pi([p_1, p_2] \,\|\, [q_1, q_2])$ denote the measure of divergence between two pairs of measures, $[p_1, p_2]$ and $[q_1, q_2]$, with the first of each pair pertaining to $Z^{(1)}$ and the second to $Z^{(2)}$. It then follows from the properties of the divergence measure that we have the following:

(i) $$\rho_\pi([\beta_*(\pi), \beta_*(\pi)], [\mu_1, \mu_2]) \begin{cases} = 0, & \text{if } H_0(\pi) \text{ true} \\ > 0, & \text{if } H_0(\pi) \text{ false,} \end{cases}$$

where

$$\beta_*(\pi) \;=\; \arg\inf_{\beta \in \mathcal{B}} \rho_\pi([\beta, \beta], [\mu_1, \mu_2])$$

for

$$\rho_\pi(\beta, \mu) \;=\; \inf_{[P_1, P_2] \in \mathbf{P}_1(\beta, \beta)} D_\pi([P_1, P_2] \,\|\, [\mu_1, \mu_2]);$$

(ii) $$\rho_\pi([\beta_{1,*}(\pi), \beta_{2,*}(\pi)], [\mu_1, \mu_2]) \begin{cases} = 0, & \text{if } H_{PV}(\pi) \text{ true} \\ > 0, & \text{if } H_{PV}(\pi) \text{ false,} \end{cases}$$

where

$$[\beta_{1,*}(\pi), \beta_{2,*}(\pi)] \;=\; \underset{[\beta_1,\beta_2]\in\mathcal{B}\times\mathcal{B}}{\arg\inf}\; \rho_\pi([\beta_1, \beta_2], [\mu_1, \mu_2]),$$

for

$$\rho_\pi([\beta_1, \beta_2], [\mu_1, \mu_2]) \;=\; \inf_{[P_1,P_2]\in\mathbf{P}_1(\beta_1,\beta_2)} D_\pi([P_1, P_2] \,\|\, [\mu_1, \mu_2]).$$

Given these properties, we can decompose $\mathcal{D}(\pi) = \rho_\pi([\beta_*(\pi), \beta_*(\pi)], [\mu_1, \mu_2])$ into two parts:

$$\mathcal{D}(\pi) \;=\; \mathcal{D}_1(\pi) \,+\, \mathcal{D}_2(\pi)$$

where

$$\begin{aligned}
\mathcal{D}_1(\pi) &= \rho_\pi([\beta_*(\pi), \beta_*(\pi)], [\mu_1, \mu_2]) \,-\, \rho_\pi(\beta_{1,*}(\pi), \beta_{2,*}(\pi)], [\mu_1, \mu_2]),\\
\mathcal{D}_2(\pi) &= \rho_\pi(\beta_{1,*}(\pi), \beta_{2,*}(\pi)], [\mu_1, \mu_2]).
\end{aligned}$$

It can be recognized that: if $H_0(\pi)$ is true then $\mathcal{D}_1(\pi) = \mathcal{D}_2(\pi) = 0$; if $H_{PV}(\pi)$ is true then $\mathcal{D}_1(\pi) \neq 0$ but $\mathcal{D}_2(\pi) = 0$; if $H_{MS}(\pi)$ is true then $\mathcal{D}_1(\pi) \neq 0$ and $\mathcal{D}_2(\pi) \neq 0$. Therefore, an examination of $\mathcal{D}(\pi)$ reveals whether the model is structurally stable, $H_0(\pi)$, or not, $H_A(\pi)$. On the other hand, an examination of $\mathcal{D}_1(\pi)$ and $\mathcal{D}_2(\pi)$ reveals whether the model is structurally stable, $H_0(\pi)$, or exhibits parameter variation, $H_{PV}(\pi)$, or is structurally unstable due to more general forms of misspecification, $H_{MS}(\pi)$. Therefore, we propose performing inference using sample analogs of $\mathcal{D}(\pi)$, $\mathcal{D}_1(\pi)$ and $\mathcal{D}_2(\pi)$.

To present these sample analogs, we need some additional notation. Replace $Z(r)$ by the time series $\{Z_t;\, t = 1, 2, \ldots, T\}$. It is assumed that the potential instability occurs at $t = [T\pi] = T_1$ say, where $[\cdot]$ denotes the integer part in this context. We refer to T_1 as the break-point. We divide the sample into two sub-samples: $\mathcal{T}_1(\pi) = \{1, 2, \ldots, T_1\}$, consisting of the observations up to and including the break-point; and $\mathcal{T}_2(\pi) = \{T_1{+}1, T_1{+}2, \ldots T\}$, consisting of the observations after the break.

It is well known that IT methods based on the assumption of independently and identically distributed data are sub-optimal if the data are weakly dependent.[4] Various approaches have been proposed for handling this dependence: we employ the kernel smoothing methods proposed by Smith (2004).[5] Within this approach, the original moment function in period t, $g(Z_t, \beta) =$

[4] See Kitamura (1997) and Kitamura and Stutzer (1997).

[5] Other possibilities for handling the dependence include blocking methods (see Kitamura, 1997 and Kitamura and Stutzer, 1997) or the use of parametric models (see Kitamura, 2006).

$g_t(\beta)$ say, is replaced by the kernel smoothed version,

$$g_t^s(\beta) = \frac{1}{h_T} \sum_{j=t-T}^{t-1} k\left(\frac{j}{h_T}\right) g_{t-j}(\beta), \tag{5}$$

where the superscript s indicates the operation of kernel smoothing and, respectively, h_T and $k(\cdot)$ denote the bandwidth and a kernel function, that are discussed in detail in Section 3. To implement IT estimation using kernel smoothing, we replace the true measures, $[\mu_1, \mu_2]$ by the empirical measures $[\hat{\mu}_1, \hat{\mu}_2]$. Notice that these measures relate to the stationary distributions of $Z^{(1)}$ and $Z^{(2)}$.[6] Since we allow for the measures to be different, $\hat{\mu}_{1,t} = T_1^{-1}$ for $t \in \mathcal{T}_1(\pi)$ and $\hat{\mu}_{2,s} = T_2^{-1}$ for $T_2 = T - T_1$ and $s \in \mathcal{T}_2(\pi)$. Following Kitamura and Stutzer (1997), we also replace the measures P_i by the probability mass functions $\hat{P}_1 = [p_{1,1}, p_{1,2} \ldots, p_{1,T_1}]$, $\hat{P}_2 = [p_{2,1}, p_{2,2} \ldots, p_{2,T_2}]$.

In our inference procedures, $\beta_{i,*}(\pi)$ and $\beta_*(\pi)$ are replaced, respectively, by the partial-sample IT estimators, $\hat{\beta}_i(\pi)$, and the restricted partial-sample IT estimator, $\hat{\beta}_R(\pi)$, defined as follows. The (unrestricted) partial-sample IT estimators are,

$$[\hat{\beta}_1(\pi), \hat{\beta}_2(\pi)] \;=\; \underset{[\beta_1,\beta_2]\in\mathcal{B}\times\mathcal{B}}{\arg\inf}\; \rho_{\pi,T}([\beta_1, \beta_2], [\hat{\mu}_1, \hat{\mu}_2]) \tag{6}$$

where

$$\rho_{\pi,T}([\beta_1, \beta_2], [\hat{\mu}_1, \hat{\mu}_2]) = \inf_{[\hat{P}_1,\hat{P}_2]\in\hat{\mathbf{P}}_1(\beta_1,\beta_2)} D_\pi([\hat{P}_1, \hat{P}_2] \,\|\, [\hat{\mu}_1, \hat{\mu}_2]) \tag{7}$$

and

$$\hat{\mathbf{P}}_1(\beta_1, \beta_2) \;=\; \left\{ (\hat{P}_1, \hat{P}_2) :\; p_{i,t} > 0, \sum_{t\in\mathcal{T}_i(\pi)} p_{i,t} = 1, \sum_{t\in\mathcal{T}_i(\pi)} p_{i,t} g_t^s(\beta_i),\; i = 1, 2 \right\}. \tag{8}$$

On the other hand, the restricted partial-sample IT estimator is,

$$\hat{\beta}_R(\pi) \;=\; \arg \inf_{[\beta,\beta]\in\mathcal{B}\times\mathcal{B}} \rho_{\pi,T}([\beta, \beta], [\hat{\mu}_1, \hat{\mu}_2]). \tag{9}$$

We propose performing inference based on scaled versions of the following analogs to $\mathcal{D}(\pi)$, $\mathcal{D}_1(\pi)$ and $\mathcal{D}_2(\pi)$,

$$\hat{\mathcal{D}}_T(\pi) \;=\; \hat{\mathcal{D}}_{1,T}(\pi) + \hat{\mathcal{D}}_{2,T}(\pi) \tag{10}$$

$$\hat{\mathcal{D}}_{1,T}(\pi) \;=\; \rho_{\pi,T}([\hat{\beta}_R(\pi), \hat{\beta}_R(\pi)], [\hat{\mu}_1, \hat{\mu}_2]) \;-\; \rho_{\pi,T}([\hat{\beta}_1(\pi), \hat{\beta}_2(\pi)], [\hat{\mu_1}, \hat{\mu_2}]) \tag{11}$$

$$\hat{\mathcal{D}}_{2,T}(\pi) \;=\; \rho_{\pi,T}([\hat{\beta}_1(\pi), \hat{\beta}_2(\pi)], [\hat{\mu_1}, \hat{\mu_2}]) \tag{12}$$

[6] See Smith (2004)[p.19].

To implement our procedures, it is necessary to choose a measure of divergence. Kitamura and Stutzer (1997) use the Kullback-Leibler information criterion (KLIC) distance. Golan (2002, 2006) considers the extension of Kitamura and Stutzer's (1997) methods to more general measures such as the generalized cross entropy and Cressie-Read (CR) divergence measure (Cressie and Read, 1984). The framework above can be applied to any of these settings, but for concreteness we focus on the CR divergence measure which is defined as follows in our context:

$$D_{\pi}^{(\alpha)}([\hat{P}_1, \hat{P}_2] \,\|\, [\hat{\mu}_1, \hat{\mu}_2]) \;=\; \frac{\alpha}{1+\alpha}\left\{\sum_{i=1}^{2}\sum_{t\in\mathcal{T}_i(\pi)} p_{i,t}\left\{\left(\frac{p_{i,t}}{\hat{\mu}_{i,t}}\right)^{\alpha} - 1\right\}\right\} \tag{13}$$

which is defined for $-\infty < \alpha < \infty$. Appropriate choices of α lead to certain familiar estimation methods: for example, $lim_{\alpha\to 0} D_{\pi}^{(\alpha)}(\cdot\|\cdot)$ yields the optimand for exponential tilting (ET) estimator of Kitamura and Stutzer (1997) in each sub-sample, and $lim_{\alpha\to -1} D_{\pi}^{(\alpha)}(\cdot\|\cdot)$ yields the empirical likelihood (EL) estimator of Owen (2001) in each sub-sample. Moreover, Newey and Smith (2004) and Anatolyev (2005) demonstrate that EL has better second order bias properties than ET and so in the following sections we develop versions of $\hat{\mathcal{D}}_i(\pi)$ based on EL estimators.

So far, we have focused on the fixed break-point case. The extension to the unknown break-point case is as follows. The null hypothesis of structural stability becomes $H_0(\Pi) : \, H_0(\pi) \; \forall \pi \in \Pi \subset (0,1)$. The difference between $H_0(\pi)$ and $H_0(\Pi)$ is that the former specifies precisely the point at which the structural break is suspected. This difference is reflected in the associated test statistics, with tests for $H_0(\pi)$ being designed to have power against a break at π and the tests for $H_0(\Pi)$ being designed to maximize power against a weighted sequence of alternatives that allows for breaks at all points in Π. These test statistics are developed in Section 4. Before that we turn to another issue that arises in the implementation of our tests. As mentioned in the introduction, there are two options regarding the sequencing of kernel smoothing and sample splitting: split the sample then kernel smooth (smooth *after* sample splitting) or kernel smooth then split the sample (smooth *before* sample splitting). The former only smooths over the moment functions for which $t \in \mathcal{T}_i(\pi)$, for all $\pi \in \Pi$, and gives rise to the following smoothed moment functions

$$g_t^{sa}(\beta) = \begin{cases} \frac{1}{h_T}\sum_{j=t-[T\pi]}^{t-1} k\left(\frac{j}{h_T}\right) g_{t-j}(\beta), & t = 1, ..., [T\pi] \\ \frac{1}{h_T}\sum_{j=t-T}^{t-[T\pi]-1} k\left(\frac{j}{h_T}\right) g_{t-j}(\beta), & t = [T\pi]+1, ..., T, \end{cases} \tag{14}$$

whilst the latter approach yields $g_t^{sb}(\beta) \equiv g_t^{s}(\beta)$, given by (5), for $t \in \mathcal{T}_i(\pi)$, and all $\pi \in \Pi$.

In the following section, we explore the impact of this sequencing on the first order asymptotic behaviour of the unrestricted and restricted partial-sample IT estimators.

3 Large sample behaviour of partial-sample IT estimators

Based on the full sample, the EL (IT) criterion function would be

$$Q_T(\beta, \lambda) = \frac{1}{T}\sum_{t=1}^{T} \ln\left(1 + k\lambda' g_t^s(\beta)\right)$$

where $g_t^s(\beta)$ is defined at (5) and $k = k_1/k_2$ with $k_j = \int_{-\infty}^{\infty} k(\omega)^j d\omega$, $j = 1, 2$. Whilst $\beta \in \mathcal{B} \subset \Re^k$, the auxiliary parameters $\lambda \in \Lambda_T$ are restricted so that w.p.a.1 (with probability approaching 1) $k\lambda' g_t^s(\beta) > -1$ for all $\left(\beta', \lambda'\right)' \in \mathcal{B} \times \Lambda_T$ and $t = 1, ..., T$. Specifically, Λ_T is defined so that bounds are placed on λ that "shrink" with T, at an appropriate rate. The full-sample EL estimator is then defined as

$$\tilde{\beta} \equiv \arg \min_{\beta \in \mathcal{B}} \sup_{\lambda \in \Lambda_T} Q_T(\beta, \lambda).$$

Estimation proceeds in two steps:

1. $Q_T\left(\beta, \lambda\right)$ is maximised over λ, for given β, yielding

$$\tilde{\lambda}\left(\beta\right) = \arg \sup_{\lambda \in \Lambda_T} Q_T\left(\beta, \lambda\right).$$

2. The EL estimator, $\tilde{\beta}$, is the minimiser of the profile EL objective function, $Q_T\left(\beta, \tilde{\lambda}(\beta)\right)$:

$$\tilde{\beta} = \arg \min_{\beta \in \mathcal{B}} Q_T\left(\beta, \tilde{\lambda}(\beta)\right).$$

Consider, now, splitting the sample according to $\mathcal{T}_i(\pi)$, $i = 1, 2$, for all $\pi \in \Pi$, which yields the (unrestricted) partial-sample EL estimators $\hat{\beta}_i(\pi)$, $i = 1, 2$, based on the two sub-samples $t \in \mathcal{T}_i(\pi)$, $i = 1, 2$, respectively. To analyse these estimators, let us (for the moment) employ the smoothed moment functions *after* the sample split. (As noted previously, it will be shown that the use of $g_t^{sa}(\beta)$ or $g_t^s(\beta)$ makes no difference, asymptotically, to the sampling results obtained for the partial-sample EL estimators.) Specifically, the (unrestricted) partial-sample EL (PSEL) estimators are defined by

$$\hat{\beta}_i^a(\pi) = \arg \min_{\beta \in \mathcal{B}} \sup_{\lambda \in \Lambda_T} \frac{1}{T} \sum_{t \in \mathcal{T}_i(\pi)} \ln\left(1 + k\lambda' g_t^{sa}(\beta)\right), \quad i = 1, 2,$$

where $g_t^{sa}(\beta)$ is given by (14), and, correspondingly,

$$\hat{\lambda}_i^a(\pi) = \arg \sup_{\lambda \in \Lambda_T} \frac{1}{T} \sum_{t \in \mathcal{T}_i(\pi)} \ln\left(1 + k\lambda' g_t^{sa}(\hat{\beta}_i^a(\pi))\right), \quad i = 1, 2.$$

To analyse these estimators for all $\pi \in \Pi \subset (0,1)$ define $\theta' = \left(\beta_1', \beta_2'\right)' \in \Phi = \mathcal{B} \times \mathcal{B}$, $\gamma' = \left(\lambda_1', \lambda_2'\right)' \in \Gamma_T = \Lambda_T \times \Lambda_T$ and the following $(2\ell \times 1)$ unsmoothed and smoothed moment functions

$$g_t(\theta, \pi) = \mathbb{I}_{t,T}(\pi) \begin{pmatrix} g_t(\beta_1) \\ 0 \end{pmatrix} + (1 - \mathbb{I}_{t,T}(\pi)) \begin{pmatrix} 0 \\ g_t(\beta_2) \end{pmatrix}$$

$$g_t^{sa}(\theta, \pi) = \mathbb{I}_{t,T}(\pi) \begin{pmatrix} g_t^{sa}(\beta_1) \\ 0 \end{pmatrix} + (1 - \mathbb{I}_{t,T}(\pi)) \begin{pmatrix} 0 \\ g_t^{sa}(\beta_2) \end{pmatrix} \tag{15}$$

where $\mathbb{I}_{t,T}(\pi)$ is an indicator variable that takes the value 1 if $t \leq [T\pi]$ and the value 0 otherwise. Let

$$Q_T^a(\theta, \gamma, \pi) = \frac{1}{T} \sum_{t=1}^{T} \ln\left(1 + k\gamma' g_t^{sa}(\theta, \pi)\right)$$

then we have $\hat{\theta}^a(\pi) = \left(\hat{\beta}_1^a(\pi)', \hat{\beta}_2^a(\pi)'\right)'$ where

$$\hat{\theta}^a(\pi) = \arg \min_{\theta \in \Theta} \sup_{\gamma \in \Gamma_T} Q_T^a(\theta, \gamma, \pi) \tag{16}$$

with

$$\hat{\gamma}^a(\pi) = \arg \sup_{\gamma \in \Gamma_T} Q_T^a(\hat{\theta}^a(\pi), \gamma, \pi). \tag{17}$$

To develop the analysis, we need to impose certain assumptions and we follow the spirit of Smith (2004). We consider behaviour under the null of no change, and assume the data satisfy the following condition.

Assumption 1 *Data are generated by a sequence of strictly stationary and strong mixing* $\boldsymbol{Z}$*-valued random vectors* $\{Z_t\}_{t=1}^{\infty}$*, with mixing coefficients,* $\alpha(j)$*, satisfying* $\sum_{j=1}^{\infty} j^2 \alpha(j)^{(\upsilon-1)/\upsilon} < \infty$*, for some* $\upsilon > 1$*, where* $\boldsymbol{Z}$ *is a Borel subset of* $\Re^d$.

As noted in the previous section, we handle the dependence in the data implied by Assumption 1 through kernel smoothing. The next assumption addresses the bandwidth, h_T, and choice of kernel, $k(.)$, such that they obey conditions similar to those laid out in Theorem 1(a) of Andrews (1991). Let

$$\bar{k}(\omega) = \begin{cases} \sup_{b \geq \omega} |k(b)|, & \omega \geq 0 \\ \sup_{b \leq \omega} |k(b)|, & \omega < 0 \end{cases}$$

and $K(\lambda) = (2\pi)^{-1} \int k(x) \exp(-\iota x \lambda) dx$, the spectral window generator of the kernel $k(.)$, with $k_j = \int_{-\infty}^{\infty} k(\omega)^j d\omega$, $j = 1, 2$.

Assumption 2 *(i) $h_T = O(T^{\frac{1}{2\delta}})$ for some $\delta > 1$; (ii) $k(.) : \Re \to [-k_{\max}, k_{\max}]$, $k_{\max} < \infty$, $k(0) \neq 0$, $k_1 \neq 0$, and $k(.)$ is continuous at 0 and almost everywhere; (iii) $\int_{-\infty}^{\infty} \bar{k}(\omega) d\omega < \infty$; (iv) $|K(x)| \geq 0$ for all $x \in \Re$.*

Assumption 2(i) is a slight adaptation of Smith (2004), as used by Guay and Lamarche (2010), which simplifies certain aspects of the proofs at no extra cost.

We must also place restrictions on the (unsmoothed) moment function $g_t(\beta) = g(Z_t, \beta)$, and these are specified in the following assumptions. Define the following quantities: $\bar{g}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\beta)$, $\Omega(\beta) = \lim_{T\to\infty} var\left(\sqrt{T}\bar{g}_T(\beta)\right)$, and $\bar{g}_{[T\pi]}(\beta) = \frac{1}{T}\sum_{t=1}^{[T\pi]} g_t(\beta)$. The smoothed counterparts of $\bar{g}_T(\beta)$ and $\bar{g}_{[T\pi]}(\beta)$ are $\bar{g}_T^{sa}(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t^{sa}(\beta)$ and $\bar{g}_{[T\pi]}^{sa}(\beta) = \frac{1}{T}\sum_{t=1}^{[T\pi]} g_t^{sa}(\beta)$, respectively.

Assumption 3 *(i) $E[\sup_{\beta\in\mathcal{B}} ||g_t(\beta)||^{\eta}] < \infty$ for some $\eta > \max\left[4\upsilon, \frac{2\delta}{\delta-1}\right]$; (ii) $\Omega(\beta)$ is finite and p.d. for all $\beta \in \mathcal{B} \subset \Re^k$, where $\mathcal{B}$ is a compact parameter set; (iii) The moment function $g(z, \beta) \subset \Re^{\ell}$ is continuous in z for all $\beta \in \mathcal{B}$, and is continuous at each $\beta \in \mathcal{B}$ w.p.a.1; (iv) $g(\beta_0) = 0$ and $\inf_{\pi\in\Pi} \|g(\theta, \pi)\| > 0$ for all $\theta \neq \theta_0 = (\beta_0', \beta_0')'$.*

The existence of $g(\beta) \equiv E[g_t(\beta)]$ and $g(\theta, \pi) \equiv (\pi g(\beta_1)', (1-\pi) g(\beta_2)')'$ is guaranteed by Assumption 3(i), whilst Assumption 3(iv) is a global identification condition. Assumptions 1- 3 ensure that an appropriate FCLT applies to both $\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)$, with $\lim_{T\to\infty} var\left(\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right) = \pi\Omega_0$, and $\sqrt{T}\bar{g}_{[T\pi]}^{sa}(\beta_0)$, with $\lim_{T\to\infty} var\left(\sqrt{T}\bar{g}_{[T\pi]}^{sa}(\beta_0)\right) = k_1\pi\Omega_0$, for all $\pi \in [0, 1]$, where $\Omega_0 = \Omega(\beta_0)$. These assumptions also ensure that a (weak) ULLN applies to $\bar{g}_T(\beta)$ and to both $\bar{g}_T(\theta, \pi) \equiv \frac{1}{T}\sum_{t=1}^{T} g_t(\theta, \pi)$ and $\bar{g}_T^{sa}(\theta, \pi) \equiv \frac{1}{T}\sum_{t=1}^{T} g_t^{sa}(\theta, \pi)$, with the latter two also being uniform over $\pi \in [0, 1]$.[7]

The following assumption restricts the bounds on λ, ensuring that they shrink to zero more slowly than the stochastic rate of convergence of $\tilde{\lambda}$,

Assumption 4 $\lambda \in \Lambda_T = \left\{\lambda : \|\lambda\| \leq B\left(T/h_T^2\right)^{-\varepsilon}\right\}$, *where* $\frac{\delta}{\eta(\delta-1)} < \varepsilon < \frac{1}{2}$, *for some finite* $B > 0$.

[7] Indeed, Andrews (1993, Proof of Theorem A1) shows that $\sup_{\pi} \sup_{\theta} \|\bar{g}_T(\theta, \pi) - \bar{g}(\theta, \pi)\| = o_p(1)$.

Under the above assumptions, we can establish the consistency of the PSEL estimator as follows:

Theorem 1 *Under Assumptions 1-4: (i)* $\sup_{\pi\in\Pi}\left\|\hat{\theta}^a(\pi)-\theta_0\right\|=o_p(1)$, *and (ii)* $\sup_{\pi\in\Pi}\|\hat{\gamma}^a(\pi)\|=o_p(1)$.

To establish asymptotic normality, the following assumptions are made regarding the (unsmoothed) derivative of the moment function $G_t(\beta)=\partial g_t(\beta)/\partial\beta'$, and it will be useful to define $G(\beta)=E\left[G_t(\beta)\right]$, which exists by Assumption 5(i), below.

Assumption 5 *(i)* $E[\sup_{\beta\in\mathcal{B}}||G_t(\beta)||^{\eta/(\eta-1)}]<\infty$ *for some* $\eta>\max[4v,\frac{2\delta}{\delta-1}]$; *(ii) The moment function* $g(z,\beta)\subset\Re^{\ell}$ *is continuously partially differentiable in* β *in a neighbourhood* $\mathcal{B}_0$ *of* $\beta_0\in int(\mathcal{B})$*, w.p.a.1; (iii)* $G_0\equiv G(\beta_0)$ *has full rank k.*

It will also be useful to define the following matrices

$$A(\pi)=\begin{bmatrix}\pi & 0\\ 0 & 1-\pi\end{bmatrix}$$

$$\Omega_0(\pi)=\lim_{T\to\infty}var\left(\sqrt{T}\bar{g}_T(\theta_0,\pi)\right)=\begin{bmatrix}\pi\Omega_0 & 0\\ 0 & (1-\pi)\,\Omega_0\end{bmatrix}=A(\pi)\otimes\Omega_0$$

$$G_0(\pi)=\begin{bmatrix}\pi G_0 & 0\\ 0 & (1-\pi)\,G_0\end{bmatrix}=A(\pi)\otimes G_0$$

and $M_0=\Omega_0^{-1/2}G_0,\ P_0=M_0\left(M_0'M_0\right)^{-1}M_0$. Under Assumptions 1 and 3, Andrews (1993, Proof of Theorem 1), shows that $\xi_T(\pi)\Longrightarrow J_\ell(\pi)$, as a process indexed by $\pi\in\Pi$, where

$$\xi_T(\pi)=\left(I_2\otimes\Omega_0^{-1/2}\right)\sqrt{T}\bar{g}_T(\theta_0,\pi)=\begin{bmatrix}\Omega_0^{-1/2}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\\ \Omega_0^{-1/2}\left\{\sqrt{T}\bar{g}_T(\beta_0)-\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right\}\end{bmatrix}$$

and

$$J_\ell(\pi)=\begin{bmatrix}B_\ell(\pi)\\ B_\ell(1)-B_\ell(\pi)\end{bmatrix}$$

with $B_\ell(\pi)$, $\pi\in[0,1]$, being a vector of ℓ mutually independent standard Brownian motions on $[0,1]$. Furthermore, Assumptions 1, 2 and 3, and arguments similar to Smith (2004, Lemma A3) establish that $h_T\bar{V}_T^a(\theta_0,\pi)\xrightarrow{p}k_2\Omega_0(\pi)$, uniformly in π, where

$$\bar{V}_T^a(\theta,\pi)=\frac{1}{T}\sum_{t=1}^{T}g_t^{sa}(\theta,\pi)\,g_t^{sa}(\theta,\pi)'.$$

Theorem 2 *Under Assumptions 1-5, every sequence of PSEL estimators defined by (16) and (17), $T\geq 1$,*

$$\begin{aligned}\sqrt{T}\left(\hat{\theta}^a(\pi)-\theta_0\right) &= -\left(A(\pi)^{-1}\otimes\left(M_0'M_0\right)^{-1}M_0'\right)\xi_T(\pi)+o_{p\pi}(1)\\ &\Longrightarrow -\left(A(\pi)^{-1}\otimes\left(M_0'M_0\right)^{-1}M_0'\right)J_\ell(\pi)\end{aligned}$$

$$\begin{aligned}(\sqrt{T}/h_T)\hat{\gamma}^a(\pi) &= \left(A(\pi)^{-1}\otimes\Omega_0^{-1/2}\left(I_\ell-P_0\right)\right)\xi_T(\pi)+o_{p\pi}(1)\\ &\Longrightarrow \left(A(\pi)^{-1}\otimes\Omega_0^{-1/2}\left(I_\ell-P_0\right)\right)J_\ell(\pi)\end{aligned}$$

where $\Longrightarrow$ denotes weak convergence to a process indexed by $\pi\in\Pi$, provided Π has closure in $(0,1)$, and $o_{p\pi}(1)$ denotes terms that are $o_p(1)$ uniformly in $\pi\in\Pi$. Further, $\hat{\theta}(\cdot)$ and $\hat{\gamma}(\cdot)$ are asymptotically uncorrelated.

Alternatively, the weak convergence results could be stated as

$$\begin{aligned}(A(\pi)\otimes I_k)\sqrt{T}\left(\hat{\theta}^a(\pi)-\theta_0\right) &\Longrightarrow -\left(I_2\otimes\left(M_0'M_0\right)^{-1}M_0'\right)J_\ell(\pi)\\ (A(\pi)\otimes I_\ell)(\sqrt{T}/h_T)\hat{\gamma}^a(\pi) &\Longrightarrow \left(I_2\otimes\Omega_0^{-1/2}\left(I_\ell-P_0\right)\right)J_\ell(\pi).\end{aligned}$$

These results ensure that, uniformly in π, $h_T\bar{V}_T^a(\hat{\theta}^a(\pi),\pi)\xrightarrow{p}k_2\Omega_0(\pi)$ and $\frac{1}{T}\sum_{t=1}^{T}\dfrac{\partial g_t^{sa}(\hat{\theta}^a(\pi),\pi)}{\partial\theta'}\xrightarrow{p}k_1G_0(\pi)$.

The following Theorem establishes that the order of sample splitting and kernel smoothing makes no difference, asymptotically, to the weak convergence results obtained in Theorems 1 and 2 or, indeed, the results of Section 4. For smoothing before the sample split, define the $(2\ell\times 1)$ smoothed moment function as

$$g_t^{sb}(\theta,\pi)=\mathbb{I}_{t,T}(\pi)\begin{pmatrix}g_t^s(\beta_1)\\ 0\end{pmatrix}+\left(1-\mathbb{I}_{t,T}(\pi)\right)\begin{pmatrix}0\\ g_t^s(\beta_2)\end{pmatrix},\tag{18}$$

where $g_t^s(\beta)$ is defined at (5), with $\bar{g}_{[T\pi]}^{sb}(\beta_0)=\frac{1}{T}\sum_{t=1}^{[T\pi]}g_t^s(\beta_0)$, $\bar{g}_T^{sb}(\theta,\pi)=\frac{1}{T}\sum_{t=1}^{T}g_t^{sb}(\theta,\pi)$ and $\bar{V}_T^b(\theta,\pi)=\frac{1}{T}\sum_{t=1}^{T}g_t^{sb}(\theta,\pi)\,g_t^{sb}(\theta,\pi)'$. This leads to the PSEL estimators

$$\hat{\theta}^b(\pi)=\arg\min_{\theta\in\Theta}\sup_{\gamma\in\Gamma_T}\frac{1}{T}\sum_{t=1}^{T}\ln\left(1+k\gamma' g_t^{sb}(\theta,\pi)\right)\tag{19}$$

and

$$\hat{\gamma}^{b}(\pi) = \arg \sup_{\gamma\in\Gamma_T} \frac{1}{T}\sum_{t=1}^{T} \ln\left(1 + k\gamma' g_t^{sb}(\hat{\theta}(\pi),\pi)\right). \tag{20}$$

Theorem 3 *Under the assumptions of Theorem 2 with $\hat{\theta}^{b}(\pi)$ and $\hat{\gamma}^{b}(\pi)$ defined by (19) and (20), respectively,*

$$\begin{aligned}
\sup_{\pi\in\Pi}\left\|\sqrt{T}\left(\hat{\theta}^{a}(\pi) - \hat{\theta}^{b}(\pi)\right)\right\| &= o_p(1)\\
\sup_{\pi\in\Pi}\left\|(\sqrt{T}/h_T)\left(\hat{\gamma}^{a}(\pi) - \hat{\gamma}^{b}(\pi)\right)\right\| &= o_p(1)
\end{aligned}$$

The next Theorem details the asymptotic distribution of the restricted PSEL estimators, whether or not the moment functions are smoothed after of before the sample split defined by π. These restricted PSEL estimators are constructed as follows. Define the restricted $(2\ell \times 1)$ smoothed moment function as

$$\dot{g}_t^{s}(\beta,\pi) = \mathbb{I}_{t,T}(\pi)\begin{pmatrix} g_t^{s}(\beta)\\ 0\end{pmatrix} + (1 - \mathbb{I}_{t,T}(\pi))\begin{pmatrix} 0\\ g_t^{s}(\beta)\end{pmatrix},$$

where smoothing can occur after or before the sample split and let $\dot{Q}_T(\beta,\gamma,\pi) = \frac{1}{T}\sum_{t=1}^{T}\ln(1 + k\gamma'\dot{g}_t^{s}(\beta,\pi))$, then

$$\begin{aligned}
\tilde{\beta}(\pi) &= \arg\min_{\beta\in\mathcal{B}} \sup_{\gamma\in\Gamma_T} \dot{Q}_T(\beta,\gamma,\pi)\\
&= \arg\min_{\beta\in\mathcal{B}}\left\{\sup_{\lambda\in\Lambda_T}\frac{1}{T}\sum_{t=1}^{[T\pi]}\ln\left(1 + k\lambda' g_t^{s}(\beta)\right) + \sup_{\lambda\in\Lambda_T}\frac{1}{T}\sum_{t=[T\pi]+1}^{T}\ln\left(1 + k\lambda' g_t^{s}(\beta)\right)\right\}
\end{aligned}$$

and

$$\tilde{\gamma}(\pi) = \arg \sup_{\gamma\in\Gamma_T}\frac{1}{T}\sum_{t=1}^{T}\ln\left(1 + k\gamma'\dot{g}_t^{s}(\tilde{\beta},\pi)\right)$$

so that

$$\begin{aligned}
\tilde{\lambda}_1(\pi) &= \arg \sup_{\lambda\in\Lambda_T}\frac{1}{T}\sum_{t=1}^{[T\pi]}\ln\left(1 + k\lambda' g_t^{s}(\tilde{\beta}(\pi))\right)\\
\tilde{\lambda}_2(\pi) &= \arg \sup_{\lambda\in\Lambda_T}\frac{1}{T}\sum_{t=[T\pi]+1}^{T}\ln\left(1 + k\lambda' g_t^{s}(\tilde{\beta}(\pi))\right).
\end{aligned}$$

Theorem 4 *Under the assumptions of Theorem 2 with $\tilde{\beta}(\pi)$ and $\tilde{\gamma}(\pi)$ equal to either the after or before sample split PSEL estimators,*

$$\begin{aligned}
\sqrt{T}\left(\tilde{\beta}(\pi) - \beta_0\right) &= -\left(M_0'M_0\right)^{-1} M_0'\left\{\Omega_0^{-1/2}\sqrt{T}\bar{g}_T(\beta_0)\right\} + o_{p\pi}(1)\\
&\Longrightarrow -\left(M_0'M_0\right)^{-1} M_0' B_\ell(1)
\end{aligned}$$

and

$$
\begin{aligned}
\left(\sqrt{T}/h_T\right)\tilde{\gamma}(\pi) &= \left(A(\pi)^{-1} - \iota_2\iota_2' \otimes \Omega_0^{-1/2}\left(I_\ell - P_0\right)\right)\xi_T(\pi) + o_{p\pi}(1) \\
&= \frac{1}{\pi(1-\pi)}\left(a(\pi) \otimes \Omega_0^{-1/2}\right)\left(I_\ell - P_0\right)\left(a(\pi)' \otimes I_\ell\right)\xi_T(\pi) + o_{p\pi}(1) \\
&\Longrightarrow \left(A(\pi)^{-1} - \iota_2\iota_2' \otimes \Omega_0^{-1/2}\left(I_\ell - P_0\right)\right)J_\ell(\pi) \\
&= \left(A(\pi)^{-1} \otimes \Omega_0^{-1/2}\right)J_\ell(\pi) - \left(\iota_2 \otimes \Omega_0^{-1/2}P_0\right)B_\ell(1)
\end{aligned}
$$

where $a(\pi)' = (1-\pi, \ -\pi)$.

4 Testing Structural Stability

In this section, we propose tests based on EL for testing the hypotheses described in Section 2. It turns out to be most convenient to present the tests in the following order: Section 4.1 presents tests for $\mathcal{D}_1(\pi) = 0$, Section 4.2 presents tests for that $\mathcal{D}_2(\pi) = 0$, and Section 4.3 presents tests for $\mathcal{D}(\pi) = 0$. Section 4.4 discusses the various tests and includes details of where percentiles of the limiting distributions are tabulated in the literature. In the presentation of the tests, we focus on the unknown break-point case; the fixed break-point case is covered as part of the discussion in Section 4.4.

Before presenting these statistics, we note, again, that in the light of Lemma 6, in the Appendix, and Theorem 3, we shall not, henceforth, distinguish between the use of $g_t^{sa}(\theta,\pi)$ or $g_t^{sb}(\theta,\pi) \equiv g_t^s(\theta,\pi)$ smoothed moment functions and shall simply refer, hereafter, to $g_t^s(\theta,\pi)$, which could be either as the difference does not influence the first order asymptotic analysis. Thus, let $Q_T(\theta,\gamma,\pi) = \frac{1}{T}\sum_{t=1}^T \ln\left(1 + k\gamma' g_t^s(\theta,\pi)\right)$ and $\bar{g}_T^s(\theta,\pi) = \frac{1}{T}\sum_{t=1}^T g_t^s(\theta,\pi)$. Further, define $\bar{V}_T^s(\theta,\pi) = \frac{1}{T}\sum_{t=1}^T g_t^s(\theta,\pi)g_t^s(\theta,\pi)'$, so that (from Smith (2004, Theorem 2.1)) it can be shown that

$$
\sup_{\pi\in\Pi}\left\| h_T \bar{V}_T^s(\hat{\theta}(\pi),\pi) - k_2\Omega_0(\pi)\right\| = o_p(1).
$$

4.1 Testing $\mathcal{D}_1(\pi) = 0$

To test $\mathcal{D}_1(\pi) = 0$ for a fixed π, the obvious statistic is the EL-likelihood ratio statistic

$$
\mathcal{LR}_T(\pi) \ = \ 2\left(k_2/k_1^2\right)\left(T/h_T\right)\left\{\dot{Q}_T\left(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi\right) - Q_T\left(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi\right)\right\}. \tag{21}
$$

In view of extant results in the EL literature on testing parametric restrictions,[8] we also consider inference based on the EL-Wald statistic for $\beta_1 = \beta_2$,

$$\mathcal{W}_T(\pi) \;=\; (k_2/k_1^2)T\;\left(\hat{\beta}_1(\pi) - \hat{\beta}_2(\pi)\right)'\left\{V_T^W(\hat{\theta}(\pi))\right\}^{-1}\left(\hat{\beta}_1(\pi) - \hat{\beta}_2(\pi)\right) \tag{22}$$

and the Lagrange Multiplier statistic, based on $\tilde{\rho}(\pi)$ the Lagrange Multiplier associated with the restriction $\beta_1 = \beta_2$,

$$\mathcal{LM}_T(\pi) \;=\; (k_2/k_1^2)(T/h_T)\tilde{\rho}(\pi)'\left\{V_T^\rho(\tilde{\beta}(\pi))\right\}^{-1}\tilde{\rho}(\pi)/\left(\pi(1-\pi)\right) \tag{23}$$

where

$$\begin{aligned}
V_T^W(\theta) &= \sum_{i=1}^{2}\bar{G}_{T_i}^s(\beta_i)'\left\{\bar{V}_{T_i}^s(\beta_i)\right\}^{-1}\bar{G}_{T_i}^s(\beta_i) \\
\bar{G}_{T_i}^s(\beta) &= \frac{1}{T}\sum_{t\in\mathcal{T}_i(\pi)}\frac{\partial g_t^s(\beta)}{\partial\beta'}, \quad \bar{V}_{T_i}^s(\beta) = \frac{h_T}{T}\sum_{t\in\mathcal{T}_i(\pi)}g_t^s(\beta)g_t^s(\beta)' \\
V_T^\rho(\beta) &= \bar{G}_T^s(\beta)'\left\{\bar{V}_T^s(\beta)\right\}^{-1}\bar{G}_T^s(\beta) \\
\bar{G}_T^s(\beta) &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_t^s(\beta)}{\partial\beta'}, \quad \bar{V}_T^s(\beta) = \frac{h_T}{T}\sum_{t=1}^{T}g_t^s(\beta)g_t^s(\beta)'.
\end{aligned}$$

Thus, from here on we use $\hat{\mathcal{D}}_{1,T}(\pi)$ to denote any one of $\mathcal{W}_T(\pi)$, $\mathcal{LM}_T(\pi)$ or $\mathcal{LR}_T(\pi)$.[9]

To test $D_1(\pi) = 0$ for all $\pi \in \Pi \in (0,1)$, we utilize results from the structural stability testing literature and consider inference based on the following functionals of $\hat{\mathcal{D}}_{1,T}(\pi)$,

$$\tau\left[\hat{\mathcal{D}}_{1,T}(\pi)\right] = \begin{cases} \sup_{\pi\in\Pi}\hat{\mathcal{D}}_{1,T}(\pi) \equiv \sup\hat{\mathcal{D}}_{1,T}(\pi) \\ \int_\Pi \hat{\mathcal{D}}_{1,T}(\pi)dN(\pi) \equiv \text{ave}\;\hat{\mathcal{D}}_{1,T}(\pi) \\ \log\left\{\int_\Pi \exp\left\{\frac{1}{2}\hat{\mathcal{D}}_{1,T}(\pi)\right\}dN(\pi)\right\} \equiv \exp\;\hat{\mathcal{D}}_{1,T}(\pi) \end{cases} \tag{24}$$

where $N(\pi)$ defines the prior distribution for the break-point $\pi \in \Pi$, which we will assume to be uniform.[10] The following Theorem shows each of these test statistics are (first order) asymptotically equivalent, for different choices of $\hat{\mathcal{D}}_{1,T}(\pi)$ and common choice of functional $\tau\,[.]$.

Theorem 5 *Under Assumptions 1-5, we have*

$$\sup_{\pi\in\Pi}\left|\hat{\mathcal{D}}_{1,T}(\pi) - \mathcal{S}_T(\pi)\right| = o_p(1),$$

[8] See Qin and Lawless (1994), Smith (2004).

[9] This involves a slight abuse of notation compared to Section 2 because the distances here are scaled.

[10] See Andrews (1993), Andrews and Ploberger (1994) and Sowell (1996).

where

$$\begin{aligned}\mathcal{S}_T(\pi) &= \frac{\xi_T(\pi)'\left(a(\pi)\otimes I_\ell\right)' P_0\left(a(\pi)\otimes I_\ell\right)\xi_T(\pi)}{\pi(1-\pi)} \\ &\Longrightarrow \frac{\left(B_k(\pi)-\pi B_k(1)\right)'\left(B_k(\pi)-\pi B_k(1)\right)}{\pi(1-\pi)} \equiv W_k(\pi),\end{aligned}$$

$B_k(\pi)-\pi B_k(1)$ is a vector of Brownian bridges and $B_k(\pi)$ is a vector of k independent standard Brownian motions, and for each functional (24)

$$\tau\left[\hat{\mathcal{D}}_{1,T}(\pi)\right] \Longrightarrow \tau\left[W_k(\pi)\right].$$

4.2 Testing $\mathcal{D}_2(\pi)$

To test $D_2(\pi)=0$, we consider inference based on the appropriate EL-likelihood ratio statistic

$$\mathcal{LR}_T^*(\pi) \;=\; 2\left(k_2/k_1^2\right)\left(T/h_T\right) Q_T\left(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi\right). \tag{25}$$

Again, motivated by results in the EL testing literature, we also consider inference based on the following alternative statistics,

$$\mathcal{O}_T(\pi) = (k_2/k_1^2)\left(T/h_T\right)\bar{g}_T^s(\hat{\theta}(\pi),\pi)'\left\{\bar{V}_T^s(\hat{\theta}(\pi),\pi)\right\}^{-1}\bar{g}_T^s(\hat{\theta}(\pi),\pi)' \tag{26}$$

$$\mathcal{LM}_T^*(\pi) = (T/h_T)\hat{\gamma}(\pi)'\left\{\bar{V}_T^s(\hat{\theta}(\pi),\pi)\right\}\hat{\gamma}(\pi)/k_2. \tag{27}$$

For a fixed π, $\mathcal{O}_T(\pi)$ is the EL counterpart of the GMM overidentifying test statistic; $\mathcal{LM}_T^*(\pi)$ is a Lagrange Multiplier statistic, based on $\hat{\gamma}(\pi)$; and, $\mathcal{LR}_T^*(\pi)$ is a Likelihood Ratio type statistic.

Letting $\hat{\mathcal{D}}_{2,T}(\pi)$ denote any one of $\mathcal{O}_T(\pi)$, $\mathcal{LM}_T^*(\pi)$ or $\mathcal{LR}_T^*(\pi)$,[11] we use similar ideas to the previous sub-section to test $\mathcal{D}_2(\pi)$ for all $\pi\in\Pi$ based on $\tau\left[\hat{\mathcal{D}}_{2,T}(\pi)\right]$. The limiting distribution of the latter statistic is given in the following theorem.

Theorem 6 *Under Assumptions 1-5, we have*

$$\sup_{\pi\in\Pi}\left|\hat{\mathcal{D}}_{2,T}(\pi)-\mathcal{S}_T^*(\pi)\right| = o_p(1),$$

where

$$\begin{aligned}\mathcal{S}_T^*(\pi) &= \xi_T(\pi)'\left(A(\pi)^{-1}\otimes(I_\ell-P_0)\right)\xi_T(\pi) \\ &\Longrightarrow J_{\ell-k}(\pi)'\left(A(\pi)\otimes I_{\ell-k}\right)^{-1} J_{\ell-k}(\pi) \equiv W_{\ell-k}^*(\pi)\end{aligned}$$

[11] Again, this involves a slight abuse of notation compared to Section 2 because the distances here are scaled.

and $J_{\ell-k}(\pi) = \begin{bmatrix} B_{\ell-k}(\pi) \\ B_{\ell-k}(1) - B_{\ell-k}(\pi) \end{bmatrix}$*, where* $B_{\ell-k}(\pi)$ *is a vector of* $\ell - k$ *independent standard Brownian motions, and for each functional in (24)*

$$\tau\left[\hat{\mathcal{D}}_{2,T}(\pi)\right] \Longrightarrow \tau\left[W^*_{\ell-k}(\pi)\right].$$

4.3 Testing $\mathcal{D}(\pi) = 0$

Given the discussion in Section 2, testing $\mathcal{D}(\pi) = 0$ can be achieved by employing statistics which are functionals of the processes, $\hat{\mathcal{D}}_{1,T}(\pi)$ and $\hat{\mathcal{D}}_{2,T}(\pi)$. Specifically, we consider the combined process $\hat{\mathcal{D}}_T(\pi) = \hat{\mathcal{D}}_{1,T}(\pi) + \hat{\mathcal{D}}_{2,T}(\pi)$ for any of the choices of $\hat{\mathcal{D}}_{1,T}(\pi)$ and $\hat{\mathcal{D}}_{2,T}(\pi)$ defined in Sections 4.1 and 4.2 respectively, and the functionals $\tau\left[\hat{\mathcal{D}}_T(\pi)\right]$ for any $\tau[.]$ defined in (24). Then, we have the following Corollary to Theorems 5 and 6:

Corollary 1 *Under Assumptions 1-5, we have*

$$\sup_{\pi\in\Pi}\left|\hat{\mathcal{D}}_T(\pi) - \mathcal{S}_T(\pi) - \mathcal{S}^*_T(\pi)\right| = o_p(1),$$

and for each functional in (24)

$$\tau\left[\hat{\mathcal{D}}_T(\pi)\right] \Longrightarrow \tau\left[W_k(\pi) + W^*_{\ell-k}(\pi)\right].$$

4.4 Discussion

Sections 4.1-4.3 present tests of the hypotheses of interest in the unknown break-point case. The corresponding results for the fixed break-point case follows directly from the proofs of Theorems 5 and 6 and so are presented in the following corollary.

Corollary 2 *Under Assumptions 1-5, and if* $H_0(\pi)$ *holds for some* $\pi \in (0,1)$ *then* $\hat{\mathcal{D}}_{1,T}(\pi) \xrightarrow{d} \chi^2_k$*,* $\hat{\mathcal{D}}_{2,T}(\pi) \xrightarrow{d} \chi^2_{2(\ell-k)}$*, and* $\hat{\mathcal{D}}_T(\pi) \xrightarrow{d} \chi^2_{2\ell-k}$*, where* $\hat{\mathcal{D}}_{1,T}(\pi)$*,* $\hat{\mathcal{D}}_{2,T}(\pi)$ *and* $\hat{\mathcal{D}}_T(\pi)$ *are defined in Sections 4.1, 4.2 and 4.3 respectively and* χ^2_ν *denotes a chi-squared distribution with* ν *degrees of freedom.*

We now consider the relationship between our statistics and others in the literature. As noted in the introduction, Guay and Lamarche (2010) derive some of our test statistics from the perspective of testing the stability of the identifying and overidentifying restrictions, a terminology that derives from Hall and Sen's (1999) framework for testing structural instability in models estimated via GMM. Comparing Guay and Lamarche's (2010) framework specialized to EL with our info-metric framework, it can be seen that their tests of the stability of the identifying restrictions are the same as our tests of $\mathcal{D}_1(\pi) = 0$, and their tests of the stability of the overidentifying restrictions are the same as our tests of $\mathcal{D}_2(\pi) = 0$.[12] While the same tests result, the info-metric approach has the advantage that it is based on the concept of minimizing the distance between the class of probability distributions restricted to satisfy the moment condition and the true probability distribution. This allows us to relate the various hypotheses of interest in structural instability testing to the distance between certain classes of probability distributions and the true distribution. We believe this is a more fundamental - and also more instructive - representation of these hypotheses than their expression in terms of identifying restrictions (parameter variation) and overidentifying restrictions as is done in both the GMM and GEL frameworks.

Guay and Lamarche (2010) observe that their GEL-based tests are first order asymptotically equivalent to their GMM counterparts under both the null of stability and local alternatives.[13] Given our previous remarks, this equivalence obviously extends to our statistics as well. One advantage of this equivalence is that the percentiles for the limiting distributions of our statistics have already been tabulated in the literature. Specifically, percentiles of $\tau[W_k(\pi)]$ are presented in Andrews (2003)[Table 1] (for $\tau[\cdot] = sup(\cdot)$) and Andrews and Ploberger (1994)[Tables 1 and 2] (for $\tau[\cdot] = ave(\cdot)$, $exp(\cdot)$); the percentiles for $\tau[W^*_{\ell-k}(\pi)]$ are presented in Hall and Sen (1999)[Table 1] and Sen (1997). Percentiles for $\tau[W_k(\pi) + W^*_{\ell-k}(\pi)]]$ are reported in Sen (1997). A second advantage of the equivalence under local alternatives is that Theorem 5 continues to hold under local alternatives to the moment condition that do not involve parameter variation, and Theorem 6 continues to hold for local alternatives to the moment condition that involve parameter variation alone. These properties suggest that the individual applications of tests based on $\hat{\mathcal{D}}_{1,T}(\pi)$ and $\hat{\mathcal{D}}_{2,T}(\pi)$ have the potential to reveal when the instability is confined to parameter variation alone.

[12] Guay and Lamarche (2010) do not consider the analog to $D(\pi) = 0$ in their framework. However, Sen (1997) does propose and analyze such a test within the GMM framework.

[13] Li (2011) establishes the same result for EL-based test statistics.

5 Monte Carlo Evidence

In this section, we report results from a simulation study that gives insights into the finite sample performance of the EL-based tests.

Following Ghysels, Guay, and Hall (1997) and Hall and Sen (1999), we consider the following data generation process

$$x_t = \beta_1 x_{t-1} + u_t + \alpha u_{t-2}, \qquad u_t \sim IN(0,1), \qquad \text{for } t = 1,2,\ldots,T/2$$
$$x_t = \beta_2 x_{t-1} + u_t + \alpha u_{t-2}, \qquad u_t \sim IN(0,1), \qquad \text{for } t = T/2+1, T/2+2,\ldots,T.$$

We suppose that the researcher estimates an AR(1) model for x_t, with AR parameter β, based on the moment condition $E[g_t(\beta)] = 0$ where

$$g_t(\beta) = \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} (x_t - \beta x_{t-1}).$$

We consider three choices of sample size: $T = 200, 400, 600$. On each replication we calculate $\tau[\hat{\mathcal{D}}_{1,T}(\pi)]$, $\tau[\hat{\mathcal{D}}_{2,T}(\pi)]$ and $\tau[\hat{\mathcal{D}}_T(\pi)]$ for the three versions of $\tau[\cdot]$ defined in (24) and: $\hat{\mathcal{D}}_{1,T}(\pi) = \mathcal{W}_T(\pi)$, $\mathcal{LM}_T(\pi)$ or $\mathcal{LR}_T(\pi)$; $\hat{\mathcal{D}}_{2,T}(\pi) = \mathcal{O}_T(\pi)$, $\mathcal{LM}_T^*(\pi)$ or $\mathcal{LR}_T^*(\pi)$; $\hat{\mathcal{D}}_T(\pi) = \mathcal{O}_T(\pi) + \mathcal{W}_T(\pi)$, $\mathcal{LM}_T(\pi) + \mathcal{LM}_T^*(\pi)$, $\mathcal{LR}_T(\pi) + \mathcal{LR}_T^*(\pi)$. All these statistics are calculated using $\Pi = [\varepsilon, 1-\varepsilon]$ for the following choices of trimming parameter, $\varepsilon = 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45$.

We consider two versions of each statistic one based on the unsmoothed moment condition and one based on the smoothed moment condition. We report both because if $H_0(\pi)$ holds in this model then $g_t(\beta)$ is a martingale difference sequence and so smoothing is actually unnecessary. Such examples provide evidence on the potential impact of smoothing on finite sample performance. In all cases, smoothing is applied before the sample split and performed using quadratic spectral kernels with,

$$k_T(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

where $x = j/h_T$, $\hat{h}_T = 1.3221[\hat{\alpha}(2)T]^{1/5}$ and

$$\hat{\alpha}(2) = \sum_{a=1}^{p} w_a \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8}\left\{\sum_{a=1}^{p} w_a \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4}\right\}^{-1}$$

where $\hat{\rho}_a$, $\hat{\sigma}_a^2$ are estimated AR(1) coefficients and error variances based on moment functions $g_t(\hat{\beta})$ ($p \times 1$; $a = 1,2,\ldots,p$), respectively.[14] For ease of exposition, we refer to the tests based

[14] This choice corresponds to the optimal bandwidth based on an AR(1) approximation to the moment function with $w_a = 1$; see Andrews (1991)[p.834-5] with $w_a = 1$ in his equation (6.4).

on the unsmoothed (smoothed) moment conditions as unsmoothed (smoothed) tests.

We consider the size properties of the tests using the following two choices of parameters for which $H_0(\pi)$ holds: $(\beta_1, \beta_2, \alpha) = (0.4, 0.4, 0)$ and $(0.8, 0.8, 0)$ - termed DGP1 and DGP2, respectively. Thus the null hypotheses of all tests considered holds for choices of parameter values. In all cases, results are based on 1000 replications.

At sample size $T = 200$, all of the tests are oversized for at least one choice of ϵ. Therefore, we omit these results here.[15] Tables 1-3 report the rejection frequencies the tests of $D_1(\pi)$, $D_2(\pi)$ and $D(\pi)$ respectively at $T = 400$, Tables 4-6 report the the empirical size of the tests of $D_1(\pi)$, $D_2(\pi)$ and $D(\pi)$ respectively at $T = 600$. In each case the nominal size of the test is 0.05. From Tables 1-3, it can be seen that for $T = 400$ the unsmoothed tests exhibit relative rejection frequencies that approximately equal the nominal sizes for all values of the trimming parameter, but the smoothed tests are oversized with rejection frequencies approximately equal to the nominal size only for trimming parameters $\varepsilon \geq 0.35$ or 0.4. From Tables 4-6, it can be seen that the asymptotic approximation is far better at $T = 600$ with the smoothed tests exhibiting rejection frequencies approximately equal to the nominal size for all $\varepsilon \geq 0.2$ or 0.25. Behind this broad summary, there are some variations in performance of various tests. For example, the $Ave-$ functional tends to yield statistics whose tail behaviour is better approximated by the asymptotic theory than either $Exp-$ or the $Sup-$ statistics. In future work, we plan to explore the power of the tests.

6 Concluding remarks

In this paper, we develop an info-metric framework for testing hypotheses about structural instability in nonlinear, dynamic models estimated from the information in population moment conditions. Our methods are designed to distinguish between three states of the world: (i) the model is structurally stable in the sense that the population moment condition holds at the same parameter value throughout the sample; (ii) the model parameters change at some point in the sample but otherwise the model is correctly specified; (iii) the model exhibits more general forms of instability than a single shift in the parameters. An advantage of the info-metric approach is that the null hypotheses concerned are formulated in terms of distances between various choices

[15] These results are available from the authors upon request.

of probability measures constrained to satisfy (i) and (ii) and the empirical measure of the sample. Under the alternative hypotheses considered, the model is assumed to exhibit structural instability at a single point in the sample, referred to as the break-point; our analysis allows for the break-point to be either fixed *a priori* or treated as occuring at some unknown point within a certain fraction of the sample. We propose various test statistics that can be thought of as sample analogs of the distances described above, and derive their limiting distributions under the appropriate null hypothesis. In principle, there are a number of possible measures of distance that can be used in this context but we focus on the measure associated with Empirical Likelihood estimation. The limiting distributions of our statistics are non-standard but coincide with various distributions that arise in the literature on structural instability testing within the Generalized Method of Moments framework. A small simulation study illustrates the finite sample performance of our test statistics under the null hpothesis.

7 Appendix

Here we collect together some intermediate Lemmas and prove the main Theorems. Following Andrews (1993), we use the following notation: $X_T(\pi) = o_{p\pi}(1)$ if $\sup_{\pi\in\Pi} \|X_T(\pi)\| = o_p(1)$ and $X_T(\pi) = O_{p\pi}(1)$ if $\sup_{\pi\in\Pi} \|X_T(\pi)\| = O_p(1)$.

The first result is a FCLT and second a generic (weak) ULLN.

Lemma 1 *Under Assumptions 1-3(i)(ii)*

$$\begin{aligned} k_1^{-1}\Omega_0^{-1/2}\sqrt{T}\bar{g}^{sa}_{[T\pi]}(\beta_0) &= \Omega_0^{-1/2}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) + o_{p\pi}(1) \\ &\Longrightarrow B_\ell(\pi) \end{aligned}$$

where $B(\pi)$ is a vector of k mutually independent standard Brownian motions on $[0,1]$, and

$$\begin{aligned} k_1^{-1}\left(I_2\otimes\Omega_0^{-1/2}\right)\sqrt{T}\bar{g}^{sa}_T(\theta_0,\pi) &= \left(I_2\otimes\Omega_0^{-1/2}\right)\sqrt{T}\bar{g}_T(\theta_0,\pi) + o_p(1) \\ &\Longrightarrow J_\ell(\pi) = \begin{bmatrix} B_\ell(\pi) \\ (B_\ell(1) - B_\ell(\pi)) \end{bmatrix}. \end{aligned}$$

Proof of Lemma 1: Firstly, by Andrews (1993), $\Omega_0^{-1/2}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) \Longrightarrow B(\pi)$. Second, and following Smith (2004, Lemma A2), we can write

$$\sqrt{T}\bar{g}^{sa}_{[T\pi]}(\beta_0) = \frac{1}{h_T}\sum_{j=1-[T\pi]}^{[T\pi]-1} k\left(\frac{j}{h_T}\right)\left\{\frac{1}{\sqrt{T}}\sum_{t=\max[1,1-j]}^{\min[[T\pi],[T\pi]-j]} g_t(\beta_0)\right\}.$$

Define $A_{[T\pi]}(j) = \{t : t \notin [\max[1, 1-j], \min[[T\pi], [T\pi]-j]]\}$. Then

$$\frac{1}{\sqrt{T}}\sum_{t=\max[1,1-j]}^{\min[[T\pi],[T\pi]-j]} g_t(\beta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[T\pi]} g_t(\beta_0) - u_{jT}(\pi),$$

where $u_{jT}(\pi) = \frac{1}{\sqrt{T}}\sum_{t=1}^{|j|} g_t(\beta_0)$, for $j < 0$, and $u_{jT}(\pi) = \frac{1}{\sqrt{T}}\sum_{t=[T\pi]-j+1}^{[T\pi]} g_t(\beta_0)$, for $j \geq 0$. Thus, in both cases, $u_{Tj}(\pi)$ consists of $|j|$ terms, uniformly in π, and so $u_{jT}(\pi) = \sqrt{\frac{|j|}{T}}O_{p\pi}(1)$, where the $O_{p\pi}(1)$ term is independent of j. This enables us to write

$$\begin{aligned}\sqrt{T}\bar{g}^{sa}_{[T\pi]}(\beta_0) &= \frac{1}{h_T}\sum_{j=1-[T\pi]}^{[T\pi]-1} k\left(\frac{j}{h_T}\right)\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) + \left\{\frac{1}{h_T}\sum_{j=1-[T\pi]}^{[T\pi]-1}\sqrt{\frac{|j|}{T}}k\left(\frac{j}{h_T}\right)\right\}O_{p\pi}(1)\\ &= \frac{1}{h_T}\sum_{j=1-T}^{T-1} k\left(\frac{j}{h_T}\right)\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) + e_T(\pi)\\ &= k_1\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) + o_{p\pi}(1)\end{aligned}$$

where we have used $\lim_{T\to\infty}\frac{1}{h_T}\sum_{j=1-T}^{T-1} k\left(\frac{j}{h_T}\right) = k_1$ (see, for example, Smith (2004, proof of Lemma A1) and $\sup_{\pi\in\Pi}\|e_T(\pi)\| = o_p(1)$). To see the latter, by repeated use of the triangle inequality we have

$$\begin{aligned}\|e_T(\pi)\| \leq\ & \left\{\left|\frac{1}{h_T}\sum_{j=1-T}^{-[T\pi]} k\left(\frac{j}{h_T}\right)\right| + \left|\frac{1}{h_T}\sum_{j=[T\pi]}^{T-1} k\left(\frac{j}{h_T}\right)\right|\right\}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\pi]} g_t(\beta_0)\right\|\\ &+\left\{\frac{1}{h_T}\sum_{j=1-T}^{T-1}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right|\right\}O_{p\pi}(1)\end{aligned}$$

where the second line follows from

$$\left|\frac{1}{h_T}\sum_{j=1-[T\pi]}^{[T\pi]-1}\sqrt{\frac{|j|}{T}}k\left(\frac{j}{h_T}\right)\right| \leq \frac{1}{h_T}\sum_{j=1-[T\pi]}^{[T\pi]-1}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right| \leq \frac{1}{h_T}\sum_{j=1-T}^{T-1}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right|.$$

Since, $\frac{1}{h_T}\sum_{j=1-T}^{T-1} k\left(\frac{j}{h_T}\right) = k_1 + o(1)$, both $\left|\frac{1}{h_T}\sum_{j=1-T}^{-[T\pi]} k\left(\frac{j}{h_T}\right)\right|$ and $\left|\frac{1}{h_T}\sum_{j=[T\pi]}^{T-1} k\left(\frac{j}{h_T}\right)\right|$ are $o(1)$, uniformly in π, whilst $\sup_{\pi\in\Pi}\left\|\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right\| = O_p(1)$, and Smith (2004, Lemma C1) can easily be extended to show that $\lim_{T\to\infty}\frac{1}{h_T}\sum_{t=1-T}^{T-1}\left\{\frac{|t|}{T}\right\}^r\left|k\left(\frac{t}{h_T}\right)\right| = 0$, for all $r > 0$. Therefore, $e_T(\pi) = o_{p\pi}(1)$. Thus

$$k_1^{-1}\Omega_0^{-1/2}\sqrt{T}\bar{g}^{sa}_{[T\pi]}(\beta_0) = \Omega_0^{-1/2}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) + o_{p\pi}(1).$$

Similar analysis shows that

$$\begin{aligned}
k_1^{-1}\Omega_0^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=[T\pi]+1}^{T} g_t^{sa}(\beta_0) &= \Omega_0^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=[T\pi]+1}^{T} g_t\left(\beta_0\right)+o_{p\pi}(1) \\
&= \Omega_0^{-1/2}\left(\sqrt{T}\bar{g}_T(\beta_0)-\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right)+o_{p\pi}(1)
\end{aligned}$$

so that

$$k_1^{-1}\left(I_2\otimes\Omega_0^{-1/2}\right)\sqrt{T}\bar{g}_T^{sa}(\theta_0,\pi)=\left(I_2\otimes\Omega_0^{-1/2}\right)\sqrt{T}\bar{g}_T(\theta_0,\pi)+o_p(1),$$

since $\bar{g}_T(\theta_0,\pi)=\left(\bar{g}_{[T\pi]}(\beta_0)',\ \bar{g}_T(\beta_0)'-\bar{g}_{[T\pi]}(\beta_0)'\right)'$, and the result follows. ■

Lemma 2 *Define $m_t(\beta)=m(Z_t;\beta)$ and $m(\beta)=E[m(Z_t;\beta)]$, with Z_t satisfying Assumption 1 and assume sufficient regularity (Assumptions 3 (i) and (iii)) so that $\sup_{\beta\in\mathcal{B}}\|\bar{m}_T(\beta)-m(\beta)\|=o_p(1)$, where $\bar{m}_T(\beta)=\frac{1}{T}\sum_{t=1}^{T}m_t(\beta)$. Let $m_t^{sa}(\beta)$ be the smoothed version of $m_t(\beta)$, defined in an analogous manner to $g_t^{sa}(\beta)$ at (14), and (following (15)), define*

$$\begin{aligned}
m_t^{sa}(\theta,\pi) &= \mathbb{I}_{t,T}(\pi)\begin{pmatrix} m_t^{sa}(\beta_1) \\ 0 \end{pmatrix}+\left(1-\mathbb{I}_{t,T}(\pi)\right)\begin{pmatrix} 0 \\ m_t^{sa}(\beta_2) \end{pmatrix} \\
\bar{m}_T^{sa}(\theta,\pi) &= \frac{1}{T}\sum_{t=1}^{T}m_t^{sa}(\theta,\pi)
\end{aligned}$$

with $m(\theta,\pi)=\left(\pi m(\beta_1)',(1-\pi)\,m(\beta_2)'\right)'$. Then, $\sup_{\pi\in\Pi}\sup_{\theta\in\Theta}\|\bar{m}_T^{sa}(\theta,\pi)-k_1 m(\theta,\pi)\|=o_p(1)$.

Proof of Lemma 2: We can write

$$\bar{m}_T^{sa}(\theta,\pi)-k_1 m(\theta,\pi)=\begin{pmatrix} \left\{\frac{1}{T}\sum_{t=1}^{[T\pi]}m_t^{sa}(\beta_1)\right\}-k_1\pi m(\beta_1) \\ \left\{\frac{1}{T}\sum_{t=[T\pi]+1}^{T}m_t^{sa}(\beta_2)\right\}-k_1(1-\pi)m(\beta_2) \end{pmatrix}$$

In particular, and by the triangle inequality with $\bar{m}_{[T\pi]}^{sa}(\beta)=\frac{1}{T}\sum_{t=1}^{[T\pi]}m_t^{sa}(\beta)$,

$$\begin{aligned}
\left\|\bar{m}_{[T\pi]}^{sa}(\beta)-k_1\pi m(\beta)\right\| &\leq \left\|\bar{m}_{[T\pi]}^{sa}(\beta)-k_1\bar{m}_{[T\pi]}(\beta)\right\|+k_1\left\|\bar{m}_{[T\pi]}(\beta)-\pi m(\beta)\right\| \\
&\leq \left\|\bar{m}_{[T\pi]}^{sa}(\beta)-\sum_{j=1-T}^{T-1}\frac{1}{h_T}k\left(\frac{j}{h_T}\right)\bar{m}_{[T\pi]}(\beta)\right\| \\
&\quad +\left|\sum_{j=1-T}^{T-1}\frac{1}{h_T}k\left(\frac{j}{h_T}\right)-k_1\right|\left\|\bar{m}_{[T\pi]}(\beta)\right\| \\
&\quad +k_1\left\|\bar{m}_{[T\pi]}(\beta)-\pi m(\beta)\right\|.
\end{aligned}$$

By Andrews (1993, Proof of Lemma A1), $\sup_{\pi\in\Pi}\sup_{\beta}\left\|\bar{m}_{[T\pi]}(\beta)-\pi m(\beta)\right\|=o_p(1)$ and since $\sum_{j=1-T}^{T-1}\frac{1}{h_T}k\left(\frac{j}{h_T}\right)=k_1+o(1)$, the second term is also $o_{p\pi}(1)$. Finally, and following the strategy employed in the proof of the Lemma 1, write

$$\bar{m}_{[T\pi]}^{sa}(\beta)=\frac{1}{h_T}\sum_{j=1-[T\pi]}^{[T\pi]-1}k\left(\frac{j}{h_T}\right)\left\{\bar{m}_{[T\pi]}(\beta)-u_{jT}(\beta,\pi)\right\}$$

where $u_{jT}(\beta,\pi)=\frac{1}{T}\sum_{t=1}^{|j|}g_t(\beta)$, for $j<0$, and $u_{jT}(\beta,\pi)=\frac{1}{T}\sum_{t=[T\pi]-j+1}^{[T\pi]}g_t(\beta)$, for $j\geq 0$, and it is then straightforward to show that

$$\sup_{\pi}\sup_{\beta\in\mathcal{B}}\left\|\bar{m}_{[T\pi]}^{sa}(\beta)-\sum_{j=1-T}^{T-1}\frac{1}{h_T}k\left(\frac{j}{h_T}\right)\bar{m}_{[T\pi]}(\beta)\right\|=o_p(1).$$

Similarly, it can be shown that $\sup_{\pi}\sup_{\beta\in\mathcal{B}}\left\|\left(\frac{1}{T}\sum_{t=[T\pi]+1}^{T}m_t^{sa}(\beta)\right)-k_1(1-\pi)\,m(\beta)\right\|=o_p(1)$, and the result follows. ■

The following three Lemmas are used to establish consistency of $\hat{\theta}(\pi)$ and $\hat{\gamma}(\pi)$.

Lemma 3 *Under Assumptions 1, 2(i), 3(i) and 4*

$$\sup_{\theta\in\Theta,\gamma\in\Gamma_T,1\leq t\leq T}\left|\gamma' g_t^{sa}(\theta,\pi)\right|=o_{p\pi}(1).$$

Proof of Lemma 3: By Cauchy-Schwartz,

$$\begin{aligned}\left|\gamma' g_t^{sa}(\theta,\pi)\right| &\leq \|\gamma\|\,\|g_t^{sa}(\theta,\pi)\| \\ &\leq \Delta\left(T/h_T^2\right)^{-\varepsilon}\max_{1\leq t\leq T}\left\{\sup_{\theta\in\Theta}\|g_t^{sa}(\theta,\pi)\|\right\}.\end{aligned}$$

Now,

$$\begin{aligned}\max_{1\leq t\leq T}\sup_{\theta\in\Theta}\|g_t^{sa}(\theta,\pi)\| &\leq \max_{1\leq t\leq[T\pi]}\sup_{\beta\in\mathcal{B}}\left\|\frac{1}{h_T}\sum_{j=t-[T\pi]}^{t-1}k\left(\frac{j}{h_T}\right)g_{t-j}(\beta)\right\| \\ &\quad+\max_{1\leq t\leq[T\pi]+1}\sup_{\beta\in\mathcal{B}}\left\|\frac{1}{h_T}\sum_{j=t-T}^{t-[T\pi]-1}k\left(\frac{j}{h_T}\right)g_{t-j}(\beta)\right\| \\ &\leq \max_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\|g_t(\beta)\|\left\{\frac{2}{h_T}\sum_{j=1-T}^{T-1}\left|k\left(\frac{j}{h_T}\right)\right|\right\},\end{aligned}$$

where the last inequality is independent of π. By Assumption 3(i), $E\left[\sup_{\beta\in\mathcal{B}}\|g_t(\beta)\|^{\eta}\right]\leq\Delta<\infty$, implying that $\max_{1\leq t\leq T}\left\{\sup_{\beta\in\mathcal{B}}\|g_t(\beta)\|\right\}=o_p\left(T^{1/\eta}\right)$. Furthermore, by previous results, $\frac{1}{h_T}\sum_{j=1-T}^{T-1}\left|k\left(\frac{j}{h_T}\right)\right|=O(1)$. Thus, uniformly in π,

$$\begin{aligned}\sup_{\theta\in\Theta,\gamma\in\Gamma_T,1\leq t\leq T}\left|\gamma' g_t^{sa}(\theta,\pi)\right| &\leq O(1)\left(T/h_T^2\right)^{-\varepsilon}o_p\left(T^{1/\eta}\right) \\ &= o_p\left(T^{\alpha}\right)=o_p(1)\end{aligned}$$

where $\alpha = \delta - \varepsilon\eta(\delta - 1) < 0$, because $\varepsilon > \frac{\delta}{\eta(\delta-1)}$. ■

The above result has the following implications, which will be of use later, as summarised in the following Lemma.

Lemma 4 *Under Assumptions 1-4, there exists a finite constants* $0 < \Delta < \infty$*, such that* $w.p.a.1$ *and for all* $\theta \in \Theta$ *and* $\gamma \in \Gamma_T$*, and for each* $\pi \in \Pi$,

$$h_T^{-1} Q_T^a(\theta_0, \gamma, \pi) \leq \gamma_T' \bar{g}_T^{sa}(\theta_0, \pi) - \Delta \gamma_T' \gamma_T \tag{28}$$

where $\gamma_T = k\gamma/h_T$*,* $k = k_1/k_2$ *and*

$$Q_T^a(\theta, \gamma, \pi) \geq k\gamma' \bar{g}_T^{sa}(\theta, \pi) - k^2 \Delta \gamma' \gamma. \tag{29}$$

Proof of Lemma 4: By a second order Taylor expansion, and exploiting Lemma 3, we have that for all $\theta \in \Theta$ and $\gamma \in \Gamma_T$, and each $\pi \in \Pi$

$$Q_T^a(\theta, \gamma, \pi) = k\gamma' \bar{g}_T^{sa}(\theta, \pi) - \frac{1}{2} k^2 \gamma' \bar{V}_T^a(\theta, \pi)\gamma + o_p(1) \tag{30}$$

where the $o_p(1)$ error is of smaller order than $k\gamma' \bar{g}_T^{sa}(\theta, \pi) - \dfrac{1}{2} k^2 \gamma' \bar{V}_T^a(\theta, \pi)\gamma$.
To establish (28), substitute θ_0 for θ in (30) to obtain, $w.p.a.1$,

$$h_T^{-1} Q_T^a(\theta_0, \gamma, \pi) = h_T^{-1} k\gamma_T' \bar{g}_T^{sa}(\theta_0, \pi) - \frac{1}{2} k^2 \gamma_T' h_T \bar{V}_T^a(\theta_0, \pi)\gamma_T$$

where, here, $\gamma_T = k\gamma/h_T \in \Gamma_T$. By arguments similar to Smith (2004, Lemma A3) it can be shown that $h_T \bar{V}_T^a(\theta_0, \pi) \equiv k_2 \Omega_0(\pi) + o_{p\pi}(1)$, we can now write

$$h_T^{-1} Q_T^a(\theta_0, \gamma, \pi) = \gamma_T' \bar{g}_T^{sa}(\theta_0, \pi) - \frac{k_2}{2} \gamma_T' \Omega_0(\pi)\gamma_T + o_p(\|\lambda_T\|^2)$$

where, again, the error term $o_p(\|\lambda_T\|^2)$ is negligible relative to $\gamma_T' \bar{g}_T^{sa}(\theta_0, \pi) - \dfrac{k_2}{2} \gamma_T' \Omega_0(\pi)\gamma_T$.
Thus, from standard eigenvalue theory, we can write that $w.p.a.1$

$$h_T^{-1} Q_T^a(\theta_0, \gamma, \pi) \leq \gamma_T' \bar{g}_T^{sa}(\theta_0, \pi) - \Delta \gamma_T' \gamma_T$$

for all $\gamma \in \Gamma_T$, and for each $\pi \in \Pi$.
More generally, however, $\bar{V}_T^a(\theta, \pi) = O_{p\pi}(1)$, uniformly in θ, so that by similar reasoning, we can write

$$Q_T^a(\theta, \gamma, \pi) \geq k\gamma' \bar{g}_T^{sa}(\theta, \pi) - k^2 \Delta \gamma' \gamma + o_p(\|\gamma\|^2)$$

and (29) follows from this. ■

Lemma 5 *Under Assumptions 1-4, there exists a finite constant, $\Delta > 0$, such that w.p.a.1*

$$h_T^{-1} \sup_{\gamma \in \Gamma_T} Q_T^a(\theta_0, \gamma, \pi) \leq \Delta \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2 = O_{p\pi}\left(T^{-1}\right).$$

Proof of Lemma 5: As in Smith (2004, Lemma A5), by equation (28) we have, $w.p.a.1$ and each $\pi \in \Pi$,

$$\sup_{\gamma \in \Gamma_T} h_T^{-1} Q_T^a(\theta_0, \gamma, \pi) \leq \Delta \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2$$

Since this holds for each $\pi \in \Pi$,

$$\sup_{\pi \in \Pi} \sup_{\gamma \in \Gamma_T} h_T^{-1} Q_T^a(\theta_0, \gamma, \pi) \leq \Delta \sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2.$$

The fact that $\sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi \right\|^2 = O_p\left(T^{-1}\right)$ follows from Lemma 1. ■

Proof of Theorem 1: By Lemma 4, equation (29) and Lemma 5, we have, $w.p.a.1$ and for all $\gamma \in \Gamma_T$ and each $\pi \in \Pi$

$$\begin{aligned} h_T^{-1}\left(k\gamma' \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) - k^2 \Delta \gamma' \gamma\right) &\leq h_T^{-1} Q_T^a(\hat{\theta}^a(\pi), \gamma, \pi) \\ &\leq \sup_{\gamma \in \Gamma_T} h_T^{-1} Q_T^a(\theta_0, \gamma, \pi) \\ &\leq \Delta \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2, \end{aligned}$$

for some finite $\Delta > 0$. Now define $\delta_T = B\left(T/h_T^2\right)^{-\varepsilon} > 0$, with B and ε as in Assumption 4 so that $\delta_T = O\left(T^{\alpha}\right)$, $\alpha = -\frac{\varepsilon(\delta-1)}{\delta} < -\frac{1}{\eta}$, and $\gamma = \frac{1}{k}\delta_T \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) / \left\| \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) \right\| \in \Gamma_T$. Making this substitution in the above yields

$$(\delta_T / h_T) \sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) \right\| - \Delta_2 \delta_T^2 / h_T \leq \Delta \sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2$$

$w.p.a.1$ or,

$$\sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) \right\| \leq \Delta_2 \delta_T + \Delta \frac{h_T}{\delta_T} \sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2 = \Delta_2 \delta_T \left\{ 1 + O(1) \frac{h_T}{\delta_T^2} \sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2 \right\}$$

which implies that $\sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) \right\| = O_p\left(\delta_T\right)$. This follows because $\sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2 = O_p\left(T^{-1}\right)$, so that

$$\frac{h_T}{\delta_T^2} \sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi \right\|^2 = h_T^{-1} \frac{h_T^2}{\delta_T^2} \sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\theta_0, \pi) \right\|^2 = h_T^{-1} O_p\left(\left(\frac{h_T^2}{T}\right)^{1-2\varepsilon}\right) = o_p\left(h_T^{-1}\right) = o_p(1),$$

because $1 - 2\varepsilon > 0$ and $h_T^2/T \to 0$. Therefore, since $\delta_T \to 0$, $\sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) \right\| \xrightarrow{p} 0$. But by Lemma 2, we know that $\sup_{\pi \in \Pi} \left\| \bar{g}_T^{sa}(\hat{\theta}^a(\pi), \pi) - k_1 g(\hat{\theta}^a(\pi), \pi) \right\| \xrightarrow{p} 0$. Thus,

$\sup_{\pi\in\Pi} g(\hat{\theta}^a(\pi),\pi) = o_p(1)$. Continuity of $g(\beta)$ and the identification Assumption 3(iv) then yields $\sup_{\pi\in\Pi} \left\|\hat{\theta}^a(\pi) - \theta_0\right\| = o_p(1)$.

In fact, a further refinement of the above argument (similar in spirit to that of Smith (2004, Lemma A5) shows that $\sup_{\pi\in\Pi} \left\|\bar{g}_T^{sa}(\hat{\theta}^a(\pi),\pi)\right\| = O_p(T^{-1/2})$, from which it can be shown that $h_T \bar{V}_T^a(\hat{\theta}^a(\pi),\pi) = k_2\Omega_0(\pi) + o_{p\pi}(1)$; c.f. Smith (2005)[Theorem 2.1]. Using, this (and arguments similar to the above) it can be shown that $\sup_{\pi\in\Pi} \|\hat{\gamma}^a(\pi)\| = O_p\left(h_T/\sqrt{T}\right)$ as follows.

Since, by definition, $Q_T^a(\hat{\theta}^a(\pi),\hat{\gamma}^a(\pi),\pi) \geq Q_T^a(\hat{\theta}^a(\pi),\gamma,\pi)$, for all $\gamma \in \Gamma_T$, setting $\gamma = 0 \in \Gamma_T$, and noting that $Q_T^a(\theta,0,\pi) \equiv 0$, for all $\theta \in \Theta$, we obtain, *w.p.a.*1,

$$0 \leq \frac{T}{h_T} Q_T^a(\hat{\theta}^a(\pi),\hat{\gamma}^a(\pi),\pi) = \frac{T}{h_T}\left\{k\hat{\gamma}^a(\pi)'\bar{g}_T^{sa}(\hat{\theta}^a(\pi),\pi) - \frac{1}{2}k^2\hat{\gamma}^a(\pi)'\bar{V}_T^a(\hat{\theta}^a(\pi),\pi)\hat{\gamma}^a(\pi)\right\} = O_p(1)$$

since $\frac{T}{h_T} Q_T^a(\hat{\theta}^a(\pi),\hat{\gamma}^a(\pi),\pi) \leq \sup_{\gamma\in\Gamma_T} \frac{T}{h_T} Q_T^a(\theta_0,\gamma,\pi) \leq \Delta \left\|\sqrt{T}\bar{g}_T^{sa}(\theta_0,\pi)\right\|^2 = O_p(1)$, *w.p.a.*1. Thus, since $\sup_{\pi\in\Pi} \left\|\bar{g}_T^{sa}(\hat{\theta}^a(\pi),\pi)\right\| = O_p(T^{-1/2})$ and $\sup_{\pi\in\Pi} \left\|h_T\bar{V}_T^a(\hat{\theta}^a(\pi),\pi)\right\| = O_p(1)$, it follows that $\sup_{\pi\in\Pi} \|\hat{\gamma}^a(\pi)\| = O_p\left(h_T/\sqrt{T}\right)$. This implies $\sup_{\pi\in\Pi} \|\hat{\gamma}^a(\pi)\| = o_p(1)$.

Proof of Theorem 2: Differentiating $Q_T^a(\theta,\gamma,\pi) = \frac{1}{T}\sum_{t=1}^T \ln\left(1 + k\gamma' g_t^{sa}(\theta,\pi)\right)$ with respect to θ and γ, yields the partial-sample first order conditions

$$\frac{\partial Q_T^a(\hat{\theta}^a(\pi),\hat{\gamma}^a(\pi),\pi)}{\partial\theta} = k\frac{1}{T}\sum_{t=1}^T \frac{G_t^{sa}(\hat{\theta}^a(\pi),\pi)'\hat{\gamma}^a(\pi)}{1 + k\hat{\gamma}^a(\pi)'g_t^{sa}(\hat{\theta}^a(\pi),\pi)} = 0 \quad (31)$$

$$\frac{\partial Q_T^a(\hat{\theta}^a(\pi),\hat{\gamma}^a(\pi),\pi)}{\partial\gamma} = k\frac{1}{T}\sum_{t=1}^T \frac{g_t^{sa}(\hat{\theta}^a(\pi),\pi)}{1 + k\hat{\gamma}^a(\pi)'g_t^{sa}(\hat{\theta}^a(\pi),\pi)} = 0 \quad (32)$$

where

$$G_t^{sa}(\theta,\pi) = \frac{\partial g_t^{sa}(\theta,\pi)}{\partial\theta'} = \mathbb{I}_{t,T}(\pi)\begin{pmatrix} \frac{\partial g_t^{sa}(\beta_1)}{\partial\beta_1'} & 0 \\ 0 & 0 \end{pmatrix} + (1 - \mathbb{I}_{t,T}(\pi))\begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial g_t^{sa}(\beta_2)}{\partial\beta_2'} \end{pmatrix}.$$

Writing $\hat{\varphi}^a(\pi) = \left(\hat{\theta}^a(\pi)', \frac{\hat{\gamma}^a(\pi)'}{h_T}\right)'$ and $\varphi_0 = \left(\beta_0', \beta_0', 0'\right)'$, and exploiting Lemma 1, a mean value expansion of (32) yields

$$0 = kk_1\sqrt{T}\bar{g}_T(\theta_0,\pi) + \bar{D}_T^{\varphi}\left(\bar{\varphi}^a(\pi),\pi\right)\sqrt{T}\left(\hat{\varphi}^a(\pi) - \varphi_0\right)$$

where

$$\bar{D}_T^{\varphi}(\varphi,\pi) = \frac{1}{T}\sum_{t=1}^T \left[\frac{\partial^2 Q_T^a(\theta,\gamma,\pi)}{\partial\gamma\partial\theta'},\ h_T\frac{\partial^2 Q_T^a(\theta,\gamma,\pi)}{\partial\gamma\partial\gamma'}\right]$$

and $\bar{\varphi}^a(\pi)$ is the usual mean value which may differ from row to row. Now

$$\begin{aligned}
\frac{\partial^2 Q_T^a(\theta,\gamma,\pi)}{\partial\gamma\partial\theta'} &= k\frac{1}{T}\sum_{t=1}^{T}\frac{G_t^{sa}(\theta,\pi)}{1+k\gamma' g_t^{sa}(\theta,\pi)} - k^2\frac{1}{T}\sum_{t=1}^{T}\frac{g_t^{sa}(\theta,\pi)\left(\gamma' G_t^{sa}(\theta,\pi)\right)}{\left(1+k\gamma' g_t^{sa}(\theta,\pi)\right)^2} \\
h_T\frac{\partial^2 Q_T^a(\theta,\gamma,\pi)}{\partial\gamma\partial\gamma'} &= -k^2\frac{h_T}{T}\sum_{t=1}^{T}\frac{g_t^{sa}(\theta,\pi)g_t^{sa}(\theta,\pi)'}{\left(1+k\gamma' g_t^{sa}(\theta,\pi)\right)^2}.
\end{aligned}$$

It follows from Theorem 1, Lemma 3, Lemma 2, as applied to $\frac{1}{T}\sum_{t=1}^{T} vec\left(G_t^{sa}(\theta,\pi)\right)$, and $\sup_{\pi\in\Pi}\left\|h_T\bar{V}_T^a(\bar{\theta}(\pi),\pi)-k_2\Omega_0(\pi)\right\| = o_p(1)$, that

$$0 = kk_1\sqrt{T}\bar{g}_T^{sa}(\theta_0,\pi) + D_0^{\varphi}(\pi)\sqrt{T}\left(\hat{\varphi}^a(\pi)-\varphi_0\right) + o_{p\pi}(1)$$

where

$$D_0^{\varphi}(\pi) = \left[kk_1 G_0(\pi),\ -k^2 k_2\Omega_0(\pi),\right].$$

Similarly, $\frac{\sqrt{T}}{h_T}\frac{\partial Q_T^a(\hat{\theta}^a(\pi),\hat{\gamma}^a(\pi),\pi)}{\partial\theta} = k\frac{1}{T}\sum_{t=1}^{T}\frac{G_t^{sa}(\hat{\theta}^a(\pi),\pi)'}{1+k\hat{\gamma}^a(\pi)' g_t^{sa}(\hat{\theta}^a(\pi),\pi)}\sqrt{T}\left(\frac{\hat{\gamma}^a(\pi)}{h_T}\right) = kk_1 G_0(\pi)'\sqrt{T}\left(\frac{\hat{\gamma}^a(\pi)}{h_T}\right) +$ $o_{p\pi}(1)$. Combining these results, we obtain

$$0 = \begin{pmatrix} 0 \\ \sqrt{T}\bar{g}_T(\theta_0,\pi) \end{pmatrix} + \begin{bmatrix} 0 & G_0(\pi)' \\ G_0(\pi) & -\Omega_0(\pi) \end{bmatrix}\sqrt{T}\left(\hat{\varphi}^a(\pi)-\varphi_0\right) + o_{p\pi}(1).$$

Solving for $\sqrt{T}\left(\hat{\varphi}^a(\pi)-\varphi_0\right)$, yields

$$\sqrt{T}\left(\hat{\varphi}^a(\pi)-\varphi_0\right) = \begin{pmatrix} -\left(A(\pi)^{-1}\otimes\left(M_0'M_0\right)^{-1}M_0\right) \\ \left(A(\pi)^{-1}\otimes\Omega_0^{-1/2}\left(I_\ell-P_0\right)\right) \end{pmatrix}\xi_T(\pi) + o_{p\pi}(1) \tag{33}$$

and the result follows. ■

The following Lemma establishes that Lemmas 1 and 2 also hold for moment functions smoothed before the sample split.

Lemma 6 *Under Assumptions 1, 2 and 3:*

1. $\sup_\pi\left\|\sqrt{T}\bar{g}_T^{sb}(\theta_0,\pi) - k_1\sqrt{T}\bar{g}_T(\theta_0,\pi)\right\| = o_p(1).$

2. *Let $m_t(\beta)$ and $m(\theta,\pi)$ be as defined in Lemma 2, and let $m_t^{sb}(\beta)$ be the smoothed version of $m_t(\beta)$, defined in an analogous manner to $g_t^{sb}(\beta)$, with $\bar{m}_T^{sb}(\theta,\pi)$ defined accordingly. Then $\sup_{\pi\in\Pi}\sup_{\theta\in\Theta}\left\|\bar{m}_T^{sb}(\theta,\pi)-k_1 m(\theta,\pi)\right\| = o_p(1)$.*

Proof of Lemma 6: By Smith (2004, Lemma A2), $\sqrt{T}\bar{g}_T^{sb}(\beta_0) = k_1\sqrt{T}\bar{g}_T(\beta_0)+o_p(1)$. Then, by the triangle inequality, it suffices to consider $\sqrt{T}\bar{g}_{[T\pi]}^{sb}(\beta) - k_1\sqrt{T}\bar{g}_{[T\pi]}(\beta)$, since $\frac{1}{\sqrt{T}}\sum_{t=[T\pi]+1}^{T} g_t^{sb}(\beta) =$

$\sqrt{T}\bar{g}_T^{sb}(\beta) - \sqrt{T}\bar{g}_{[T\pi]}^{sb}(\beta)$, where

$$\bar{g}_{[T\pi]}^{sb}(\beta) = \sum_{j=1-T}^{[T\pi]-1} \frac{1}{h_T} k\left(\frac{j}{h_T}\right) \frac{1}{T} \sum_{t=\max[1,1-j]}^{\min[T,[T\pi]-j]} g_t(\beta).$$

Now, when $j \geq 0$, $\max[1, 1-j] = 1$ and $\min[T, [T\pi] - j] = [T\pi] - j$. On the other hand when $j < 0$, $\max[1, 1-j] = 1 + |j|$ when $j > [T\pi] - T$, whilst $\max[1, 1-j] = 1 + |j| = T$ when $j \leq [T\pi] - T$. Thus, we can write

$$\begin{aligned} \bar{g}_{[T\pi]}^{sb}(\beta) &= \sum_{j=0}^{[T\pi]-1} \frac{1}{h_T} k\left(\frac{j}{h_T}\right) \left\{ \frac{1}{T} \sum_{t=1}^{[T\pi]} g_t(\beta) - e_{1T}(\beta, \pi) \right\} \\ &+ \sum_{j=1-T+[T\pi]}^{-1} \frac{1}{h_T} k\left(\frac{j}{h_T}\right) \left\{ \frac{1}{T} \sum_{t=1}^{[T\pi]} g_t(\beta) - e_{2T}(\beta, \pi) \right\} \\ &+ \sum_{j=1-T}^{-T+[T\pi]} \frac{1}{h_T} k\left(\frac{j}{h_T}\right) \left\{ \frac{1}{T} \sum_{t=1}^{[T\pi]} g_t(\beta) - e_{3T}(\beta, \pi) \right\} \end{aligned}$$

where

$$\begin{aligned} e_{1T}(\beta, \pi) &= \frac{1}{T} \sum_{t=[T\pi]+1-j}^{[T\pi]} g_t(\beta) \\ e_{2T}(\beta, \pi) &= \frac{1}{T} \sum_{t=1}^{|j|} g_t(\beta) - \frac{1}{T} \sum_{t=[T\pi]+1}^{[T\pi]+|j|} g_t(\beta) \\ e_{3T}(\beta, \pi) &= \frac{1}{T} \sum_{t=1}^{|j|} g_t(\beta) - \frac{1}{T} \sum_{t=[T\pi]+1}^{T} g_t(\beta) \end{aligned}$$

Therefore

$$\bar{g}_{[T\pi]}^{sb}(\beta) = \sum_{j=1-T}^{T-1} \frac{1}{h_T} k\left(\frac{j}{h_T}\right) \frac{1}{T} \sum_{t=1}^{[T\pi]} g_t(\beta) - \sum_{j=0}^{3} A_{jT}(\beta, \pi)$$

where

$$\begin{aligned} A_{0T}(\beta, \pi) &= \frac{1}{h_T} \sum_{j=[T\pi]}^{T-1} k\left(\frac{j}{h_T}\right) \frac{1}{T} \sum_{t=1}^{[T\pi]} g_t(\beta) \\ A_{1T}(\beta, \pi) &= \frac{1}{h_T} \sum_{j=0}^{[T\pi]-1} k\left(\frac{j}{h_T}\right) \frac{1}{T} \sum_{t=[T\pi]+1-j}^{[T\pi]} g_t(\beta) \\ A_{2T}(\beta, \pi) &= \frac{1}{h_T} \sum_{j=1-T+[T\pi]}^{-1} k\left(\frac{j}{h_T}\right) \left\{ \frac{1}{T} \sum_{t=1}^{|j|} g_t(\beta) - \frac{1}{T} \sum_{t=[T\pi]+1}^{[T\pi]+|j|} g_t(\beta) \right\} \\ A_{3T}(\beta, \pi) &= \frac{1}{h_T} \sum_{j=1-T}^{-T+[T\pi]} k\left(\frac{j}{h_T}\right) \left\{ \frac{1}{T} \sum_{t=1}^{|j|} g_t(\beta) - \frac{1}{T} \sum_{t=[T\pi]+1}^{T} g_t(\beta) \right\}. \end{aligned}$$

1. $\sup_\pi \left\|\sqrt{T}\bar{g}^{sb}_{[T\pi]}(\beta_0) - k_1\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right\| \xrightarrow{p} 0$.

   We show that $\sup_\pi \left\|\sqrt{T}A_{jT}(\beta_0,\pi)\right\| = o_p(1)$, for $j = 0,1,2,3$.

   (a) $\sup_\pi \left\|\sqrt{T}A_{0T}(\beta_0,\pi)\right\| = o_p(1)$: Firstly $\sup_\pi \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\pi]} g_t(\beta_0)\right\| = O_p(1)$, by Lemma 1 and the Continuous Mapping Theorem. Second, $\lim_{T\to\infty}\frac{1}{h_T}\sum_{j=1-T}^{T-1}\left|k\left(\frac{j}{h_T}\right)\right| = O(1)$, implies $\lim_{T\to\infty}\sup_\pi \frac{1}{h_T}\sum_{j=[T\pi]}^{T-1}\left|k\left(\frac{j}{h_T}\right)\right| = 0$.

   (b) $\sup_\pi \left\|\sqrt{T}A_{1T}(\beta_0,\pi)\right\| = o_p(1)$: $\left\|\frac{1}{\sqrt{|j|}}\sum_{t=[T\pi]+1-j}^{[T\pi]} g_t(\beta_0)\right\| = O_p(1)$, uniformly in j and π so that

   $$\sup_\pi \|A_{1T}(\beta_0,\pi)\| \le \left\{\frac{1}{h_T}\sum_{j=0}^{T-1}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right|\right\} O_p(1) = o_p(1)$$

   since $\lim_{T\to\infty}\frac{1}{h_T}\sum_{j=1-T}^{T-1}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right| = 0$.

   (c) $\sup_\pi \left\|\sqrt{T}A_{2T}(\beta_0,\pi)\right\| = o_p(1)$: $\frac{1}{\sqrt{|j|}}\sum_{t=1}^{|j|} g_t(\beta_0)$ and $\frac{1}{\sqrt{|j|}}\sum_{t=[T\pi]+1}^{[T\pi]+|j|} g_t(\beta_0)$ are both $O_p(1)$, uniformly in j and π, so that

   $$\sup_\pi \|A_{2T}(\beta_0,\pi)\| \le \left\{\frac{1}{h_T}\sum_{j=1-T}^{-1}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right|\right\} O_p(1) = o_p(1)$$

   as above.

   (d) $\sup_\pi \|A_{3T}(\beta_0,\pi)\| = o_p(1)$: $\frac{1}{\sqrt{|j|}}\sum_{t=1}^{|j|} g_t(\beta_0) = O_p(1)$, uniformly in j and π, and $\frac{1}{\sqrt{T}}\sum_{t=[T\pi]+1}^{T} g_t(\beta_0) = \sqrt{T}\bar{g}_T(\beta_0) - \sqrt{T}\bar{g}_{[T\pi]}(\beta_0)$ is $O_p(1)$ uniformly in π. Thus

   $$\begin{aligned}\sup_\pi \|A_{3T}(\beta_0,\pi)\| &\le \sup_\pi \frac{1}{h_T}\sum_{j=1-T}^{-T+[T\pi]}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right| O_p(1) + \sup_\pi \frac{1}{h_T}\sum_{j=1-T}^{-T+[T\pi]}\left|k\left(\frac{j}{h_T}\right)\right| O_p(1)\\ &= o_p(1)\end{aligned}$$

   since both $\frac{1}{h_T}\sum_{j=1-T}^{-T+[T\pi]}\sqrt{\frac{|j|}{T}}\left|k\left(\frac{j}{h_T}\right)\right| = o(1)$ and $\sup_\pi \frac{1}{h_T}\sum_{j=1-T}^{-T+[T\pi]}\left|k\left(\frac{j}{h_T}\right)\right| = o(1)$.

   Therefore, $\sqrt{T}\bar{g}^{s}_{[T\pi]}(\beta_0,\pi) = \sum_{j=1-T}^{T-1}\frac{1}{h_T}k\left(\frac{j}{h_T}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\pi]} g_t(\beta_0) + o_{p\pi}(1)$. The result follows from the fact that $\sum_{j=1-T}^{T-1}\frac{1}{h_T}k\left(\frac{j}{h_T}\right) = k_1 + o(1)$.

2. By Smith (2004, Lemma A1), it can be show that $\sup_{\beta\in\mathcal{B}}\left\|\bar{m}^{sb}_T(\beta) - k_1 m(\beta)\right\| = o_p(1)$. Then, by the triangle inequality, it suffices to show that

   $$\sup_\pi \sup_{\beta\in\mathcal{B}}\left\|\bar{m}^{sb}_{[T\pi]}(\beta) - \sum_{j=1-T}^{T-1}\frac{1}{h_T}k\left(\frac{j}{h_T}\right)\bar{m}_{[T\pi]}(\beta)\right\| = o_p(1),$$

where $\bar{m}^{sb}_{[T\pi]}(\beta) = \frac{1}{T}\sum_{t=1}^{[T\pi]} m_t^{sb}(\beta)$, since $\frac{1}{T}\sum_{t=[T\pi]+1}^{T} m_t^{sb}(\beta) = \bar{m}^{sb}_{T}(\beta) - \bar{m}^{sb}_{[T\pi]}(\beta)$. From Part (1), above it is clear that

$$\bar{m}^{sb}_{[T\pi]}(\beta) = \sum_{j=1-T}^{T-1} \frac{1}{h_T} k\left(\frac{j}{h_T}\right) \frac{1}{T}\sum_{t=1}^{[T\pi]} m_t(\beta) - \sum_{j=0}^{3} A_{jT}(\beta,\pi)$$

where the $A_{jT}(\beta,\pi)$ are as before but defined in terms of $m_t(\beta)$, rather than $g_t(\beta)$. It is then straightforward to show that $\sup_\pi \sup_\beta \|A_{jT}(\beta,\pi)\| = o_p(1)$, for $j = 0,1,2,3$, and the result follows. ■

Proof of Theorem 3: Define $Q_T^b(\theta,\gamma,\pi) = \frac{1}{T}\sum_{t=1}^{T} \ln\left(1 + k\gamma' g_t^s(\theta,\pi)\right)$.

1. **Consistency:** As in Lemma 3, it is straightforward to show that $\sup_{\theta\in\Theta,\gamma\in\Gamma_T,1\le t\le T} \left|\gamma' g_t^{sb}(\theta,\pi)\right| = o_{p\pi}(1)$. Therefore (30) holds with $g_t^{sb}(\theta\,\pi)$ replacing $g_t^{sa}(\theta,\pi)$ everywhere, with the error being of smaller order than the leading two terms. By arguments similar to Smith (2004, Lemma A3) it can be shown that $h_T \bar{V}_T^b(\theta_0,\pi) = k_2\Omega_0(\pi) + o_{p\pi}(1)$, but in general, $\bar{V}_T^b(\theta,\pi) = O_{p\pi}(1)$, uniformly in θ. This yields Lemma 4 but where we can write

$$h_T^{-1} Q_T^b(\theta_0,\gamma,\pi) \le \gamma_T' \bar{g}_T^{sb}(\theta_0,\pi) - \Delta\gamma_T'\gamma_T \tag{34}$$

and

$$Q_T^b(\theta,\gamma,\pi) \ge k\gamma'\bar{g}_T^{sb}(\theta,\pi) - k^2\Delta\gamma'\gamma. \tag{35}$$

From this, Lemma 5 gives

$$\sup_{\pi\in\Pi}\sup_{\gamma\in\Gamma_T} h_T^{-1} Q_T^b(\theta_0,\gamma,\pi) \le \Delta \sup_{\pi\in\Pi}\left\|\bar{g}_T^{sb}(\theta_0,\pi)\right\|^2 = O_p(T^{-1}) \tag{36}$$

where the fact that $\sup_{\pi\in\Pi}\left\|\bar{g}_T^{sb}(\theta_0,\pi\right\|^2 = O_p\left(T^{-1}\right)$ follows from Lemmas 6(part 1) and 1. Consistency for $\hat{\theta}^b(\pi)$ then follows the from the arguments of Theorem 1, but using equations (34)-(36), rather than (28),(29) and Lemma 5, respectively, and Lemma 6. A similar refinement then establishes $\sup_{\pi\in\Pi}\left\|\bar{g}_T^{sb}(\hat{\theta}^b(\pi),\pi)\right\| = O_p(T^{-1/2})$, from which it can be shown that $h_T\bar{V}_T^b(\hat{\theta}^b(\pi),\pi) = k_2\Omega_0(\pi) + o_{p\pi}(1)$; c.f. Smith (2005, Theorem 2.1). Using, this (and arguments similar to the above) it can be shown that $\sup_{\pi\in\Pi}\left\|\hat{\gamma}^b(\pi)\right\| = O_p\left(h_T/\sqrt{T}\right)$.

2. **Asymptotic Normality:** This follows the same arguments as Theorem 2, and using Lemma 6 which shows that $\sup_\pi \sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_t^{sb}(\theta,\pi)}{\partial\theta'} - k_1\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_t(\theta,\pi)}{\partial\theta'}\right\| \xrightarrow{p} 0$.

We thus obtain, by Lemma 6(part 1),

$$\sqrt{T}\left(\hat{\varphi}^{b}(\pi)-\varphi_{0}\right)=\left(\begin{array}{c}-\left(A(\pi)^{-1}\otimes\left(M_{0}^{\prime}M_{0}\right)^{-1}M_{0}\right)\\ \left(A(\pi)^{-1}\otimes\Omega_{0}^{-1/2}\left(I_{\ell}-P_{0}\right)\right)\end{array}\right)\xi_{T}(\pi)+o_{p\pi}(1)$$

and the result follows.

Proof of Theorem 4: Consistency of the estimators follows from the general arguments employed in the proof of Theorem 1, and Theorem 3. Differentiating $\dot{Q}_T(\beta,\gamma,\pi)=\frac{1}{T}\sum_{t=1}^{T}\ln\left(1+k\gamma'\dot{g}_t^s(\beta,\pi)\right)$ with respect to β and $\gamma=\left(\lambda_1',\lambda_2'\right)'$, yields the partial-sample first order conditions

$$\begin{aligned}\frac{\partial\dot{Q}_T\left(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi\right)}{\partial\beta}&=k\frac{1}{T}\sum_{t=1}^{[T\pi]}\frac{G_t^s(\tilde{\beta}(\pi))'\tilde{\lambda}_1(\pi)}{1+k\tilde{\lambda}_1(\pi)'g_t^s(\tilde{\beta}(\pi))}+k\frac{1}{T}\sum_{t=[T\pi]+1}^{T}\frac{G_t^s(\tilde{\beta}(\pi))'\tilde{\lambda}_2(\pi)}{1+k\tilde{\lambda}_2(\pi)'g_t^s(\tilde{\beta}(\pi))}=0\\ \frac{\partial\dot{Q}_T\left(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi\right)}{\partial\lambda_1}&=k\frac{1}{T}\sum_{t=1}^{[T\pi]}\frac{g_t^s(\tilde{\beta}(\pi))}{1+k\tilde{\lambda}_1(\pi)'g_t^s(\tilde{\beta}(\pi))}=0\\ \frac{\partial\dot{Q}_T\left(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi\right)}{\partial\lambda_2}&=k\frac{1}{T}\sum_{t=[T\pi]+1}^{T}\frac{g_t^s(\tilde{\beta}(\pi))}{1+k\tilde{\lambda}_2(\pi)'g_t^s(\tilde{\beta}(\pi))}=0.\end{aligned}$$

Using similar arguments to those employed in the proof of Theorem 2, a mean value expansion of $\frac{\partial\dot{Q}_T\left(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi\right)}{\partial\lambda_i}=0$ about $\left(\beta_0',0'\right)'$, $i=1,2$, yields, exploiting Lemmas 1 and 6,

$$\begin{aligned}0&=kk_1\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)+kk_1\pi G_0\sqrt{T}\left(\tilde{\beta}(\pi)-\beta_0\right)-k^2k_2\pi\Omega_0\left(\sqrt{T}/h_T\right)\tilde{\lambda}_1(\pi)+o_{p\pi}(1)\\ 0&=kk_1\sqrt{T}\bar{g}_T(\beta_0)-kk_1\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)+kk_1(1-\pi)G_0\sqrt{T}\left(\tilde{\beta}(\pi)-\beta_0\right)\\ &\quad -k^2k_2(1-\pi)\Omega_0\left(\sqrt{T}/h_T\right)\tilde{\lambda}_2(\pi)+o_{p\pi}(1)\end{aligned}$$

respectively, or

$$\begin{aligned}\pi\left(\sqrt{T}/h_T\right)\tilde{\lambda}_1(\pi)&=\Omega_0^{-1}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)+\pi\Omega_0^{-1}G_0\sqrt{T}\left(\tilde{\beta}(\pi)-\beta_0\right)+o_{p\pi}(1)\\ (1-\pi)\left(\sqrt{T}/h_T\right)\tilde{\lambda}_2(\pi)&=\Omega_0^{-1}\left(\sqrt{T}\bar{g}_T(\beta_0)-\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right)\\ &\quad+(1-\pi)\Omega_0^{-1}G_0\sqrt{T}\left(\tilde{\beta}(\pi)-\beta_0\right)+o_{p\pi}(1).\end{aligned}$$

Similarly,

$$
\begin{aligned}
\frac{\sqrt{T}}{h_T}\frac{\partial \dot{Q}_T\left(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi\right)}{\partial\beta} &= k\frac{1}{T}\sum_{t=1}^{[T\pi]}\frac{G_t^s(\tilde{\beta}(\pi))'}{1+k\tilde{\lambda}_1(\pi)'g_t^s(\tilde{\beta}(\pi))}\sqrt{T}\left(\frac{\tilde{\lambda}_1(\pi)}{h_T}\right) \\
&\quad +k\frac{1}{T}\sum_{t=[T\pi]+1}^{T}\frac{G_t^s(\tilde{\beta}(\pi))')}{1+k\tilde{\lambda}_2(\pi)'g_t^s(\tilde{\beta}(\pi))}\sqrt{T}\left(\frac{\tilde{\lambda}_2(\pi)}{h_T}\right) \\
&= kk_1\pi G_0(\pi)'\sqrt{T}\left(\frac{\tilde{\lambda}_1(\pi)}{h_T}\right)+kk_1\left(1-\pi\right)G_0(\pi)'\sqrt{T}\left(\frac{\tilde{\lambda}_2(\pi)}{h_T}\right)+o_{p\pi}(1) \\
&= 0
\end{aligned}
$$

Combining these results, we obtain

$$
\begin{aligned}
0 &= \pi G_0(\pi)'\sqrt{T}\left(\frac{\tilde{\lambda}_1(\pi)}{h_T}\right)+\left(1-\pi\right)G_0(\pi)'\sqrt{T}\left(\frac{\tilde{\lambda}_2(\pi)}{h_T}\right)+o_{p\pi}(1) \\
&= G_0'\Omega_0^{-1}\sqrt{T}\bar{g}_T^s(\beta_0)+G_0'\Omega_0^{-1}G_0\sqrt{T}\left(\tilde{\beta}(\pi)-\beta_0\right)+o_{p\pi}(1)
\end{aligned}
$$

so that

$$
\sqrt{T}\left(\tilde{\beta}(\pi)-\beta_0\right)=-\left(M_0'M_0\right)^{-1}M_0\left\{\Omega_0^{-1/2}\sqrt{T}\bar{g}_T(\beta_0)\right\}+o_{p\pi}(1),
$$

and

$$
\begin{aligned}
\pi\left(\sqrt{T}/h_T\right)\tilde{\lambda}_1(\pi) &= \Omega_0^{-1/2}\left\{\Omega_0^{-1/2}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right\} \\
&\quad -\pi\Omega_0^{-1/2}P_0\left\{\Omega_0^{-1/2}\sqrt{T}\bar{g}_T(\beta_0)\right\}+o_{p\pi}(1) \\
\left(1-\pi\right)\left(\sqrt{T}/h_T\right)\tilde{\lambda}_2(\pi) &= \Omega_0^{-1/2}\left\{\Omega_0^{-1/2}\left(\sqrt{T}\bar{g}_T(\beta_0)-\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\right)\right\} \\
&\quad -\left(1-\pi\right)\Omega_0^{-1/2}P_0\left\{\Omega_0^{-1/2}\sqrt{T}\bar{g}_T(\beta_0)\right\}+o_{p\pi}(1)
\end{aligned}
$$

or

$$
\begin{aligned}
\left(\sqrt{T}/h_T\right)\tilde{\gamma}(\pi) &= \left(A(\pi)^{-1}\otimes\Omega_0^{-1/2}\right)\xi_T(\pi) \\
&\quad -\left(\iota_2\otimes\Omega_0^{-1/2}P_0\right)\Omega_0^{-1/2}\sqrt{T}\bar{g}_T(\beta_0)+o_{p\pi}(1) \\
&= \left(A(\pi)^{-1}\otimes\Omega_0^{-1/2}\right)\xi_T(\pi)-\left(\iota_2\iota_2'\otimes\Omega_0^{-1/2}P_0\right)\xi_T(\pi)+o_{p\pi}(1) \\
&= \left(A(\pi)^{-1}-\iota_2\iota_2'\otimes\Omega_0^{-1/2}\left(I_\ell-P_0\right)\right)\xi_T(\pi)+o_{p\pi}(1) \\
&= \frac{1}{\pi(1-\pi)}\left(a(\pi)a(\pi)'\otimes\Omega_0^{-1/2}\left(I_\ell-P_0\right)\right)\xi_T(\pi)+o_{p\pi}(1) \\
&= \frac{1}{\pi(1-\pi)}\left(a(\pi)\otimes\Omega_0^{-1/2}\right)\left(I_\ell-P_0\right)\left(a(\pi)'\otimes I_\ell\right)\xi_T(\pi)+o_{p\pi}(1)
\end{aligned}
$$

where $\iota_2=(1,1)'$, $a(\pi)'=(1-\pi,\ -\pi)$ and the result follows by Lemma 1. ■

Proof of Theorem 5: Consider, first, $\mathcal{W}_T(\pi)$. Previous results, exploiting $\sqrt{T}$-consistency of $\hat{\beta}_i(\pi)$, show that

$$\left(k_1^2/k_2\right) V_T^W(\hat{\theta}(\pi)) = \frac{1}{\pi(1-\pi)} \left(M_0' M_0\right)^{-1} + o_{p\pi}(1)$$

and, combining this with (33), we obtain

$$-\left\{\left(k_1^2/k_2\right) V_T^W(\hat{\theta}(\pi))\right\}^{-1/2} \sqrt{T}\left(\hat{\beta}_1(\pi\right) - \hat{\beta}_2(\pi)) = \frac{1}{\sqrt{\pi(1-\pi)}} \left(M_0' M_0\right)^{-1/2} M_0' \left(a(\pi)' \otimes I_\ell\right) \xi_T(\pi) + o_{p\pi}(1)$$

where $a(\pi)' = (1-\pi, \ -\pi)$,so that

$$\begin{aligned}\mathcal{W}_T(\pi) &= \frac{\xi_T(\pi)' \left(a(\pi) \otimes I_\ell\right)' P_0 \left(a(\pi)' \otimes I_\ell\right) \xi_T(\pi)}{\pi(1-\pi)} + o_{p\pi}(1) \\ &= \mathcal{S}_T(\pi) + o_{p\pi}(1).\end{aligned}$$

For $\mathcal{LM}_T(\pi)$, it can be shown that

$$\begin{aligned}(\sqrt{T}/h_T)\tilde{\rho}(\pi) &= -\bar{G}_T^s(\tilde{\beta}(\pi))' \left\{\bar{V}_T^s(\tilde{\beta}(\pi))\right\}^{-1} \sqrt{T}\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi)) + o_{p\pi}(1) \\ &= \bar{C}_T^s(\tilde{\beta}(\pi))' \sqrt{T}\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi)) + o_{p\pi}(1),\end{aligned}$$

say, where $\bar{g}_{[T\pi]}^s(\beta) = \frac{1}{T}\sum_{t=1}^{[T\pi]} g_t^s(\beta)$, so that an asymptotically equivalent variant of $\mathcal{LM}_T(\pi)$ is

$$\mathcal{LM}_T(\pi) = (k_2/k_1^2)T \ \bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi))\bar{C}_T^s(\tilde{\beta}(\pi)) \left\{V_T^\rho(\tilde{\beta}(\pi))\right\}^{-1} \bar{C}_T^s(\tilde{\beta}(\pi))' \bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi)).$$

An expansion of $\sqrt{T}\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi))$ yields

$$\begin{aligned}\sqrt{T}\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi)) &= k_1\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) + k_1\pi G_0\sqrt{T}\left(\tilde{\beta}(\pi) - \beta_0\right) + o_{p\pi}(1) \\ &= k_1\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) - k_1\pi G_0 \left(M_0' M_0\right)^{-1} M_0 \left\{\Omega_0^{-1/2}\sqrt{T}\bar{g}_T(\beta_0)\right\} + o_{p\pi}(1).\end{aligned}$$

Furthermore, $\bar{G}_T^s(\tilde{\beta}(\pi)) = k_1 G_0 + o_{p\pi}(1)$ and $\bar{V}_T^s(\tilde{\beta}(\pi)) = k_2\Omega_0 + o_{p\pi}(1)$. so that

$$\left\{\left(k_1^2/k_2\right) V_T^\rho(\tilde{\beta}(\pi))\right\}^{-1/2} \bar{C}_T^s(\tilde{\beta}(\pi))' \sqrt{T}\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi)) = (M_0' M_0)^{-1/2} M_0' \left(a(\pi)' \otimes I_\ell\right) \xi_T(\pi) + o_{p\pi}(1)$$

and it immediately follows that $\sup_{\pi\in\Pi} |\mathcal{LM}_T(\pi) - \mathcal{S}_T(\pi)| = o_p(1)$.

For $\mathcal{LR}_T(\pi)$, a key expansion is that of $\sqrt{T}\bar{g}_T^s(\hat{\theta}(\pi),\pi) = \frac{1}{\sqrt{T}}\sum_{t=1}^T g_t^s(\theta,\pi)$ about θ_0, yielding

$$\begin{aligned}\sqrt{T}\bar{g}_T^s(\hat{\theta}(\pi),\pi) &= \sqrt{T}\bar{g}_T^s(\theta_0\,,\pi) + k_1 G_0(\pi)\sqrt{T}(\hat{\theta}(\pi) - \theta_0) + o_{p\pi}(1) \\ &= k_1\sqrt{T}\bar{g}_T(\theta_0\,,\pi) - k_1\left(I_2 \otimes \Omega_0^{1/2} P_0\right)\xi_T(\pi) + o_{p\pi}(1), \qquad (37)\end{aligned}$$

where (33) is exploited. Therefore, and again exploiting (33), we have

$$k_1^{-1}\left(I_2\otimes\Omega_0^{-1/2}\right)\sqrt{T}\bar{g}_T^s(\hat{\theta}(\pi),\pi) = \left(I_2\otimes(I_\ell-P_0)\right)\xi_T(\pi)+o_{p\pi}(1) \quad (38)$$

$$= \left(A(\pi)\otimes\Omega_0^{1/2}\right)\left(\sqrt{T}/h_T\right)\hat{\gamma}(\pi)+o_{p\pi}(1). \quad (39)$$

Now, noting that $Q_T(\theta,0,\pi)\equiv 0$ and $\partial Q_T(\theta,0,\pi)/\partial\gamma = k\bar{g}_T^s(\theta\,,\pi)$, for all $\theta\in\Theta$, a two term expansion of $Q_T(\hat{\theta}\,(\pi)\,,\hat{\gamma}(\pi),\pi)$ about $\hat{\gamma}(\pi)=0$, yields

$$\begin{aligned}2\left(k_2/k_1^2\right)(T/h_T)\,Q_T\left(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi\right) &= 2\left(k_2/k_1^2\right)k(\sqrt{T}/h_T)\hat{\gamma}(\pi)'\sqrt{T}\bar{g}_T^s(\hat{\theta}(\pi)\,,\pi)\\ &\quad+\left(k_2/k_1^2\right)(\sqrt{T}/h_T)\hat{\gamma}(\pi)'\left(h_T\frac{\partial^2 Q_T(\hat{\theta}(\pi),\bar{\gamma}(\pi),\pi)}{\partial\gamma\partial\gamma'}\right)(\sqrt{T}/h_T)\hat{\gamma}(\pi)\\ &= T\ \bar{g}_T^s(\hat{\theta}(\pi)\,,\pi)'\left(A(\pi)\otimes\Omega_0\right)^{-1}\bar{g}_T^s(\hat{\theta}(\pi),\pi)/k_1^2+o_{p\pi}(1) \quad (40)\end{aligned}$$

where $\bar{\gamma}(\pi)$ is the usual mean value and the third equality uses (39) and Lemma 3, which ensures that $h_T\dfrac{\partial^2 Q_T(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi)}{\partial\gamma\partial\gamma'}\overset{p}{\to}-k^2k_2\Omega_0(\pi)=-k^2k_2\left(A(\pi)\otimes\Omega_0\right)$, uniformly in π. Similarly,

$$2\left(k_2/k_1^2\right)(T/h_T)\,\dot{Q}_T\left(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi\right)=T\ \bar{g}_T^s(\tilde{\theta}(\pi),\pi)'\left(A(\pi)\otimes\Omega_0\right)^{-1}\bar{g}_T^s(\tilde{\theta}(\pi),\pi)/k_1^2+o_{p\pi}(1),$$

where $\tilde{\theta}(\pi)=\left(\tilde{\beta}(\pi)',\tilde{\beta}(\pi)'\right)'$. Furthermore, an expansion of $\sqrt{T}\bar{g}_T^s(\tilde{\theta}(\pi),\pi)$ yields

$$\begin{aligned}\sqrt{T}\bar{g}_T^s(\tilde{\theta}(\pi),\pi) &= \sqrt{T}\bar{g}_T^s(\theta_0,\pi)-k_1\left(A(\pi)\iota_2\iota_2'\otimes\Omega_0^{1/2}P_0\right)\xi_T(\pi)+o_{p\pi}(1)\\ &= \sqrt{T}\bar{g}_T^s(\hat{\theta}(\pi),\pi)+k_1\left(I_2-A(\pi)\iota_2\iota_2'\otimes\Omega_0^{1/2}P_0\right)\xi_T(\pi)+o_{p\pi}(1)\end{aligned}$$

where the second equality follows from (37). Notice that, by (38),

$$\begin{aligned}&k_1\sqrt{T}\bar{g}_T^s(\hat{\theta}(\pi),\pi)'\left(A(\pi)\otimes\Omega_0\right)^{-1}\left(I_2-A(\pi)\iota_2\iota_2'\otimes\Omega_0^{1/2}P_0\right)\xi_T(\pi)\\ &\qquad=k_1\xi_T(\pi)\left(A(\pi)^{-1}-\iota_2\iota_2'\otimes(I_\ell-P_0)P_0\right)\xi_T(\pi)+o_{p\pi}(1)\\ &\qquad=o_{p\pi}(1)\end{aligned}$$

so that

$$\begin{aligned}\mathcal{LR}_T(\pi) &= \xi_T(\pi)'\left(I_2-\iota_2\iota_2'A(\pi)\otimes\Omega_0^{1/2}P_0\right)\left(A(\pi)\otimes\Omega_0\right)^{-1}\left(I_2-A(\pi)\iota_2\iota_2'\otimes\Omega_0^{1/2}P_0\right)\xi_T(\pi)+o_{p\pi}(1)\\ &= \xi_T(\pi)'\left(I_2-\iota_2\iota_2'A(\pi)\otimes P_0\right)\left(A(\pi)\otimes I_\ell\right)^{-1}\left(I_2-A(\pi)\iota_2\iota_2'\otimes P_0\right)\xi_T(\pi)+o_{p\pi}(1)\\ &= \xi_T(\pi)'\left(A(\pi)^{-1}-\iota_2\iota_2'\otimes P_0\right)\xi_T(\pi)+o_{p\pi}(1)\\ &= \frac{\xi_T(\pi)'\left(a(\pi)a(\pi)'\otimes P_0\right)\xi_T(\pi)}{\pi(1-\pi)}+o_{p\pi}(1)\\ &= \frac{\xi_T(\pi)'\left(a(\pi)\otimes I_\ell\right)'P_0\left(a(\pi)'\otimes I_\ell\right)\xi_T(\pi)}{\pi(1-\pi)}+o_{p\pi}(1)\\ &= \mathcal{S}_T(\pi)+o_{p\pi}(1).\end{aligned}$$

using $\left(A(\pi)^{-1}-\iota_2\iota_2'\right)\left(I_2-A(\pi)\iota_2\iota_2'\right)=A(\pi)^{-1}-\iota_2\iota_2'=a(\pi)a(\pi)'/\pi(1-\pi)$.

As in Sowell (1996) and Hall and Sen (1999), we can always write $P_0=H'\Xi H$, where Ξ is the diagonal matrix of eigenvalues of P_0 and $H=[H_1',H_2']'$ is a $(\ell\times\ell)$ orthonormal matrix, so that $H'H=I_\ell=H_1'H_1+H_2'H_2$, with $H_1H_1'=I_k$ and $H_2H_2'=I_{\ell-k}$. From the properties of Ξ, $P_0=H_1'H_1$, and

$$H_1\left(a(\pi)'\otimes I_\ell\right)\xi_T(\pi)\Longrightarrow H_1\left(B_\ell(\pi)-\pi B_\ell(1)\right)=B_k(\pi)-\pi B_k(1)$$

from which we conclude that $\mathcal{S}_T(\pi)\Longrightarrow\dfrac{\left(B_k(\pi)-\pi B_k(1)\right)'\left(B_k(\pi)-\pi B_k(1)\right)}{\pi(1-\pi)}$, and $\tau\left[\hat{\mathcal{D}}_{1,T}(\pi)\right]\Longrightarrow$ $\tau\left[\dfrac{\left(B_k(\pi)-\pi B_k(1)\right)'\left(B_k(\pi)-\pi B_k(1)\right)}{\pi(1-\pi)}\right]$ by applying the continuous mapping theorem.■

Proof of Theorem 6: Since $\sup_{\pi\in\Pi}\left\|h_T\bar{V}_T^s(\hat{\theta}(\pi),\pi)-k_2\Omega_0(\pi)\right\|=o_p(1)$, we immediately have that

$$\begin{aligned}\mathcal{O}_T(\pi) &= \left(k_2/k_1^2\right)\left(T/h_T\right)\bar{g}_T^s(\hat{\theta}(\pi),\pi)'\left\{\bar{V}_T^s(\hat{\theta}(\pi),\pi)\right\}^{-1}\bar{g}_T^s(\hat{\theta}(\pi),\pi)' \\ &= T\ \bar{g}_T^s(\hat{\theta}(\pi),\pi)'\left(A(\pi)\otimes\Omega_0\right)^{-1}\bar{g}_T^s(\hat{\theta}(\pi),\pi)/k_1^2+o_{p\pi}(1)\end{aligned}$$

and

$$\begin{aligned}\mathcal{LM}_T^*(\pi) &= (T/h_T)\hat{\gamma}(\pi)'\left\{\bar{V}_T^s(\hat{\theta}(\pi),\pi)\right\}\hat{\gamma}(\pi)/k_2 \\ &= T\ \bar{g}_T^s(\hat{\theta}(\pi),\pi)'\left(A(\pi)\otimes\Omega_0\right)^{-1}\bar{g}_T^s(\hat{\theta}(\pi),\pi)/k_1^2+o_{p\pi}(1) \\ &= \mathcal{O}_T(\pi)+o_{p\pi}(1).\end{aligned}$$

By (40) it is immediate that

$$\begin{aligned}\mathcal{LR}_T^*(\pi) &= T\ \bar{g}_T^s(\hat{\theta}(\pi),\pi)'\left(A(\pi)\otimes\Omega_0\right)^{-1}\bar{g}_T^s(\hat{\theta}(\pi),\pi)/k_1^2+o_{p\pi}(1) \\ &= \mathcal{O}_T(\pi)+o_{p\pi}(1).\end{aligned}$$

This demonstrates the asymptotic equivalence of all three statistics. From (38) we also obtain

$$\begin{aligned}\mathcal{O}_T(\pi) &= \xi_T(\pi)'\left(A(\pi)^{-1}\otimes(I_\ell-P_0)\right)\xi_T(\pi)+o_{p\pi}(1) \\ &= \mathcal{S}_T^*(\pi)+o_{p\pi}(1)\end{aligned}$$

Following the arguments in the proof of Theorem 6, $I - P_0 = H_2' H_2$ so that

$$\begin{aligned} \mathcal{S}_T^*(\pi) &= \xi_T(\pi)' \left(A(\pi)^{-1} \otimes (I_\ell - P_0)\right) \xi_T(\pi) \\ &= \xi_T(\pi)' \left(A(\pi)^{-1} \otimes H_2' H_2\right) \xi_T(\pi) \\ &= \xi_T(\pi)' \left(I_2 \otimes H_2\right)' \left(A(\pi)^{-1} \otimes I_{\ell-k}\right) \left(I_2 \otimes H_2\right) \xi_T(\pi). \end{aligned}$$

Since $H_2 H_2' = I_{\ell-k}$, it follows that $H_2 B_\ell(\pi) = B_{\ell-k}(\pi)$, a $(\ell - k)$-dimensional vector of independent standard Brownian motions and

$$\left(I_2 \otimes H_2\right) \xi_T(\pi) \Longrightarrow \left(I_2 \otimes H_2\right) J_\ell(\pi) = \begin{bmatrix} B_{\ell-k}(\pi) \\ B_{\ell-k}(1) - B_{\ell-k}(\pi) \end{bmatrix}$$

implying

$$\mathcal{S}_T^*(\pi) \Longrightarrow J_{\ell-k}(\pi)' \left(A(\pi) \otimes I_{\ell-k}\right)^{-1} J_{\ell-k}(\pi).$$

Finally, $\tau \left[\hat{\mathcal{D}}_{2,T}(\pi)\right] \Longrightarrow \tau \left[J_{\ell-k}(\pi)' \left(A(\pi) \otimes I_{\ell-k}\right)^{-1} J_{\ell-k}(\pi)\right]$ by applying the continuous mapping theorem. ■

Table 1: Testing $\mathcal{D}_1(\pi) = 0$ for $T = 400$; ε =trimming parameter

		Test	unsmoothed			smoothed		
DGP	ε	*Variant*	W	LM	LR	W	LM	LR
		Ave	5.0	5.6	5.2	6.6	8.7	7.0
1	0.15	Exp	7.3	7.4	5.5	9.9	14.4	9.3
		Sup	7.9	8.4	5.7	12.2	17.2	11.0
		Ave	5.2	5.4	5.3	6.4	7.2	6.1
1	0.20	Exp	7.6	6.8	6.1	8.7	10.9	8.3
		Sup	7.7	6.7	5.2	9.8	12.1	8.2
		Ave	5.2	5.3	5.0	5.6	6.2	5.6
1	0.25	Exp	6.5	5.8	5.7	8.1	8.3	7.1
		Sup	7.4	6.6	5.0	8.4	8.7	6.4
		Ave	5.4	4.9	4.8	5.6	5.9	5.7
1	0.30	Exp	5.9	5.3	5.0	7.2	6.9	6.4
		Sup	6.4	5.4	5.3	7.2	6.6	6.0
		Ave	4.9	4.7	4.8	5.4	5.3	4.8
1	0.35	Exp	5.3	4.9	4.9	6.9	6.0	6.0
		Sup	6.2	4.9	5.0	6.5	5.9	5.7
		Ave	5.2	5.0	4.9	5.5	4.9	4.9
1	0.40	Exp	5.3	4.9	5.1	6.2	5.6	5.6
		Sup	5.6	4.3	4.6	5.9	5.1	5.1
		Ave	5.6	5.0	5.1	5.9	5.3	5.2
1	0.45	Exp	5.9	5.0	5.2	5.9	5.4	5.4
		Sup	6.1	5.0	4.8	5.4	4.6	5.0
		Ave	3.9	4.0	4.3	5.9	9.1	7.4
2	0.15	Exp	5.2	4.7	5.5	9.0	15.5	11.7
		Sup	6.7	6.9	6.3	10.7	18.0	11.7
		Ave	3.7	3.7	4.4	6.2	6.8	6.6
2	0.20	Exp	5.8	4.7	5.5	8.2	11.0	9.5
		Sup	6.2	5.3	5.8	8.8	12.4	8.2
		Ave	3.9	3.7	4.3	5.7	6.2	6.5
2	0.25	Exp	5.2	3.4	5.3	7.2	8.3	7.6
		Sup	6.5	4.1	5.6	8.3	8.5	7.2
		Ave	4.4	3.5	4.0	6.4	5.4	6.5
2	0.30	Exp	4.9	3.6	4.6	7.2	7.3	6.8
		Sup	6.1	3.3	5.1	7.1	7.0	6.4
		Ave	4.4	3.5	4.0	5.7	5.0	5.4
2	0.35	Exp	4.9	3.8	4.5	6.8	5.7	6.1
		Sup	6.3	3.7	5.3	6.1	5.4	5.4
		Ave	4.0	3.4	4.0	5.4	5.0	5.1
2	0.40	Exp	4.1	3.4	4.3	6.3	5.6	5.8
		Sup	5.4	3.4	4.4	5.6	4.5	5.0
		Ave	4.2	3.3	4.0	5.9	4.9	5.3
2	0.45	Exp	4.3	3.3	4.0	6.2	4.8	5.7
		Sup	4.7	3.1	3.7	5.6	3.7	4.5

Notes: Test Variant refers to functionals $\tau[.]$ defined in (24) and W, LM, LR, denote EL-Wald, Lagrange Multiplier and EL-likelihood ratio statitstics defined in (22), (23) and (21), respectively (Section 4.1).

Table 2: Testing $\mathcal{D}_2(\pi) = 0$ for $T = 400$; ε =trimming parameter

		Test	unsmoothed			smoothed		
DGP	ε	*Variant*	*O*	*LM*	*LR*	*O*	*LM*	*LR*
		Ave	4.7	4.7	5.5	7.2	7.5	7.4
1	0.15	Exp	5.9	5.9	5.8	10.9	11.1	10.2
		Sup	5.5	5.5	5.4	11.6	11.7	11.2
		Ave	4.7	4.7	5.5	7.0	7.2	7.5
1	0.20	Exp	5.1	5.1	5.1	9.0	9.0	9.1
		Sup	5.1	5.1	5.4	8.7	8.7	8.8
		Ave	5.2	5.2	5.7	7.2	7.4	7.4
1	0.25	Exp	4.9	4.9	5.1	8.7	8.7	9.1
		Sup	5.2	5.2	5.5	7.3	7.3	8.1
		Ave	5.2	5.2	5.4	6.9	7.1	7.1
1	0.30	Exp	4.9	4.9	4.9	7.6	7.7	8.3
		Sup	4.8	4.8	4.8	7.2	7.2	7.7
		Ave	5.3	5.3	5.6	6.8	6.9	7.3
1	0.35	Exp	4.9	4.9	5.2	7.2	7.2	7.6
		Sup	4.7	4.7	5.3	6.5	6.5	7.3
		Ave	5.3	5.3	6.0	6.8	6.8	7.2
1	0.40	Exp	5.4	5.4	5.8	7.0	7.0	7.5
		Sup	5.1	5.1	5.3	6.7	6.7	7.0
		Ave	5.3	5.3	5.8	6.7	6.7	6.6
1	0.45	Exp	5.2	5.2	5.6	6.4	6.4	6.8
		Sup	5.5	5.5	5.7	6.2	6.2	6.3
		Ave	6.0	6.0	6.2	6.6	6.6	6.5
2	0.15	Exp	6.5	6.5	6.7	9.3	9.5	8.4
		Sup	6.5	6.5	6.7	10.0	10.2	9.8
		Ave	6.1	6.1	6.5	6.1	6.1	6.5
2	0.20	Exp	6.6	6.6	6.5	7.7	7.8	7.6
		Sup	6.2	6.2	6.3	8.0	8.1	8.4
		Ave	6.5	6.5	6.8	5.9	5.9	6.6
2	0.25	Exp	6.1	6.1	6.3	6.8	6.8	7.1
		Sup	5.7	5.7	6.0	6.8	6.8	7.1
		Ave	6.2	6.2	6.6	5.8	5.8	6.0
2	0.30	Exp	5.9	5.9	6.2	6.2	6.2	6.3
		Sup	5.2	5.2	5.8	6.5	6.5	6.7
		Ave	6.2	6.2	6.9	6.0	6.0	6.0
2	0.35	Exp	5.7	5.7	6.3	6.2	6.2	6.3
		Sup	5.3	5.3	5.5	5.9	5.9	5.7
		Ave	6.5	6.5	7.2	5.5	5.5	5.7
2	0.40	Exp	6.1	6.1	7.0	5.6	5.6	6.0
		Sup	4.9	4.9	5.2	5.2	5.2	5.8
		Ave	6.9	6.9	7.2	5.6	5.6	6.0
2	0.45	Exp	6.1	6.1	6.9	5.8	5.8	6.0
		Sup	5.4	5.4	5.7	4.7	4.7	5.2

Notes: Test Variant refers to functionals $\tau[.]$ defined in (24) and statistics O, LM, LR are defined in (26), (27) and (25), respectively (Section 4.2).

Table 3: Testing $\mathcal{D}(\pi) = 0$ for $T = 400$; ε =trimming parameter

		Test	unsmoothed			smoothed		
DGP	ε	*Variant*	W	LM	LR	W	LM	LR
1	0.15	Ave	4.3	4.8	4.6	8.4	10.6	9.5
		Exp	6.5	6.2	5.7	14.0	17.4	13.3
		Sup	7.3	7.1	6.1	15.1	18.4	14.2
1	0.20	Ave	4.1	4.3	4.5	8.3	8.7	9.0
		Exp	6.1	5.3	5.7	10.9	12.7	11.3
		Sup	7.2	6.1	5.2	11.5	13.8	11.0
1	0.25	Ave	4.2	4.4	4.4	7.7	8.4	8.4
		Exp	5.5	4.9	5.2	9.9	10.7	10.4
		Sup	6.2	4.9	4.6	9.5	10.6	9.3
1	0.30	Ave	4.3	4.2	4.9	7.9	8.3	8.5
		Exp	5.4	5.2	5.3	8.9	9.3	9.3
		Sup	5.3	3.9	4.4	8.1	8.0	7.4
1	0.35	Ave	4.6	4.4	4.8	7.9	7.5	7.6
		Exp	5.3	5.0	5.2	8.1	8.1	8.7
		Sup	5.0	3.9	4.1	7.7	6.6	7.1
1	0.40	Ave	4.7	4.4	4.7	7.4	6.9	7.1
		Exp	5.2	4.7	5.1	8.0	7.1	7.5
		Sup	4.9	4.4	4.5	7.0	6.2	6.8
1	0.45	Ave	5.1	4.8	5.1	7.3	7.2	7.1
		Exp	5.4	4.9	5.5	7.5	7.4	7.1
		Sup	5.5	5.0	5.7	6.4	5.8	6.2
2	0.15	Ave	5.5	5.5	6.4	7.5	8.8	7.9
		Exp	6.6	7.1	7.2	12.4	17.3	13.7
		Sup	7.1	7.7	7.2	13.7	19.6	15.3
2	0.20	Ave	5.4	5.6	6.0	6.9	7.8	7.5
		Exp	6.3	6.2	6.6	10.4	12.5	10.6
		Sup	6.5	6.2	6.2	11.0	13.0	11.1
2	0.25	Ave	5.9	5.7	6.0	7.0	6.5	6.6
		Exp	5.5	5.2	5.9	9.8	9.4	8.7
		Sup	5.9	4.8	5.9	9.9	9.9	9.4
2	0.30	Ave	6.2	6.0	6.4	6.8	6.5	6.9
		Exp	5.4	4.7	5.9	8.6	7.4	8.3
		Sup	5.3	4.3	4.7	8.4	7.7	8.4
2	0.35	Ave	5.8	5.2	6.2	6.5	5.9	6.7
		Exp	5.2	4.7	5.6	7.5	6.2	7.6
		Sup	4.9	4.2	4.6	7.0	6.3	6.9
2	0.40	Ave	5.3	4.7	5.6	6.0	5.9	6.3
		Exp	5.3	4.2	5.6	6.5	6.1	6.8
		Sup	5.0	3.6	4.6	6.2	5.6	6.1
2	0.45	Ave	5.4	4.5	5.5	6.0	5.6	5.9
		Exp	5.8	4.6	5.8	6.2	5.4	6.2
		Sup	5.1	3.4	4.8	5.5	4.8	5.6

Notes: Test Variant refers to functionals $\tau[.]$ defined in (24) and W, LM, LR denote the conbined statitstics defined in Section 4.3.

Table 4: Tests of $\mathcal{D}_1(\pi) = 0$ for $T = 600$; ε =trimming parameter

		Test	unsmoothed			smoothed		
DGP	ε	*Variant*	W	LM	LR	W	LM	LR
1	0.15	Ave	3.8	4.7	4.6	5.3	7.5	6.4
		Exp	5.3	5.4	5.6	7.3	10.7	7.7
		Sup	6.0	6.6	6.0	7.8	12.3	7.1
1	0.20	Ave	3.7	4.7	4.8	5.6	6.5	5.9
		Exp	5.3	5.7	5.6	6.8	8.5	7.4
		Sup	5.6	4.9	5.1	7.2	9.3	7.1
1	0.25	Ave	4.3	4.9	4.5	5.6	5.7	5.8
		Exp	4.8	5.2	5.1	6.8	7.3	6.9
		Sup	5.2	4.4	4.9	6.6	6.9	5.9
1	0.30	Ave	4.2	3.9	4.0	5.7	5.2	5.6
		Exp	4.6	4.2	4.5	6.4	5.9	6.2
		Sup	5.0	4.2	4.3	6.1	5.4	5.7
1	0.35	Ave	4.5	4.0	4.4	5.3	5.0	5.2
		Exp	5.0	4.1	4.2	5.5	4.8	5.1
		Sup	4.6	3.8	4.0	5.6	4.3	4.5
1	0.40	Ave	4.6	4.4	4.5	5.6	5.0	5.2
		Exp	4.5	3.9	4.3	5.8	5.0	5.3
		Sup	4.9	3.8	3.9	5.3	4.6	4.8
1	0.45	Ave	4.1	3.9	3.9	5.6	4.8	4.7
		Exp	4.1	3.9	4.0	5.8	4.7	4.8
		Sup	4.3	4.1	4.1	5.7	4.9	5.1
2	0.15	Ave	4.1	4.3	4.5	4.5	5.9	5.7
		Exp	4.6	4.7	5.0	6.5	8.9	6.8
		Sup	5.1	5.2	5.9	7.7	10.0	6.9
2	0.20	Ave	4.2	3.8	4.2	4.5	5.3	5.3
		Exp	4.7	4.3	4.9	5.9	6.5	5.9
		Sup	5.0	5.1	5.0	6.6	6.6	5.2
2	0.25	Ave	3.8	3.3	4.1	4.6	4.9	4.9
		Exp	4.5	3.2	4.7	5.7	5.4	5.6
		Sup	5.0	3.5	4.5	6.1	5.3	5.1
2	0.30	Ave	3.6	3.9	4.1	4.5	4.8	4.7
		Exp	4.2	3.1	4.1	5.0	5.1	5.4
		Sup	5.2	3.1	4.1	5.6	4.5	4.9
2	0.35	Ave	3.6	3.4	4.0	5.0	4.6	4.8
		Exp	3.9	3.1	3.3	4.8	4.5	4.7
		Sup	4.8	3.6	4.2	4.9	4.4	4.5
2	0.40	Ave	3.9	3.5	4.0	5.2	4.5	4.9
		Exp	4.0	3.5	3.9	5.6	4.8	4.9
		Sup	4.6	3.8	4.3	5.0	4.2	4.2
2	0.45	Ave	4.2	3.6	4.2	4.9	4.9	4.6
		Exp	4.1	3.4	3.9	4.5	4.5	4.5
		Sup	4.6	3.6	4.1	4.1	4.5	4.2

Notes: See Table 1 for definitions.

Table 5: Testing $\mathcal{D}_2(\pi) = 0$ for $T = 600$; ε =trimming parameter

		Test	unsmoothed			smoothed		
DGP	ε	*Variant*	*O*	*LM*	*LR*	*O*	*LM*	*LR*
1	0.15	Ave	4.7	4.7	4.7	4.7	4.7	5.1
		Exp	4.8	4.8	4.9	6.5	6.6	6.9
		Sup	5.4	5.4	5.4	6.7	6.7	6.7
1	0.20	Ave	4.4	4.4	4.6	4.6	4.6	4.8
		Exp	4.5	4.5	4.6	5.3	5.3	5.6
		Sup	4.6	4.6	5.1	5.1	5.1	5.3
1	0.25	Ave	4.7	4.7	4.8	4.6	4.6	5.2
		Exp	4.4	4.4	4.4	5.1	5.1	5.6
		Sup	4.4	4.4	4.4	4.5	4.5	4.8
1	0.30	Ave	4.5	4.5	4.6	4.8	4.8	5.1
		Exp	4.3	4.3	4.6	4.6	4.6	5.2
		Sup	3.7	3.7	3.9	4.0	4.0	4.9
1	0.35	Ave	4.3	4.3	4.4	4.7	4.7.	5.4
		Exp	4.3	4.3	4.3	4.8	4.8	5.2
		Sup	3.7	3.7	3.8	3.9	3.9	5.0
1	0.40	Ave	4.0	4.0	4.1	4.4	4.4	4.9
		Exp	4.5	4.5	4.7	4.7	4.7	5.8
		Sup	3.8	3.8	4.0	4.0	4.0	4.8
1	0.45	Ave	4.2	4.2	4.4	4.3	4.3	5.0
		Exp	4.1	4.1	4.4	4.7	4.7	5.2
		Sup	3.7	3.7	3.9	3.7	3.7	4.1
2	0.15	Ave	4.4	4.4	4.5	5.0	5.0	5.2
		Exp	4.5	4.5	5.3	6.3	6.4	6.7
		Sup	5.0	5.0	5.2	6.4	6.5	6.8
2	0.20	Ave	4.6	4.6	4.5	4.7	4.7	5.2
		Exp	3.9	3.9	4.7	5.3	5.3	5.8
		Sup	4.4	4.4	4.6	5.1	5.1	5.6
2	0.25	Ave	4.8	4.8	5.0	4.7	4.7	5.2
		Exp	4.7	4.7	5.0	5.1	5.1	5.6
		Sup	3.9	3.9	4.4	4.5	4.5	5.6
2	0.30	Ave	5.2	5.2	5.2	4.8	4.8	5.2
		Exp	5.0	5.0	5.3	5.0	5.0	5.5
		Sup	4.1	4.1	4.5	4.4	4.4	5.4
2	0.35	Ave	5.2	5.2	5.3	4.9	4.9	5.3
		Exp	5.0	5.0	5.3	5.0	5.0	5.2
		Sup	4.5	4.5	4.9	4.2	4.2	5.6
2	0.40	Ave	5.1	5.1	5.3	4.9	4.9	5.1
		Exp	5.1	5.1	5.2	5.1	5.1	5.3
		Sup	4.9	4.9	4.9	4.5	4.5	5.5
2	0.45	Ave	4.7	4.7	4.7	4.5	4.5	5.2
		Exp	4.8	4.8	5.1	4.7	4.7	5.2
		Sup	4.2	4.2	4.5	4.7	4.7	4.8

Notes: See Table 2 for definitions.

Table 6: Testing $\mathcal{D}(\pi) = 0$ for $T = 600$; ε =trimming parameter

		Test	unsmoothed			smoothed		
DGP	ε	*Variant*	*W*	*LM*	*LR*	*W*	*LM*	*LR*
1	0.15	Ave	4.0	4.0	4.5	5.4	5.6	5.6
		Exp	5.6	5.9	5.5	8.1	10.4	7.7
		Sup	5.9	6.7	5.8	8.7	10.4	8.6
1	0.20	Ave	4.3	4.4	4.9	5.1	5.5	5.4
		Exp	5.4	5.2	5.1	7.2	8.1	6.8
		Sup	5.5	5.5	5.2	7.1	8.2	7.0
1	0.25	Ave	4.2	4.4	4.5	5.4	5.3	5.6
		Exp	4.5	4.6	4.5	6.3	6.9	6.2
		Sup	5.1	4.4	4.6	6.5	7.1	6.3
1	0.30	Ave	4.0	4.2	4.7	5.5	5.4	5.8
		Exp	4.6	4.5	4.8	6.1	5.8	6.0
		Sup	4.7	3.4	3.5	5.5	5.7	5.4
1	0.35	Ave	3.9	4.0	4.4	5.7	5.3	5.7
		Exp	4.4	4.2	4.5	5.9	5.6	5.9
		Sup	4.4	3.4	3.7	5.3	4.4	5.0
1	0.40	Ave	3.8	3.7	3.9	5.2	5.1	5.3
		Exp	4.2	4.0	3.9	5.5	5.2	5.9
		Sup	4.0	3.5	3.8	4.9	4.4	4.9
1	0.45	Ave	4.1	3.8	3.9	5.8	5.8	5.9
		Exp	4.2	3.9	4.1	5.7	5.4	6.3
		Sup	4.4	4.0	4.2	5.0	4.2	4.9
2	0.15	Ave	3.9	4.2	4.5	5.3	5.7	5.6
		Exp	4.8	4.6	4.5	7.2	9.1	7.7
		Sup	5.5	5.0	5.5	8.7	10.6	9.5
2	0.20	Ave	3.5	4.0	4.2	5.0	5.4	5.3
		Exp	4.1	4.1	4.5	6.4	6.7	6.3
		Sup	5.0	4.6	4.6	6.9	7.4	7.3
2	0.25	Ave	3.5	3.8	4.3	4.7	4.5	5.2
		Exp	4.0	3.5	3.6	6.1	6.0	6.4
		Sup	4.2	3.8	4.0	6.3	6.0	6.1
2	0.30	Ave	4.2	4.2	4.4	5.0	4.8	5.1
		Exp	4.4	3.7	3.8	5.8	5.8	6.1
		Sup	4.3	3.4	3.9	6.1	4.7	5.3
2	0.35	Ave	3.7	3.8	4.2	4.6	4.5	5.0
		Exp	4.1	3.4	3.6	5.1	5.1	5.4
		Sup	4.5	3.5	3.5	4.7	4.2	4.5
2	0.40	Ave	3.9	3.6	4.0	4.9	4.9	5.1
		Exp	3.7	3.3	3.9	4.9	4.8	5.0
		Sup	3.8	3.0	3.6	4.4	4.1	4.3
2	0.45	Ave	4.2	4.2	4.1	4.9	4.8	5.0
		Exp	4.3	4.0	4.2	5.1	4.9	4.7
		Sup	4.0	3.5	4.2	4.2	3.8	4.0

Notes: See Table 3 for definitions.

References

Anatolyev, S. (2005). 'GMM, GEL, serial correlation, and asymptotic bias', *Econometrica*, 73: 983–1002.

Andrews, D. W. K. (1991). 'Heteroscedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica*, 59: 817–858.

——— (1993). 'Tests for parameter instability and structural change with unknown change point', *Econometrica*, 61: 821–856.

——— (2003). 'Tests for parameter instability and structural change with unknown change point: a corrigendum', *Econometrica*, 71: 395–398.

Andrews, D. W. K., and Fair, R. (1988). 'Inference in econometric models with structural change', *Review of Economic Studies*, 55: 615–640.

Andrews, D. W. K., and Ploberger, W. (1994). 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica*, 62: 1383–1414.

Cressie, N., and Read, T. R. C. (1984). 'Multinomial goodness-of-fit tests', *Journal of the Royal Statistical Society, series B*, 46: 440–464.

Ghysels, E., Guay, A., and Hall, A. R. (1997). 'Predictive test for structural change with unknown breakpoint', *Journal of Econometrics*, 82: 209–233.

Ghysels, E., and Hall, A. R. (1990). 'A test for structural stability of Euler condition parameters estimated via the Generalized Method of Moments', *International Economic Review*, 31: 355–364.

Golan, A. (2002). 'Information and entropy econometrics - editor's view', *Journal of Econometrics*, 107: 1–15.

——— (2006). 'Information and entropy econometrics - a review and synthesis', *Foundations and Trends in Econometrics*, 2: 1–145.

Guay, A., and Lamarche, J.-F. (2010). 'Structural change tests for GEL criteria', Discussion paper, Département de sciences économiques, Université du Québec à Montréal, Québec, Canada.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press, Oxford, U.K.

Hall, A. R., and Sen, A. (1999). 'Structural stability testing in models estimated by Generalized Method of Moments', *Journal of Business and Economic Statistics*, 17: 335–348.

Hansen, L. P. (1982). 'Large sample properties of Generalized Method of Moments estimators', *Econometrica*, 50: 1029–1054.

Kitamura, Y. (1997). 'Empirical likelihood methods for weakly dependent processes', *Annals of Statistics*, 25: 2084–2102.

——— (2006). 'Empirical likelihood methods in econometrics: theiry and practice', Discussion paper, Cowles Foundation Discussion Paper no. 1569, Yale University, New Haven, CT, USA.

Kitamura, Y., and Stutzer, M. (1997). 'An information-theoretic alternative to generalized method of moments estimation', *Econometrica*, 65: 861–874.

Li, Y. (2011). 'Empirical Likelihood with Applications in Time Series', Ph.D. thesis, Economics, School of Social Sciences, University of Manchester, Manchester, UK.

Newey, W. K., and Smith, R. J. (2004). 'Higher order properties of GMM and generalized empirical likelihood estimators', *Econometrica*, 72: 219–255.

Owen, A. (2001). *Empirical Likelihood*. Chapman Hall/CRC, New York, NY, USA.

Qin, J., and Lawless, J. (1994). 'Empirical likelihood and general estimating equations', *Annals of Statistics*, 22: 300–325.

Sen, A. (1997). 'New tests of structural stability and applications to consumption based asset pricing models', Ph.D. thesis, Department of Economics, North Carolina State University, Raleigh, NC.

Smith, R. J. (1997). 'Alternative semi-parametric likelihood approaches to generalized method of moments estimation', *Economics Journal*, 107: 503–519.

——— (2004). 'GEL criteria for moment condition models', Discussion paper, CENMAP working paper CWP19/04 1569, The Institute of Fiscal Studies, London, UK.

——— (2005). 'Automatic Positive Semidefinite HAC Covariance Matrix and GMM Estimation', *Econometric Theory*, 21: 158–170.

Sowell, F. (1996). 'Optimal tests of parameter variation in the Generalized Method of Moments framework', *Econometrica*, 64: 1085–1108.