Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Testing for Structural Instability in Moment Restriction Models: An Info-Metric Approach

Alastair R. Hall , Yuyi Li , Chris D. Orme & Arthur Sinko

To cite this article: Alastair R. Hall , Yuyi Li , Chris D. Orme & Arthur Sinko (2015) Testing for Structural Instability in Moment Restriction Models: An Info-Metric Approach, Econometric Reviews, 34:3, 286-327, DOI: 10.1080/07474938.2014.944477

To link to this article: https://doi.org/10.1080/07474938.2014.944477

Econometric Reviews, 34(3):286–327, 2015
Published with license by Taylor & Francis
ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2014.944477

Testing for Structural Instability in Moment Restriction Models: An Info-Metric Approach

Alastair R. Hall, Yuyi Li, Chris D. Orme, and Arthur Sinko
University of Manchester, Manchester, UK

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; and (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. The limiting distributions of our statistics are nonstandard 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 Generalized empirical likelihood; Moment condition models; Parameter variation; Structural instability.

JEL Classification C12; C32; C52.

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

Address correspondence to Alastair R. Hall, Economics, SoSS, University of Manchester, Manchester M13 9PL, UK; E-mail: alastair.hall@manchester.ac.uk

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 known break point case, Andrews and Fair (1988) introduced tests for parameter variation, and Ghysels and Hall (1990) introduced so-called predictive tests that Ghysels et al. (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 et al. (1997) propose extensions of the predictive test to this setting. Building from these earlier results, 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 that 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 can yield unreliable inferences 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 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

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

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 Cressie–Read (CR) distance measure (Cressie and Read, 1984). Like Guay and Lamarche (2010), we assume the data to be weakly dependent and account for this dependence in estimation using the kernel-smoothing methods advocated by Smith (2011).

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 derives the required the first order asymptotic properties of the partial info-metric – estimators under null of structural stability – and are employed in Section 4, which 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 summarizes results from a small simulation study that indicates the finite sample performance of our methods. Section 6 concludes the paper. 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 $d \times 1$

[2]Our results are based on Li's (2011) Ph.D. thesis, which considered only the EL framework. This work was performed independently of and contemporaneously to Guay and Lamarche (2010).

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 $\Re^d$, with

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

and

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

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

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

in which $D(\cdot \,\|\, \cdot)$ is a distance, or 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

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

where $\pi \in (0, 1)$ is referred to as the break-fraction. 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 first treat π as fixed and then to discuss the extension to the unknown break point case at the end of the section.

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

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

$$\mathbf{P}_0 = \bigcup_{\beta \in \mathscr{B}} \mathbf{P}_0(\beta),$$

where

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

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 subsamples. 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

$$\mathbf{P}_1 = \bigcup_{(\beta_1, \beta_2) \in \mathscr{B} \times \mathscr{B}} \mathbf{P}_1(\beta_1, \beta_2),$$

where

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

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 subsamples 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 $(Z^{(1)}, Z^{(2)})$, with the null being

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

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 subsamples 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 subsamples – 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 situations:

(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

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

(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

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

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 $\mathscr{D}(\pi) = \rho_\pi([\beta_*(\pi), \beta_*(\pi)], [\mu_1, \mu_2])$ into two parts

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

where

$$\mathscr{D}_1(\pi) = \rho_\pi([\beta_*(\pi), \beta_*(\pi)], [\mu_1, \mu_2]) - \rho_\pi([\beta_{1,*}(\pi), \beta_{2,*}(\pi)], [\mu_1, \mu_2]),$$
$$\mathscr{D}_2(\pi) = \rho_\pi([\beta_{1,*}(\pi), \beta_{2,*}(\pi)], [\mu_1, \mu_2]).$$

It can be recognized that: if $H_0(\pi)$ is true, then $\mathscr{D}_1(\pi) = \mathscr{D}_2(\pi) = 0$; if $H_{PV}(\pi)$ is true, then $\mathscr{D}_1(\pi) \neq 0$ but $\mathscr{D}_2(\pi) = 0$; if $H_{MS}(\pi)$ is true, then $\mathscr{D}_1(\pi) \neq 0$ and $\mathscr{D}_2(\pi) \neq 0$. Therefore, an examination of $\mathscr{D}(\pi)$ reveals whether the model is structurally stable, $H_0(\pi)$, or not, $H_A(\pi)$. On the other hand, an examination of $\mathscr{D}_1(\pi)$ and $\mathscr{D}_2(\pi)$ 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 $\mathscr{D}(\pi)$, $\mathscr{D}_1(\pi)$, and $\mathscr{D}_2(\pi)$.

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 subsamples of T_1 and T_2 observations, respectively, where $\mathscr{T}_1(\pi) = \{1, 2, \ldots, T_1\}$, denotes the set of T_1 observations up to and including the break point and $\mathscr{T}_2(\pi) = \{T_1 + 1, T_1 + 2, \ldots T\}$, the set of T_2 observations after the break with $T_2 = T - T_1$.

It is well known that IT methods based on the assumption of independently and identically distributed data are asymptotically inefficient if the data are weakly dependent.[4] Various approaches have been proposed for handling this dependence: we employ quite general kernel smoothing methods as developed by Smith (2011).[5] Within

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

[5]Kitamura and Stutzer (1997) handle dependency via smoothing using a rectangular kernel, as well as blocking methods (see also Kitamura, 1997); Kitamura (2006) uses parametric models.

this approach, the original moment function in period t, $g(Z_t, \beta) = g_t(\beta)$ 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 with h_T and $k(\cdot)$ denoting the bandwidth and a kernel function, respectively, details of which are given 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 $\widehat{P}_1 = [p_{1,1}, p_{1,2} \ldots, p_{1,T_1}]$, $\widehat{P}_2 = [p_{2,1}, p_{2,2} \ldots, p_{2,T_2}]$.

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

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

and

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

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 $\mathcal{D}_2(\pi)$,

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

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

$$\widehat{\mathcal{D}}_{2,T}(\pi) = \rho_{\pi,T}([\hat{\beta}_1(\pi), \hat{\beta}_2(\pi)], [\hat{\mu}_1, \hat{\mu}_2]). \tag{12}$$

[6] See Smith (2011, p. 1195).

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 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 given as follows in our context:

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

and which is defined for $-\infty < \alpha < \infty$. Appropriate choices of α lead to certain familiar estimation methods: for example, $\lim_{\alpha \to 0} D_{\pi}^{(\alpha)}(\cdot \parallel \cdot)$ yields the optimand for the ET estimator of Kitamura and Stutzer (1997) in each subsample, and $\lim_{\alpha \to -1} D_{\pi}^{(\alpha)}(\cdot \parallel \cdot)$ yields the EL estimator of Owen (2001) in each subsample.

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, and their asymptotic properties under the null hypothesis, are developed in Section 4.

In the following section, we first derive the first order asymptotic behavior of the unrestricted and restricted partial-sample IT estimators under the null hypothesis.

3. LARGE SAMPLE BEHAVIOR OF PARTIAL-SAMPLE IT ESTIMATORS

For the purposes of developing the asymptotic theory underpinning the partial-sample IT estimators, it is convenient to exploit the equivalence between GEL estimation and that of an IT approach based on the CR divergence measure. That is, any such IT estimator has a GEL equivalent; see Newey and Smith (2004). As discussed in Newey and Smith (2004), and also Smith (2011), let $\rho(v)$ be a continuous, twice differentiable and concave function on its domain $\mathcal{V}$, an open interval containing 0. Let $\rho_j(v) \equiv \partial^j \rho(v)/\partial v^j$, $\rho_j = \rho_j(0)$ for $j = 0, 1, 2, \ldots$, and impose the normalisation that $\rho_1 = \rho_2 = -1$. Then, based on the full sample, the GEL (IT) criterion function would be[7]

$$Q_T(\beta, \lambda) = \frac{1}{T} \sum_{t=1}^{T} [\rho(k\lambda' g_t^s(\beta)) - \rho_0],$$

[7] We adopt the notation $Q_T(\beta, \lambda)$ rather than $\widehat{P}(\beta, \lambda)$, employed by Smith (2011), to avoid confusion with P as discussed in Section 2.

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 $\beta \in \mathscr{B} \subset \Re^k$, the auxiliary GEL parameters $\lambda \in \Lambda_T$ are restricted so that w.p.a. 1 (with probability approaching 1) $k\lambda' g_t^s(\beta) \in \mathscr{V}$, for all $(\beta', \lambda')' \in \mathscr{B} \times \Lambda_T$ and $t = 1, \ldots, T$. Specifically, Λ_T imposes bounds on λ that "shrink" with T, but at a slower rate than $h_T/\sqrt{T}$ (see Assumption 4) which is the convergence rate of both the GEL and partial-sample GEL estimator for λ.

The (full-sample) GEL estimator is then defined as

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

Estimation proceeds in the following two steps:

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

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

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

$$\tilde{\beta} = \arg\min_{\beta \in \mathscr{B}} Q_T(\beta, \tilde{\lambda}(\beta)),$$

and $\tilde{\lambda} \equiv \tilde{\lambda}(\tilde{\beta})$.

Whilst still employing $g_t^s(\beta)$, consider, now, splitting the sample according to $\mathscr{T}_i(\pi)$, $i = 1, 2$, for all $\pi \in \Pi$, to obtain the (unrestricted) partial-sample GEL (PSGEL) estimators $\hat{\beta}_i(\pi)$, $i = 1, 2$, based on the two subsamples $t \in \mathscr{T}_i(\pi)$, $i = 1, 2$, respectively.[8] Specifically,

$$\hat{\beta}_i(\pi) = \arg\min_{\beta \in \mathscr{B}} \sup_{\lambda \in \Lambda_T} \frac{1}{T} \sum_{t \in \mathscr{T}_i(\pi)} [\rho(k\lambda' g_t^s(\beta)) - \rho_0], \qquad i = 1, 2,$$

and, correspondingly,

$$\hat{\lambda}_i(\pi) = \arg\sup_{\lambda \in \Lambda_T} \frac{1}{T} \sum_{t \in \mathscr{T}_i(\pi)} [\rho(k\lambda' g_t^s(\hat{\beta}_i(\pi))) - \rho_0], \qquad i = 1, 2.$$

[8] To present the main results, the moment functions are smoothed before splitting the sample according to π. Another possible avenue is to smooth the moment functions after splitting the sample. Indeed, the latter might be viewed as more natural and this is pursued in the Monte Carlo study, Section 5. However, whilst there is no difference asymptotically between the two approaches, the proofs are more straightforward in the former case.

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

$$g_t(\theta, \pi) = \mathbb{I}_{t,T}(\pi) \begin{pmatrix} g_t(\beta_1) \\ 0 \end{pmatrix} + (1 - \mathbb{I}_{t,T}(\pi)) \begin{pmatrix} 0 \\ g_t(\beta_2) \end{pmatrix},$$
$$g_t^s(\theta, \pi) = \mathbb{I}_{t,T}(\pi) \begin{pmatrix} g_t^s(\beta_1) \\ 0 \end{pmatrix} + (1 - \mathbb{I}_{t,T}(\pi)) \begin{pmatrix} 0 \\ g_t^s(\beta_2) \end{pmatrix}, \tag{14}$$

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

$$Q_T(\theta, \gamma, \pi) = \frac{1}{T} \sum_{t=1}^{T} [\rho(k\gamma' g_t^s(\theta, \pi)) - \rho_0].$$

Then we have $\hat{\theta}(\pi) = (\hat{\beta}_1(\pi)', \hat{\beta}_2(\pi)')'$, where

$$\hat{\theta}(\pi) = \arg\min_{\theta \in \Theta} \sup_{\gamma \in \Gamma_T} Q_T(\theta, \gamma, \pi) \tag{15}$$

with

$$\hat{\gamma}(\pi) = \arg\sup_{\gamma \in \Gamma_T} Q_T(\hat{\theta}(\pi), \gamma, \pi). \tag{16}$$

Throughout this paper, the asymptotic analysis addresses behavior under the null hypothesis, only, and requires certain assumptions that follow the spirit of Smith (2011). The data satisfy the following condition.

Assumption 1. *Data are generated by a sequence of strictly stationary and strong mixing* $\mathbf{Z}$*-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* $\mathbf{Z}$ *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, $k(\cdot)$, 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 $k(\cdot)$, 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)* $k(\cdot) : \Re \to [-k_{\max}, k_{\max}]$, $k_{\max} < \infty$, $k(0) \neq 0$, $k_1 \neq 0$, *and* $k(\cdot)$ *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 (2011), 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: $\bar{g}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\beta)$, $\Omega(\beta) = \lim_{T\to\infty} var(\sqrt{T}\bar{g}_T(\beta))$, 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 $\bar{g}^s_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} g^s_t(\beta)$ and $\bar{g}^s_{[T\pi]}(\beta) = \frac{1}{T}\sum_{t=1}^{[T\pi]} g^s_t(\beta)$, respectively.

Assumption 3. *(i)* $E[\sup_{\beta\in\mathscr{B}} ||g_t(\beta)||^{\eta}] < \infty$ *for some* $\eta > \max[4\upsilon, \frac{2\delta}{\delta-1}]$. *(ii)* $\Omega(\beta)$ *is finite and p.d. for all* $\beta \in \mathscr{B} \subset \Re^k$, *where* $\mathscr{B}$ *is a compact parameter set. (iii) The moment function* $g(z, \beta) \subset \Re^{\ell}$ *is continuous in* z *for all* $\beta \in \mathscr{B}$, *and is continuous at each* $\beta \in \mathscr{B}$ *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) ensures the population moment condition is satisfied at β_0 and also provides a global identification condition. Assumptions 1–3 ensure that an appropriate Functional Central Limit Theorem (FCLT) applies to both $\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)$, with $\lim_{T\to\infty} var(\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)) = \pi\Omega_0$, and $\sqrt{T}\bar{g}^s_{[T\pi]}(\beta_0)$, with $\lim_{T\to\infty} var(\sqrt{T}\bar{g}^s_{[T\pi]}(\beta_0)) = k_1\pi\Omega_0$, for all $\pi \in [0, 1]$, where $\Omega_0 = \Omega(\beta_0)$. These assumptions also ensure that a (weak) Uniform Law of Large Numbers (ULLN) not only applies to $\bar{g}_T(\beta)$, but also to both $\bar{g}_T(\theta, \pi) \equiv \frac{1}{T}\sum_{t=1}^{T} g_t(\theta, \pi)$ and $\bar{g}^s_T(\theta, \pi) \equiv \frac{1}{T}\sum_{t=1}^{T} g^s_t(\theta, \pi)$, with the latter two being uniform over $\pi \in [0, 1]$.[9]

The following assumption formally imposes the restrictions on $\rho(\cdot)$ and also restricts the bounds on λ, ensuring that they shrink to zero more slowly than the stochastic rate of convergence of both $\tilde{\lambda}$ and $\hat{\gamma}(\pi)$.

Assumption 4. *(a)* $\rho(v)$ *is a continuous, twice differentiable and concave function on its domain* $\mathscr{V}$, *an open interval containing 0, such that* $\rho_1 = \rho_2 = -1$. *(b)* $\lambda \in \Lambda_T = \{\lambda : \|\lambda\| \leq B(T/h_T^2)^{-\varepsilon}\}$, *where* $\frac{\delta}{\eta(\delta-1)} < \varepsilon < \frac{1}{2}$, *for some finite* $B > 0$.

Under the above assumptions, we can establish the consistency of the PSGEL estimator as follows.

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

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

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 $G(\beta) = E[G_t(\beta)]$, 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 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 $M_0 = \Omega_0^{-1/2}G_0$, $P_0 = M_0(M_0'M_0)^{-1}M_0'$. 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

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

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 (2011, Lemma A3) establish that $h_T\bar{V}_T^s(\theta_0,\pi) \xrightarrow{p} k_2\Omega_0(\pi)$, uniformly in π, where

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

Theorem 2. *Under Assumptions 1–5, every sequence of PSGEL estimators defined by (15) and (16),* $T \geq 1$, *satisfies*

$$\begin{aligned} \sqrt{T}(\hat{\theta}(\pi) - \theta_0) &= -(A(\pi)^{-1} \otimes (M_0'M_0)^{-1}M_0')\xi_T(\pi) + o_{p\pi}(1) \\ &\Longrightarrow -(A(\pi)^{-1} \otimes (M_0'M_0)^{-1}M_0')J_\ell(\pi), \\ \left(\sqrt{T}/h_T\right)\hat{\gamma}(\pi) &= -(A(\pi)^{-1} \otimes \Omega_0^{-1/2}(I_\ell - P_0))\xi_T(\pi) + o_{p\pi}(1) \\ &\Longrightarrow -(A(\pi)^{-1} \otimes \Omega_0^{-1/2}(I_\ell - P_0))J_\ell(\pi), \end{aligned}$$

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

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

These results ensure that, from Smith (2005, Theorem 2.1),

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

and

$$\sup_{\pi\in\Pi} \left\| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial g_t^s(\hat{\theta}(\pi), \pi)}{\partial \theta'} - k_1 G_0(\pi) \right\| = o_p(1).$$

The next Theorem details the asymptotic distribution of the restricted PSGEL estimators, which are constructed as follows. Define the restricted $(2\ell \times 1)$ smoothed moment function as

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

so that, from (14), $g_t^s((\beta', \beta')', \pi) \equiv \dot{g}_t^s(\beta, \pi)$, and let $\dot{Q}_T(\beta, \gamma, \pi) = \frac{1}{T}\sum_{t=1}^{T}[\rho(k\gamma\dot{g}_t^s(\beta, \pi)) - \rho_0]$, then the restricted PSGEL estimators are defined by

$$\begin{aligned}\tilde{\beta}(\pi) &= \arg\min_{\beta\in\mathscr{B}} \sup_{\gamma\in\Gamma_T} \dot{Q}_T(\beta, \gamma, \pi) \\ &= \arg\min_{\beta\in\mathscr{B}} \left\{ \sup_{\lambda\in\Lambda_T} \frac{1}{T}\sum_{t=1}^{[T\pi]}[\rho(k\lambda' g_t^s(\beta)) - \rho_0] \right. \\ &\qquad \left. + \sup_{\lambda\in\Lambda_T} \frac{1}{T}\sum_{t=[T\pi]+1}^{T}[\rho(k\lambda' g_t^s(\beta)) - \rho_0] \right\}\end{aligned}$$

and

$$\tilde{\gamma}(\pi) = \arg\sup_{\gamma\in\Gamma_T} \frac{1}{T}\sum_{t=1}^{T}[\rho(k\gamma\dot{g}_t^s(\tilde{\beta}(\pi), \pi)) - \rho_0],$$

so that

$$\tilde{\lambda}_1(\pi) = \arg\sup_{\lambda\in\Lambda_T} \frac{1}{T}\sum_{t=1}^{[T\pi]}[\rho(k\lambda' g_t^s(\tilde{\beta}(\pi))) - \rho_0],$$

$$\tilde{\lambda}_2(\pi) = \arg\sup_{\lambda\in\Lambda_T} \frac{1}{T}\sum_{t=[T\pi]+1}^{T}[\rho(k\lambda' g_t^s(\tilde{\beta}(\pi))) - \rho_0].$$

Theorem 3. *Under Assumptions 1–5, every sequence of restricted PSGEL estimators, $T \geq 1$, satisfies*

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

and

$$\begin{aligned}\left(\sqrt{T}/h_T\right)\tilde{\gamma}(\pi) &= -(A(\pi)^{-1} - \imath_2\imath_2' \otimes \Omega_0^{-1/2}(I_\ell - P_0))\xi_T(\pi) + o_{p\pi}(1)\\ &= -\frac{1}{\pi(1-\pi)}(a(\pi)\otimes\Omega_0^{-1/2})(I_\ell - P_0)(a(\pi)'\otimes I_\ell)\xi_T(\pi) + o_{p\pi}(1)\\ &\Longrightarrow -(A(\pi)^{-1} - \imath_2\imath_2' \otimes \Omega_0^{-1/2}(I_\ell - P_0))J_\ell(\pi)\\ &= (\imath_2 \otimes \Omega_0^{-1/2}P_0)B_\ell(1) + (A(\pi)^{-1}\otimes\Omega_0^{-1/2})J_\ell(\pi),\end{aligned}$$

where $a(\pi)' = (1-\pi, -\pi)$, and $\imath_2 = (1,1)'$.

4. TESTING STRUCTURAL STABILITY

In this section, we propose tests based on GEL 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 $\mathscr{D}_1(\pi) = 0$, Section 4.2 presents tests for that $\mathscr{D}_2(\pi) = 0$, and Section 4.3 presents tests for $\mathscr{D}(\pi) = 0$. 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.

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

To test $\mathscr{D}_1(\pi) = 0$ for a fixed π, the obvious statistic is the GEL-likelihood ratio statistic (c.f. Smith, 2011, p. 1208)

$$\mathscr{LR}_T(\pi) = 2(k_2/k_1^2)(T/h_T)\{\dot{Q}_T(\tilde{\beta}(\pi), \tilde{\gamma}(\pi), \pi) - Q_T(\hat{\theta}(\pi), \hat{\gamma}(\pi), \pi)\}. \quad (17)$$

In view of extant results in the GEL literature on testing parametric restrictions,[10] we also consider inference based on the GEL-Wald statistic for testing $\beta_1 = \beta_2$,

$$\mathscr{W}_T(\pi) = (k_2/k_1^2)(T/h_T)(\hat{\beta}_1(\pi) - \hat{\beta}_2(\pi))'\{V_T^W(\hat{\theta}(\pi))\}^{-1}(\hat{\beta}_1(\pi) - \hat{\beta}_2(\pi)), \tag{18}$$

where

$$V_T^W(\theta) = \sum_{i=1}^{2}\{\bar{G}_{T_i}^s(\beta_i)'\{\bar{V}_{T_i}^s(\beta_i)\}^{-1}\bar{G}_{T_i}^s(\beta_i)\}^{-1},$$
$$\bar{G}_{T_i}^s(\beta) = \frac{1}{T}\sum_{t\in\mathcal{T}_i(\pi)}\frac{\partial g_t^s(\beta)}{\partial\beta'}, \qquad \bar{V}_{T_i}^s(\beta) = \frac{1}{T}\sum_{t\in\mathcal{T}_i(\pi)} g_t^s(\beta)g_t^s(\beta)',$$

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

$$\mathscr{LM}_T(\pi) = (k_2/k_1^2)(T/h_T)\tilde{\zeta}(\pi)'\{V_T^{\zeta}(\tilde{\beta}(\pi))\}^{-1}\tilde{\zeta}(\pi)/(\pi(1-\pi)), \tag{19}$$

where

$$V_T^{\zeta}(\beta) = \bar{G}_T^s(\beta)'\{\bar{V}_T^s(\beta)\}^{-1}\bar{G}_T^s(\beta),$$
$$\bar{G}_T^s(\beta) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_t^s(\beta)}{\partial\beta'}, \qquad \bar{V}_T^s(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t^s(\beta)g_t^s(\beta)'.$$

Henceforth, let $\widehat{\mathscr{D}}_{1,T}(\pi)$ denote any one of the statistics in (17), (18), or (19).[11]

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 $\widehat{\mathscr{D}}_{1,T}(\pi)$,

$$\tau[\widehat{\mathscr{D}}_{1,T}(\pi)] = \begin{cases} \sup_{\pi\in\Pi}\widehat{\mathscr{D}}_{1,T}(\pi) \equiv \sup\widehat{\mathscr{D}}_{1,T}(\pi) \\ \int_{\Pi}\widehat{\mathscr{D}}_{1,T}(\pi)dN(\pi) \equiv \text{ave}\,\widehat{\mathscr{D}}_{1,T}(\pi) \\ \log\left\{\int_{\Pi}\exp\left\{\frac{1}{2}\widehat{\mathscr{D}}_{1,T}(\pi)\right\}dN(\pi)\right\} \equiv \exp\widehat{\mathscr{D}}_{1,T}(\pi), \end{cases} \tag{20}$$

where $N(\pi)$ defines the prior distribution for the break point $\pi \in \Pi$, which we will assume to be uniform.[12] The following theorem shows each of these test statistics are (first order) asymptotically equivalent, for different choices of $\widehat{\mathscr{D}}_{1,T}(\pi)$ and common choice of functional $\tau[\cdot]$.

[10] See Qin and Lawless (1994) and Smith (2011).

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

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

Theorem 4. *Under the null of $\mathscr{D}_1(\pi) = 0$ and Assumptions 1–5, we have*

$$\sup_{\pi \in \Pi} |\widehat{\mathscr{D}}_{1,T}(\pi) - \mathscr{S}_T(\pi)| = o_p(1),$$

where

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

$B_k(\pi) - \pi B_k(1)$ is a vector of Brownian bridges and $B_k(\pi)$ is a vector of k independent standard Brownian motions

An immediate consequence of the Continuous Mapping Theorem (CMT) is that

$$\tau[\widehat{\mathscr{D}}_{1,T}(\pi)] \Longrightarrow \tau[W_k(\pi)]$$

for each functional (20).

4.2. Testing $\mathscr{D}_2(\pi)$

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

$$\mathscr{LR}_T^*(\pi) = 2(k_2/k_1^2)(T/h_T)Q_T(\hat{\theta}(\pi), \hat{\gamma}(\pi), \pi). \tag{21}$$

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

$$\mathscr{O}_T(\pi) = (k_2/k_1^2)(T/h_T)\bar{g}_T^s(\hat{\theta}(\pi), \pi)'\{\bar{V}_T^s(\hat{\theta}(\pi), \pi)\}^{-1}\ \bar{g}_T^s(\hat{\theta}(\pi), \pi) \tag{22}$$

$$\mathscr{LM}_T^*(\pi) = (T/h_T)\hat{\gamma}(\pi)'\{\bar{V}_T^s(\hat{\theta}(\pi), \pi)\}\hat{\gamma}(\pi)/k_2. \tag{23}$$

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

Letting $\widehat{\mathscr{D}}_{2,T}(\pi)$ denote any one of (21), (22) or (23),[13] we use similar ideas to the previous sub-section to test $\mathscr{D}_2(\pi)$ for all $\pi \in \Pi$ based on $\tau[\widehat{\mathscr{D}}_{2,T}(\pi)]$. The limiting distribution of the latter statistic is given in the following theorem.

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

Theorem 5. *Under the null of* $\mathscr{D}_2(\pi) = 0$ *and Assumptions 1–5, we have*

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

where

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

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.*

Again, the CMT implies that $\tau[\widehat{\mathscr{D}}_{2,T}(\pi)] \Longrightarrow \tau[W_{\ell-k}^*(\pi)]$.

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

Given the discussion in Section 2, testing $\mathscr{D}(\pi) = 0$ can be achieved by employing statistics which are functionals of the processes, $\widehat{\mathscr{D}}_{1,T}(\pi)$ and $\widehat{\mathscr{D}}_{2,T}(\pi)$. Specifically, we consider the combined process $\widehat{\mathscr{D}}_T(\pi) = \widehat{\mathscr{D}}_{1,T}(\pi) + \widehat{\mathscr{D}}_{2,T}(\pi)$ for the choices of $\widehat{\mathscr{D}}_{1,T}(\pi)$ and $\widehat{\mathscr{D}}_{2,T}(\pi)$ defined in Sections 4.1 and 4.2, respectively, and the functionals $\tau[\widehat{\mathscr{D}}_T(\pi)]$, defined by (20). Then, we have the following Corollary to Theorems 4 and 5, which implies that $\tau[\widehat{\mathscr{D}}_T(\pi)] \Longrightarrow \tau[W_k(\pi) + W_{\ell-k}^*(\pi)]$.

Corollary 1. *Under the null of* $\mathscr{D}(\pi) = 0$ *and Assumptions 1–5, we have*

$$\sup_{\pi\in\Pi} |\widehat{\mathscr{D}}_T(\pi) - \mathscr{S}_T(\pi) - \mathscr{S}_T^*(\pi)| = o_p(1).$$

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 4 and 5 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* $\widehat{\mathscr{D}}_{1,T}(\pi) \xrightarrow{d} \chi_k^2$, $\widehat{\mathscr{D}}_{2,T}(\pi) \xrightarrow{d} \chi_{2(\ell-k)}^2$, *and* $\widehat{\mathscr{D}}_T(\pi) \xrightarrow{d} \chi_{2\ell-k}^2$, *where* $\widehat{\mathscr{D}}_{1,T}(\pi)$, $\widehat{\mathscr{D}}_{2,T}(\pi)$ *and* $\widehat{\mathscr{D}}_T(\pi)$ *are defined in Sections 4.1, 4.2 and 4.3, respectively, and* χ_v^2 *denotes a chi-squared distribution with* v *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 $\mathcal{D}_2(\pi) = 0$.[14] 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. Furthermore, this advantage extends to the partial sum estimators which also have an informational interpretation within our IT framework for structural change.

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.[15] 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 4 continues to hold under local alternatives to the moment condition that do not involve parameter variation, and Theorem 5 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 $\widehat{\mathcal{D}}_{1,T}(\pi)$ and $\widehat{\mathcal{D}}_{2,T}(\pi)$ have the potential to reveal when the instability is confined to parameter variation alone.

Finally we note that the assumption of strict stationarity (Assumption 1) is sufficient but not necessary for the limiting distributions stated in Theorems 4 and 5. These results would still apply provided the Jacobian and the long run variance are homogenous across the sub-samples and we can apply an FCLT to the sample moment and ULLN to certain functions of data. However, if the Jacobian, say, changes at some point in the sample then the limiting distributions are not anticipated to hold for the same reasons as those

[14]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.

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

diagnosed in Hansen's (2000) analysis of the sup-test in the linear regression model when there is a shift in the marginal distribution of the regressors.

5. MONTE CARLO EVIDENCE

In this section, we report simulation results that give insight into the finite sample performance of the IT-based tests for the special cases of EL[16] and ET.

Following Ghysels et al. (1997) and Hall and Sen (1999), we consider the slightly modified data generation process

$$\begin{aligned} x_t &= \beta_1 x_{t-1} + u_t + \alpha u_{t-1}, \quad u_t \sim IN(0,1), \quad &&\text{for } t = 1, 2, \dots, T/2 \\ x_t &= \beta_2 x_{t-1} + u_t + \alpha u_{t-1}, \quad u_t \sim IN(0,1), \quad &&\text{for } t = T/2+1, T/2+2, \dots, T, \end{aligned}$$

and corresponding 2×1 vector of "instruments," $z_t = (z_{t,1}, z_{t,2})'$. We suppose that the researcher estimates an $AR(1)$ model for x_t based on the moment condition $E[g_t(\beta_0)] = 0$, where

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

Eight Data Generation Processes (DGPs) are employed defined by the choice of parameter values $(\beta_1, \beta_2, \alpha)$ and instruments z_t. They are as follows: DGP_1, DGP_2, and DGP_3 model a situation with no breaks ($\beta_1 = \beta_2 = 0.4$) and valid instruments (x_{t-2}, x_{t-3}); DGP_4, DGP_5, DGP_6 model a structural break in the data through parameter variation ($\beta_1 = 0.4$, $\beta_2 = 0.8$), but the instruments (x_{t-2}, x_{t-3}) remain valid; whilst DGP_7 and DGP_8 model situations when there is misspecification through both parameter variation ($\beta_1 = 0.4$, $\beta_2 = 0.8$) and invalid instruments (x_{t-1}, x_{t-2}). The remaining difference is through the value of α: for DGP_i, $\alpha = \alpha_i$ where $\alpha_i = 0$ for $i = 1, 4$, $\alpha_i = 0.4$ for $i = 2, 5, 7$, and $\alpha_i = 0.8$ for $i = 3, 6, 8$. Although we discuss results from all eights DGP's below, we only explictly report results for the DGP's with $\alpha = 0.4$; the remaining results are available in the working paper version of this paper, see Hall et al. (2013).

The sampling experiments consider four different sample sizes of $T = 200, 400, 800$, 1,600, where in each case the various test statistics are constructed employing the following estimation procedures: (i) EL; (ii) kernel-smoothed empirical likelihood (*ELk*); (iii) kernel-smoothed exponential tilting (*ETk*); and, (iv) asymptotically efficient (kernel-smoothed) GMM, exploiting kernel-smoothed HAC estimation (*GMMk*). For each of the IT estimators (models (i)–(iii)), we calculate the following statistics: $\tau[\widehat{\mathscr{D}}_{1,T}(\pi)]$, $\tau[\widehat{\mathscr{D}}_{2,T}(\pi)]$ and $\tau[\widehat{\mathscr{D}}_T(\pi)]$ for the three functionals $\tau[\cdot]$ defined in (20) and $\widehat{\mathscr{D}}_{1,T}(\pi)$ given by (17)–(19), $\widehat{\mathscr{D}}_{2,T}(\pi)$ given by (21)–(23) and $\widehat{\mathscr{D}}_T(\pi) = \widehat{\mathscr{D}}_{1,T}(\pi) + \widehat{\mathscr{D}}_{2,T}(\pi)$, being $\mathscr{LR}_T(\pi) + \mathscr{LR}_T^*(\pi)$, $\mathscr{W}_T(\pi) + \mathscr{O}_T(\pi)$, or $\mathscr{LM}_T(\pi) + \mathscr{LM}_T^*(\pi)$, respectively. For the *GMMk* estimator only the

[16]To speed up the simulation process, we adopt a modified version of EL estimator proposed by Owen. To avoid $-\infty$, $\log(x)$ for $x < 1/T$ is replaced by a second degree polynomial.

Wald statistic is considered. All these statistics are calculated using $\Pi = [\varepsilon, 1-\varepsilon]$, for a trimming parameter $\varepsilon = 0.20$, and, for each DGP and sample size, sampling results are obtained from 1,000 replications employing a 5% nominal significance level for each test procedure.[17]

We report unsmoothed (*EL*) and smoothed (*ELk*, *ETk*, and *GMMk*) versions of the test statistics. In the latter case and exploiting Lemma 3 in the Appendix, the moment condition is smoothed separately in each sub-sample defined by π, but with common bandwidth h_T.[18] That is

$$g_t^s(\beta) = \begin{cases} \dfrac{1}{h_T} \displaystyle\sum_{j=t-[T\pi]}^{t-1} k\left(\dfrac{j}{h_T}\right) g_{t-j}(\beta), & t = 1, \ldots, [T\pi] \\ \dfrac{1}{h_T} \displaystyle\sum_{j=t-T}^{t-[T\pi]-1} k\left(\dfrac{j}{h_T}\right) g_{t-j}(\beta), & t = [T\pi]+1, \ldots, T \end{cases}.$$

For *ELk* and *ETk* a moment-smoothing counterpart of quadratic spectral kernels is employed (Smith, 2011),

$$k(x) = \left(\frac{5\pi}{8}\right)^{1/2} \frac{1}{x} J_1\left(\frac{6\pi x}{5}\right),$$
$$J_v(z) = \frac{z^v}{2^v} \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{2^{2k}\Gamma(k+1)\Gamma(v+k+1)},$$

yielding $k_1 = (5\pi/2)^{1/2}$ and $k_2 = 2\pi$. For *GMMk*, the following quadratic-spectral kernel is employed [19]

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

The bandwidth employed, when smoothing, is "estimated" by $\hat{h}_T = 1.3221[\hat{\alpha}(2)T]^{1/5}$, where

$$\hat{\alpha}(2) = \sum_{a=1}^{p} w_a \frac{4\hat{\rho}_a^2 \hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8} \left\{ \sum_{a=1}^{p} w_a \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4} \right\}^{-1} \tag{24}$$

[17]Results for 1% and 10% nominal significance levels and trimmimg parameter values of $\varepsilon = 0.15, 0.25, 0.30, 0.35, 0.40, 0.45$ are available upon request.

[18]Results for the case with two different bandwidth windows for the two subsamples perform consistently worse, see discussion later in this section.

[19]Simulation results for Bartlett and Parzen implied kernels (Smith, 2011) are available upon request.

and $\hat{\rho}_a$ and $\hat{\sigma}_a^2$ are estimated $AR(1)$ coefficients and error variances, respectively, based on moment functions $g_t(\hat{\beta})$ ($\ell \times 1$; $a = 1, 2, \ldots, \ell$).[20] In particular, for *ELk* and *ETk* the unsmoothed version of the objective function is initially optimized to yield $\hat{\beta}$. Then, $\hat{\beta}$ is used to compute $\hat{\rho}_a$ and $\hat{\sigma}_a^2$ and then to estimate $\hat{h}_T$ (Eq. 24). The process repeats up to 5 times or until $\hat{h}_T^{(i)} = \hat{h}_T^{(i-1)}$, $i = 2, \ldots, 5$.

Tables 1–3 summarize the sampling results for DGP_2, DGP_5, and DGP_7 and are structured in the following way. Each table consists of four vertical panels, for *EL*, *ELk*, *ETk*, and *GMMk*, respectively, with each panel reporting results for sample sizes $T = 200, 400, 800, 1{,}600$. Horizontally, the results are divided into three big blocks for each of the $\widehat{\mathscr{D}}_{1,T}(\pi)$, $\widehat{\mathscr{D}}_{2,T}(\pi)$, and $\widehat{\mathscr{D}}_T(\pi)$ test procedures, within which sampling results for each of the $sup(\cdot)$, $exp(\cdot)$, and $ave(\cdot)$ functionals are reported. Each of these "functional" blocks consists of $\mathscr{LR}_T(\pi)$, $\mathscr{W}_T(\pi)$ ($\mathscr{O}_T(\pi)$ for $\widehat{\mathscr{D}}_{2,T}(\pi)$), and $\mathscr{LM}_T(\pi)$ test statistics.

We first consider the empirical significance levels of the tests when there is no structural break: DGP_1, DGP_2 (see Table 1), and DGP_3. Thus the null hypothesis for each test procedure is correct. For DGP_1, which is the case where kernel-smoothing is redundant, tests based on *EL* exhibit empirical significance levels which converge quite quickly to the nominal 5% level, but slightly over-reject at $T = 200$. For larger T and each functional, the $\mathscr{LR}$ and $\mathscr{W}$ variants have better finite sample properties than that of $\mathscr{LM}$. The Wald test based on *GMMk* is slightly undersized, in all its forms. For tests based on (smoothed) *ELk* and *ETk* criteria, convergence of empirical significance levels appears much slower, however, with the sup functional of all tests exhibiting empirical significance levels of 6.2% to 10.8%, at $T = 1{,}600$. The ave functional seems to be preferable for all test statistics with empirical rejection frequencies in the range 4.7% to 6.5% for $T = 800$ and 4.4% to 5.6% for $T = 1{,}600$. However, for *ELk* and *ETk* criteria, all tests for $T = 200$ and most of the tests for $T = 400$ exhibit much larger empirical significance levels than the nominal 5%.

For DGP_2 (Table 1) and DGP_3, and as might be expected, the *EL*-based tests reject the null too often since moment conditions are serially correlated ($\alpha = 0.4$ and 0.8, respectively). However, for *ELk* and *ETk*, although all the sup-tests now perform slightly better the previous qualitative features remain the same, with tests based on the ave functional yielding rejection rates in the range 3.5% to 8.7% for $T = 800$ and 3.8% to 6.8% for $T = 1{,}600$, under DGP_2. The finite sample performance deteriorates a little under DGP_3, $\alpha = 0.8$.

For DGP_4, DGP_5 (Table 2), and DGP_6 (parameter variation, with $\alpha = 0$, 0.4 and 0.8, respectively) $\widehat{\mathscr{D}}_{1,T}(\pi)$ and, consequently, $\widehat{\mathscr{D}}_T(\pi)$ are designed to exhibit some power whilst $\widehat{\mathscr{D}}_{2,T}(\pi)$ tests should remain relatively insensitive since its null distribution continues to hold under local parameter variation, and it is useful to see if this is reflected in the finite sample behaviour. As expected, empirical rejection rates for all $\widehat{\mathscr{D}}_{1,T}(\pi)$ and $\widehat{\mathscr{D}}_T(\pi)$ tests increases rapidly towards 100% as the sample size increases, across all the *DGP*s considered. However, those for the $\widehat{\mathscr{D}}_{T,2}(\pi)$ do not so and, indeed, remain fairly stable as

[20]This choice corresponds to optimal bandwidth based on an AR(1) approximation to the moment function with $w_a = 1$; see Andrews (1991)[pp. 834–835] with $w_a = 1$ in his Eq. (6.4).

TABLE 1
DGP_2 Results

	EL				*ELk*				*ETk*				*GMMk*			
	200	400	800	1600	200	400	800	1600	200	400	800	1600	200	400	800	1600
$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Sup*																
LR	0.124	0.121	0.111	0.123	0.237	0.115	0.080	0.070	0.135	0.094	0.079	0.072	0.114	0.099	0.086	0.077
W	0.112	0.106	0.110	0.119	0.192	0.127	0.103	0.077	0.193	0.130	0.103	0.079	0.114	0.099	0.086	0.077
LM	0.151	0.132	0.107	0.127	0.255	0.137	0.090	0.080	0.102	0.079	0.082	0.076	0.114	0.099	0.086	0.077
$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Ave*																
LR	0.098	0.103	0.086	0.097	0.185	0.094	0.060	0.064	0.090	0.079	0.058	0.065	0.084	0.090	0.065	0.061
W	0.095	0.101	0.087	0.095	0.125	0.098	0.064	0.068	0.140	0.103	0.060	0.068	0.084	0.090	0.065	0.061
LM	0.097	0.103	0.080	0.091	0.150	0.089	0.064	0.064	0.057	0.070	0.054	0.061	0.084	0.090	0.065	0.061
$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Exp*																
LR	0.122	0.110	0.101	0.104	0.225	0.110	0.075	0.067	0.130	0.088	0.069	0.067	0.103	0.101	0.081	0.068
W	0.108	0.111	0.105	0.102	0.170	0.118	0.087	0.068	0.182	0.119	0.087	0.069	0.103	0.101	0.081	0.068
LM	0.131	0.122	0.094	0.108	0.228	0.120	0.081	0.074	0.085	0.078	0.069	0.069	0.103	0.101	0.081	0.068
$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Sup*																
LR	0.157	0.135	0.108	0.105	0.163	0.089	0.065	0.051	0.142	0.089	0.063	0.054	0.024	0.028	0.040	0.037
W	0.094	0.095	0.092	0.093	0.104	0.048	0.039	0.035	0.063	0.043	0.036	0.035	0.024	0.028	0.040	0.037
LM	0.240	0.164	0.118	0.105	0.361	0.214	0.116	0.073	0.375	0.265	0.161	0.107	0.024	0.028	0.040	0.037
$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Ave*																
LR	0.123	0.109	0.080	0.094	0.107	0.079	0.052	0.045	0.105	0.079	0.048	0.044	0.034	0.050	0.040	0.042
W	0.096	0.091	0.075	0.089	0.058	0.059	0.040	0.038	0.058	0.060	0.035	0.038	0.034	0.050	0.040	0.042
LM	0.159	0.124	0.083	0.093	0.229	0.115	0.068	0.054	0.222	0.139	0.085	0.068	0.034	0.050	0.040	0.042
$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Exp*																
LR	0.156	0.125	0.095	0.101	0.144	0.084	0.058	0.048	0.126	0.088	0.057	0.052	0.032	0.045	0.042	0.039
W	0.105	0.095	0.085	0.093	0.088	0.054	0.044	0.038	0.061	0.047	0.039	0.036	0.032	0.045	0.042	0.039
LM	0.227	0.160	0.101	0.097	0.325	0.174	0.094	0.062	0.326	0.216	0.123	0.087	0.032	0.045	0.042	0.039
$\widehat{\mathscr{D}}_{T}(\pi)$, *Sup*																
LR	0.189	0.164	0.139	0.127	0.207	0.117	0.077	0.049	0.165	0.106	0.080	0.059	0.071	0.075	0.069	0.053
W	0.132	0.140	0.128	0.123	0.186	0.105	0.079	0.046	0.163	0.109	0.077	0.050	0.071	0.075	0.069	0.053
LM	0.272	0.203	0.144	0.124	0.411	0.236	0.135	0.081	0.362	0.238	0.157	0.107	0.071	0.075	0.069	0.053
$\widehat{\mathscr{D}}_{T}(\pi)$, *Ave*																
LR	0.146	0.131	0.107	0.104	0.129	0.092	0.058	0.048	0.108	0.089	0.059	0.047	0.066	0.064	0.053	0.047
W	0.120	0.118	0.105	0.104	0.097	0.077	0.055	0.045	0.102	0.077	0.053	0.045	0.066	0.064	0.053	0.047
LM	0.173	0.138	0.105	0.106	0.275	0.131	0.071	0.052	0.218	0.129	0.087	0.061	0.066	0.064	0.053	0.047
$\widehat{\mathscr{D}}_{T}(\pi)$, *Exp*																
LR	0.191	0.169	0.126	0.118	0.196	0.114	0.074	0.055	0.145	0.095	0.075	0.059	0.075	0.075	0.058	0.047
W	0.135	0.144	0.119	0.118	0.167	0.102	0.065	0.050	0.147	0.101	0.068	0.052	0.075	0.075	0.058	0.047
LM	0.258	0.185	0.134	0.123	0.377	0.208	0.116	0.073	0.329	0.213	0.138	0.085	0.075	0.075	0.058	0.047

The table consists of four vertical panels: unsmoothed Empirical Likelihood (EL), kernel-smoothed Empirical Likelihood (ELk), kernel-smoothed Exponential Tilting (ETk), and kernel-smoothed GMM (GMMk). Each vertical panel report results for sample sizes $T = 200, 400, 800, 1600$. Horizontally, the results are divided into three big blocks for $\widehat{\mathscr{D}}_{1,T}(\pi)$, $\widehat{\mathscr{D}}_{2,T}(\pi)$, and $\widehat{\mathscr{D}}_{T}(\pi)$ tests. For each test we report $sup(\cdot)$, $exp(\cdot)$, and $ave(\cdot)$ statistics. Each statistics block consists of $\mathscr{LR}_T(\pi)$, $\mathscr{W}_T(\pi)$ ($\mathscr{O}_T(\pi)$ for $\widehat{\mathscr{D}}_{2,T}(\pi)$), and $\mathscr{LM}_T(\pi)$.

TABLE 2
DGP_5 Results

	EL				*ELk*				*ETk*				*GMMk*			
	200	400	800	1600	200	400	800	1600	200	400	800	1600	200	400	800	1600
								$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Sup*								
LR	0.739	0.958	0.999	1.00	0.688	0.900	0.990	1.00	0.690	0.921	0.997	1.00	0.650	0.922	0.998	1.00
W	0.723	0.959	0.999	1.00	0.657	0.921	0.998	1.00	0.736	0.928	0.998	1.00	0.650	0.922	0.998	1.00
LM	0.590	0.885	0.995	1.00	0.810	0.941	0.997	1.00	0.468	0.858	0.994	1.00	0.650	0.922	0.998	1.00
								$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Ave*								
LR	0.779	0.959	0.998	1.00	0.707	0.900	0.990	1.00	0.728	0.926	0.997	1.00	0.719	0.929	0.998	1.00
W	0.778	0.956	0.998	1.00	0.677	0.915	0.997	1.00	0.772	0.935	0.998	1.00	0.719	0.929	0.998	1.00
LM	0.611	0.910	0.995	1.00	0.795	0.942	0.997	1.00	0.397	0.869	0.997	1.00	0.719	0.929	0.998	1.00
								$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Exp*								
LR	0.801	0.966	0.999	1.00	0.707	0.907	0.991	1.00	0.741	0.929	0.997	1.00	0.723	0.942	0.998	1.00
W	0.795	0.964	0.999	1.00	0.697	0.935	0.998	1.00	0.783	0.944	0.998	1.00	0.723	0.942	0.998	1.00
LM	0.634	0.922	0.995	1.00	0.816	0.954	0.997	1.00	0.466	0.886	0.997	1.00	0.723	0.942	0.998	1.00
								$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Sup*								
LR	0.194	0.224	0.316	0.496	0.196	0.177	0.211	0.315	0.178	0.129	0.176	0.316	0.036	0.072	0.170	0.323
W	0.119	0.170	0.300	0.493	0.227	0.149	0.135	0.255	0.049	0.046	0.095	0.234	0.036	0.072	0.170	0.323
LM	0.291	0.267	0.316	0.482	0.567	0.372	0.332	0.386	0.507	0.394	0.371	0.453	0.036	0.072	0.170	0.323
								$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Ave*								
LR	0.127	0.135	0.189	0.294	0.123	0.107	0.122	0.186	0.114	0.087	0.110	0.179	0.046	0.072	0.114	0.195
W	0.100	0.123	0.182	0.293	0.123	0.090	0.102	0.158	0.053	0.057	0.081	0.147	0.046	0.072	0.114	0.195
LM	0.163	0.145	0.188	0.291	0.434	0.233	0.179	0.207	0.305	0.189	0.194	0.244	0.046	0.072	0.114	0.195
								$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Exp*								
LR	0.179	0.194	0.277	0.441	0.175	0.164	0.186	0.286	0.156	0.123	0.157	0.288	0.055	0.089	0.162	0.293
W	0.133	0.161	0.266	0.435	0.218	0.148	0.141	0.239	0.056	0.051	0.100	0.225	0.055	0.089	0.162	0.293
LM	0.258	0.221	0.278	0.435	0.526	0.330	0.285	0.337	0.456	0.330	0.314	0.395	0.055	0.089	0.162	0.293
								$\widehat{\mathscr{D}}_{T}(\pi)$, *Sup*								
LR	0.676	0.917	0.998	1.00	0.426	0.789	0.980	1.00	0.580	0.834	0.993	1.00	0.504	0.821	0.994	1.00
W	0.617	0.906	0.998	1.00	0.590	0.842	0.993	1.00	0.626	0.846	0.993	1.00	0.504	0.821	0.994	1.00
LM	0.596	0.820	0.991	1.00	0.843	0.908	0.994	1.00	0.643	0.824	0.993	1.00	0.504	0.821	0.994	1.00
								$\widehat{\mathscr{D}}_{T}(\pi)$, *Ave*								
LR	0.637	0.896	0.994	1.00	0.400	0.777	0.983	1.00	0.539	0.825	0.994	1.00	0.501	0.796	0.990	1.00
W	0.602	0.886	0.994	1.00	0.522	0.819	0.991	1.00	0.555	0.817	0.992	1.00	0.501	0.796	0.990	1.00
LM	0.494	0.771	0.988	1.00	0.766	0.882	0.987	1.00	0.427	0.767	0.993	1.00	0.501	0.796	0.990	1.00
								$\widehat{\mathscr{D}}_{T}(\pi)$, *Exp*								
LR	0.700	0.923	0.997	1.00	0.452	0.813	0.985	1.00	0.598	0.862	0.994	1.00	0.534	0.843	0.995	1.00
W	0.649	0.913	0.997	1.00	0.607	0.862	0.995	1.00	0.634	0.865	0.995	1.00	0.534	0.843	0.995	1.00
LM	0.595	0.828	0.991	1.00	0.836	0.914	0.992	1.00	0.592	0.826	0.994	1.00	0.534	0.843	0.995	1.00

The table consists of four vertical panels: unsmoothed Empirical Likelihood (EL), kernel-smoothed Empirical Likelihood (ELk), kernel-smoothed Exponential Tilting (ETk) and kernel-smoothed GMM (GMMk). Each vertical panel report results for sample sizes $T = 200, 400, 800, 1600$. Horizontally, the results are divided into three big blocks for $\widehat{\mathscr{D}}_{1,T}(\pi)$, $\widehat{\mathscr{D}}_{2,T}(\pi)$, and $\widehat{\mathscr{D}}_{T}(\pi)$ tests. For each test we report $sup(\cdot)$, $exp(\cdot)$, and $ave(\cdot)$ statistics. Each statistics block consists of $\mathscr{LR}_T(\pi)$, $\mathscr{W}_T(\pi)$ ($\mathscr{O}_T(\pi)$ for $\widehat{\mathscr{D}}_{2,T}(\pi)$), and $\mathscr{LM}_T(\pi)$.

TABLE 3
DGP_7 Results

	EL				*ELk*				*ETk*				*GMMk*			
	200	400	800	1600	200	400	800	1600	200	400	800	1600	200	400	800	1600
								$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Sup*								
LR	0.686	0.930	0.995	1.00	0.768	0.875	0.966	0.999	0.758	0.949	0.996	1.00	0.631	0.917	0.998	1.00
W	0.717	0.956	0.998	1.00	0.646	0.872	0.975	0.998	0.777	0.944	0.998	1.00	0.631	0.917	0.998	1.00
LM	0.589	0.863	0.993	1.00	0.834	0.922	0.988	1.00	0.349	0.692	0.962	1.00	0.631	0.917	0.998	1.00
								$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Ave*								
LR	0.773	0.955	0.998	1.00	0.764	0.882	0.954	0.994	0.653	0.892	0.986	0.999	0.722	0.938	0.999	1.00
W	0.770	0.957	0.998	1.00	0.619	0.844	0.951	0.992	0.785	0.956	0.997	1.00	0.722	0.938	0.999	1.00
LM	0.659	0.918	0.996	1.00	0.821	0.924	0.985	1.00	0.277	0.678	0.970	1.00	0.722	0.938	0.999	1.00
								$\widehat{\mathscr{D}}_{1,T}(\pi)$, *Exp*								
LR	0.758	0.954	0.998	1.00	0.769	0.890	0.969	0.999	0.724	0.946	0.997	1.00	0.713	0.944	0.998	1.00
W	0.779	0.968	0.998	1.00	0.663	0.882	0.977	0.997	0.809	0.962	0.998	1.00	0.713	0.944	0.998	1.00
LM	0.651	0.914	0.996	1.00	0.841	0.933	0.989	1.00	0.343	0.737	0.976	1.00	0.713	0.944	0.998	1.00
								$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Sup*								
LR	0.888	0.989	0.995	1.00	0.754	0.932	0.987	0.998	0.947	0.999	1.00	1.00	0.773	0.997	1.00	1.00
W	0.876	0.999	1.00	1.00	0.901	0.997	1.00	1.00	0.813	0.997	1.00	1.00	0.773	0.997	1.00	1.00
LM	0.951	0.999	1.00	1.00	0.990	0.999	1.00	1.00	0.992	0.999	1.00	1.00	0.773	0.997	1.00	1.00
								$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Ave*								
LR	0.925	0.989	0.995	1.00	0.773	0.932	0.987	0.998	0.976	0.999	1.00	1.00	0.943	0.999	1.00	1.00
W	0.950	0.999	1.00	1.00	0.964	0.999	1.00	1.00	0.941	0.999	1.00	1.00	0.943	0.999	1.00	1.00
LM	0.971	0.999	1.00	1.00	0.988	0.999	1.00	1.00	0.991	0.999	1.00	1.00	0.943	0.999	1.00	1.00
								$\widehat{\mathscr{D}}_{2,T}(\pi)$, *Exp*								
LR	0.912	0.989	0.995	1.00	0.773	0.932	0.987	0.998	0.970	0.999	1.00	1.00	0.919	0.998	1.00	1.00
W	0.937	0.999	1.00	1.00	0.956	0.999	1.00	1.00	0.919	0.999	1.00	1.00	0.919	0.998	1.00	1.00
LM	0.967	0.999	1.00	1.00	0.990	0.999	1.00	1.00	0.992	0.999	1.00	1.00	0.919	0.998	1.00	1.00
								$\widehat{\mathscr{D}}_{T}(\pi)$, *Sup*								
LR	0.935	0.991	0.997	1.00	0.508	0.697	0.905	0.995	0.985	1.00	1.00	1.00	0.961	1.00	1.00	1.00
W	0.974	1.00	1.00	1.00	0.941	1.00	1.00	1.00	0.974	1.00	1.00	1.00	0.961	1.00	1.00	1.00
LM	0.984	1.00	1.00	1.00	0.996	1.00	1.00	1.00	0.999	1.00	1.00	1.00	0.961	1.00	1.00	1.00
								$\widehat{\mathscr{D}}_{T}(\pi)$, *Ave*								
LR	0.942	0.991	0.997	1.00	0.524	0.698	0.905	0.995	0.993	1.00	1.00	1.00	0.989	1.00	1.00	1.00
W	0.990	1.00	1.00	1.00	0.993	1.00	1.00	1.00	0.990	1.00	1.00	1.00	0.989	1.00	1.00	1.00
LM	0.985	1.00	1.00	1.00	0.997	1.00	1.00	1.00	0.997	1.00	1.00	1.00	0.989	1.00	1.00	1.00
								$\widehat{\mathscr{D}}_{T}(\pi)$, *Exp*								
LR	0.942	0.991	0.997	1.00	0.525	0.698	0.905	0.995	0.997	1.00	1.00	1.00	0.979	1.00	1.00	1.00
W	0.986	1.00	1.00	1.00	0.987	1.00	1.00	1.00	0.987	1.00	1.00	1.00	0.979	1.00	1.00	1.00
LM	0.990	1.00	1.00	1.00	0.999	1.00	1.00	1.00	0.999	1.00	1.00	1.00	0.979	1.00	1.00	1.00

The table consists of four vertical panels: unsmoothed Empirical Likelihood (EL), kernel-smoothed Empirical Likelihood (ELk), kernel-smoothed Exponential Tilting (ETk) and kernel-smoothed GMM (GMMk). Each vertical panel report results for sample sizes $T = 200, 400, 800, 1600$. Horizontally, the results are divided into three big blocks for $\widehat{\mathscr{D}}_{1,T}(\pi)$, $\widehat{\mathscr{D}}_{2,T}(\pi)$, and $\widehat{\mathscr{D}}_{T}(\pi)$ tests. For each test we report *sup*$(\cdot)$, *exp*$(\cdot)$, and *ave*$(\cdot)$ statistics. Each statistics block consists of $\mathscr{LR}_T(\pi)$, $\mathscr{W}_T(\pi)$ ($\mathscr{O}_T(\pi)$ for $\widehat{\mathscr{D}}_{2,T}(\pi)$), and $\mathscr{LM}_T(\pi)$.

the sample increases. For example among the *ELk* and *ETk* based tests the ave($\mathscr{LR}_T^*(\pi)$) seems least sensitive with rejections rates the range 6.6% to 22.5% across all sample sizes and *DGP*s. Tests derived from the *GMMk* criteria exhibit similar behavior.[21]

For DGP_7 (Table 3) and DGP_8 all tests should have power with rejection frequencies approaching 100% as the sample size grows. However, there are some caveats associated with kernel-based tests $\widehat{\mathscr{D}}_{1,T}(\pi)$ and as a result with $\widehat{\mathscr{D}}_T(\pi)$. Since $\widehat{\mathscr{D}}_{1,T}(\pi)$ is based on restricted models *and* $\hat{h}_T$ is evaluated for each value for π, occasional departures from the quasi-optimum lead to non-convergence issues and associated numerical problems when constructing $\hat{h}_T$, covariance matrices and test statistics. A manifestation of this is observing falling rejection frequencies to somewhat less than 100% as T increases; this indicates problems with convergence rather than "falling power" per se. For DGP_8, $\alpha = 0.8$, this problem is most pronounced. The observed power of the $\widehat{\mathscr{D}}_{2,T}(\pi)$ test is very close to 100% from $T = 200$, for all tests save $\mathscr{LR}_T^*(\pi)$ which implicitly involves estimation of the restricted model. The observed power of the $\widehat{\mathscr{D}}_{1,T}(\pi)$ tests are lower due to the non-convergence problems mentioned above.

Finally, we consider the calculation the bandwidth parameter employed with kernel-smoothing methods. The sampling results described above are based on reevaluating $\hat{h}_T$ for each value of π, however we restrict it to be the same for each of the subsamples that are then used to smooth the moment function. Two alternative strategies would be (i) reestimate $\hat{h}_{\pi T}$ and $\hat{h}_{(1-\pi)T}$ for each of the two subsamples; or, (ii) estimate $\hat{h}_T$ only once using restricted model for parameter estimation. In additional simulations, we compared these two alternative strategies in the context of the statistics based on the sup functional, under DGP_1.[22] Such statistics had relatively inferior finite sample behavior, as reported in Table 1. We find the first strategy demonstrates very poor Sup-test performance: even for $T = 1{,}600$ the empirical significance level of the test is from two to four times larger than the nominal one. However, we find the second strategy performs much better: for $T = 800$ to 1,600 the empirical significance level is close to the nominal one and comparable with Ave- and Exp-tests. This, admittedly, limited evidence suggests that choice of bandwidth is critically important for finite sample behaviour when considering Information-Theoretic approaches to structural stability testing.

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

[21] Hall and Sen (1999) propose a strategy in which the break point is estimated by the argument that yields the supremum of the parameter variation test, and then the fixed break point version of the overidentifying restrictions test is applied for that estimated break point. They find this approach reduces the sensitivity of the overidentifying restrictions test to parameter variation. We conjecture a similar approach could be taken using the IT tests.

[22] Reported in Hall et al. (2013).

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; and (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 occurring 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. 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 employed EL and ET methods and illustrates the finite sample performance of our test statistics under both the null of stability and alternatives of structural instability. This study revealed that the finite sample size properties of the IT tests are sensitive to the bandwidth used in filtering the sample moment. In particular, estimation of subsample specific bandwidths—arguably the most intuitively natural approach—leads to the worst performance. The issue of how best to calculate the bandwidths in this context remains to be resolved and is an interesting topic for future research.

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),*

$$
\begin{aligned}
k_1^{-1}\Omega_0^{-1/2}\sqrt{T}\bar{g}^{s}_{[T\pi]}(\beta_0) &= \Omega_0^{-1/2}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) + o_{p\pi}(1) \qquad (25)\\
&\Longrightarrow B_\ell(\pi),
\end{aligned}
$$

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

$$k_1^{-1}(I_2 \otimes \Omega_0^{-1/2})\sqrt{T}\bar{g}_T^s(\theta_0,\pi) = (I_2 \otimes \Omega_0^{-1/2})\sqrt{T}\bar{g}_T(\theta_0,\pi) + o_{p\pi}(1) \Longrightarrow J_\ell(\pi) = \begin{bmatrix} B_\ell(\pi) \\ (B_\ell(1) - B_\ell(\pi)) \end{bmatrix}. \quad (26)$$

Proof of Lemma 1. Following Smith (2011, Lemma A2), we can write

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

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 $j \leq [T\pi]-T$. Exploiting this, some straightforward (but tedious) algebra reveals that

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

where

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

$$A_{1T}(\beta,\pi) = \frac{1}{h_T}\sum_{j=0}^{[T\pi]-1} k\left(\frac{j}{h_T}\right)\frac{1}{T}\sum_{t=[T\pi]+1-j}^{[T\pi]} g_t(\beta),$$

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

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

Smith (2011, Lemma A1), shows that $\sum_{j=1-T}^{T-1}\frac{1}{h_T}k(\frac{j}{h_T}) = k_1 + o(1)$ and $\Omega_0^{-1/2}\sqrt{T}\,\bar{g}_{[T\pi]}(\beta_0) \Longrightarrow B_\ell(\pi)$, by Andrews (1993); thus, $\sqrt{T}\bar{g}^s_{[T\pi]}(\beta_0) = k_1\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) - \sqrt{T}\sum_{j=0}^{3} A_{jT}(\beta,\pi) + o_{p\pi}(1)$ and (25) follows if $\|\sqrt{T}A_{jT}(\beta_0,\pi)\| = o_{p\pi}(1)$, for $j = 0,1,2,3$. First, $\lim_{T\to\infty}\frac{1}{h_T}\sum_{j=1-T}^{T-1}|k(\frac{j}{h_T})| = O(1)$, implies $\lim_{T\to\infty}\sup_\pi \frac{1}{h_T}\sum_{j=[T\pi]}^{T-1}|k(\frac{j}{h_T})| = 0$, and thus, since $\|\sqrt{T}\bar{g}_{[T\pi]}(\beta_0)\| = O_{p\pi}(1)$, $\|\sqrt{T}A_{0T}(\beta_0,\pi)\| = o_{p\pi}(1)$. Second, $\|\frac{1}{\sqrt{|j|}}\sum_{t=[T\pi]+1-j}^{[T\pi]} g_t(\beta_0)\| = O_p(1)$, uniformly in j and π and Smith (2011, Lemma C1) is easily extended to show that $\lim_{T\to\infty}\frac{1}{h_T}\sum_{j=1-T}^{T-1}\sqrt{\frac{|j|}{T}}|k(\frac{j}{h_T})| = 0$, so that

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

The results for $\sqrt{T}A_{2T}(\beta_0,\pi)$ and $\sqrt{T}A_{3T}(\beta_0,\pi)$ follow in a similar fashion so that (25) holds.

Similarly,

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

so that

$$k_1^{-1}(I_2\otimes\Omega_0^{-1/2})\sqrt{T}\bar{g}_T^s(\theta_0,\pi) = (I_2\otimes\Omega_0^{-1/2})\sqrt{T}\bar{g}_T(\theta_0,\pi) + o_{p\pi}(1),$$

since $\bar{g}_T(\theta_0,\pi) = (\bar{g}_{[T\pi]}(\beta_0)', \bar{g}_T(\beta_0)' - \bar{g}_{[T\pi]}(\beta_0)')'$, and (26) 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 $\sup_{\beta\in\mathscr{B}}\|\bar{m}_T(\beta) - m(\beta)\| = o_p(1)$, where $\bar{m}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} m_t(\beta)$. Let $m_t^s(\beta)$ be the smoothed version of $m_t(\beta)$, defined in an analogous manner to $g_t^s(\beta)$ at (5), and (following (14)), define*

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

$$\bar{m}_T^s(\theta,\pi) = \frac{1}{T}\sum_{t=1}^{T} m_t^s(\theta,\pi),$$

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

Proof of Lemma 2. We can write

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

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

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

By Andrews (1993, Proof of Lemma A1), $\sup_\beta \|\bar{m}_{[T\pi]}(\beta) - \pi m(\beta)\| = o_{p\pi}(1)$, and since $\sum_{j=1-T}^{T-1} \frac{1}{h_T} k(\frac{j}{h_T}) = k_1 + o(1)$, the second term is also $o_{p\pi}(1)$. Then, by the triangle inequality, it remains to show that

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

since $\frac{1}{T}\sum_{t=[T\pi]+1}^{T} m^s_t(\beta) = \bar{m}^s_T(\beta) - \bar{m}^s_{[T\pi]}(\beta)$. From the proof of Lemma 1, above, it is clear that

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

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 $\sup_\beta \|A_{jT}(\beta, \pi)\| = o_{p\pi}(1)$, for $j = 0, 1, 2, 3$, and the result follows. □

The technical analysis undertaken in this paper employs (5), which assumes that smoothing is undertaken before the sample separation. Alternatively, the moment function could be smoothed after sample separation yielding

$$\bar{m}^{s*}_{[T\pi]}(\beta) = \frac{1}{T}\sum_{t=1}^{[T\pi]} \frac{1}{h_T} \sum_{j=t-[T\pi]}^{t-1} k\left(\frac{j}{h_T}\right) m_{t-j}(\beta)$$

for some $m_t(\beta)$ as defined in Lemma 2. This makes no difference asymptotically, as described in the following Lemma. (The proof is omitted as it follows similar arguments to those used in the proofs of Lemmas 1 and 2.)

Lemma 3. *Define* $\bar{e}^s_{[T\pi]} = \bar{m}^s_{[T\pi]}(\beta) - \bar{m}^{s*}_{[T\pi]}(\beta)$, *as above.*

(i) *Under the assumptions of Lemma 1, with* $m_t(\beta) \equiv g_t(\beta)$, $\sqrt{T}\bar{e}^s_{[T\pi]}(\beta_0) = o_{p\pi}(1)$.
(ii) *Under the assumptions of Lemma 1,* $\sup_{\beta \in \mathscr{B}} \|\bar{e}^s_{[T\pi]}(\beta)\| = o_{p\pi}(1)$.

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

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

$$\sup_{\theta\in\Theta,\gamma\in\Gamma_T,1\le t\le T}|\gamma' g_t^s(\theta,\pi)| = o_{p\pi}(1),$$

*so that w.p.a.*1, $k\gamma' g_t^s(\theta,\pi)\in\mathcal{V}$, *for all* $\theta\in\Theta,\gamma\in\Gamma_T$ *and* $\pi\in\Pi$.

Proof of Lemma 4. By Cauchy–Schwartz,

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

Now,

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

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

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

where $\alpha=\delta-\varepsilon\eta(\delta-1)<0$, because $\varepsilon>\frac{\delta}{\eta(\delta-1)}$, and thus *w.p.a.*1, $k\gamma' g_t^{sa}(\theta,\pi)\in\mathcal{V}$, for all $\theta\in\Theta$, $\gamma\in\Gamma_T$, and $\pi\in\Pi$. □

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

Lemma 5. *Under Assumptions 1–4, there exists a finite constant* $0<\Delta<\infty$, *such that w.p.a.*1 *and for all* $\theta\in\Theta$ *and* $\gamma\in\Gamma_T$, *and for each* $\pi\in\Pi$,

$$h_T^{-1}Q_T(\theta_0,\gamma,\pi)\le-\gamma_T'\bar{g}_T^s(\theta_0,\pi)-\Delta\gamma_T'\gamma_T, \tag{27}$$

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

$$Q_T(\theta,\gamma,\pi) \geq -k\gamma'\bar{g}_T^s(\theta,\pi) - k^2\Delta\gamma'\gamma. \tag{28}$$

Proof of Lemma 5. By a second order Taylor expansion about $\gamma = 0$, and exploiting Lemma 4, we have that for all $\theta \in \Theta$ and $\gamma \in \Gamma_T$, and each $\pi \in \Pi$

$$\begin{aligned} Q_T(\theta,\gamma,\pi) &= \gamma'\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\rho(k\bar{\gamma}'g_T^s(\theta,\pi))}{\partial\gamma} + \frac{1}{2}\gamma'\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\rho(k\bar{\gamma}'g_t^s(\theta,\pi))}{\partial\gamma\partial\gamma'}\gamma \\ &\equiv k\gamma'\frac{1}{T}\sum_{t=1}^{T}\rho_1(k\bar{\gamma}'g_t^s(\theta,\pi))g_t^s(\theta,\pi) \\ &\quad + \frac{k^2}{2}\gamma'\frac{1}{T}\sum_{t=1}^{T}\rho_2(k\bar{\gamma}'g_t^s(\theta,\pi))g_t^s(\theta,\pi)g_t^s(\theta,\pi)'\gamma, \end{aligned}$$

where $\bar{\gamma}$ is the usual "mean value" vector. Then by Lemma 4 and the normalization $\rho_1 = \rho_2 = -1$, we can write

$$Q_T(\theta,\gamma,\pi) = -k\gamma'\bar{g}_T^s(\theta,\pi) - \frac{1}{2}k^2\gamma'\bar{V}_T^s(\theta,\pi)\gamma + o_p(1), \tag{29}$$

where the $o_p(1)$ error is of smaller order than -$k\gamma'\bar{g}_T^s(\theta,\pi) - \frac{1}{2}k^2\gamma'\bar{V}_T^s(\theta,\pi)\gamma$.

To establish (27), substitute θ_0 for θ in (29) to obtain, $w.p.a.1$,

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

where, here, $\gamma_T = k\gamma/h_T \in \Gamma_T$. By arguments similar to Smith (2011, Lemma A3) it can be shown that $h_T\bar{V}_T^s(\theta_0,\pi) \equiv k_2\Omega_0(\pi) + o_{p\pi}(1)$, and we can now write

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

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

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

for all $\gamma \in \Gamma_T$, and for each $\pi \in \Pi$.

More generally, however, $\bar{V}_T(\theta,\pi) = O_{p\pi}(1)$, uniformly in θ, so that by similar reasoning, we can write

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

and (28) follows from this. □

Lemma 6. *Under Assumptions 1–4, there exists a finite constant, $\Delta > 0$, such that w.p.a.*1

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

Proof of Lemma 6. As in Smith (2011, Lemma A5), by Eq. (27) we have *w.p.a.*1 and each $\pi \in \Pi$

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

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

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

since $\sup_{\pi \in \Pi} \|\bar{g}_T^s(\theta_0, \pi\|^2 = O_p(T^{-1})$, from Lemma 1, the result then follows. □

Proof of Theorem 1. By Lemma 5, Eq. (28) and Lemma 6, we have *w.p.a.*1, and for all $\gamma \in \Gamma_T$ and each $\pi \in \Pi$,

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

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

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

*w.p.a.*1 or

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

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

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

because $1-2\varepsilon > 0$ and $h_T^2/T \to 0$. Therefore, since $\delta_T \to 0$, $\sup_{\pi\in\Pi} \|\bar{g}_T^s(\hat{\theta}(\pi),\pi)\| \xrightarrow{p} 0$. But by Lemma 2, we know that $\sup_{\pi\in\Pi} \|\bar{g}_T^s(\hat{\theta}(\pi),\pi) - k_1 g(\hat{\theta}(\pi),\pi)\| \xrightarrow{p} 0$. Thus, $\sup_{\pi\in\Pi} g(\hat{\theta}(\pi),\pi) = o_p(1)$. Continuity of $g(\beta)$ and the identification Assumption 3(iv) then yields $\sup_{\pi\in\Pi} \|\hat{\theta}(\pi) - \theta_0\| = o_p(1)$.

In fact, a further refinement of the above argument (similar in spirit to that of Smith, 2011, Lemma A7) shows that $\sup_{\pi\in\Pi} \|\bar{g}_T^s(\hat{\theta}(\pi),\pi)\| = O_p(T^{-1/2})$, implying that $\sup_{\pi\in\Pi} \|\hat{\theta}(\pi) - \theta_0\| = O_p(T^{-1/2})$. It then follows that that $h_T \bar{V}_T^s(\hat{\theta}(\pi),\pi) = k_2\Omega_0(\pi) + o_{p\pi}(1)$; c.f. (Smith, 2005, Theorem 2.1). Using this (and arguments similar to the above), it can then be shown that $\sup_{\pi\in\Pi} \|\hat{\gamma}(\pi)\| = O_p(h_T/\sqrt{T})$ as follows.

By definition, $Q_T(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi) \geq Q_T(\hat{\theta}(\pi),\gamma,\pi)$, for all $\gamma \in \Gamma_T$. Then, setting $\gamma = 0 \in \Gamma_T$, and noting that $Q_T(\theta,0,\pi) \equiv 0$, for all $\theta \in \Theta$, and exploiting Lemma 4, a second-order mean value expansion yields, *w.p.a.*1,

$$
\begin{aligned}
0 &\leq \frac{T}{h_T} Q_T(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi) \\
&= \frac{T}{h_T}\left\{-k\hat{\gamma}(\pi)'\bar{g}_T^s(\hat{\theta}(\pi),\pi) - \frac{1}{2}k^2\hat{\gamma}(\pi)'\bar{V}_T^s(\hat{\theta}(\pi),\pi)\hat{\gamma}(\pi)\right\}.
\end{aligned}
$$

Then, since $\frac{T}{h_T} Q_T(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi) \leq \sup_{\gamma\in\Gamma_T} \frac{T}{h_T} Q_T(\theta_0,\gamma,\pi) \leq \Delta\|\sqrt{T}\bar{g}_T^s(\theta_0,\pi)\|^2 = O_{p\pi}(1)$, *w.p.a.*1, by Lemma 6, and the fact that $\sup_{\pi\in\Pi} \|\bar{g}_T^s(\hat{\theta}(\pi),\pi)\| = O_p(T^{-1/2})$ and $\sup_{\pi\in\Pi} \|h_T\bar{V}_T(\hat{\theta}(\pi),\pi)\| = O_p(1)$, it follows that $\sup_{\pi\in\Pi} \|\hat{\gamma}(\pi)\| = O_p(h_T/\sqrt{T})$. This implies $\sup_{\pi\in\Pi} \|\hat{\gamma}(\pi)\| = o_p(1)$.

Proof of Theorem 2. Differentiating $Q_T(\theta,\gamma,\pi) = \frac{1}{T}\sum_{t=1}^T[\rho(k\lambda' g_t^s(\theta,\pi)) - \rho_0]$ with respect to θ and γ yields the partial-sample first order conditions

$$\frac{\partial Q_T(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi)}{\partial\theta} = k\frac{1}{T}\sum_{t=1}^T \rho_1(k\hat{\gamma}(\pi)'g_t^s(\theta,\pi))G_t^s(\hat{\theta}(\pi),\pi)'\hat{\gamma}(\pi) = 0, \tag{30}$$

$$\frac{\partial Q_T(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi)}{\partial\gamma} = k\frac{1}{T}\sum_{t=1}^T \rho_1(k\hat{\gamma}(\pi)'g_t^s(\theta,\pi))g_t^s(\theta,\pi) = 0, \tag{31}$$

where

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

Writing $\hat{\varphi}(\pi) = (\hat{\theta}(\pi)', \frac{\hat{\gamma}(\pi)'}{h_T})'$ and $\varphi_0 = (\beta_0', \beta_0', 0')'$, and exploiting Lemma 1, a mean value expansion of (31) yields

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

since $\rho_1 = -1$, where

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

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

$$\begin{aligned}\frac{\partial^2 Q_T(\theta,\gamma,\pi)}{\partial\gamma\partial\theta'} &= k\frac{1}{T}\sum_{t=1}^{T}\rho_2(k\hat{\gamma}(\pi)'g_t^s(\theta,\pi))G_t^s(\theta,\pi)\\ &\quad + k^2\frac{1}{T}\sum_{t=1}^{T}\rho_2(k\hat{\gamma}(\pi)'g_t^s(\theta,\pi))g_t^s(\theta,\pi)(\hat{\gamma}(\pi)'G_t^s(\theta,\pi)),\\ h_T\frac{\partial^2 Q_T(\theta,\gamma,\pi)}{\partial\gamma\partial\gamma'} &= k^2\frac{h_T}{T}\sum_{t=1}^{T}\rho_2(k\hat{\gamma}(\pi)'g_t^s(\theta,\pi))g_t^s(\theta,\pi)g_t^s(\theta,\pi)'.\end{aligned}$$

Noting that $\rho_2 = -1$, it follows from Theorem 1, Lemma 4, Lemma 2, as applied to $\frac{1}{T}\sum_{t=1}^{T} vec(G_t^s(\theta,\pi))$, and $\sup_{\pi\in\Pi}\|h_T\bar{V}_T(\bar{\theta}(\pi),\pi) - k_2\Omega_0(\pi)\| = o_p(1)$, with $k^2k_2 = kk_1$, that

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

where

$$D_0^{\varphi}(\pi) = kk_1[G_0(\pi), \Omega_0(\pi),].$$

Similarly, $\frac{\sqrt{T}}{h_T}\frac{\partial Q_T(\hat{\theta}(\pi),\hat{\gamma}(\pi),\pi)}{\partial\theta} = -kk_1G_0(\pi)'\sqrt{T}(\frac{\hat{\gamma}(\pi)}{h_T}) + o_{p\pi}(1)$. Combining these results, we obtain

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

Solving for $\sqrt{T}(\hat{\varphi}(\pi) - \varphi_0)$ yields

$$\sqrt{T}(\hat{\varphi}(\pi) - \varphi_0) = -\begin{pmatrix}(A(\pi)^{-1}\otimes(M_0'M_0)^{-1}M_0') \\ (A(\pi)^{-1}\otimes\Omega_0^{-1/2}(I_\ell - P_0))\end{pmatrix}\xi_T(\pi) + o_{p\pi}(1), \tag{32}$$

and the result follows. □

Proof of Theorem 3. Consistency of the estimators follows from the general arguments employed in the proof of Theorems 1, and 2. Differentiating $\dot{Q}_T(\beta,\gamma,\pi) = \frac{1}{T}\sum_{t=1}^{T}[\rho(k\lambda'\dot{g}_t^s(\beta,\pi)) - \rho_0]$ with respect to β and $\gamma = (\lambda_1', \lambda_2')'$, yields the partial-sample first order conditions

$$\begin{aligned}
\frac{\partial \dot{Q}_T(\tilde{\beta}(\pi), \tilde{\gamma}(\pi), \pi)}{\partial \beta} &= k\frac{1}{T}\sum_{t=1}^{[T\pi]} \rho_1(k\tilde{\lambda}_1(\pi)' g_t^s(\tilde{\beta}(\pi)))G_t^s(\tilde{\beta}(\pi))'\tilde{\lambda}_1(\pi) \\
&\quad + k\frac{1}{T}\sum_{t=[T\pi]+1}^{T} \rho_1(k\tilde{\lambda}_2(\pi)' g_t^s(\tilde{\beta}(\pi)))G_t^s(\tilde{\beta}(\pi))'\tilde{\lambda}_2(\pi) \\
&= 0, \\
\frac{\partial \dot{Q}_T(\tilde{\beta}(\pi), \tilde{\gamma}(\pi), \pi)}{\partial \lambda_1} &= k\frac{1}{T}\sum_{t=1}^{[T\pi]} \rho_1(k\tilde{\lambda}_1(\pi)' g_t^s(\tilde{\beta}(\pi)))g_t^s(\tilde{\beta}(\pi)) = 0, \\
\frac{\partial \dot{Q}_T(\tilde{\beta}(\pi), \tilde{\gamma}(\pi), \pi)}{\partial \lambda_2} &= k\frac{1}{T}\sum_{t=[T\pi]+1}^{T} \rho_1(k\tilde{\lambda}_2(\pi)' g_t^s(\tilde{\beta}(\pi)))g_t^s(\tilde{\beta}(\pi)) = 0.
\end{aligned}$$

Using similar arguments to those employed in the proof of Theorem 2, a Taylor expansion of $\sqrt{T}\frac{\partial \dot{Q}_T(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi)}{\partial \lambda_i} = 0$ about $(\beta_0', 0')'$, $i = 1, 2$, yields, exploiting Lemma 1,

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

respectively, or

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

from which we note

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

Similarly, we have

$$\begin{aligned}
&\frac{\sqrt{T}}{h_T}\frac{\partial \dot{Q}_T(\tilde{\beta}(\pi),\tilde{\gamma}(\pi),\pi)}{\partial \beta}\\
&= k\frac{1}{T}\sum_{t=1}^{[T\pi]}\rho_1(k\tilde{\lambda}_1(\pi)'g_t^s(\tilde{\beta}(\pi)))G_t^s(\tilde{\beta}(\pi))'\sqrt{T}\left(\frac{\tilde{\lambda}_1(\pi)}{h_T}\right)\\
&\quad + k\frac{1}{T}\sum_{t=[T\pi]+1}^{T}\rho_1(k\tilde{\lambda}_2(\pi)'g_t^s(\tilde{\beta}(\pi)))G_t^s(\tilde{\beta}(\pi))'\sqrt{T}\left(\frac{\tilde{\lambda}_2(\pi)}{h_T}\right)\\
&= -kk_1\pi G_0{}'\sqrt{T}\left(\frac{\tilde{\lambda}_1(\pi)}{h_T}\right) - kk_1(1-\pi)G_0{}'\sqrt{T}\left(\frac{\tilde{\lambda}_2(\pi)}{h_T}\right) + o_{p\pi}(1)\\
&= 0.
\end{aligned}$$

Combining these results, we obtain

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

so that

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

and

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

or

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

and the result follows by Lemma 1. □

Proof of Theorem 4. Consider, first, $\mathscr{W}_T(\pi)$. Previous results, exploiting $\sqrt{T}$-consistency of $\hat{\beta}_i(\pi)$, show that

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

and, combining this with (32), we obtain

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

so that

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

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

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

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 $\mathscr{LM}_T(\pi)$ is

$$\begin{aligned} \mathscr{LM}_T(\pi) = (k_2/k_1^2)T\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi))\bar{C}_T^s(\tilde{\beta}(\pi))\left\{h_T^{-1}V_T^\zeta(\tilde{\beta}(\pi))\right\}^{-1} \\ \times \bar{C}_T^s(\tilde{\beta}(\pi))'\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi))/(\pi(1-\pi)). \end{aligned}$$

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

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

Furthermore, $\bar{G}_T^s(\tilde{\beta}(\pi)) = k_1G_0 + o_{p\pi}(1)$ and $h_T\bar{V}_T^s(\tilde{\beta}(\pi)) = k_2\Omega_0 + o_{p\pi}(1)$, so that

$$\begin{aligned} \bar{C}_T^s(\tilde{\beta}(\pi))'\sqrt{T}\bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi)) &= -\frac{k_1^2}{k_2}M_0'\left\{\Omega_0^{-1/2}\sqrt{T}\bar{g}_{[T\pi]}(\beta_0) - \pi\Omega_0^{-1/2}\sqrt{T}\bar{g}_T(\beta_0)\right\} + o_{p\pi}(1) \\ &= -\frac{k_1^2}{k_2}M_0'\left(a(\pi)' \otimes I_\ell\right)\xi_T(\pi) + o_{p\pi}(1), \end{aligned}$$

and, since $h_T^{-1} V_T^{\zeta}(\tilde{\beta}(\pi)) = \frac{k_1^2}{k_2} M_0' M_0 + o_{p\pi}(1)$,

$$\begin{aligned}&\{(k_1^2/k_2) h_T^{-1} V_T^{\zeta}(\tilde{\beta}(\pi))\}^{-1/2} \bar{C}_T^s(\tilde{\beta}(\pi))' \sqrt{T} \bar{g}_{[T\pi]}^s(\tilde{\beta}(\pi))\\ &\quad = -(M_0' M_0)^{-1/2} M_0' (a(\pi)' \otimes I_\ell) \xi_T(\pi) + o_{p\pi}(1),\end{aligned}$$

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

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

$$\begin{aligned}\sqrt{T} \bar{g}_T^s(\hat{\theta}(\pi), \pi) &= \sqrt{T} \bar{g}_T^s(\theta_0, \pi) + k_1 G_0(\pi) \sqrt{T} (\hat{\theta}(\pi) - \theta_0) + o_{p\pi}(1)\\ &= k_1 \sqrt{T} \bar{g}_T(\theta_0, \pi) - k_1 (I_2 \otimes \Omega_0^{1/2} P_0) \xi_T(\pi) + o_{p\pi}(1), \qquad (33)\end{aligned}$$

where (32) is exploited. Therefore, and again exploiting (32), we have

$$\begin{aligned}k_1^{-1} (I_2 \otimes \Omega_0^{-1/2}) \sqrt{T} \bar{g}_T^s(\hat{\theta}(\pi), \pi) &= (I_2 \otimes (I_\ell - P_0)) \xi_T(\pi) + o_{p\pi}(1) \qquad (34)\\ &= -(A(\pi) \otimes \Omega_0^{1/2}) \left(\sqrt{T}/h_T\right) \hat{\gamma}(\pi) + o_{p\pi}(1). \qquad (35)\end{aligned}$$

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

$$\begin{aligned}&2(k_2/k_1^2)(T/h_T) Q_T(\hat{\theta}(\pi), \hat{\gamma}(\pi), \pi)\\ &\quad = -2(k_2/k_1^2) k (\sqrt{T}/h_T) \hat{\gamma}(\pi)' \sqrt{T} \bar{g}_T^s(\hat{\theta}(\pi), \pi)\\ &\qquad + (k_2/k_1^2) \left(\sqrt{T}/h_T\right) \hat{\gamma}(\pi)' \left(h_T \frac{\partial^2 Q_T(\hat{\theta}(\pi), \bar{\gamma}(\pi), \pi)}{\partial \gamma \partial \gamma'}\right) \left(\sqrt{T}/h_T\right) \hat{\gamma}(\pi)\\ &\quad = T \bar{g}_T^s(\hat{\theta}(\pi), \pi)' (A(\pi) \otimes \Omega_0)^{-1} \bar{g}_T^s(\hat{\theta}(\pi), \pi)/k_1^2 + o_{p\pi}(1), \qquad (36)\end{aligned}$$

where $\bar{\gamma}(\pi)$ is the usual mean value and the third equality uses (35) and Lemma 4, which ensures that $h_T \frac{\partial^2 Q_T(\hat{\theta}(\pi), \bar{\gamma}(\pi), \pi)}{\partial \gamma \partial \gamma'} \xrightarrow{p} -k^2 k_2 \Omega_0(\pi) = -k^2 k_2 (A(\pi) \otimes \Omega_0)$, uniformly in π. Similarly,

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

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

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

where the second equality follows from (33). Notice that, by (34),

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

so that

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

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

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

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

from which we conclude that $\mathscr{S}_T(\pi) \Longrightarrow \frac{(B_k(\pi)-\pi B_k(1))'(B_k(\pi)-\pi B_k(1))}{\pi(1-\pi)}$. □

Proof of Theorem 5. Since $\sup_{\pi\in\Pi}\|h_T\bar{V}_T^s(\hat{\theta}(\pi),\pi)-k_2\Omega_0(\pi)\| = o_p(1)$ and $\sqrt{T}\bar{g}_T^s(\hat{\theta}(\pi),\pi) = O_{p\pi}(1)$, we immediately have that

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

and, by (35),

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

By (36), it is immediate that

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

This demonstrates the asymptotic equivalence of all three statistics. From (34), we also obtain

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

Following the arguments in the proof of Theorem 4, $I - P_0 = H_2'H_2$, so that

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

Since $H_2H_2' = I_{\ell-k}$, it follows that $H_2B_\ell(\pi) = B_{\ell-k}(\pi)$, a $(\ell-k)$-dimensional vector of independent standard Brownian motions and

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

implying

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

ACKNOWLEDGMENTS

This paper is based on results developed as part of the thesis for which Li was awarded a Ph.D. in Economics at the University of Manchester, UK in 2011. The computational element of this research was achieved using the High Throughput Computing facility of the Faculty of Engineering and Physical Sciences, University of Manchester. We would like to the thank the Editors, Esfandiar Maasoumi and Amos Golan, and an anonymous referee for constructive comments on an earlier version of this paper. The usual caveat applies, of course.

FUNDING

The first author acknowledges the support of the ESRC grant RES-062-23-1351.

REFERENCES

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

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* 61:821–856.

Andrews, D. W. K. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica* 71:395–398.

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

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

Cressie, N., 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., Hall, A. R. (1997). Predictive test for structural change with unknown breakpoint. *Journal of Econometrics* 82:209–233.

Ghysels, E., 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.

Golan, A. (2006). Information and entropy econometrics: A review and synthesis. *Foundations and Trends in Econometrics* 2:1–145.

Guay, A., 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, U.K.: Oxford University Press.

Hall, A. R., Li, Y., Orme, C. D., and Sinko, A. (2013). Testing for Structural Instability in Moment Restriction Models: an Info-metric Approach. Discussion paper, University of Manchester Economics Discussion Paper Series: EDP-1326, http://www.socialsciences.manchester.ac.uk/disciplines/economics/research/discussionpapers/pdf/EDP-1326.pdf.

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

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics* 97:93–115.

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.

Kitamura, Y. (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., 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., Smith, R. J. (2004). Higher order properties of GMM and Generalized Empirical Likelihood estimators. *Econometrica* 72:219–256.

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

Qin, J., 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.

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

Smith, R. J. (2011). GEL criteria for moment condition models. *Econometric Theory* 27:1192–1235.

Sowell, F. (1996). Optimal tests of parameter variation in the generalized method of moments framework. *Econometrica* 64:1085–1108.