Structural change tests for GEL criteria

Alain Guay[a], Jean-François Lamarche[b,†]

[a] Département de sciences économiques, Université du Québec à Montréal, Québec, Canada, CIRPÉE and CIREQ

[b] Department of Economics, Brock University, St. Catharines, Ontario, Canada

February 17, 2010

Abstract

This paper examines structural change tests based on generalized empirical likelihood methods in the time series context. Standard structural change tests for GMM with strongly identified parameters are adapted to the GEL context. We show that when moment conditions are properly smoothed, these test statistics converge to the same asymptotic distribution as in GMM, in cases with known and unknown breakpoints. New test statistics specific to GEL methods are also introduced. Finally, we propose stability tests in the GEL framework that are robust to weak identification and dependent data. A simulation study examines the small sample properties of the tests.

Keywords: Generalized empirical likelihood, generalized method of moments, parameter instability, structural change, weak identification.
JEL codes: C12, C32

† The first author gratefully acknowledges financial support of Fonds pour la Formation de Chercheurs et l'aide à la Recherche (FCAR). The first and second authors are thankful for the financial support of the Social Sciences and Humanities Research Council of Canada (SSHRC). Corresponding author: Alain Guay, Department of Economics, Université du Québec à Montréal, Case postale 8888, succursale Centre-ville, Montréal (Québec), Canada, H3C 3P8, E-mail: guay.alain@uqam.ca.

1 Introduction

Recently, a number of alternative estimators to GMM have been proposed. Hansen, Heaton and Yaron (1996) suggested the continuous updated estimator (CUE) which shares the same objective function as GMM but with a weighting matrix that depends on the parameters of interest. The empirical likelihood (EL) (see Qin and Lawless (1994)) and the exponential tilting (ET) estimators (see Kitamura and Stutzer (1997)) have also been proposed. Newey and Smith (2004) showed (in an i.i.d. setting) that, although estimators based on the GMM, EL, ET or that are CUE have the same asymptotic distribution, they have different higher order asymptotic properties. Amongst their findings it is shown that the expression for the second order asymptotic bias of GEL has fewer components than the one for GMM (with EL having the fewest). Anatolyev (2005) extended the Newey and Smith setting to allow for serial correlation and showed that smoothing the moment conditions reduces bias even further. These alternative estimators are special cases of the generalized empirical likelihood (GEL) class considered by Smith (1997).

An important aspect of the validation of an estimation strategy is the stability of the parameters of interest and the respective objective function used. In particular, GMM and GEL suppose that the parameters of interest and moment restrictions are stable across time. To our knowledge, structural change tests based on GEL have not yet been proposed. In view of the importance of detecting structural changes and given the recent developments of GEL methods as an alternative to GMM, it appears important to study structural change tests for these methods of estimation.

In that respect, a class of partial-sample GEL (PS-GEL) estimators is introduced and we establish the weak convergence of the resulting parameter vector to a function of Brownian motions. We also show that the PS-GEL estimators of the Lagrange multiplier parameters also weakly converge to a function of Brownian motions uncorrelated to the asymptotic distribution of the vector of parameters. These asymptotic distributions are derived under the null hypothesis of stability and general alternatives of structural change (see Sowell, 1996) for an unknown breakpoint. These results allow us to derive the asymptotic distributions of structural change tests in the GEL context. We consider cases of structural change which can occur in the parameters of interest or in the overidentifying restrictions used to estimate these parameters. As in Andrews (1993), we study standard Wald, LM anf LR types test statistics for parameters instability in cases of pure structural change test when the entire parameter vector is subject to structural change and partial structural change where only a subset of the parameter vector is subject to structural change. We show that these statistics when computed with smoothed moment conditions follow the same asymptotic distribution than in the GMM context (Andrews, 1993). Second, we examine tests for the stability of overidentifying restrictions. An equivalent test statistics to Hall and Sen's (1999) statistics in the GMM context is adapted to the GEL for smoothed moment conditions. Two new tests specific to the GEL framework are proposed to detect instability of overidentifying restrictions.

We show that these new statistics have the same asymptotic distribution at first order than the one derived by Hall and Sen (1999).

Recently, weak identification has received a large amount of attention (Stock and Wright, 2000). When the presence of weak instruments is suspected, robust structural change tests to weak identification need to be implemented. This paper proposes test statistics of structural change in the context of weakly identified or completely unidentified cases for the GEL framework. The first one is based on a renormalized criterion function of GEL evaluated at a restricted PS estimator. The second is asymptotically equivalent to the first and is based on the Lagrange multiplier of the restricted partial-sample estimator. The second group of tests includes a test statistic derived from a GEL criterion that uses moment conditions corresponding to the first-order conditions of the restricted PS-GEL estimator whose dimension is identical to the number of parameters and a statistic based on the corresponding Lagrange multiplier. Under weak identification or completely unidentified case, these test statistics are not asymptotically pivotal. As in Caner (2007), we show that their limits are bounded by a distribution which is nuisance parameter free and robust to identification problems. For the first group, the asymptotic bound is a function of the number of moment conditions while for the second group, the asymptotic bound depends on the number of parameters. The derivation of the bound under general local alternatives shows that the first group can have power against instability of parameter values or overidentifying restrictions while the second group is specifically designed to detect parameters instability.

The main findings of the simulation study are summarized as follows. We find that the average and exponential versions of the Lagrange multiplier-based test, one of the new test proposed, has very good rejection frequencies under the null hypothesis. Further this newly proposed test has the highest power. Given that this test statistic targets structural changes in the overidentifying restrictions and that standard tests for a structural change in the parameter vector (e.g., Wald and LR tests) have lower power, it appears that testing for an unknown change in the overidentifying restrictions is important in empirical applications.

The paper is organized as follows. Section 2 presents formally the full-sample and partial-sample GMM and GEL estimators. Section 3 presents the test statistics proposed and their respective asymptotic distributions. The simulation results are in Section 4 while the proofs are in the Appendix.

2 Full and partial-samples GMM and GEL estimators

To establish the asymptotic distribution theory of tests for structural change we need to elaborate on the specification of the parameter vector in our generic setup. We will consider parametric models indexed by parameters (β, δ) where $\beta \in B$, with $B \subset R^r$ and $\delta \in \Delta \subset R^\nu$. Following Andrews (1993) we make a distinction between pure structural change when no subvector δ appears and the entire parameter vector is subject to structural change under the alternative and partial structural change which corresponds to cases where only a subvector β is subject to structural change under the alternative. The generic null can be written as follows (the S stands for *stability*):

$$H_0^S : \beta_t = \beta_0 \qquad \forall t = 1, \ldots, T. \tag{1}$$

The tests considered assume as alternative that at some point in the sample there is a single structural break, like for instance:

$$\beta_t = \begin{cases} \beta_1(s) & t = 1, ..., [sT] \\ \beta_2(s) & t = [sT] + 1, ..., T \end{cases}$$

where s determines the fraction of the sample before and after the assumed breakpoint and $[.]$ denotes the greatest integer function. The separation $[Ts]$ represents a possible breakpoint which is governed by an unknown parameter s. Hence, we will consider a setup with a parameter vector which encompasses any kind of partial or pure structural change involving a single breakpoint. In particular, we consider a p dimensional parameter vector $\theta = (\beta_1', \beta_2', \delta')'$ where β_1 and $\beta_2 \in B \subset R^r$ and $\theta \in \Theta = B \times B \times \Delta \subset R^p$ where $p = 2r + \nu$. The parameters β_1 and β_2 apply to the samples before and after the presumed breakpoint and the null implies that:

$$H_0^S : \beta_1 = \beta_2 = \beta_0.. \tag{2}$$

Thus, under the null, $\theta_0 = (\beta_0', \beta_0', \delta_0')'$.

We will formulate all of our models in terms of θ. Special cases could be considered whenever restrictions are imposed in the general parametric formulation. One such restriction would be that $\theta_0 = (\beta_0', \beta_0')'$, which would correspond to the null of a pure structural change hypothesis. Once we have defined the moment conditions we will also translate this into overidentifying restrictions and relate it to structural change tests, following the analysis of Sowell (1996b) and of Hall and Sen (1999).

2.1 Definitions

We assume a triangular array of random variables $\{x_{Tt} : 1 \leq t \leq T, T \geq 1\}$. Triangular arrays of random variables are required to study local power of the structural change tests. However, to simplify the notation x_{Tt} is denoted x_t hereafter. Suppose a $q \times 1$ vector function of data $g(x_t, \beta, \delta)$ which depends on some unknown $r + \nu$-vector of parameters $(\beta', \delta')' \in B \times \Delta \subset R^{r+\nu}$ and that in the population their expected value is 0. That is,

$$E[g(x_t, \beta_0, \delta_0)] = 0.$$

Definition 2.1. *The full-sample General Method of Moments estimator $\{\tilde{\beta}_T, \tilde{\delta}_T\}$ is a sequence of random vectors such that:*

$$\left(\tilde{\beta}_T', \tilde{\delta}_T'\right)' = arg \min_{(\beta,\delta) \in B \times \Delta} g_T(\beta, \delta)' \hat{W}_T g_T(\beta, \delta)$$

where $\hat{W}_T$ is a random positive definite symmetric $q \times q$ matrix and $g_T(\beta, \delta) = \frac{1}{T}\sum_{t=1}^{T} g(x_t, \beta, \delta)$. The optimal weighting matrix is defined to be the inverse of the covariance matrix of the moment conditions, $W_T = \Omega_T^{-1}$ where Ω_T is a consistent estimator of

$$\Omega = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(x_t, \beta_0, \delta_0)\right).$$

The optimal weighting matrix can be estimated consistently using methods developed by Gallant (1987), Andrews and Monahan (1992) and Newey and West (1994), among several others. We call $\tilde{\theta}_T = \left(\tilde{\beta}_T', \tilde{\beta}_T', \tilde{\delta}_T'\right)'$ the full sample estimator of θ.

Several tests for structural change involve partial-sample GMM estimators defined by Andrews (1993). We consider two subsamples, the first is based on observations $t = 1, \ldots, [Ts]$ and the second covers $t = [Ts] + 1, \ldots, T$ where $s \in S \subset (0, 1)$. The partial-sample GMM estimators based on the first and the second subsamples are formally defined as:

Definition 2.2. *A partial-sample General Method of Moments estimator $\{\hat{\theta}_T(s)\}$ is a sequence of random vectors such that:*

$$\hat{\theta}_T(s) = arg \min_{\theta\in\Theta} g_T(\theta, s)'\hat{W}_T(s) g_T(\theta, s)$$

for all $s \in S$. Define $g_t(\theta, s) = (g(x_t, \beta_1, \delta)', 0')' \in R^{2q\times 1}$ for $t = [Ts]+1, \ldots, T$ and $g_t(\theta, s) = (g(x_t, \beta_2, \delta)', 0')' \in R^{2q\times 1}$ for $t = [Ts] + 1, \ldots, T$ such that

$$g_T(\theta, s) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta, s) = \frac{1}{T}\sum_{t=1}^{[Ts]}\begin{bmatrix} g(x_t, \beta_1, \delta) \\ 0 \end{bmatrix} + \frac{1}{T}\sum_{t=[Ts]+1}^{T}\begin{bmatrix} 0 \\ g(x_t, \beta_2, \delta) \end{bmatrix}$$

and $\hat{W}_T(s)$ is a random positive definite symmetric $2q \times 2q$ matrix.

According to this definition, $\hat{\theta}_T(s) = \left(\hat{\beta}_{1T}(s)', \hat{\beta}_{2T}(s)', \hat{\delta}_T(s)'\right)'$ is a $2r + \nu$-vector with an estimator $\hat{\beta}_{1T}(s)$ that uses the first subsample $t = 1, \ldots, [Ts]$, an estimator $\hat{\beta}_{2T}(s)$ that uses the second subsample $t = [Ts] + 1, \ldots, T$ and an estimator $\hat{\delta}_T(s)$ that uses the entire sample.

The partial-sample optimal weighting matrix is defined as the inverse of $\Omega(s)$ where

$$\Omega(s) = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\begin{bmatrix} \sum_{t=1}^{[Ts]} g(x_t, \beta_0, \delta_0) \\ \sum_{t=[Ts]+1}^{T} g(x_t, \beta_0, \delta_0) \end{bmatrix}\right)$$

which under the null (2) is asymptotically equal to

$$\Omega(s) = \begin{bmatrix} s\Omega & 0 \\ 0 & (1-s)\Omega \end{bmatrix}.$$

To characterize the asymptotic distribution we define the following matrices:

$$G^{\beta} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\partial g(x_t, \beta_0, \delta_0)/\partial\beta' \in R^{q\times r},$$

$$G^{\delta} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\partial g(x_t, \beta_0, \delta_0)/\partial\delta' \in R^{q\times\nu},$$

$$G(s) = \begin{bmatrix} sG^{\beta} & 0 & sG^{\delta} \\ 0 & (1-s)G^{\beta} & (1-s)G^{\delta} \end{bmatrix} \in R^{2q\times(2r+\nu)}.$$

In the GEL setting, the parameter vector is augmented by a vector of auxiliary parameters λ where each element of this vector is associated with an element of the smoothed moment conditions $g_{tT}(\theta)$ to be defined below. The generic null hypothesis of no structural change for this vector of auxiliary parameters is written as follows:

$$H_0^S : \lambda_t = \lambda_0 = 0 \qquad \forall t = 1, \ldots, T. \tag{3}$$

As for the parameter vector β, the tests we will consider assume as alternative that at some point in the sample there is a single structural break, namely:

$$\lambda_t = \begin{cases} \lambda_1(s) & t = 1, ..., [sT] \\ \lambda_2(s) & t = [sT]+1, ..., T. \end{cases}$$

Thus, under the null $H_0^S = \lambda_1 = \lambda_2 = \lambda_0 = 0$. We will show later that a structural change in λ is associated with instability in the overidentifying restrictions.

As in the GMM context an adjustment for the dynamic structure of $g(x_t, \theta)$ is also required in the GEL context (see Kitamura and Stutzer (1997), Smith (2000), Smith (2004) and Guggenberger and Smith (2008)). The adjustment consists of smoothing the original moment conditions $g(x_t, \theta)$. Defining the smoothed moment conditions as

$$g_{tT}(\beta, \delta) = \frac{1}{M_T}\sum_{m=t-T}^{t-1} k\left(\frac{m}{M_T}\right) g\left(x_{Tt-m}, \beta, \delta\right)$$

for $t = 1, \ldots, T$ and M_T is a bandwidth parameter, $k(\cdot)$ a kernel function and we define where $k_j = \int_{-\infty}^{\infty} k(a)^j da$. The GEL criteria is then given by:

$$\sum_{t=1}^{T} \frac{[\rho(k\lambda' g_{tT}(\theta)) - \rho_0]}{T}$$

where $k = \frac{k_1}{k_2}$ (see Smith (2004)).

We now formally define the restricted Generalized Empirical Likelihood (GEL) estimator using the entire sample $\tilde{\theta}_T = \left(\tilde{\beta}_T', \tilde{\delta}_T'\right)'$.

Definition 2.3. *Let $\rho(\phi)$ be a function of a scalar ϕ that is concave on its domain, an open interval Φ that contains 0. Also, let $\widetilde{\Lambda}_T(\beta, \delta) = \{\lambda : k\lambda' g_{tT}(\beta, \delta) \in \Phi, t = 1, \ldots, T\}$ with $k = \frac{k_1}{k_2}$. Then, the full-sample GEL estimator $\{\tilde{\theta}_T\}$ is a sequence of random vectors such that:*

$$\left(\tilde{\beta}_T', \tilde{\delta}_T'\right)' = \arg\min_{(\beta,\delta)\in B\times\Delta} \sup_{\lambda\in\widetilde{\Lambda}_T(\beta,\delta)} \sum_{t=1}^{T} \frac{[\rho(k\lambda' g_{tT}(\beta, \delta)) - \rho_0]}{T}$$

where $\rho_j() = \partial^j \rho()/\partial \phi^j$ *and* $\rho_j = \rho_j(0)$ *for* $j = 0, 1, 2, \ldots$.

The criteria is normalized so that $\rho_1 = \rho_2 = -1$ (see Smith (2004)). As mentioned earlier, the GEL estimator admits a number of special cases recently proposed in the econometrics literature. The CUE of Hansen, Heaton and Yaron (1996) corresponds to the following quadratic function $\rho(\phi) = -(1+\phi)^2/2$. The EL estimator (Qin and Lawless, 1994) is a GEL estimator with $\rho(\phi) = \ln(1-\phi)$. The ET estimator (Kitamura and Stutzer, 1997) is obtained with $\rho(\phi) = -\exp(\phi)$.

More precisely, the GEL estimator is obtained as the solution to a saddle point problem. Firstly, the criterion is maximized for given vector (β, δ). Thus,

$$\tilde{\lambda}_T(\beta,\delta) = \arg \sup_{\lambda \in \widetilde{\Lambda}(\beta,\delta)} \sum_{t=1}^{T} \frac{[\rho(k\lambda' g_{tT}(\beta,\delta)) - \rho_0]}{T}.$$

Secondly, the GEL estimator $\left(\tilde{\beta}'_T, \tilde{\delta}'_T\right)'$ is given by the following minimization problem:

$$\left(\tilde{\beta}'_T, \tilde{\delta}'_T\right)' = \arg \min_{(\beta,\delta)\in B\times\Delta} \sum_{t=1}^{T} \frac{\left[\rho\left(k\tilde{\lambda}_T(\beta,\delta)' g_{tT}(\beta,\delta)\right) - \rho_0\right]}{T}.$$

From now on, following Kitamura and Stutzer (1997) and Guggenberger and Smith (2008) we focus on the truncated kernel defined by

$$k(x) = 1 \text{ if } |x| \leq 1 \text{ and } k(x) = 0 \text{ otherwise}$$

to obtain the following smoothed moment conditions[1]

$$g_{tT}(\beta,\delta) = \frac{1}{2K_T+1} \sum_{j=-K_T}^{K_T} g(x_{t-j}, \beta, \delta)$$

where K_T is related to the bandwidth parameter M_T (see section 4). To handle the endpoints in the smoothing we use the approach of Smith (2004) and Guggenberger and Smith (2008) which sets

$$g_{tT}(\beta,\delta) = \frac{1}{2K_T+1} \sum_{j=\max\{t-T,-K_T\}}^{\min\{t-1,K_T\}} g(x_{t-j}, \beta, \delta).$$

We can easily show for this kernel that $k = \frac{k_1}{k_2} = 1$. A consistent estimator of the long run covariance matrix is then given by:

$$\tilde{\Omega}_T = \frac{2K_T+1}{T} \sum_{t=1}^{T} g_{tT}(\tilde{\beta},\tilde{\delta}) g_{tT}(\tilde{\beta}_T,\tilde{\delta}_T)'.$$

[1]We focus on the truncated kernel to simplify the notation and the proofs. Results derived in the following also holds for the class of kernels K_1 considered in Andrews (1991). Moreover Anatolyev (2005) establishes that among positive kernels, only the uniform truncated kernel proposed by Kitamura and Stutzer (1997) removes the bias component involved by the third moments of the moment conditions.

The weighting matrix thus obtained using this type of kernel is similar to the one obtained with the Bartlett kernel estimator of the long run covariance matrix of the moment conditions (see Smith (2004)). Define also the derivatives of the smoothed moment conditions as:

$$G_{tT}(\beta,\delta) = \frac{1}{2K_T+1}\sum_{j=-K_T}^{K_T}\frac{\partial g(x_{t-j},\beta,\delta)}{\partial\left(\beta',\delta'\right)}.$$

Now consider a possible breakpoint $[Ts]$. Define the vector of auxiliary parameters $\lambda(s) = (\lambda_1', \lambda_2')'$ where λ_1 is the vector of the auxiliary parameters for the first part of the sample (e.g., $t = 1, \ldots, [Ts]$) and λ_2 for the second part of the sample ($t = [Ts]+1, \ldots, T$). The partial-sample GEL estimators for $s \in S$ based on the first and the second subsamples are formally defined as:

Definition 2.4. *Let $\rho(\phi)$ be a function of a scalar ϕ that is concave on its domain, an open interval Φ that contains 0. Also, let $\widehat{\Lambda}_T(\theta,s) = \{\lambda(s) = (\lambda_1',\lambda_2')' : \lambda(s)'g_{tT}(\theta,s) \in \Phi, t = 1,\ldots,T\}$ for all $s \in S$, where $g_{tT}(\theta,s) = (g_{tT}(\beta_1,\delta)',0')' \in R^{2q\times 1}$ for $t = 1,\ldots,[Ts]$ and $g_{tT}(\theta,s) = (0', g_{tT}(\beta_2,\delta)')' \in R^{2q\times 1}$ for $t = [Ts]+1,\ldots,T$ with $\lambda(s) = (\lambda_1',\lambda_2')' \in R^{2q\times 1}$. A partial-sample Generalized Empirical Likelihood (PS-GEL) estimator $\{\hat{\theta}_T(s)\}$ is a sequence of random vectors such that:*

$$\begin{aligned}\hat{\theta}_T(s) &= \arg\min_{\theta\in\Theta}\sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta,s)}\sum_{t=1}^{T}\frac{[\rho(\lambda(s)'g_{tT}(\theta,s))-\rho_0]}{T}\\ &= \arg\min_{\theta\in\Theta}\sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta,s)}\left(\sum_{t=1}^{[Ts]}\frac{[\rho(\lambda_1'g_{tT}(\beta_1,\delta))-\rho_0]}{T}+\sum_{t=[Ts]+1}^{T}\frac{[\rho(\lambda_2'g_{tT}(\beta_2,\delta))-\rho_0]}{T}\right).\end{aligned}$$

To be more precise, the first-order conditions corresponding to the Lagrange multiplier λ are obtained from the maximization of the partial-sample GEL criterion for a given β_1, β_2, δ. Thus for a given s

$$\begin{aligned}\hat{\lambda}_{1T}(\beta_1,\delta,s) &= \arg\sup_{\lambda_1\in\widehat{\Lambda}_{1T}(\beta_1,\delta,s)}\sum_{t=1}^{[Ts]}\frac{[\rho(k\lambda_1'g_{tT}(\beta_1,\delta))-\rho_0]}{T},\\ \hat{\lambda}_{2T}(\beta_2,\delta,s) &= \arg\sup_{\lambda_2\in\widehat{\Lambda}_{2T}(\beta_2,\delta,s)}\sum_{t=[Ts]+1}^{T}\frac{[\rho(k\lambda_2'g_{tT}(\beta_2,\delta))-\rho_0]}{T}\end{aligned}$$

with $\widehat{\Lambda}_{1T}(\beta_1,\delta,s) = \{\lambda_1 : k\lambda_1'g_{tT}(\beta_1,\delta) \in \Phi, t = 1,\ldots,[Ts]\}$ and $\widehat{\Lambda}_{2T}(\beta_2,\delta,s) = \{\lambda_2 : k\lambda_2'g_{tT}(\beta_2,\delta) \in \Phi, t = [Ts]+1,\ldots,T\}$. The corresponding first-order conditions are given by:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{[Ts]}\rho_1\left(\hat{\lambda}_{1T}(\beta_1,\delta,s)'g_{tT}(\beta_1,\delta)\right)g_{tT}(\beta_1,\delta) &= 0\\ \frac{1}{T}\sum_{t=[Ts]+1}^{T}\rho_1\left(\hat{\lambda}_{2T}(\beta_2,\delta,s)'g_{tT}(\beta_1,\delta)\right)g_{tT}(\beta_2,\delta) &= 0.\end{aligned}$$

The partial-sample GEL estimators $\hat{\theta}_T(s) = \left(\hat{\beta}_{1T}(s)', \hat{\beta}_{2T}(s)', \hat{\delta}_T(s)'\right)'$ are the minimizer of the partial-sample GEL criterion. By writing $\hat{\lambda}_{1T}(s) = \hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), \hat{\delta}_T(s), s)$ and $\hat{\lambda}_{2T}(s) = \hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), \hat{\delta}_T(s), s)$, the corresponding first-order conditions are:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{[Ts]}\rho_1\left(\hat{\lambda}_{1T}(s)' g_{tT}(\hat{\beta}_{1T}(s),\hat{\delta}_T(s))\right)G_{tT}^{\beta}(\hat{\beta}_{1T}(s),\hat{\delta}_T))'\hat{\lambda}_{1T}(s) &= 0,\\
\frac{1}{T}\sum_{t=[Ts]+1}^{T}\rho_1\left(\hat{\lambda}_{2T}(s)' g_{tT}(\hat{\beta}_{2T}(s),\hat{\delta}_T(s))\right)G_{tT}^{\beta}(\hat{\beta}_{2T}(s),\hat{\delta}_T(s))'\hat{\lambda}_{2T}(s) &= 0,
\end{aligned}$$

and writing $\hat{\lambda}_T(\hat{\theta}_T(s),s)=\hat{\lambda}_T(s)$, the first-order conditions for δ are

$$\frac{1}{T}\sum_{t=1}^{T}\rho_1\left(\hat{\lambda}_T(s)' g_{tT}(\hat{\theta}_T(s),s)\right)G_{tT}^{\delta}(\hat{\theta}_T(s),s)'\hat{\lambda}_T(s)=0.$$

The next Theorem shows the convergence in probability of $\{\hat{\theta}_T(s),\hat{\lambda}_T(s),T\geq 1\}$ and the corresponding rate of convergence.

Theorem 2.1. *If Assumptions 6.1, 6.2, 6.3, 6.5, 6.6 and 6.7 are satisfied then for every sequence of PS-GEL estimators* $\{\hat{\theta}_T(s),\hat{\lambda}_T(s),T\geq 1\}$, $\sup_{s\in S}\left\|\hat{\theta}_T(s)-\theta_0\right\|\xrightarrow{p}0$ *and* $\sup_{s\in S}\left\|\hat{\lambda}_T(s)\right\|\xrightarrow{p}0$. *Moreover* $\sup_{s\in S}\left\|\hat{\lambda}_T(s)\right\|=O_p\left[\left(T/(2K_T+1)^2\right)^{-1/2}\right]$ *and* $\sup_{s\in S}\left\|\frac{1}{T}\sum_{t=1}^{T}g_{tT}(\hat{\theta}_T(s),s)\right\|=O_p(T^{-1/2})$.

Proof: See the Appendix.

Now we define the estimator

$$\widehat{\Omega}_T(s)=\begin{bmatrix} s\widehat{\Omega}_{1T}(s) & 0\\ 0 & (1-s)\widehat{\Omega}_{2T}(s)\end{bmatrix},$$

with

$$\widehat{\Omega}_{1T}(s) = \frac{2K_T+1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\beta}_{1T}(s),\hat{\delta}_T(s))g_{tT}(\hat{\beta}_{1T}(s),\hat{\delta}_T(s))'$$

and

$$\widehat{\Omega}_{2T}(s) = \frac{2K_T+1}{T-[Ts]}\sum_{t=[Ts]+1}^{T}g_{tT}(\hat{\beta}_{2T}(s),\hat{\delta}_T(s))g_{tT}(\hat{\beta}_{2T}(s),\hat{\delta}_T(s))'.$$

We denote $\{B(s):s\in[0,1]\}$ as q-dimensional vectors of mutually independent Brownian motions on $[0,1]$ and define

$$J(s)=\begin{bmatrix}\Omega^{1/2}B(s)\\ \Omega^{1/2}(B(1)-B(s))\end{bmatrix}$$

where $B(\pi)$ is a q-dimensional vector of standard Brownian motions.

The next Theorem shows the weak convergence of $\{\hat{\theta}_T(s),\hat{\lambda}_T(s),T\geq 1\}$.

Theorem 2.2. *Under Assumptions 6.1 to 6.12 and the null of no structural change, every sequence of PS-GEL estimators* $\{\hat{\theta}_T(\cdot), \hat{\lambda}_T(\cdot), T \geq 1\}$ *satisfies*

$$\begin{aligned} \sqrt{T}\left(\hat{\theta}_T(\cdot) - \theta_0\right) &\Rightarrow \left(G(\cdot)'\Omega(\cdot)^{-1}G(\cdot)\right)^{-1} G(\cdot)'\Omega(\cdot)^{-1}J(\cdot), \\ \frac{\sqrt{T}}{2K_T+1}\hat{\lambda}_T(\cdot) &\Rightarrow \left(\Omega(\cdot)^{-1} - \Omega(\cdot)^{-1}\left(G(\cdot)'\Omega(\cdot)^{-1}G(\cdot)\right)^{-1} G(\cdot)'\Omega(\cdot)^{-1}\right) J(\cdot) \end{aligned}$$

as a process indexed by $s \in S$, *where* S *has closure in (0,1) and the sequence GEL estimators* $\hat{\theta}_T(\cdot)$ *and the auxiliary sequence parameter estimator* $\hat{\lambda}_T(\cdot)$ *are asymptotically uncorrelated.*

Proof: See the Appendix.

The purpose of the next subsection is to refine the null hypothesis of no structural change. Such a refinement will enable us to construct various tests for structural change in the spirit of Sowell (1996a) and Hall and Sen (1999). Next, we present tests for parameter constancy, tests for stability of overidentifying restrictions and finally tests that are robust to some form of unidentification.

2.2 Refining the null hypothesis

The moment conditions for the full sample under the null can be written as: $Eg_t(\beta_0, \delta_0) = 0, \forall t = 1, \ldots, T$. Following Sowell (1996a), we can project the moment conditions on the subspace identifying the parameters and the subspace of overidentifying restrictions. In particular, considering the (standardized) moment conditions for the full-sample GMM estimator, such a decomposition corresponds to:

$$\Omega^{-1/2}Eg_t(\beta_0, \delta_0) = P_G\Omega^{-1/2}Eg_t(\beta_0, \delta_0) + (I_q - P_G)\Omega^{-1/2}Eg_t(\beta_0, \delta_0),$$

where $P_G = \Omega^{-1/2}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1/2}$. The first term is the projection identifying the parameter vector and the second term is the projection for the overidentifying restrictions. The projection argument enables us to refine (split) the null hypothesis (2), H_0^S. For instance, following Hall and Sen (1999) we can consider the null, $H_0^I(s)$, for the case of a single breakpoint in β by the projection on the space corresponding to G^β, which separates the identifying restrictions across the two subsamples:

$$H_0^I(s) = \begin{cases} P_{G^\beta}\Omega^{-1/2}E[g_t(\beta_0, \delta_0)] = 0 & \forall t = 1, \ldots, [Ts] \\ P_{G^\beta}\Omega^{-1/2}E[g_t(\beta_0, \delta_0)] = 0 & \forall t = [Ts]+1, \ldots, T. \end{cases}$$

Moreover, the overidentifying restrictions are stable if they hold before and after the breakpoint. This is formally stated as $H_0^O(s) = H_0^{O1}(s) \cap H_0^{O2}(s)$ with:

$$\begin{aligned} H_0^{O1}(s) : (I_q - P_G)\Omega^{-1/2}E[g_t(\beta_0, \delta_0)] &= 0 \qquad \forall t = 1, \ldots, [Ts] \\ H_0^{O2}(s) : (I_q - P_G)\Omega^{-1/2}E[g_t(\beta_0, \delta_0)] &= 0 \qquad \forall t = [Ts]+1, \ldots, T. \end{aligned}$$

We can then write the null hypothesis as $H_0^S = H_0^I(s)$ & $H_0^O(s)$. The projection reveals that instability must be a result of a violation of at least one of the three hypotheses: $H_0^I(s), H_0^{O1}(s)$ or $H_0^{O2}(s)$. Note that because the overidentifying restrictions are not used in estimation we can test their stability in each subsample. In contrast, because the identifying restrictions are used in estimation we can always find parameter values that satisfy them in each subsample. Hence we can not split H_0^I. Various tests can be constructed with local power properties against any particular one of these three null hypotheses (and typically no power against the others). To elaborate further on this we consider a sequence of local alternatives based on the moment conditions:

Assumption 2.1. *A sequence of local alternatives is specified as:*

$$Eg_t(\beta_0, \delta_0) = h(\eta, \tau, \frac{t}{T})/\sqrt{T} \tag{4}$$

where $h(\eta, \tau, r)$, for $r \in [0, 1]$, is a q-dimensional function. The parameter τ locates structural changes as a fraction of the sample size and the vector η defines the local alternatives[2]. These local alternatives are chosen to show that the structural change tests presented in this paper have non trivial power against a large class of alternatives. Also, our asymptotic results can be compared with Sowell's results for the GMM framework.

Now define

$$J^*(s) = \begin{bmatrix} \Omega^{1/2} B(s) - H(s) \\ \Omega^{1/2}(B(1) - B(s)) - (H(1) - H(s)) \end{bmatrix}$$

where $H(s) = \int_0^s h(\eta, \tau, r)dr$.

Theorem 2.3. *Under Assumptions 6.1 to 6.12 and the alternative (4), every sequence of PS-GEL estimators* $\{\hat{\theta}_T(\cdot), \hat{\lambda}_T(\cdot), T \geq 1\}$ *satisfies*

$$\begin{aligned} \sqrt{T}\left(\hat{\theta}_T(\cdot) - \theta_0\right) &\Rightarrow \left(G(\cdot)'\Omega(\cdot)^{-1}G(\cdot)\right)^{-1} G(\cdot)'\Omega(\cdot)^{-1}J^*(\cdot), \\ \frac{\sqrt{T}}{2K_T + 1}\hat{\lambda}_T(\cdot) &\Rightarrow \left(\Omega(\cdot)^{-1} - \Omega(\cdot)^{-1}\left(G(\cdot)'\Omega(\cdot)^{-1}G(\cdot)\right)^{-1} G(\cdot)'\Omega(\cdot)^{-1}\right) J^*(\cdot) \end{aligned}$$

as a process indexed by $s \in S$, where S has closure in (0,1).

Proof: See the Appendix.

3 Tests for structural change

3.1 Tests for parameter constancy

In this section we introduce several tests for structural change for parameter stability and establish their asymptotic distribution. The null hypothesis is (2), or more precisely $H_0^I(s)$. We present Wald, Lagrange multiplier

[2]The function $h(\cdot)$ allows for a wide range of alternative hypotheses (see Sowell (1996a)). In its generic form it can be expressed as the uniform limit of step functions, $\eta \in R^i$, $\tau \in R^j$ such that $0 < \tau_1 < \tau_2 < \ldots < \tau_j < 1$ and θ^* is in the interior of Θ. Therefore it can accommodate multiple breaks.

and likelihood ratio-type statistics based on smoothed moment conditions. The first is the usual Wald statistic which is given by:

$$Wald_T(s) = T\left(\hat{\beta}_{1T}(s) - \hat{\beta}_{2T}(s)\right)'(\hat{V}_\Omega(s))^{-1}\left(\hat{\beta}_{1T}(s) - \hat{\beta}_{2T}(s)\right),$$

where $\hat{V}_\Omega(s) = \left(\hat{V}_1(s)/s + \hat{V}_2(s)/(1-s)\right)$ and $\hat{V}_i(s) = \left(\hat{G}^\beta_{i,tT}(s)'\hat{\Omega}^{-1}_{i,T}(s)\hat{G}^\beta_{i,tT}(s)\right)^{-1}$ for $i = 1, 2$ corresponding to the first and the second part of the sample. For the first and the second subsamples:

$$\begin{aligned}
\widehat{G}^\beta_{1,tT}(s) &= \frac{1}{[Ts]}\sum_{t=1}^{[Ts]}\frac{\partial g_{tT}(\hat{\beta}_{1T}(s), \hat{\delta}_T(s))}{\partial \beta_1'}, \\
\widehat{G}^\beta_{2,tT}(s) &= \frac{1}{T-[Ts]}\sum_{t=[Ts]+1}^{T}\frac{\partial g_{tT}(\hat{\beta}_{2T}(s), \hat{\delta}_T(s))}{\partial \beta_2'}.
\end{aligned}$$

The Lagrange multiplier statistic does not involve estimators obtained from subsamples, rather it relies on full-sample parameter estimates. The $LM_T(s)$ simplifies to (see Andrews, 1993) :

$$LM_T(s) = \frac{T}{s(1-s)}\hat{g}_{1T}(\tilde{\theta}_T, s)'\widetilde{\Omega}_T^{-1}\widetilde{G}^\beta_{tT}\left[(\widetilde{G}^\beta_{tT})'\widetilde{\Omega}_T^{-1}\widetilde{G}^\beta_{tT}\right]^{-1}(\widetilde{G}^\beta_{tT})'\widetilde{\Omega}_T^{-1}\hat{g}_{1T}(\tilde{\theta}_T, s).$$

where

$$\begin{aligned}
\hat{g}_{1T}(\tilde{\theta}_T, s) &= \frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T), \\
\widetilde{G}^\beta_{tT} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T)}{\partial \beta'}.
\end{aligned}$$

Thus, the $LM_T(s)$ test corresponds to the projection of the smoothed moment conditions evaluated at the full-sample estimator on the subspace identifying the parameter vector β.

The LR-type statistic is defined as the difference between the GEL objective function for the full sample evaluated at the restricted estimator and the partial-sample GEL function evaluated at the unrestricted estimator:

$$LR_T(s) = \frac{2T}{2K+1}\left[\sum_{t=1}^{T}\frac{\left[\rho(\hat{\lambda}_T(\tilde{\theta}_T, s)'g_{tT}(\tilde{\theta}_T, s)) - \rho_0\right]}{T} - \sum_{t=1}^{T}\frac{\left[\rho(\hat{\lambda}_T(\hat{\theta}_T(s), s)'g_{tT}(\hat{\theta}_T(s), s)) - \rho_0\right]}{T}\right]$$

where $\hat{\lambda}_T(\tilde{\theta}_T, s) = \left(\tilde{\lambda}_{1T}(\tilde{\beta}_T, \tilde{\delta}_T, s)', \tilde{\lambda}_{2T}(\tilde{\beta}_T, \tilde{\delta}_T, s)'\right)'$ is the solution of the respective following maximization problem:

$$\hat{\lambda}_{1T}(\beta, \delta, s) = \arg\sup_{\lambda_1 \in \widehat{\Lambda}_{1T}(\beta,\delta,s)}\sum_{t=1}^{[Ts]}\frac{[\rho(\lambda_1' g_{tT}(\beta, \delta)) - \rho_0]}{T}$$

and

$$\hat{\lambda}_{2T}(\beta, \delta, s) = \arg\sup_{\lambda_2 \in \widehat{\Lambda}_{2T}(\beta,\delta,s)}\sum_{t=[Ts]+1}^{T}\frac{[\rho(\lambda_2' g_{tT}(\beta, \delta)) - \rho_0]}{T}$$

evaluated at the restricted estimator $\tilde{\theta}_T = \left(\tilde{\beta}_T', \tilde{\delta}_T'\right)'$ with $\widehat{\Lambda}_{1T}(\beta, \delta, s) = \{\lambda_1 : k\lambda_1' g_{tT}(\beta, \delta) \in \Phi, t = 1, \ldots, [Ts]\}$ and $\widehat{\Lambda}_{2T}(\beta, \delta, s) = \{\lambda_2 : k\lambda_2' g_{tT}(\beta, \delta) \in \Phi, t = [Ts]+1, \ldots, T\}$.

We state now the main Theorem which establishes the asymptotic distribution of the Wald, LM and LR-type test statistics under the null and the local alternative (4).

Theorem 3.1. *Under the null hypothesis H_0 in (2) and Assumptions 6.1 to 6.12, the following processes indexed by s for a given set S whose closure lies in $(0,1)$ satisfy:*

$$Wald_T(s) \Rightarrow Q_r(s), LM_T(s) \Rightarrow Q_r(s), LR_T(s) \Rightarrow Q_r(s),$$

with

$$Q_r(s) = \frac{BB_r(s)'BB_r(s)}{s(1-s)}$$

and under the alternative (4)

$$Q_r(s) = \frac{BB_r(s)'BB_r(s)}{s(1-s)} + \frac{(H(s) - sH(1))'\,\Omega^{-1/2} P_{G^\beta} \Omega^{-1/2}\,(H(s) - sH(1))}{s(1-s)},$$

where $BB_r(s) = B_r(s) - sB_r(1)$ is a Brownian bridge, B_r is r-vector of independent Brownian motions and $P_{G^\beta} = \Omega^{-1/2} G^\beta \left[(G^\beta)'\Omega^{-1}G^\beta\right]^{-1} (G^\beta)'\Omega^{-1/2}$.

Proof: See the Appendix.

Theorem 3.1 tells us that the asymptotic distributions under the null of the Wald, LR-type and LM statistics are the same as those obtained by Andrews (1993) for the GMM estimator. The asymptotic distribution under the null and the alternative given in the previous Theorem is only valid for smoothed moment conditions. Indeed, smoothing the moment conditions is necessary to obtain test statistics whose asymptotic distributions does not depend on nuisance parameters (except s). This also holds for other results in this paper.

When s is unknown, i.e. the case of unknown breakpoint, we can use the above result to construct statistics across $s \in S$. In the context of maximum likelihood estimation, Andrews and Ploberger (1994) derived asymptotic optimal tests which are characterized by an average exponential form. The Sowell (1996a) optimal tests are a generalization of the Andrews and Ploberger approach to the case of two measures that do not admit densities. The most powerful test is given by the Radon-Nikodym derivative of the probability measure implied by the local alternative with respect to the probability measure implied by the null hypothesis.

The optimal average exponential form is the following:

$$Exp - Q_T = (1+c)^{-r/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c} Q_T(s)\right) dH(s)$$

where various choices of c determine power against close or more distant alternatives and $H(\cdot)$ is the weight function over the value of $s \in S$. In the case of close alternatives ($c = 0$), the optimal test statistic takes the average form, $aveQ_T = \int_S Q_T(s)dH(s)$. For a distant alternative ($c = \infty$), the optimal test statistics takes the exponential form, $expQ_T = log\left(\int_S \exp[\frac{1}{2}Q_T(s)]dJ(s)\right)$. The supremum form often used in the literature corresponds to the case where $c/(1+c) \to \infty$. The sup test is given by $\sup Q_T = \sup_{s\in S} Q_T(s)$.

The following Theorem gives the asymptotic distribution for the exponential mapping for Q_T when Q_T corresponds to the Wald, LM and LR ratio-type tests under the null.

Theorem 3.2. *Under the null hypothesis H_0 in (2) and Assumptions 6.1 to 6.12, the following processes indexed by s for a given set S whose closure lies in (0,1) satisfy:*

$$\sup Q_T \Rightarrow \sup_{s\in S} Q_r(s), \quad aveQ_T(s) \Rightarrow \int_S Q_r(s)dJ(s), \quad expQ_T \Rightarrow log\left(\int_S \exp[\frac{1}{2}Q_r(s)]dJ(s)\right),$$

with

$$Q_r(s) = \frac{BB_r(s)'BB_r(s)}{s(1-s)}.$$

Proof: See Andrews (1993) and Andrews and Ploberger (1994).

This result is obtained through the application of the continuous mapping theorem (see Pollard (1984)). This implies that we can rely on the critical values tabulated for the case of GMM-based tests. For example, the critical values for the statistics defined by the supremum over all breakpoints $s \in S$ of $Wald_T(s)$, $LM_T(s)$ or $LR_T(s)$ can be found in the original paper by Andrews (1993). The same is true for the Sowell (1996a) and Andrews and Ploberger (1994) asymptotic optimal tests.

3.2 Tests for the stability of the overidentifying restrictions

The tests presented in the previous section are based on the projection of the moment conditions on the subspace of identifying restrictions. In this section we are interested with testing against violations of $H_0^{O1}(s)$ or $H_0^{O2}(s)$. The local alternatives are given by the projection of the moment condition on the subspace orthogonal to the identifying restrictions. For instance, in the case of a single breakpoint, the local alternatives by Assumption 2.1 correspond to:

$$\begin{aligned} H_A^{O1}(s) : (I_q - P_G)\Omega^{-1/2}E[g_t(\theta_0)] &= (I_q - P_G)\Omega^{-1/2}\frac{\eta_1}{\sqrt{T}} \qquad t = 1,\ldots,[Ts] \\ H_A^{O2}(s) : (I_q - P_G)\Omega^{-1/2}E[g_t(\theta_0)] &= (I_q - P_G)\Omega^{-1/2}\frac{\eta_2}{\sqrt{T}} \qquad t = [Ts]+1,\ldots,T \end{aligned}$$

Sowell (1996b) introduced optimal tests for the violation of the overidentifying restrictions when the violation occurs before the breakpoint corresponding to the alternative H_A^{O1}. The statistic is based on the projection of the partial sum of the full-sample estimator on the appropriate subspace. Hall and Sen (1999) introduce a test for the case where the violation can occurs before or after the breakpoint i.e. H_A^{O1} or H_A^{O2}. The statistic is based on the overidentifying restriction test for the sample before and after the considered breakpoint s.

We propose here statistics specially designed to detect instability before and after the possible breakpoint that are equivalent to Hall and Sen's statistics. In these tests, the entire parameter vector is allowed to vary for both subsamples. Thus $\theta = (\beta_1', \beta_2')'$. The first statistic is based on the same statistic as the one of Hall and Sen (1999) except that it is computed with smoothed moment conditions. The $O_T(s)$ statistic is the sum of the GMM-type criterion function for smoothed moment conditions in each subsample

$$O_T(s) = O1_T(s) + O2_T(s)$$

where

$$O1_T(s) = \left[\frac{1}{\sqrt{[Ts]}} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_{1T}(s)) \right]' \widehat{\Omega}_{1T}^{-1}(s) \left[\frac{1}{\sqrt{[Ts]}} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_{1T}(s)) \right]$$

and

$$O2_T(s) = \left[\frac{1}{\sqrt{(T-[Ts])}} \sum_{t=[Ts]+1}^{[T]} g_{tT}(\hat{\beta}_{2T}(s)) \right]' \widehat{\Omega}_{2T}^{-1}(s) \left[\frac{1}{\sqrt{(T-[Ts])}} \sum_{t=[Ts]+1} g_{tT}(\hat{\beta}_{2T}(s)) \right].$$

A new test for the GEL counterparts of $O_T(s)$ is based on the sum of its objective function for both subsamples, namely:

$$O_T^{GEL}(s) = O1_T^{GEL}(s) + O2_T^{GEL}(s)$$

where

$$O1_T^{GEL}(s) = \frac{2[Ts]}{2K+1} \sum_{t=1}^{[Ts]} \frac{\left[\rho(\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s)' g_{tT}(\hat{\beta}_{1T}(s))) - \rho_0\right]}{[Ts]}$$

and

$$O2_T^{GEL}(s) = \frac{2(T-[Ts])}{2K+1} \sum_{t=[Ts]+1}^{T} \frac{\left[\rho(\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s)' g_{tT}(\hat{\beta}_{2T}(s))) - \rho_0\right]}{T-[Ts]}.$$

The duality between overidentifying restrictions and the auxiliary Lagrange multiplier parameters $\lambda(\cdot)$ for the partial-sample estimation allows us to propose a new structural change test for overidentifying restrictions based on $\lambda(\cdot)$. This statistic is defined as following:

$$LM_T^O(s) = LM1_T^O(s) + LM2_T^O(s)$$

where

$$LM1_T^O(s) = \frac{[Ts]}{(2K_T+1)^2} \hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s)' \widehat{\Omega}_{1T}(s) \hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s)$$

and

$$LM2_T^O(s) = \frac{[T-Ts]}{(2K_T+1)^2} \hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s)' \widehat{\Omega}_{2T}(s) \hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s).$$

The equivalence with the overidentifying test of instability results from the fact that $\sqrt{[Ts]}/(2K_T+1)\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s)$ is asymptotically equivalent at the first order to $\widehat{\Omega}_{1T}(s)^{-1}\frac{1}{\sqrt{[Ts]}}\sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_{1T}(s))$ and $\sqrt{T-[Ts]}/(2K_T+1)\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s)$ to $\widehat{\Omega}_{2T}(s)^{-1}\frac{1}{\sqrt{T-[Ts]}}\sum_{t=[Ts]+1}^{T} g_{tT}(\hat{\beta}_{2T}(s))$.

The following Theorem provides the asymptotic distribution of $Q_T^O(s)$ which equals $O_T(s)$, $O_T^{GEL}(s)$ and $LM_T^O(s)$ under the null and the alternative hypotheses for the supremum, the average mapping and exponential mapping.

Theorem 3.3. *Under the null of no structural change and Assumptions 6.1 to 6.12, the following processes indexed by s for a given set S whose closure lies in (0,1) satisfy:*

$$\sup Q_T^O \Rightarrow \sup_{s\in s} Q_{q-r}(s), \quad aveQ_T^O \Rightarrow \int_S Q_{q-r}(s)dJ(s), \quad expQ_T^O \Rightarrow log\left(\int_S \exp[\frac{1}{2}Q_{q-r}(s)]dJ(s)\right),$$

with

$$Q_{q-r}(s) = \frac{B_{q-r}(s)'B_{q-r}(s)}{s} + \frac{[B_{q-r}(1)-B_{q-r}(s)]'[B_{q-r}(1)-B_{q-r}(s)]}{(1-s)}$$

and under the alternative (4)

$$\begin{aligned} Q_{q-r}(s) &= \frac{B_{q-r}(s)'B_{q-r}(s)}{s} + \frac{[B_{q-r}(1)-B_{q-r}(s)]'[B_{q-r}(1)-B_{q-r}(s)]}{(1-s)} \\ &+ \frac{H(s)'\Omega^{-1/2}(I-P_G)\Omega^{-1/2}H(s)}{(1-s)} + \frac{(H(1)-H(s))'\Omega^{-1/2}(I-P_G)\Omega^{-1/2}(H(1)-H(s))}{(1-s)} \end{aligned}$$

where $B_{q-r}(s)$ is a $q-r$-dimensional vector of independent Brownian motion.

Proof: See the Appendix.

The last two terms in the asymptotic distribution under the alternative given in Theorem 3.3 show that the test statistics have non trivial power to detect overidentifying restrictions instability before and after the possible breakpoint point. Note also that the asymptotic distributions under the null and the alternative are only valid for smoothed moment conditions. The asymptotic critical values for the interval $S = [.15, .85]$ can be found in Hall and Sen (1999). For other symmetric interval $[s_0, 1-s_0]$, critical values can be obtained in Guay (2003), Tables 1 to 3 for a number of overidentifying restrictions divided by 2 (in those tables). To see this, note that the critical values for the supremum, the average and the log exponential mappings applied to $\frac{B_{2q-2r}(s)'B_{2q-2r}(s)}{s}$ are equivalent to ones corresponding to $\frac{B_{q-r}(s)'B_{q-r}(s)}{s} + \frac{(B_{q-r}(1)-B_{q-r}(s))'(B_{q-r}(1)-B_{q-r}(s))}{1-s}$ for a symmetric interval S[3].

3.3 Structural change tests robust to weak identification or completely unidentified cases

We propose in this section test statistics robust to the context of weak identification as defined by Stock and Wright (2000) or to the completely unidentified case. Consider the pure structural change case, namely:

[3]This is verified by comparing the critical values in Hall and Sen (1999) and Guay (2003). The critical values in Table 1 in Hall and Sen for $q-r$ in our notation are the same than the critical values in Guay (2003) but for $2q-2r$.

$\theta = (\beta', \beta')'$. We first consider the null hypothesis (2) of a one time structural break in the parameter values presented in Section 2, i.e.

$$H_0 : \beta_1 = \beta_2 = \beta_0. \tag{5}$$

In this case, under the null $\theta_0 = (\beta_0', \beta_0')'$. To perform structural change tests, the parameters must be estimated under the null and/or under the alternative. The dependence of structural change test statistics on a parameter estimator complicates the derivation of the limit distribution in weakly identified case. In the presence of weak identification, some of the parameters are not consistent so we can not assume the existence of partial derivatives of the moment conditions with respect to the whole parameter vector. Consequently, traditional structural change test statistics are not asymptotically pivotal. To solve this problem, Caner (2007) proposed structural change tatistics in the continuous updating GMM framework for which the asymptotic distributions under the null are bounded. The corresponding asymptotic bound is robust to weak identification or completely unidentified cases and is free of nuisance parameters (except the interval for the breakpoint, as usual). We follow here the same strategy as Caner (2007) but in the GEL framework.

As aforementioned, we need to replace θ_0 by an estimator in order to perform stability tests. In that respect, let us introduce a restricted estimator $\tilde{\theta}_T(s) = \left(\tilde{\beta}_T(s)', \tilde{\beta}_T(s)'\right)'$ obtained with the partial-sample GEL objective function. A restricted partial-sample GEL estimator $\{\tilde{\theta}_T(s)\}$ is a sequence of random vectors such that:

$$\begin{aligned}
\tilde{\theta}_T(s) &= \arg\min_{\theta\in\Theta} \sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta,s)} \widehat{P}(\theta(s), \lambda(s), s) \\
&= \arg\min_{\theta\in\Theta} \sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta,s)} \left(\sum_{t=1}^{[Ts]} \frac{[\rho(\lambda_1' g_{tT}(\beta)) - \rho_0]}{T} + \sum_{t=[Ts]+1}^{T} \frac{[\rho(\lambda_2' g_{tT}(\beta)) - \rho_0]}{T} \right) \\
&= \arg\min_{\theta\in\Theta} \left[\sup_{\lambda_1\in\widehat{\Lambda}_{1T}(\beta,s)} \sum_{t=1}^{[Ts]} \frac{[\rho(\lambda_1' g_{tT}(\beta)) - \rho_0]}{T} + \sup_{\lambda_2\in\widehat{\Lambda}_{2T}(\beta,s)} \sum_{t=[Ts]+1}^{T} \frac{[\rho(\lambda_2' g_{tT}(\beta)) - \rho_0]}{T} \right]
\end{aligned}$$

for all $s \in S$ with $\lambda(s) = (\lambda_1', \lambda_2')' \in R^{2q\times 1}$, $\widehat{\Lambda}_T(\theta, s) = \{\lambda(s) = (\lambda_1', \lambda_2')' : \lambda(s)' g_{tT}(\theta, s)\}$ where $g_{tT}(\theta, s) = (g_{tT}(\beta)', 0')' \in R^{2q\times 1}$ for $t = 1, \ldots, [Ts]$ and $g_{tT}(\theta, s) = (0', g_{tT}(\beta)')' \in R^{2q\times 1}$ for $t = [Ts]+1, \ldots, T$. Thus for a given s

$$\begin{aligned}
\hat{\lambda}_{1T}(\beta, s) &= \arg \sup_{\lambda_1\in\widehat{\Lambda}_{1T}(\beta,s)} \sum_{t=1}^{[Ts]} \frac{[\rho(\lambda_1' g_{tT}(\beta)) - \rho_0]}{T}, \\
\hat{\lambda}_{2T}(\beta, s) &= \arg \sup_{\lambda_2\in\widehat{\Lambda}_{2T}(\beta,s)} \sum_{t=[Ts]+1}^{T} \frac{[\rho(\lambda_2' g_{tT}(\beta)) - \rho_0]}{T}
\end{aligned}$$

with $\widehat{\Lambda}_{1T}(\beta, s) = \{\lambda_1 : \lambda_1' g_{tT}(\beta) \in \Phi, t = 1, \ldots, [Ts]\}$ and $\widehat{\Lambda}_{2T}(\beta, s) = \{\lambda_2 : \lambda_2' g_{tT}(\beta) \in \Phi, t = [Ts]+1, \ldots, T\}$. For this restricted partial-sample GEL, the parameter vector β is restricted to be stable across the sample while the Lagrange multiplier parameters are allowed to vary across subsamples in contrast to the full-sample GEL.

A robust test based on the GEL is composed of the sum of its renormalized objective function for both subsamples, namely:

$$GELR_T(s) = GELR1_T(s) + GELR2_T(s) = \frac{2T}{2K_T+1}\widehat{P}(\tilde{\theta}_T(s), \hat{\lambda}_T(\tilde{\theta}_T(s), s), s)$$

where

$$GELR1_T(s) = \frac{2[Ts]}{2K+1}\sum_{t=1}^{[Ts]}\frac{\left[\rho(\hat{\lambda}_{1T}(\tilde{\beta}_T, s)' g_{tT}(\tilde{\beta}_T(s))) - \rho_0\right]}{[Ts]}$$

and

$$GELR2_T(s) = \frac{2(T-[Ts])}{2K+1}\sum_{t=[Ts]+1}^{T}\frac{\left[\rho(\hat{\lambda}_{2T}(\tilde{\beta}_T, s)' g_{tT}(\tilde{\beta}_T(s))) - \rho_0\right]}{T-[Ts]}.$$

A similar statistic was introduced by Guggenberger and Smith (2008) and Otsu (2006) for testing $H_0 : \theta = \theta_0$ without considering structural change. In their cases, the derivation is facilitated because θ_0 is known.

The GEL framework allows us to propose an asymptotically equivalent statistic based on the Lagrange multiplier parameters $\lambda(\cdot)$ evaluated at $\tilde{\theta}_T(s)$[4]. The statistic is defined as:

$$LM_T^R(s) = LM1_T^R(s) + LM2_T^R(s)$$

where

$$LM1_T^R(s) = \frac{[Ts]}{(2K_T+1)^2}\hat{\lambda}_{1T}(\tilde{\beta}_T(s), s)'\widehat{\Omega}_{1T}(\tilde{\beta}_T(s), s)\hat{\lambda}_{1T}(\tilde{\beta}_T(s), s)$$

and

$$LM2_T^R(s) = \frac{[T-Ts]}{(2K_T+1)^2}\hat{\lambda}_{2T}(\tilde{\beta}_T(s), s)'\widehat{\Omega}_{2T}(\tilde{\beta}_T(s), s)\hat{\lambda}_{2T}(\tilde{\beta}_T(s), s).$$

We show in the Appendix that both test statistics are asymptotically equivalent at the first order to the S-based test statistic in Caner (2007). The test statistic is not asymptotically pivotal but asymptotically boundedly pivotal. The bound is then nuisance parameters free and robust to identification problems under the null. The following Theorem gives this asymptotic bound under the null of no structural change and the local alternative (4).

Theorem 3.4. *Suppose that Assumptions 6.1 to 6.5 and 6.7 to 6.12 hold at the true value of the parameters θ_0, the processes $GELR_T(s)$ and $LM_T^R(s)$ indexed by s for a given set S whose closure lies in $(0,1)$ are asymptotically boundedly pivotal and the asymptotic bound distribution is given by:*

$$Q_q^R(s) \Rightarrow \frac{B_q(s)'B_q(s)}{s} + \frac{[B_q(1)-B_q(s)]'[B_q(1)-B_q(s)]}{1-s}$$

[4] We can also propose a LR-type test statistic as in Caner (2007) but for GEL framework. However, Caner (2007) shows that the LR-type statistic can be very conservative when the number of moment conditions is large compared to the number of parameters. We can show that this result holds also in the GEL framework for smoothed moment conditions. Moreover, simulation results provided by Caner (2007) confirm this and his S-based statistic clearly outperforms the LR-type statistic. For this reason, we do not present the GEL version of the LR-type statistic.

under the null of no structural change and under the alternative (4)

$$
\begin{aligned}
Q_q^R(s) \quad \Rightarrow \quad & \frac{B_q(s)'B_q(s)}{s} + \frac{H(s)'\Omega(\theta_0)^{-1}H(s)}{s} \\
& \frac{[B_q(1)-B_q(s)]'\,[B_q(1)-B_q(s)]}{(1-s)} + \frac{[H(1)-H(s)]'\,\Omega(\theta_0)^{-1}\,[H(1)-H(s)]}{(1-s)},
\end{aligned}
$$

where $B_q(s)$ is a q-vector of standard Brownian motion.

Proof: See the Appendix.

The asymptotic bound derived in this Theorem depends on the number of moment conditions and the derivation under the alternative shows that both test statistics can have no trivial power against instability of parameters and overidentifying restrictions. Since the asymptotic bound is valid for $\forall s \in S$, the supremum, the average and the exponential mappings of both statistics are also asymptotic bounded by the respective mapping of the bound. Critical values under the null for the different mappings are given in the same tables than those in the subsection 3.2.

Now we propose a second set of tests based on the first-order conditions evaluated at the restricted partial sample GEL estimator. The first statistic is similar to the one proposed by Caner (2007) for the GMM-CUE which is a Kleibergen (2005)-type statistic but adapted here for the GEL context. To introduce the statistic, we need to define the following matrices:

$$
\begin{aligned}
\hat{D}_{1T}(\beta, s) &= \frac{1}{T}\sum_{t=1}^{[Ts]} \rho_1(\hat{\lambda}_{1T}(\beta, s)'g_{tT}(\beta))G_{tT}(\beta), \\
\hat{D}_{2T}(\beta, s) &= \frac{1}{T}\sum_{t=[Ts]+1}^{T} \rho_1(\hat{\lambda}_{2T}(\beta, s)'g_{tT}(\beta))G_{tT}(\beta).
\end{aligned}
$$

For $t = 1, \ldots, [Ts]$, we also define

$$K_{1tT}(\beta, s) = \hat{D}_{1T}(\beta, s)'\hat{\Omega}_{1T}(\beta, s)^{-1}g_{tT}(\beta)$$

and for $t = [Ts] + 1, \ldots, T$

$$K_{2tT}(\beta, s) = \hat{D}_{2T}(\beta, s)'\hat{\Omega}_{2T}(\beta, s)^{-1}g_{tT}(\beta)$$

with

$$\hat{\Omega}_{1T}(\beta, s) = \frac{2K_T + 1}{[Ts]}\sum_{t=1}^{[Ts]} g_{tT}(\beta)g_{tT}(\beta)'$$

and

$$\hat{\Omega}_{2T}(\beta, s) = \frac{2K_T + 1}{T - [Ts]}\sum_{t=[Ts]+1}^{T} g_{tT}(\beta)g_{tT}(\beta)'.$$

We now need to introduce another restricted estimator $\tilde{\theta}_{K,T}(s) = \left(\tilde{\beta}_{K,T}(s)', \tilde{\beta}_{K,T}(s)'\right)'$ obtained with the restricted partial-sample GEL objective function with $K_{1t}(\beta, s)$ and $K_{2t}(\beta, s)$ as moment conditions, namely

$$\begin{aligned}
\tilde{\theta}_{K,T}(s) &= \arg\min_{\theta\in\Theta} \sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta,s)} \widehat{P}_K(\theta(s), \upsilon(s), s) \\
&= \arg\min_{\theta\in\Theta} \sup_{\upsilon(s)\in\widehat{\Upsilon}_T(\theta,s)} \left(\sum_{t=1}^{[Ts]} \frac{[\rho(\upsilon_1' K_{1tT}(\beta, s)) - \rho_0]}{T} + \sum_{t=[Ts]+1}^{T} \frac{[\rho(\upsilon_2' K_{2tT}(\beta, s)) - \rho_0]}{T} \right) \\
&= \arg\min_{\theta\in\Theta} \left[\sup_{\upsilon_1\in\hat{\Upsilon}_{1T}(\beta,s)} \left(\sum_{t=1}^{[Ts]} \frac{[\rho(\upsilon_1' K_{1tT}(\beta, s)) - \rho_0]}{T} \right) + \sup_{\upsilon_2\in\hat{\Upsilon}_{2T}(\beta,s)} \left(\sum_{t=[Ts]+1}^{T} \frac{[\rho(\upsilon_2' K_{2tT}(\beta, s)) - \rho_0]}{T} \right) \right]
\end{aligned}$$

for all $s \in S$ where $K_{1tT}(\beta, s) \in R^{r\times 1}$ and $K_{2tT}(\beta, s) \in R^{r\times 1}$ with $\upsilon(s) = (\upsilon_1', \upsilon_2')' \in R^{2r\times 1}$ and $\widehat{\Upsilon}_T(\theta, s) = \{\upsilon(s) = (\upsilon_1', \upsilon_2')' : \upsilon(s)' K_{tT}(\theta, s) \in \Phi, t = 1, \ldots, T\}$ where $K_{tT}(\theta, s) = (K_{1tT}(\beta, s)', 0')' \in R^{2r\times 1}$ for $t = 1, \ldots, [Ts]$ and $K_{tT}(\theta, s) = (0', K_{2tT}(\beta, s)')' \in R^{2r\times 1}$ for $t = [Ts] + 1, \ldots, T$.

The $KGEL_T(s)$-statistic for testing the null hypothesis of parameter stability defined in (5) is, for a given $s \in S$:

$$KGEL_T(s) = KGEL1_T(s) + KGEL2_T(s)$$

where

$$KGEL1_T(s) = \frac{2[Ts]}{2K+1} \sum_{t=1}^{[Ts]} \frac{\left[\rho(\hat{\upsilon}_{1T}(\tilde{\beta}_{K,T}(s), s)' K_{1tT}(\tilde{\beta}_{K,T}(s), s))) - \rho_0\right]}{[Ts]}$$

and

$$KGEL2_T(s) = \frac{2(T-[Ts])}{2K+1} \sum_{t=[Ts]+1}^{T} \frac{\left[\rho(\hat{\upsilon}_{2T}(\tilde{\beta}_{K,T}(s), s)' K_{2tT}(\tilde{\beta}_{K,T}(s))) - \rho_0\right]}{T-[Ts]}.$$

The GEL framework also allows us to propose an asymptotically equivalent test statistic based on the Lagrange multiplier parameters $\upsilon(\cdot)$ for the moment conditions $K_{1t}(\beta, s)$ and $K_{2t}(\beta, s)$ evaluated at $\tilde{\beta}_{K,T}(s)$. The statistic is defined as:

$$KLM_T^R(s) = KLM1_T^R(s) + KLM2_T^R(s)$$

where

$$KLM1_T^R(s) = \frac{[Ts]}{(2K_T+1)^2} \hat{\upsilon}_{1T}(\tilde{\beta}_{K,T}(s), s)' \left(\hat{D}_{1T}(\tilde{\beta}_{K,T}(s), s)' \hat{\Omega}_{1T}(\tilde{\beta}_{K,T}(s), s)^{-1} \hat{D}_{1T}(\tilde{\beta}_{K,T}(s), s)\right) \hat{\upsilon}_{1T}(\tilde{\beta}_{K,T}(s), s)$$

and

$$KLM2_T^R(s) = \frac{[T-Ts]}{(2K_T+1)^2} \hat{\upsilon}_{2T}(\tilde{\beta}_{K,T}(s), s)' \left(\hat{D}_{2T}(\tilde{\beta}_{K,T}(s), s)' \hat{\Omega}_{2T}(\tilde{\beta}_{K,T}(s), s)^{-1} \hat{D}_{2T}(\tilde{\beta}_{K,T}(s), s)\right) \hat{\upsilon}_{2T}(\tilde{\beta}_{K,T}(s), s).$$

Theorem 3.5. *The $KGEL_T(s)$ and $KLM_T^R(s)$ processes indexed by s for a given set S whose closure lies in $(0,1)$ are asymptotically boundedly pivotal and the asymptotic bound distribution is given by:*

$$Q_p(s) \Rightarrow \frac{B_r(s)'B_r(s)}{s} + \frac{[B_r(1)-B_r(s)]'[B_r(1)-B_r(s)]}{1-s}$$

under the null of no structural change and under the alternative (4)

$$\begin{aligned} Q_p(s) \quad \Rightarrow \quad & \frac{B_r(s)'B_r(s)}{s} + \frac{H(s)'\Omega(\beta_0)^{-1/2}P_{G(\beta_0)}\Omega(\beta_0)^{-1/2}H(s)}{s} \\ & \frac{[B_r(1)-B_r(s)]'[B_r(1)-B_r(s)]}{1-s} + \frac{[H(1)-H(s)]'\Omega(\beta_0)^{-1/2}P_{G(\beta_0)}\Omega(\beta_0)^{-1/2}[H(1)-H(s)]}{(1-s)}, \end{aligned}$$

where $B_r(s)$ is a r-vector of standard Brownian motion and

$$P_{G(\beta_0)} = \Omega(\beta_0)^{-1/2}G(\beta_0)\left(G(\beta_0)'\Omega(\beta_0)^{-1}G(\beta_0)\right)^{-1}G(\beta_0)'\Omega(\beta_0)^{-1/2}$$

with $G(\beta_0) = \lim_{T\to\infty}\left[T^{-1}\sum_{t=1}^{T}\partial g_t(\beta_0)/\partial\beta'\right]$.

Proof: See the Appendix.

The asymptotic bound depends on the number of parameters rather than the number of moment conditions. The asymptotic bound under the alternative shows that these test statistics are specifically designed to detect instability in parameter values. Critical values under the null for the different mappings are also given in the same tables than those in the subsection 3.2. The asymptotic bound under the local alternative allows us to examine the power of the test statistic under different assumptions with respect to identification. Consider the following decomposition of the alternative:

$$\frac{h(\eta,\tau,\frac{t}{T})}{\sqrt{T}} = P_{G(\beta_0)}\Omega(\beta_0)^{-1/2}\frac{h(\eta,\tau,\frac{t}{T})}{\sqrt{T}} + \left(I_q - P_{G(\beta_0)}\right)\Omega(\beta_0)^{-1/2}\frac{h(\eta,\tau,\frac{t}{T})}{\sqrt{T}}.$$

This decomposition and the asymptotic bound under the alternative show that the ability of the test statistic to detect a structural change in the parameter values depends on the Jacobian matrix $G(\beta_0)$. Under weak identification, as defined by Stock and Wright (2000), $G_T(\beta_0)$ has a weak value which means that $G_T(\beta_0) = \frac{C}{\sqrt{T}}$ for a C matrix of dimension $q \times p$. In this case, the test statistic has trivial power equals to the size. Obviously, it is also the case under unidentification since $G(\beta_0) = 0$. In fact, the test statistic will detect instability in parameter values for alternatives such that $\frac{h(\eta,\tau,\frac{t}{T})}{T^{\alpha}}$ for $\alpha \geq 1$ in the weak identification case. For instance, the test statistic will detect structural change in the parameter values with no trivial power for the following fixed alternative:

$$H_A^I(s) = \begin{cases} \beta_1(s) &= \beta_0 & \forall t = 1,\ldots,[Ts] \\ \beta_2(s) &= \beta_0 + \eta & \forall t = [Ts]+1,\ldots,T. \end{cases}$$

The discussion above also holds for the S-based test statistic proposed by Caner (2007) who derived the bound only under the null.

4 Simulation evidence

In this section, we investigate the small sample properties of structural change tests in the GEL context for well identified parameters. We leave for future investigation the study of small sample properties of the proposed structural change tests for weakly identified and unidentified parameters in nonlinear models estimated by GEL methods.

The tests for an unknown structural change presented in this paper fall in two categories, all considering structural stability based on GEL estimators. The two categories are the result of splitting the null hypothesis into two components to explore alternative sources of instability. The first set of tests, the $Wald$, LM and LR tests, consider a structural change in the parameters of our model while the second set of tests, the O, O^{GEL} and LM^O tests, focus on a structural change in the overidentifying restrictions imposed on our model. We evaluate the performance of these tests in a simulated environment identical to the one found in Guay and Lamarche (2009) who proposed test statistics to detect structural change that are based on the estimated weights of a GEL problem. This environment was also used by Ghysels *et al.* (1997) and in Hall and Sen (1999) and consists of an autoregressive process of order one for a time series x_t. Only one parameter is estimated (denoted by β in the expression below) and two moment conditions formed with the lagged values of x_t are used. We therefore have one overidentifying restriction.

The data generating process is given by

$$x_t = \beta_i x_{t-1} + u_t$$

for $t = 1, \ldots, T$. Structural change in the identifying restrictions (in the parameter) is studied by considering different values of β_i where the index $i = 1, 2$ denotes the first or second subsamples. Structural stability in the overidentifying restrictions is studied by allowing for an ARMA(1,2) model

$$x_t = \beta_i x_{t-1} + u_t + \alpha u_{t-2}$$

and considering nonzero values of α in the second subsample. The change is set at $T/2$. In the above, $u_t \sim N(0, 1)$. The sample size was set to 200 observations and the number of Monte Carlo replications was 2000.

Table 1 summarizes the different parametrization and is adapted from Hall and Sen (1999). The null hypothesis of structural stability is denoted by H_0^S (DGP 1 to 3). For those DGPs we vary the magnitude of the autoregressive parameter β. The alternatives of instability in the parameters or in the overidentifying restrictions are denoted by H_A^I (DGP 4 to 6), where we vary the magnitude of the change in the autoregressive parameter, and H_A^O (DGP 7 to 10) where we consider various values of the moving average parameter, respectively. In this situation only one parameter is estimated using two moment conditions created with the first two lags of x_t. Under H_0^S, where $\alpha = 0$, the instruments are appropriate. Under the first class of alternative hypothesis (H_A^I)

the two instruments are also valid while they no longer are for the second part of the sample with the second class of alternative hypothesis (H_A^O).

Under the null hypothesis (DGP 1 to 3) and for DGP 4 to 6, we fix $K_T = 0$ since the error term u_t are uncorrelated. For DGPs 7, 8, 9 and 10, we select a bandwidth parameter M_T by the automatic, data-driven, procedure proposed by Newey and West (1994) and K_T is fixed as $K_T = [(M_T - 1)/2]$. The average K_T, taken over the Monte Carlo replications, was found to vary between 1.6 and 2.3, increasing with the moving average component. Lastly, a trimming rule of 0.15 was used, namely $S = [.15, .85]$.

Table 2 contains the rejection frequencies for the test statistics designed to have power against a structural change in the parameters while Table 3 presents the results for test statistics which are designed to have power against a structural change in the overidentifying restrictions. All test statistics were computed in the GEL setting and the supremum, exponential and average versions of the test statistics are presented.

Focusing first on size we find that the average and exponential versions of the Lagrange multiplier-based test, LM^O, is very accurate for all data generating processes under the null hypothesis. The O-test proposed by Hall and Sen (1999) is ranked second with some underrejection while the GEL counterpart of the O test, the O^{GEL} test which is based on the sum of the GEL objective function in both subsamples, overrejects the null hypothesis. The average mapping of the O^{GEL} test performs best, having overrejection magnitudes similar to the underrejection magnitudes of the O test. The standard test (but computed in a GEL setting) for an unknown structural change, the $Wald$ and LR tests, also overreject the null hypothesis with the average and exponential LR mappings performing best. The LM test statistics (supremum, average and exponential mappings) significantly underreject for all DGPs (1 to 3) under the null and have poor power for other DGPs. For these reasons, the rejection frequencies for the LM test statistics are not reported in the tables here.

The study of power is divided into two cases. In case 1 structural change occurs in the parameter values while in case 2 structural change occurs in the overidentifying restrictions. Under the alternative of instability in the parameter, H_A^I (DGP 4 to 6), we see that the O, O^{GEL} and LM^O tests have no useful power as they are not geared towards these type of deviations from the null hypothesis. The $Wald$ and LR have good power properties but one has to keep in mind that their power is inflated due to their overrejections under the null hypothesis.

Under the alternative of instability in the overidentifying restrictions, e.g. H_A^O, (DGP 7 to 10), we see that the test statistics (the $Wald$ and LR tests) specially designed to detect a change in the parameter have much less power than O, O^{GEL} and LM^O tests. An important result in this paper is that the LM^O (and O) tests, tests that suffer from very little size distortions, have the largest power, and particularly so for the average mapping. As in Hall and Sen (1999), the power of the O, O^{GEL} and LM^O tests is greater than the power for the $Wald$ and LR tests indicating that testing for an unknown change in the overidentifying restrictions is important in empirical applications.

The increase in the autoregressive coefficient from 0 to 0.8 does not impact greatly on the rejection frequencies under the null hypothesis but under the alternative hypotheses the magnitude of the change is important. Under H_A^I, for example, we see that power is close to unity when the change in the autoregressive parameter is quite extreme (0 to 0.8, see DGP 5). Under H_A^O, which captures a change in the overidentifying restrictions, an increase (in absolute terms) in the moving average coefficient increases power.

The tests statistics presented in this paper, in particular the LM^O tests, should then be added to the Pearson-type statistics based on implied probabilities to detect structural change presented by Guay and Lamarche (2009) to complement the specification and testing arsenal of the practitioners. All of these test statistics are computed using generalized empirical likelihood estimators and are are computed in a single step which eliminates the need to compute the weighting matrix required for GMM estimation.

5 Conclusion

In this paper we have examined tests for structural change that are based on generalized empirical likelihood methods and applicable to a time series context. Given the recent developments of generalized empirical likelihood methods as an alternative to GMM, it appears important to study structural change tests for these methods of estimation. Test statistics were considered for cases in which the parameters are fully identified as well as for cases of weak identification and complete unidentification.

We introduced a class of partial-sample GEL estimators and showed that estimators of the Lagrange multiplier parameters weakly converge to a function of Brownian motions uncorrelated to the asymptotic distribution of the vector of parameters. These asymptotic distributions are derived under the null hypothesis of stability and general alternatives of structural change for an unknown breakpoint. These results allowed us to derive the asymptotic distributions of structural change tests in the GEL context. Specifically targeted tests, either to a structural change in the parameters or a structural change in the overidentifying restrictions used to estimate them, were considered. For the former, we showed that, in a time series context, our test statistics based on the GEL method followed the same asymptotic distribution than in the GMM context (Andrews, 1993). For the latter, test statistics equivalent to Hall and Sen's (1999) statistics in the GMM context were adapted to the GEL method for smoothed moment conditions. Further, we proposed two new tests specific to the GEL framework to detect instability in the overidentifying restrictions. We showed that these new statistics have the same asymptotic distribution at first order than the one derived by Hall and Sen (1999). This paper also proposed test statistics of structural change in the context of weakly identified or completely unidentified cases for the GEL framework. The first one is based on a renormalized criterion function of GEL evaluated at a restricted PS estimator. The second is asymptotically equivalent to the first and is based on the Lagrange multiplier of the restricted partial-sample estimator. These test statistics are not asymptotically pivotal and we show that their limits are bounded by a distribution which is nuisance parameter free and robust to identification problems. For the first group, the asymptotic bound is a function of the number of moment conditions while for the second group, the asymptotic bound depends on the number of parameters. Our simulation study revealed that one of the newly proposed test, the Lagrange multiplier-based test, has very good finite samples properties, both in terms of size and power. We are currently investigating the small sample properties of tests for a structural change when the parameters are not fully identified. This will be accomplished using nonlinear models estimated by GEL methods.

6 Appendix

6.1 Assumptions

We consider triangular arrays because they are required to derive asymptotic results under the Pitman drift alternatives. Define $\mathbf{X}$ to be the domain of $g(\cdot, \theta)$ to include the support of $x_{Tt}, \forall t, \forall T$. Let B_0 and Δ_0 denote compact subsets of R^r and R^ν that contains neighborhoods of β_0 and δ_0 in the parameter spaces B and Δ. Finally, let μ_{Tt} denote the distribution of x_{Tt} and let $\bar{\mu}_T = (1/T)\sum_{t=1}^{T}\mu_{Tt}$. Throughout the Appendix, w.p.a.1 means with probability approaching one; p.s.d. denotes positive semi-definite; $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix; $\overset{p}{\to}$ and $\overset{d}{\to}$ denote respectively convergence in probability and in distribution and $\Rightarrow$ denotes weak convergence as defined by Pollard (1984, pp. 64-66). Finally, C denotes a generic positive constant that may differ according to its use.

Assumption 6.1. *$\{x_{Tt} : t \leq T, T \geq 1\}$ is a triangular array of $\mathbf{X}$-valued rv's that is L^0-near epoch dependent (NED) on a strong mixing base $\{Y_{Tt} : t = \dots, 0, 1, \dots; T \geq 1\}$, where $\mathbf{X}$ is a Borel subset of R^k, and $\{\mu_{Tt} : T \geq 1\}$ is tight on $\mathbf{X}$*[5].

Define the smoothed moment conditions as:[6]

$$g_{tT}(\beta, \delta) = \frac{1}{M_T}\sum_{m=t-T}^{t-1} k\left(\frac{m}{M_T}\right) g\left(x_{Tt-m}, \beta, \delta\right)$$

for an appropriate kernel and M_T is a bandwidth parameter. From now on, we consider the uniform kernel proposed by Kitamura and Stutzer (1997):

$$g_{tT}(\beta, \delta) = \frac{1}{2K_T + 1}\sum_{m=-K_T}^{K_T} g\left(x_{Tt-m}, \beta, \delta\right).$$

Assumption 6.2. *$K_T/T^2 \to 0$ and $K_T \to \infty$ as $T \to \infty$ and $K_T = O_p\left(T^{\frac{1}{2\eta}}\right)$ for some $\eta > 1$*[7].

Assumption 6.3. *For some $d > \max\left(2, \frac{2\eta}{\eta-1}\right)$, $\{g\left(x_{Tt}, \beta, \delta\right) : t \leq T, T \geq 1\}$ is a triangular array of mean zero R^q-valued rv's that is L^2-near epoch dependent of size $-\frac{1}{2}$ on a strong mixing base $\{Y_{Tt} : t = \dots, 0, 1, \dots; T \geq 1\}$, of size $-d/(d-2)$ and $\sup \|g\left(x_{Tt}, \beta, \delta\right)\|^d < \infty$.*

Assumption 6.4. *$Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{Ts} g\left(x_{Tt}, \beta, \delta\right)\right) \to s\Omega \ \ \forall s \in [0, 1]$ for some positive definite $q \times q$ matrix Ω.*

The above assumptions are sufficient to yield weak convergence of the standardized partial sum of the smoothed moment conditions under the null and the alternatives (see Lemmas 6.1 and 6.2). In the following, x_t is used to denote x_{Tt} for notational simplicity.

[5]For a definition of L^p-near epoch dependence and tightness, see Andrews (1993, p. 829-830). For a presentation of the concept of near epoch dependence, we refer the reader to Gallant and White (1988) (chapters 3 and 4).

[6]Note here that g_{tT} denotes the smoothed moment conditions and x_{Tt} a triangular array of random variables.

[7]This assumption is slightly different than that in Smith (2004) but facilitates the proofs at no real cost.

Assumption 6.5. *$\tilde{g}(\beta_0, \delta_0) = 0$ with $(\beta_0, \delta_0) \in B \times \Delta$ where $\tilde{g}(\beta_0, \delta_0) = \lim_{T\to\infty} \sum_{t=1}^{T} Eg(x_t, \beta, \delta)$ and B and Δ are bounded subsets of R^r and R^ν, $g(x_t, \beta, \delta)$ is continuous in x for all $(\beta, \delta) \in B \times \Delta$ and is continuous in (β, δ) uniformly over $(\beta, \delta, x) \in B \times \Delta \times C$ for all compact sets $C \subset \mathbf{X}$;*

Assumption 6.6. *For every neighborhood $\Theta_0 \subset \Theta$ of θ_0, $\inf_{s\in S}\left(\inf_{\theta\in\Theta/\Theta_0} \|g(\theta, s)\|\right) > 0$ where $g(\theta, s) = (s\tilde{g}(\beta_1, \delta)', (1-s)\tilde{g}(\beta_2, \delta)')'$.*

Assumption 6.7. ***(a)** $\rho(\cdot)$ is twice continuously differentiable and concave on its domain, an open interval Φ containing 0, $\rho_1 = \rho_2 = -1$; **(b)** $\lambda(s) \in \widehat{\Lambda}_T(s)$ where $\widehat{\Lambda}_T(s) = \{\lambda(s) : \|\lambda(s)\| \leq D\left(T/((2K_T+1)^2\right)^{-\zeta}\}$ for some $D > 0$ with $\frac{1}{2} > \zeta > \frac{1}{d(1-1/\eta)}$.*

Assumption 6.7 (b) parallels the assumption in Newey and Smith (2004) and Smith (2004) but for $\lambda(s) = (\lambda_1', \lambda_2')'$. It specifies bounds on $\lambda(s)$ and with the existence of higher than second moments in Assumption 6.3 leads to the arguments $\lambda(s)' g_{tT}(\theta, s)$ being in the domain Φ of $\rho(\cdot)$ w.p.a.1 in the first subsample for all β_1, δ and $1 \leq t \leq [Ts]$ and in the second subsample for all β_2, δ and $[Ts]+1 \leq t \leq T$ (see Lemma 6.3).

Under Assumptions 6.1, 6.2, 6.3, 6.5, 6.6 and 6.7, we show for the partial-sample GEL estimator that $\sup_{s\in S} \|\hat{\theta}_T(s) - \theta_0\| \xrightarrow{p} 0$, $\sup_{s\in S} \|\hat{\lambda}_T(s)\| \xrightarrow{p} 0$, $\|\hat{\lambda}_T(s)\| = O_p\left(T/(2K_T+1)^2\right)^{-1/2}$ and $\sup_{s\in S} \|\frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T(s), s)\| = O_p(T^{-1/2})$.

The consistency of the full-sample GEL estimator is obtained by slight modifications of Assumptions 6.6 and 6.7 (b). Assumption 6.6 must be modified by a simplified version with $\tilde{g}(\beta, \delta)$ instead of $g(\theta, s)$. Assumption 6.7 (b) holds but for the full-sample Lagrange multiplier λ. The consistency result that $\tilde{\theta}_T \xrightarrow{p} \theta_0$ is then derived under weaker conditions than in Smith (2004).

The following high level assumptions are sufficient to derive the weak convergence under the null of the PS-GEL estimators $\hat{\theta}_T(s)$ and $\hat{\lambda}_T(s)$. These assumptions are similar to the ones in Andrews (1993).

Assumption 6.8. $\sup_{s\in S} \|\hat{\Omega}_{iT}(s) - \Omega\| \xrightarrow{p} 0$ *where Ω is defined in Section 2.1 and S whose closure lies in* $(0, 1)$ *for $i = 1, 2$.*

Assumption 6.8 holds under conditions given in Andrews (1991) and Lemma A.3 in Smith (2004). To respect these conditions, Assumption 6.3 can be replaced by the following assumption:

Assumption 6.3′. $\{g(x_{Tt}, \beta, \delta) : t \leq T, T \geq 1\}$ is a triangular array of mean zero R^q-valued rv's that is α-mixing with mixing coefficients $\sum_{j=1}^{\infty} j^2 \alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$ and $sup_{t\leq T, T\geq 1} E\|g(x_{Tt}, \beta, \delta)\|^d < \infty$ for some $d > \max\left(4\nu, \frac{2\eta}{\eta-1}\right)$.

Assumptions 6.3′ and 6.8 guarantee for the full-sample and partial-sample GEL that

$$\tilde{\Omega}_T = \frac{2K+1}{T} \sum_{t=1}^{T} g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T)' \xrightarrow{p} \Omega$$

and

$$\hat{\Omega}_T(s) = \frac{2K+1}{T} \sum_{t=1}^{T} g_{tT}(\hat{\theta}_T(s), s) g_{tT}(\hat{\theta}_T(s), s)' \xrightarrow{p} \Omega(s).$$

Now, let $G(\beta,\delta)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left[\partial g(x_t,\beta,\delta)/\partial\left(\beta',\delta'\right)\right]$ and $G=G(\beta_0,\delta_0)$.

Assumption 6.9. *$g(x,\beta,\delta)$ is differentiable in $(\beta,\delta)\,,\forall\,(\beta,\delta)\in B_0\times\Delta_0\ \forall x\in\mathbf{X_0}\subset\mathbf{X}$ for a Borel measurable set $\mathbf{X_0}$ that satisfies $P\left(x_t\in\mathbf{X_0}\right)=1\forall t\leq T,T\geq 1$, $g(x,\beta,\delta)$ is Borel measurable in $x\ \forall\,(\beta,\delta)\in B_0\times\Delta_0$, $\partial g(x_t,\beta,\delta)/\partial\left(\beta',\delta'\right)$ is continuous in (x,β,δ) on $\mathbf{X}\times B_0\times\Delta_0$,*

$$\sup_{1\leq t\leq T}E\left[\sup_{(\beta,\delta)\in B_0\times\Delta_0}\|\partial g(x_t,\beta,\delta)/\partial\left(\beta',\delta'\right)\|^{d/(d-1)}\right]<\infty$$

and $rank(G)=r+\nu$.

Assumption 6.10. *$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{Ts}Eg_{tT}(\beta,\delta)$ exists uniformly over $(\beta,\delta,s)\in B\times\Delta\times S$ and equals $s\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}Eg(x_t,\beta,\delta)=s\tilde{g}(\beta,\delta)$.*

Assumption 6.11. *$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{Ts}E\partial g_{tT}(\beta_0,\delta_0)/\partial\left(\beta',\delta'\right)$ exists uniformly over $s\in S$ and equals $sG\ \forall s\in S$ and S whose closure lies in $(0,1)$.*

Assumption 6.12. *$G(s)'\Omega(s)^{-1}G(s)$ is nonsingular $\forall s\in S$ and has eigenvalues bounded away from zero $\forall s\in S$ and S whose closure lies in $(0,1)$.*

Assumptions 6.10 and 6.11 are asymptotic covariance stationary conditions and follow directly from $Eg_{tT}(\beta,\delta)=Eg_t(\beta,\delta)+o_p(1)$ and $E\partial g_{tT}(\beta_0,\delta_0)/\partial\left(\beta',\delta'\right)=E\partial g_t(\beta_0,\delta_0)/\partial\left(\beta',\delta'\right)+o_p(1)$ for the uniform kernel. Assumption 6.12 guarantees that the partial-sample GEL estimators $\hat{\theta}_T(s)$ has a well defined asymptotic variance $\forall s\in S$ and holds if G^β and G^δ are full rank.

6.2 Lemmas

Lemma 6.1. *Under Assumptions 6.1 to 6.4, the asymptotic distribution of the smoothed moment conditions under the null is given by:*

$$\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0)\Rightarrow B(s),$$

where $B(s)$ is a q-dimensional vector of standard Brownian motion.

Proof of Lemma 6.1

First, under Assumptions 6.1, 6.3 and 6.4 , Lemma A4 in Andrews (1993) implies:

$$\Omega^{-1/2}\sum_{t=1}^{[Ts]}g_t\left(\beta_0,\delta_0\right)\Rightarrow B(s)$$

where $B(s)$ is a q-vector of standard Brownian motion.

Second, the smoothed moment condition are defined as:

$$\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\frac{1}{2K_T+1}\sum_{j=-K_T}^{K_T}g_{t-j}\left(\beta_0,\delta_0\right).$$

Considering the "endpoint effect" introduced by the extra K_T terms, we have:

$$\begin{aligned}\sum_{t=1}^{[Ts]}\sum_{j=-K_T}^{K_T}\frac{1}{2K_T+1}g_{t-j}\left(\beta_0,\delta_0\right) &= \sum_{t=1}^{[Ts]}\frac{1}{2K_T+1}\sum_{j=\max\{t-[Ts],-K_T\}}^{\min\{t-1,K_T\}}g_{t-j}\left(\beta_0,\delta_0\right)\\ &= \sum_{t=K_T+1}^{[Ts]-K_T}g_t(\theta_0)+\sum_{t=1}^{K_T}\frac{t+K_T}{2K_T+1}g_t(\beta_0,\delta_0)+\sum_{t=[Ts]-K_T+1}^{[Ts]}\frac{[Ts]-t+K_T+1}{2K_T+1}g_t(\beta_0,\delta_0)\\ &= \sum_{t=1}^{[Ts]}g_t(\beta_0,\delta_0)+\sum_{t=1}^{K_T}\frac{t-K_T-1}{2K_T+1}g_t(\beta_0,\delta_0)+\sum_{t=[Ts]-K_T+1}^{[Ts]}\frac{[Ts]-t-K_T}{2K_T+1}g_t(\beta_0,\delta_0)\end{aligned}$$

which implies that

$$\sum_{t=1}^{[Ts]}g_t\left(\beta_0,\delta_0\right)=\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0)+O_p\left(\frac{K_T^2}{2K_T+1}\right).$$

Under the Assumptions that $\max_{1\le t\le T}\|g_t(\beta_0,\delta_0)\|=o_p(T^{1/2})$ and $K_T^2/T\to 0$, we get

$$\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_t\left(\beta_0,\delta_0\right)=\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0)+o_p(1)$$

which yields the asymptotic equivalence.

The following Lemma provides the asymptotic distribution of the smoothed moment condition under the general sequence of local alternatives appearing in (4).

Lemma 6.2. *Under the alternative (4) and let Assumptions 6.1, 6.2, 6.4 and replacing $g(x_t,\beta,\delta)$ by $g(x_t,\beta,\delta)-h(\eta,\tau,\frac{t}{T})/\sqrt{T}$ in Assumption 6.3, then*

$$\frac{1}{\sqrt{T}}\Omega^{-1/2}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0)\Rightarrow B(s)+\Omega^{-1/2}H(s)$$

where $H(s)=\int_0^s h(\eta,\tau,u)du$ and $B(s)$ is a q-dimensional vectors of standard Brownian motion.

Proof of Lemma 6.2

Under the alternative (4), by the Lemma A4 of Andrews (1993), the sample smoothed moments satisfy:

$$\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\frac{1}{2K_T+1}\sum_{j=-K_T}^{K_T}\left(g_{t-j}\left(\beta_0,\delta_0\right)-\frac{h\left((t-j)/T\right)}{\sqrt{T}}\right)\Rightarrow B(s),$$

where $h(t/T)\equiv h(\eta,\tau,\frac{t}{T})$ to reduce the notation. The left-hand side term can be rewritten as:

$$\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\sum_{j=-K_T}^{K_T}g_{t-j}\left(\beta_0,\delta_0\right)-\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\frac{1}{2K_T+1}\sum_{j=-K_T}^{K_T}\frac{h\left((t-j)/T\right)}{\sqrt{T}}.$$

Let us now examine the last term,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\frac{1}{2K_T+1}\sum_{j=-K_T}^{K_T}\frac{h\left((t-j)/T\right)}{\sqrt{T}}=\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\frac{1}{2K_T+1}\sum_{j=\max\{t-[Ts],-K_T\}}^{\min\{t-1,K_T\}}\frac{h\left((t-j)/T\right)}{\sqrt{T}}$$

which equals

$$\begin{aligned}&\frac{1}{T}\sum_{t=K_T+1}^{[Ts]-K_T}h\left(t/T\right)+\frac{1}{T}\sum_{t=1}^{K_T}\frac{t+K_T}{2K_T+1}h\left(t/T\right)+\frac{1}{T}\sum_{t=[Ts]-K_T+1}^{[Ts]}\frac{[Ts]-t+K_T+1}{2K_T+1}h\left(t/T\right)\\&=\frac{1}{T}\sum_{t=1}^{[Ts]}\frac{h\left(t/T\right)}{\sqrt{T}}+\frac{1}{T}\sum_{t=1}^{K_T}\frac{t-K_T-1}{2K_T+1}h\left(t/T\right)+\frac{1}{T}\sum_{t=[Ts]-K_T+1}^{[Ts]}\frac{[Ts]-t-K_T}{2K_T+1}h\left(t/T\right).\end{aligned}$$

The first term of the last equality converges to $\int_0^s h(\nu)d\nu$. Under the assumption that $\frac{K_T^2}{T}\to 0$, the last two terms converge to zero. The result follows.

Lemma 6.3. *Under the null and Assumptions 6.3 and 6.7,*

$$\sup_{s\in S}\sup_{\theta\in\Theta,\lambda(s)\in\widehat{\Lambda}_T(s),1\le t\le T}|\lambda(s)'g_{tT}\left(\theta,s\right)|\xrightarrow{p}0.$$

Also w.p.a.1 $\widehat{\Lambda}_T(s)\subseteq\widehat{\Lambda}_T\left(\theta,s\right)$ *where* $\widehat{\Lambda}_T\left(\theta,s\right)=\{\lambda(s)=(\lambda_1',\lambda_2')' \;:\; \lambda(s)'g_{tT}\left(\theta,s\right)\in\Phi,(t=1,\dots,T)\}$.

Proof of Lemma 6.3

We first show that the results hold for both subsambles for a given s. Let $\widehat{\Lambda}_{1T}(s)=\widehat{\Lambda}_T(s)$ for $t=1,\dots,[Ts]$) and $\widehat{\Lambda}_{2T}(s)=\widehat{\Lambda}_T(s)$ for $t=[Ts]+1,\dots,T$). So, we have

$$\begin{aligned}\sup_{\theta\in\Theta_T,\lambda\in\widehat{\Lambda}_T(s),1\le t\le T}|\lambda(s)'g_{tT}\left(\theta,s\right)| \;&\le\; \sup_{\beta\in B,\delta\in\Delta,\lambda_1\in\widehat{\Lambda}_{1T}(s),1\le t\le[Ts]}|\lambda_1'g_{tT}\left(\beta,\delta\right)|\\&+\sup_{\beta\in B,\delta\in\Delta,\lambda_2\in\widehat{\Lambda}_{2T}(s),[Ts]+1\le t\le T}|\lambda_2'g_{tT}\left(\beta,\delta\right)|.\end{aligned}$$

Consider the first subsample, by the Cauchy-Schwarz inequality and Assumption 6.7 (b)

$$\sup_{\beta\in B,\lambda_1\in\widehat{\Lambda}_{1T}(s),1\le t\le[Ts]}|\lambda_1'g_{tT}\left(\beta,\delta\right)| \;\le\; D\left(T/(2K_T+1)^2\right)^{-\zeta}\sup_{\beta\in B,\delta\in\Delta,1\le t\le[Ts]}\|g_{tT}\left(\beta,\delta\right)\|$$

For the last term on the RHS, we get

$$\begin{aligned}\sup_{\beta\in B,\delta\in\Delta,1\le t\le[Ts]}\|g_{tT}\left(\beta,\delta\right)\| \;&\le\; \frac{1}{2K+1}\sup_{\beta\in B,\delta\in\Delta,1\le t\le[Ts]}\left\|\sum_{m=\max\{t-[Ts],-K_T\}}^{\min\{t-1,K_T\}}g_{t-m}\left(\beta,\delta\right)\right\|\\&\le\; \sup_{\beta\in B,\delta\in\Delta,1\le t\le[Ts]}\|g_t\left(\beta,\delta\right)\|\end{aligned}$$

uniformly in t. Using Assumption 6.3 and by Markov's inequality:

$$\sup_{\beta\in B,\delta\in\Delta,1\leq t\leq T} \|g_t(\beta,\delta)\| = O_p\left(T^{1/d}\right).$$

Hence

$$\sup_{\beta\in B,\delta\in\Delta,\lambda_1\in\widehat{\Lambda}_{1T}(s),1\leq t\leq[Ts]} |\lambda_1' g_{tT}(\beta,\delta)| \leq D\left(T/(2K_T+1)^2\right)^{-\zeta} O_p\left(T^{1/d}\right) \xrightarrow{p} 0$$

by Assumption 6.7 (b). This also holds for the second subsample.

Therefore under the null

$$\sup_{\beta\in B,\delta\in\Delta,\lambda_1\in\widehat{\Lambda}_{1T}(s),1\leq t\leq[Ts]} |\lambda_1' g_{tT}(\beta,\delta)| \xrightarrow{p} 0$$

and $\lambda_1' g_{tT} \in \Phi$ for $t = 1,\ldots,[Ts]$ w.p.a.1 for all $\beta \in B$, $\delta \in \Delta$ which implies that $\lambda_1 \in \widehat{\Lambda}_{1T}(\beta,\delta,s)$. For the second subsample,

$$\sup_{\beta\in B,\delta\in\Delta,\lambda_2\in\widehat{\Lambda}_{2T}(s),[Ts]+1\leq t\leq T} |\lambda_2' g_{tT}(\beta,\delta)| \xrightarrow{p} 0$$

and $\lambda_2' g_{tT} \in \Phi$ for $t = [Ts]+1,\ldots,T$ w.p.a.1 for all $\beta \in B$, $\delta \in \Delta$ which implies that $\lambda_2 \in \widehat{\Lambda}_{2T}(\beta,\delta,s)$. Finally, these results holds uniformly $\forall s \in S$.

Lemma 6.4. *Under Assumptions 6.1, 6.2, 6.3, 6.5 and 6.10*

$$\sup_{s\in S}\sup_{\theta\in\Theta} \|g_T(\theta,s) - g(\theta,s)\| \xrightarrow{p} 0$$

where $g(\theta,s) = (s\tilde{g}(\beta,\delta)', (1-s)\tilde{g}(\beta,\delta)')'$.

Proof of Lemma 6.4

Using $\sum_{[Ts]+1}^{T} = \sum_1^T - \sum_1^{[Ts]}$, the result of the Lemma holds if

$$\sup_{\beta\in B,\delta\in\Delta}\ \sup_{T\underline{s}\leq R\leq T} \left|\frac{1}{T}\sum_{t=1}^{R}[g_{tT}(\beta,\delta) - g(\beta,\delta)]\right| \xrightarrow{p} 0$$

where $\underline{s} = inf\{s\ :\ s\in S\}$.

By the triangular inequality

$$\begin{aligned}\sup_{\theta\in\Theta}\ \sup_{T\underline{s}\leq R\leq T}\left\|\frac{1}{T}\sum_{t=1}^{R}[g_{tT}(x_t,\beta,\delta) - g(\beta,\delta)]\right\| \leq{}& \sup_{\theta\in\Theta}\ \sup_{T\underline{s}\leq R\leq T}\left\|\frac{1}{T}\sum_{t=1}^{R}[g_{tT}(\beta,\delta) - Eg_{tT}(\beta,\delta)]\right\| \\ &+ \sup_{\theta\in\Theta}\ \sup_{T\underline{s}\leq R\leq T}\left\|\frac{1}{T}\sum_{t=1}^{R}[Eg_{tT}(\beta,\delta) - g(\beta,\delta)]\right\|.\end{aligned}$$

We show that both terms on the right-hand side converge in probability to zero. For the first term, we first show that $\frac{1}{T}\sum_{t=1}^{R}\left[g_{tT}(\beta,\delta)-Eg_{tT}(\beta,\delta)\right]=\frac{1}{T}\sum_{t=1}^{R}\left[g_t(\beta,\delta)-Eg_t(\beta,\delta)\right]+o_p(1)$. By the proof similar to the one in Lemma 6.1, we can show that:

$$\frac{1}{T}\sum_{t=1}^{R}g_t(\beta,\delta)=\frac{1}{T}\sum_{t=1}^{R}g_{tT}(\beta,\delta)+o_p(1).$$

This also holds for the partial sum of the expectation, the result follows. Now using Lemma A3 in Andrews with Assumptions 6.1 and 6.7 guarantees the UWL for $\sup_{R\leq T}\left\|\frac{1}{T}\sum_{t=1}^{R}\left[g_t(\beta,\delta)-Eg_t(\beta,\delta)\right]\right\|$. This yields

$$\sup_{\theta\in\Theta}\sup_{T\underline{S}\leq R\leq T}\left\|\frac{1}{T}\sum_{t=1}^{R}\left[g_t(\beta,\delta)-Eg_t(\beta,\delta)\right]\right\|\xrightarrow{p}0$$

which directly implies

$$\sup_{\theta\in\Theta}\sup_{T\underline{S}\leq R\leq T}\left\|\frac{1}{T}\sum_{t=1}^{R}\left[g_{tT}(\beta,\delta)-Eg_{tT}(\beta,\delta)\right]\right\|\xrightarrow{p}0.$$

For the second term, by a similar argument $\frac{1}{T}\sum_{t=1}^{R}Eg_{tT}(\beta,\delta)=\frac{1}{T}\sum_{t=1}^{R}Eg_t(\beta,\delta)+o_p(1)$ and the convergence in probability to zero holds by Assumption 6.10.

Now define

$$\begin{aligned}\widehat{P}\left(\theta(s),\lambda(s),s\right) &= \sum_{t=1}^{T}\frac{\left[\rho(k\lambda(s)'g_{tT}(\theta,s))-\rho_0\right]}{T}\\ &= \sum_{t=1}^{[Ts]}\frac{\left[\rho(k\lambda_1'g_{tT}(\beta_1,\delta))-\rho_0\right]}{T}+\sum_{t=[Ts]+1}^{T}\frac{\left[\rho(k\lambda_2'g_{tT}(\beta_2,\delta))-\rho_0\right]}{T}\end{aligned}$$

and $\hat{g}_T(\theta_0,s)=\frac{1}{T}\sum_{t=1}^{T}g_{tT}(\theta_0,s)$.

Lemma 6.5. *Under Assumptions 6.3, 6.7 and 6.8, there is a constant C such that w.p.a.1.*

$$\frac{1}{2K_T+1}\sup_{s\in S}\sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta_0,s)}\widehat{P}\left(\theta_0,\lambda(s),s\right)=\sup_{s\in S}C\|\hat{g}_T(\theta_0,s)\|^2.$$

Proof of Lemma 6.5

By a proof similar to the one of Lemma A.5 in Smith (2004)[8], we can show that

$$\frac{1}{2K_T+1}\sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta_0,s)}\widehat{P}\left(\theta_0,\lambda(s),s\right)=C\|\hat{g}_T(\theta_0,s)\|^2$$

for a given $s\in S$ w.p.a.1. Since this holds for all $s\in S$, this holds for s which achieves the supremum.

[8]In his proof, Smith (2004) uses the fact that $(2K_T+1)\sum_{t=1}^{T}\rho_2\left(\dot{\lambda}'g_{tT}(\beta_0,\delta_0)\right)g_{tT}(\beta_0,\delta_0)g_{tT}(\beta_0,\delta_0)'/T\xrightarrow{p}-\Omega$ in our notation. This needs more restrictive assumptions than those imposed here. In fact, we only need that $(2K_T+1)\sum_{t=1}^{T}\rho_2\left(\dot{\lambda}'g_{tT}(\beta_0,\delta_0)\right)g_{tT}(\beta_0,\delta_0)g_{tT}(\beta_0,\delta_0)'/T\leq -CI_q$ in the p.s.d. sense w.p.a.1 which holds by the fact that the outer product of smoothed moment conditions is automatically positive semi-definite.

6.3 Proofs of Theorems

Proof of Theorem 2.1

The outline of the proof is similar to that of Lemma A.6 and Theorem 2.2 in Smith (2004) except that the results have to be established uniformly in $s \in S$ and by taking into account of the differences in Assumptions 6.2, 6.3 and 6.7 with respect to the corresponding assumptions in Smith (2004).

First, we show that $\sup_{s\in S} \|\hat{g}_T(\hat{\theta}_T(s), s)\|^2 = O_p(T^{-1})$ which allows us to show that $\sup_{s\in S} \|\hat{\theta}(s) - \theta_0\| \xrightarrow{p} 0$.

By arguments similar to Smith (2004), we can show that $\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T(s), s) g_{tT}(\hat{\theta}_T(s), s)'/T = O_p(1)$. Following Newey and Smith (2001) and Smith (2004), let

$$\bar{\lambda}_T(s) = \left(\bar{\lambda}_{1T}', \bar{\lambda}_{2T}'\right)' = -\frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T(s), s)\kappa_T / \|g_T(\hat{\theta}_T(s), s)\|$$

with $\kappa_T = D\left(T/(2K_T+1)^2\right)^{-\zeta}$ and

$$\begin{aligned}
\bar{\lambda}_{1T} &= -\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_{1T}, \delta_T)\kappa_T / \|g_T(\hat{\theta}_T(s), s)\|, \\
\bar{\lambda}_{2T} &= -\frac{1}{T}\sum_{t=[Ts]+1}^{T} g_{tT}(\hat{\beta}_{2T}, \delta_T)\kappa_T / \|g_T(\hat{\theta}_T(s), s)\|.
\end{aligned}$$

By Lemma 6.3, $\sup_{s\in S} \max_{1\le t\le T} |\bar{\lambda}(s)' g_{tT}\left(\hat{\theta}_T(s), s\right)| \xrightarrow{p} 0$ and $\bar{\lambda}_T(s) \in \Lambda_T(\hat{\theta}_T(s), s)$ w.p.a.1. Thus, for a given s, $\dot{\lambda}_T(s) = \left(\dot{\lambda}_{1T}', \dot{\lambda}_{2T}'\right)'$ with $\dot{\lambda}_{1T} = \tau_1\bar{\lambda}_{1T}, 0 \le \tau_1 \le 1$ and $\dot{\lambda}_{2T} = \tau_2\bar{\lambda}_{2T}, 0 \le \tau_2 \le 1$,

$$\sup_{s\in S} \sum_{t=1}^{T} \left[\rho_2(\dot{\lambda}_T(s)' g_{tT}(\hat{\theta}_T(s), s)) - \rho_2(0)\right] g_{tT}(\hat{\theta}_T(s), s) g_{tT}(\hat{\theta}_T(s), s)'/T \xrightarrow{p} 0$$

and therefore $\sup_{s\in S}(2K_T+1)\sum_{t=1}^{T} \rho_2(\dot{\lambda}_T(s)' g_{tT}(\hat{\theta}_T(s), s)) g_{tT}(\hat{\theta}_T(s), s) g_{tT}(\hat{\theta}_T(s), s)'/T \ge -CI_{2q}$ in the p.s.d. sense w.p.a.1. Hence, by a second-order Taylor expansion

$$\begin{aligned}
\frac{1}{2K_T+1}\widehat{P}(\hat{\theta}_T(s), \bar{\lambda}_T(s), s) &= -\left(\frac{\bar{\lambda}_T(s)}{2K_T+1}\right)' \hat{g}_T(\hat{\theta}_T(s), s) \\
&+ \left(\frac{\bar{\lambda}_T(s)}{2K_T+1}\right)' \left(\sum_{t=1}^{T} \rho_2(\dot{\lambda}_T(s)' g_{tT}(\hat{\theta}_T(s), s)\hat{g}_{tT}(\hat{\theta}_T(s), s) g_{tT}(\hat{\theta}_T(s), s)'/T\right) \bar{\lambda}_T(s)/2 \\
&\ge -\left(\frac{\bar{\lambda}_T(s)}{2K_T+1}\right)' \hat{g}_T(\hat{\theta}_T(s), s) - C\left(\frac{\bar{\lambda}_T(s)}{2K_T+1}\right)'\left(\frac{\bar{\lambda}_T(s)}{2K_T+1}\right) \\
&= \|\hat{g}_T(\hat{\theta}_T(s), s)\|\left(\frac{\kappa_T}{2K_T+1}\right) - C\left(\frac{\kappa_T}{2K_T+1}\right)^2
\end{aligned}$$

w.p.a.1 and this holds $\forall s \in S$. Now using Lemma 6.5, we get w.p.a.1

$$\begin{aligned}
\sup_{s\in S}\|\hat{g}_T(\hat{\theta}_T(s),s)\|\left(\frac{\kappa_T}{2K_T+1}\right)-C\left(\frac{\kappa_T}{2K_T+1}\right)^2 &\leq \sup_{s\in S}\frac{1}{2K_T+1}\widehat{P}\left(\hat{\theta}_T(s),\bar{\lambda}_T(s),s\right)\\
&\leq \sup_{s\in S}\sup_{\lambda(s)\in\widehat{\Lambda}_T(\hat{\theta}_T(s),s)}\frac{1}{2K_T+1}\widehat{P}\left(\hat{\theta}_T(s),\lambda(s),s\right)\\
&\leq \sup_{s\in S}\sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta_0,s)}\frac{1}{2K_T+1}\widehat{P}\left(\theta_0,\lambda(s),s\right)\\
&\leq \sup_{s\in S}C\|\hat{g}_T(\theta_0,s)\|^2=O_p\left(T^{-1}\right)
\end{aligned}$$

as $\|\hat{g}_T(\theta_0,s)\|=O_p(T^{-1/2})$ by CLT (Corollary 3.1 of Wooldridge and White (1988)). This yields

$$\sup_{s\in S}\|\hat{g}_T(\hat{\theta}_T(s),s)\|\leq C\left(\frac{\kappa_T}{2K_T+1}\right)+\sup_{s\in S}C\|\hat{g}_T(\theta_0,s)\|^2\left(\frac{\kappa_T}{2K_T+1}\right)=O_p\left(\kappa_T/(2K_T+1)\right),$$

which implies $\sup_{s\in S}\|\hat{g}_T(\hat{\theta}_T(s),s)\|=O_p(T^{-1/2})$ by Assumption 6.2 for all $\eta>1$. By the result that $\sup_{s\in S}\|g_T(\hat{\theta}_T(s),s\|=O_p(T^{-1/2})$ we have that $\sup_{s\in S}g_T(\hat{\theta}_T(s),s)\xrightarrow{p}0$. By Lemma 6.4, $\sup_{s\in S}\sup_{\theta\in\Theta}\|g_T(\theta,s)-g(\theta,s)\|\xrightarrow{p}0$ and $g(\beta,\delta)$ is continuous by Assumption 6.5. The triangular inequality then gives that $\sup_{s\in S}g(\hat{\theta}_T(s),s)\xrightarrow{p}0$. Since $g(\beta,\delta)=0$ has a unique zero at β_0 and δ_0 (by Assumption 6.6), for every neighborhood $\Theta_0(\in\Theta)$ of θ_0, $\inf_{s\in S}\left(\inf_{\theta\in\Theta/\Theta_0}\|g(\theta,s)\|\right)>0$, then $\sup_{s\in S}\|\hat{\theta}_T(s)-\theta_0\|\xrightarrow{p}0$.

Now we need to show $\sup_{s\in S}\|\hat{\lambda}_T(s)\|=O_p\left(\left(T/(2K+1)^2\right)^{-1/2}\right)$and $\sup_{s\in S}\|\hat{\lambda}_T(s)\|\xrightarrow{p}0$. By a second-order Taylor expansion around $\lambda(s)=0$, for a given $s\in S$ and for any $\dot{\lambda}_T(s)=\left(\dot{\lambda}_{1T}',\dot{\lambda}_{2T}'\right)'$ with $\dot{\lambda}_{1T}=\tau_1\hat{\lambda}_{1T},0\leq\tau_1\leq 1$ and $\dot{\lambda}_{2T}=\tau_2\hat{\lambda}_{2T},0\leq\tau_2\leq 1$

$$\begin{aligned}
(2K_T+1)\widehat{P}\left(\hat{\theta}_T(s),0,s\right) &\leq \sup_{\lambda(s)\in\widehat{\Lambda}_T(\hat{\theta}_T(s),s)}(2K_T+1)\widehat{P}\left(\hat{\theta}_T(s),\lambda(s),s\right)\\
&= (2K_T+1)\widehat{P}(\hat{\theta}_T(s),\hat{\lambda}_T(s),s)\\
&\leq -(2K_T+1)\hat{\lambda}_T(s)'\hat{g}_T(\hat{\theta}_T(s),s)\\
&+ \hat{\lambda}_T(s)'\left((2K_T+1)\sum_{t=1}^{T}\rho_2(\dot{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s))\hat{g}_{tT}(\hat{\theta}_T(s),s)g_{tT}(\hat{\theta}_T(s),s)'/T\right)\hat{\lambda}_T(s)/2\\
&\leq -(2K_T+1)\hat{\lambda}_T(s)'\hat{g}_T(\hat{\theta}_T(s),s)-C\hat{\lambda}_T(s)'\hat{\lambda}_T(s)\\
&\leq (2K_T+1)\|\hat{\lambda}_T(s)\|\|\hat{g}_T(\hat{\theta}_T(s),s)\|-C\|\hat{\lambda}_T(s)\|^2
\end{aligned}$$

w.p.a.1. Since $\widehat{P}\left(\hat{\theta}_T(s),0,s\right)=0,\forall s\in S$, this implies directly that $C\|\hat{\lambda}_T(s)\|\leq(2K_T+1)\|\hat{g}_T(\hat{\theta}_T(s),s)\|$ and this holds for all $s\in S$ which implies that $\sup_{s\in S}C\|\hat{\lambda}_T(s)\|\leq\sup_{s\in S}(2K_T+1)\|\hat{g}_T(\hat{\theta}_T(s),s)\|$. Finally, considering that $\sup_{s\in S}\|\hat{g}_T(\hat{\theta}_T(s),s)\|=O_p(T^{-1/2})$ directly yields that $\sup_{s\in S}\|\hat{\lambda}_T(s)\|=O_p\left[\left(T/(2K_T+1)^2\right)^{-1/2}\right]$ and $\sup_{s\in S}\|\hat{\lambda}_T(s)\|\xrightarrow{p}0$ by Assumption 6.2.

Proof of Theorem 2.2

The first-order conditions of the partial-sample GEL with respect to $\lambda(s)$ and $\theta(s)$ are:

$$\frac{1}{T}\sum_{t=1}^{T}\rho_1(\hat{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s))g_{tT}(\hat{\theta}_T(s),s)=0.$$

$$\frac{1}{T}\sum_{t=1}^{T}\rho_1(\hat{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s))G_{tT}(\hat{\theta}_T(s),s)'\hat{\lambda}_T(s)=0.$$

By a mean-value expansion of the former first-order conditions for the partial-sample GEL where $\Xi_T=\left(\hat{\beta}_{1T}(s)',\hat{\beta}_{2T}(s)',\hat{\delta}_T(s)',\frac{\hat{\lambda}_{1T}(s)'}{2K_T+1},\frac{\hat{\lambda}_{2T}(s)'}{2K_T+1}\right)'$ and $\Xi_0=(\beta_0',\beta_0',\delta_0',0,0)'$ with the latter first-order conditions yields:

$$0=-T^{1/2}\left(\begin{array}{c}0\\ \frac{1}{T}\sum_{t=1}^{T}g_{tT}(\theta_0,s)\end{array}\right)+\bar{M}(s)T^{1/2}\left(\hat{\Xi}_T(s)-\Xi_0\right)$$

where

$$\frac{1}{T}\sum_{t=1}^{T}g_{tT}(\theta_0,s)=\frac{1}{T}\sum_{t=1}^{[Ts]}\left[\begin{array}{c}g_{tT}(\beta_0,\delta_0)\\ 0\end{array}\right]+\frac{1}{T}\sum_{t=[Ts]+1}^{T}\left[\begin{array}{c}0\\ g_{tT}(\beta_0,\delta_0)\end{array}\right]$$

and

$$\bar{M}(s)=\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{cc}0 & \bar{M}_{12}(s)\\ \bar{M}_{21}(s) & \bar{M}_{22}(s)\end{array}\right]$$

with $\bar{M}_{12}(s)=\rho_1\left(\hat{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)G_{tT}(\hat{\theta}_T(s),s)'$, $\bar{M}_{21}(s)=\rho_1\left(\bar{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)G_{tT}(\bar{\theta}_T(s),s)'$ and $\bar{M}_{22}(s)=(2K_T+1)\,\rho_2\left(\bar{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)g_{tT}(\bar{\theta}_T(s),s)g_{tT}(\hat{\theta}_T(s),s)'$ and $\bar{\theta}_T(s)$ is a random vector on the line segment joining $\hat{\theta}_T(s)$ and θ_0 and $\bar{\lambda}_T(s)$ is a random vector joining $\hat{\lambda}_T(s)$ to $(0',0')'$ that may differ from row to row.

Now, we need to show that $\bar{M}(s)\xrightarrow{p}M(s)$ where

$$M(s)=-\left[\begin{array}{cc}0 & G(s)'\\ G(s) & \Omega(s)\end{array}\right].$$

By Lemma 6.3:

$$\sup_{s\in S}\sup_{1\le t\le T}|\hat{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)|\xrightarrow{p}0$$

and

$$\sup_{s\in S}\sup_{1\le t\le T}|\bar{\lambda}_T(s)'g_{tT}(\bar{\theta}_T(s),s)|\xrightarrow{p}0$$

which implies

$$\begin{aligned}\sup_{s\in S}\max_{1\le t\le T}|\rho_1\left(\hat{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)-\rho_1(0)| &\xrightarrow{p} 0\\ \sup_{s\in S}\max_{1\le t\le T}|\rho_1\left(\bar{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)-\rho_1(0)| &\xrightarrow{p} 0\end{aligned}$$

and

$$\sup_{s\in S}\max_{1\le t\le T}\left|\rho_2\left(\hat{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)-\rho_2(0)\right|\xrightarrow{p}0.$$

To show that

$$\sup_{s\in S}\frac{1}{T}\sum_{t=1}^{T}\rho_1\left(\hat{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)G_{tT}(\hat{\theta}_T(s),s)\xrightarrow{p}-G(s)$$

and

$$\sup_{s\in S}\frac{1}{T}\sum_{t=1}^{T}\rho_1\left(\bar{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)G_{tT}(\hat{\theta}_T(s),s)\xrightarrow{p}-G(s),$$

it remains to show that

$$\sup_{s\in S}\left\|\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)-G(s)\right\|\xrightarrow{p}0.$$

By the triangular inequality

$$\begin{aligned}\sup_{s\in S}\left\|\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)-G(s)\right\| \le\ & \sup_{s\in S}\left\|\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)-E\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)\right\|\\ &+\sup_{s\in S}\left\|E\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)-E\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\theta_0,s)\right\|\\ &+\sup_{s\in S}\left\|E\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\theta_0,s)-G(s)\right\|.\end{aligned}$$

The first term on the right-hand side $\xrightarrow{p}0$ by an application of UWL given by Lemma A3 in Andrews (1993) which implies

$$\sup_{s\in S}\sup_{\theta\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)-E\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)\right\|\xrightarrow{p}0.$$

The second term $\xrightarrow{p}0$ under the tightness of $\{\bar{\mu}_T;\ T\ge 1\}$ (Assumption 6.1), Assumption 6.9 and $(\beta,\delta)\xrightarrow{p}(\beta_0,\delta_0)$ (see equations A.13 and A.14 in Andrews 1993). Finally, the third term $\xrightarrow{p}0$ by Assumption 6.11 and by $\sum_{t=1}^{[Ts]}G_t(\theta_0,s)=\sum_{t=1}^{[Ts]}G_{tT}(\theta_0,s)+o_p(1)$.

Assumptions 6.8 implies that

$$\frac{2K_T+1}{T}\sum_{t=1}^{[Ts]}g_{tT}(\bar{\beta}_{1T},\bar{\delta}_T)g_{tT}(\hat{\beta}_{1T},\hat{\delta}_T)'\xrightarrow{p}s\Omega$$

and

$$\frac{2K_T+1}{T}\sum_{t=[Ts]+1}^{T}g_{tT}(\bar{\beta}_{2T},\bar{\delta}_T)g_{tT}(\hat{\beta}_{2T},\hat{\delta}_T)'\xrightarrow{p}(1-s)\Omega$$

which yields

$$\frac{2K_T+1}{T}\sum_{t=1}^{T}\rho_2\left(\bar{\lambda}_T(s)'g_{tT}(\hat{\theta}_T(s),s)\right)g_{tT}(\bar{\theta}_T(s),s)g_{tT}(\hat{\theta}_T(s),s)' \xrightarrow{p} -\Omega(s).$$

Moreover, under Assumption 6.12:

$$M(s)^{-1}=\left[\begin{array}{cc} -\Sigma(s) & H(s) \\ H(s)' & P(s) \end{array}\right]$$

where $\Sigma(s)=\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1}$, $H(s)=\Sigma(s)G(s)'\Omega(s)^{-1}$ and $P(s)=\Omega(s)^{-1}-\Omega(s)^{-1}G(s)\Sigma(s)G(s)'\Omega(s)^{-1}$.

As $\bar{M}(s)$ is positive definite w.p.a.1, we obtain:

$$\begin{aligned}\sqrt{T}\left(\Xi_T(s)-\Xi_0\right) &= -\bar{M}^{-1}(s)\left(0,-\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}g_{tT}(\theta_0,s)'\right)+o_p(1)\\ &= -\left(H(s)',P(s)\right)'\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}g_{tT}(\theta_0,s)+o_p(1).\end{aligned}$$

We also have by Lemma 6.1,

$$\Omega^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}(\theta_0,s)\Rightarrow J(s)$$

for $s\in S$. Combining the results above yields:

$$\begin{aligned}\sqrt{T}\left(\hat{\theta}_T(s)-\theta_0\right) &= -\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1}G(s)'\Omega(s)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}(\theta_0,s)+o_p(1)\\ &\Rightarrow -\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1}G(s)'\Omega(s)^{-1}J(s)\end{aligned}$$

and

$$\begin{aligned}\frac{\sqrt{T}}{2K_T+1}\hat{\lambda}_T(s) &= -\left(\Omega^{-1}(s)-\Omega^{-1}(s)G(s)\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1}G(s)'\Omega(s)^{-1}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}(\theta_0,s)+o_p(1)\\ &\Rightarrow -\left(\Omega^{-1}(s)-\Omega^{-1}(s)G(s)\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1}G(s)'\Omega(s)^{-1}\right)J(s).\end{aligned}$$

Proof of Theorem 2.3

This is a direct implication of Lemma 6.2 and the proof of Theorem 2.2.

Proof of Theorem 3.1

The results for the $Wald_T(s)$ and $LM_T(s)$ under the null can be directly derived from similar arguments to those used in the proof of Theorem 3 in Andrews (1993). For the Wald statistic:

$$\begin{aligned}
\widehat{G}^{\beta}_{1,tT}(s) &= \frac{1}{[Ts]}\sum_{t=1}^{[Ts]}\frac{\partial g_t(\hat{\beta}_{1T}(s),\hat{\delta}_T(s))}{\partial\beta_1'}+o_p(1),\\
\widehat{G}^{\beta}_{2,tT}(s) &= \frac{1}{T-[Ts]}\sum_{t=[Ts]+1}^{T}\frac{\partial g_t(\hat{\beta}_{2T}(s),\hat{\delta}_T(s))}{\partial\beta_2'}+o_p(1),\\
\widehat{\Omega}_{1T}(s) &\xrightarrow{p} \Omega_1(s),\;\; \widehat{\Omega}_{2T}(s)\xrightarrow{p}\Omega_2(s)
\end{aligned}$$

and for the LM statistic:

$$\begin{aligned}
\hat{g}_{1T}(\tilde{\theta}_T,s) &= \frac{1}{T}\sum_{t=1}^{[Ts]}g_t(\tilde{\beta}_T,\tilde{\delta}_T)+o_p(1),\\
\widetilde{G}^{\beta}_{tT} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_t(\tilde{\beta}_T,\tilde{\delta}_T)}{\partial\beta'}+o_p(1),\\
\widetilde{\Omega}_T &\xrightarrow{p} \Omega.
\end{aligned}$$

The asymptotic distribution under the alternative is a direct implication of Theorem 2.3. For the $LR_T(s)$ statistic, expanding the partial-sample GEL objective function evaluated at the unrestricted estimator about $\lambda=0$ yields,

$$\begin{aligned}
\frac{2T}{(2K_T+1)}\frac{1}{T}\sum_{t=1}^{T}\rho(\hat{\lambda}_T(\hat{\theta}_T(s),s)'g_{tT}(\hat{\theta}_T(s),s) &= -\frac{2T}{(2K_T+1)}\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_T(\hat{\theta}_T(s),s)'g_{tT}(\hat{\theta}_T(s),s)-\\
&\quad \frac{T}{(2K_T+1)}\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_T(\hat{\theta}_T(s),s)'g_{tT}(\hat{\theta}_T(s),s)g_{tT}(\hat{\theta}_T(s),s)'\hat{\lambda}_T(\hat{\theta}_T(s),s)\\
&+ o_p(1)
\end{aligned}$$

since $\rho_1(\cdot)\xrightarrow{p}-1$ and $\rho_2(\cdot)\xrightarrow{p}-1$.

By the fact that $\hat{\Omega}_T(s)=\frac{2K_T+1}{T}\sum_{t=1}^{T}g_{tT}(\hat{\theta}_T(s),s)g_{tT}(\hat{\theta}_T(s),s)'$ is a consistent estimator of $\Omega(s)$ and by $\sqrt{T}/(2K_T+1)\hat{\lambda}_T(s)=-\Omega(s)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}(\hat{\theta}_T(s),s))+o_p(1)$, we get

$$\frac{2T}{(2K_T+1)}\frac{1}{T}\sum_{t=1}^{T}\rho(\hat{\lambda}_T(\hat{\theta}_T(s),s)'g_{tT}(\hat{\theta}_T(s),s))=Tg_T(\hat{\theta}_T(s),s)'\Omega(s)^{-1}g_T(\hat{\theta}_T(s),s)+o_p(1).$$

Similarly, the expansion of the partial-sample GEL objective function but evaluated at the restricted estimator yields:

$$\frac{2T}{(2K_T+1)}\frac{1}{T}\sum_{t=1}^{T}\rho(\hat{\lambda}_T(\tilde{\theta}_T,s)'g_{tT}(\tilde{\theta}_T,s))=Tg_T(\tilde{\theta}_T,s)'\Omega(s)^{-1}g_T(\tilde{\theta}_T,s)+o_p(1)$$

since that $\tilde{\Omega}_T(s) = \frac{2K_T+1}{T}\sum_{t=1}^{T} g_{tT}(\tilde{\theta}_T, s) g_{tT}(\tilde{\theta}_T, s)'$ is a consistent estimator of $\Omega(s)$ under the null. The $LR_T(s)$ is then asymptotically equivalent to the LR statistic defined in Andrews (1993) for the standard GMM.

Proof of Theorem 3.3

First, for the statistic $O_T(s)$, the asymptotic equivalence between $\sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_{1T}(s))$ with $\sum_{t=1}^{[Ts]} g(\hat{\beta}_{1T}(s))$ and $\sum_{t=[Ts]+1}^{T} g_{tT}(\hat{\beta}_{1T}(s))$ with $\sum_{t=[Ts]+1}^{T} g(x_t, \hat{\beta}_{1T}(s))$ is direct implication of the Lemmas 6.1 and 6.2 and by the asymptotic consistency of the estimator $\hat{\Omega}_{1T}(s)$ and $\hat{\Omega}_{2T}(s)$ for Ω, the result under the null and alternative follows directly from the proofs for Theorems 2.2 and 2.3 and subsection A.2 in Hall and Sen (1999).

Second, for the statistic $O_T(s)^{GEL}$, as in the proof of Theorem 3.2, we can show that:

$$\frac{2[Ts]}{2K_T+1}\sum_{t=1}^{[Ts]} \frac{\left[\rho(\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s)' g_{tT}(\hat{\beta}_{1T}(s))) - \rho_0\right]}{[Ts]} = O1_T(s) + o_p(1).$$

and

$$\frac{2(T-[Ts])}{2K_T+1}\sum_{t=[Ts]+1}^{T} \frac{\left[\rho(\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s)' g_{tT}(\hat{\beta}_{2T}(s))) - \rho_0\right]}{T-[Ts]} = O2_T(s) + o_p(1)$$

The asymptotic distribution under the null and the alternative follows directly.

Finally, for the statistic $LM_T^O(s)$, the have the following asymptotic equivalences:

$$\begin{aligned}
\frac{\sqrt{[Ts]}}{(2K_T+1)}\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s) &= -\Omega(s)^{-1}([Ts])^{-1/2}\sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_{1T}(s)) + o_p(1) \\
\frac{\sqrt{T-[Ts]}}{(2K_T+1)}\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s) &= -\Omega(s)^{-1}(T-[Ts])^{1/2}\sum_{t=[Ts]+1}^{T} g_{tT}(\hat{\beta}_{2T}(s)) + o_p(1)
\end{aligned}$$

which implies directly the asymptotic distribution of this statistic under the null and the alternative.

Proof of Theorem 3.4

Since $\tilde{\theta}_T(s)$ minimizes the restricted partial sample GEL for all $s \in S$, this implies for all $s \in S$ and all T,

$$\widehat{P}(\tilde{\theta}_T(s), \hat{\lambda}_T(\tilde{\theta}_T(s), s) \leq \widehat{P}(\theta_0, \hat{\lambda}_T(\tilde{\theta}_T(s), s), s).$$

The limit for $\widehat{P}(\tilde{\theta}_T(s), \hat{\lambda}_T(\tilde{\theta}_T(s), s), s)$ is then bounded by the limit of $\widehat{P}(\theta_0, \hat{\lambda}_T(\tilde{\theta}_T(s), s), s)$. Let $\hat{\lambda}_T(\theta_0, s) = \arg\max_{\lambda_s \in \hat{\Lambda}_T(\theta_0, s)} \widehat{P}(\theta_0, \lambda(s), s)$ and $\dot{\lambda}_T(s) = \tau\hat{\lambda}_T(s)$, $0 \leq \tau \leq 1$. Thus, $\widehat{P}(\tilde{\theta}_T(s), \hat{\lambda}_T(\tilde{\theta}_T(s), s), s) \leq \widehat{P}(\theta_0, \hat{\lambda}_T(\tilde{\theta}_T(s), s), s) \leq \widehat{P}(\theta_0, \hat{\lambda}_T(\theta_0, s), s)$. By a second-order Taylor expansion with Lagrange remainder and using $(2K_T+1)\sum_{t=1}^{T}\rho_2(\dot{\lambda}(s)'g_{tT}(\theta_0,s))g_{tT}(\theta_0,s)g_{tT}\theta_0,s)'/T \xrightarrow{p} -\Omega(s)$,

$$\begin{aligned}\frac{1}{2K_T+1}\widehat{P}(\theta_0,\hat{\lambda}_T(\theta_0,s),s) &= -\left(\frac{\hat{\lambda}_T(\theta_0,s)}{2K_T+1}\right)'\hat{g}_T(\theta_0,s)\\ &+ \left(\frac{\hat{\lambda}_T(\theta_0,s)}{2K_T+1}\right)'\left(\sum_{t=1}^{T}\rho_2\left(\dot{\lambda}_T(s)'g_{tT}(\theta_0,s)\right)g_{tT}(\theta_0,s)g_{tT}(\theta_0,s)'/T\right)\hat{\lambda}_T(\theta_0,s)/2\\ &= \hat{g}_T(\theta_0,s)'\Omega(s)^{-1}\hat{g}_T((\theta_0,s)-\hat{g}_T(\theta_0,s)'\Omega(s)^{-1}\hat{g}_T(\theta_0,s)/2+o_p(1)\\ &= \hat{g}_T(\theta_0,s)'\Omega(s)^{-1}\hat{g}_T(\theta_0,s)/2+o_p(1)\end{aligned}$$

w.p.a.1 where the second equality holds by $\frac{1}{2K_T+1}\hat{\lambda}_T(\theta_0,s) = -\Omega(s)^{-1}\hat{g}_T(\theta_0,s)+o_p(1)$. The asymptotic distribution of the statistic $\frac{2T}{2K_T+1}\widehat{P}(\tilde{\theta}_T(s),\hat{\lambda}_T(\tilde{\theta}_T(s),s),s)$ is then asymptotically bounded for all $s\in S$ by the asymptotic distribution of $T\hat{g}_T(\theta_0,s)'\Omega(s)^{-1}\hat{g}_T(\theta_0,s)$. By using Lemma 6.1, the result under the null follows. Lemma 6.2 yields the asymptotic distribution under the alternative. The equivalence for the statistic $LM_T^R(\tilde{\theta}_T(s),s)$ is straightforward to show.

Proof of Theorem 3.5

To prove this theorem, additional assumptions are needed. Let

$$\Sigma(\beta_0) = \lim_{T\to\infty} var\left(\frac{1}{T}\sum_{t=1}^{T}(g_t(\beta_0)', vec(G_t(\beta_0))')\right)'$$

a $(q+qr)\times(q+qr)$ positive semi-definite symmetric matrix and

$$\Sigma(\beta_0) = \begin{bmatrix}\Omega(\beta_0) & \Omega_{gG}(\beta_0)\\ \Omega_{Gg}(\beta_0) & \Omega_{GG}(\beta_0)\end{bmatrix}$$

where $\Omega_{gG}(\beta_0)=\Omega_{Gg}(\beta_0)'$ is a $(q\times qr)$ matrix and $\Omega_{GG}(\beta_0)$ is a $(qr\times qr)$ matrix.

We define the estimators under the null of no structural change

$$\hat{\Sigma}_{1T}(\beta_0,s) = \frac{2K_T+1}{[Ts]}\sum_{t=1}^{[Ts]}(g_t(\beta_0)', vec(G_t(\beta_0))')'\,(g_t(\beta_0)', vec(G_t(\beta_0)-EG_{tT}(\beta_0))')$$

$$\hat{\Sigma}_{2T}(\beta_0,s) = \frac{2K_T+1}{T-[Ts]}\sum_{t=[Ts]+1}^{T}(g_t(\beta_0)', vec(G_t(\beta_0))')'\,(g_t(\theta_0)', vec(G_t(\beta_0)-EG_{tT}(\beta_0))')\,.$$

Assumption 6.8′. Under the true value of the parameters θ_0, $\sup_{s\in S}\|\hat{\Sigma}_{iT}(\beta_0,s)-\Sigma(\beta_0\|\xrightarrow{p} 0$ with S whose closure lies in $(0,1)$ for $i=1,2$.

Assumption 6.3″. Under the true value of the parameters θ_0, $\{g\left(x_{Tt},\beta_0\right), vec\left(G\left(x_{Tt},\beta_0\right)-EG\left(x_{Tt},\beta_0\right)\right):$ $t\leq T, T\geq 1\}$ is a triangular array of mean zero R^q-valued rv's that is α-mixing with mixing coefficients

$\sum_{j=1}^{\infty} j^2\alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$ with $sup_{t\leq T,T\geq 1}E\|g\left(x_{Tt},\beta_0\right)\|^d < \infty$ and $sup_{t\leq T,T\geq 1}E\|G\left(x_{Tt},\beta_0\right)\|^d < \infty$ for some $d > \max\left(4\nu, \frac{2\eta}{\eta-1}\right)$.

Assumptions 6.3$''$ and 6.8$'$ guarantee for the restricted partial-sample GEL that

$$\hat{\Omega}_{Gg,1T}(\beta_0,s) = \frac{2K+1}{T}\sum_{t=1}^{[Ts]} vec\left(G_{tT}(\beta_0)\right) g_{tT}(\beta_0)' \xrightarrow{p} s\Omega_{Gg}(\beta_0), \tag{6}$$

$$\hat{\Omega}_{Gg,2T}(\beta_0,s) = \frac{2K+1}{T}\sum_{t=[Ts]+1}^{T} vec\left(G_{tT}(\beta_0)\right) g_{tT}(\theta_0)' \xrightarrow{p} (1-s)\Omega_{Gg}(\beta_0), \tag{7}$$

and

$$\hat{\Omega}_{GG,1T}(\beta_0,s) = \frac{2K+1}{T}\sum_{t=1}^{[Ts]} vec\left(G_{tT}(\beta_0)\right) vec\left(G_{tT} - EG_{tT}(\beta_0)\right)' \xrightarrow{p} s\Omega_{GG}(\beta_0),$$

$$\hat{\Omega}_{GG,2T}(\beta_0,s) = \frac{2K+1}{T}\sum_{t=[Ts]+1}^{T} vec\left(G_{tT}(\beta_0)\right) vec\left(G_{tT}(\beta_0) - EG_{tT}(\beta_0)\right)' \xrightarrow{p} (1-s)\Omega_{GG}(\beta_0).$$

Lemma 6.1 can be shown for the derivatives of the smoothed moment conditions under Assumptions 6.1, 6.2, 6.3$''$ and 6.8$'$ as shown for the smoothed moment conditions. Thus, the asymptotic distribution of the derivatives of the centered smoothed moment conditions under the null is given by:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} vec\left(G_{tT}(\beta_0) - EG_{tT}(\beta_0)\right) \Rightarrow \Omega_{GG}(\beta_0)^{1/2}B_{qr}(s) \tag{8}$$

where $B_{qr}(s)$ is a qr-dimensional vector of standard Brownian motion. Using Lemma 6.1, this yields for the whole vector $(g_{tT}(\beta_0)', (vec(G_{tT}(\beta_0) - EG_{tT}(\beta_0)))')'$

$$T^{-1/2}\sum_{t=1}^{[Ts]} (g_{tT}(\beta_0)', (vec(G_{tT}(\beta_0) - EG_{tT}(\beta_0)))')' \Rightarrow \Sigma(\beta_0)^{1/2}B_{q+qr}(s) \tag{9}$$

where $B_G(s)$ is a $((q+qr)\times 1)$-vector of standard Brownian motion.

We also need the following assumptions:

Assumption 6.13. *$\upsilon(s) \in \widehat{\Upsilon}_T(s)$ where $\widehat{\Upsilon}_T(s) = \{\upsilon(s) : \|\upsilon(s)\| \leq D\left(T/((2K_T+1)^2\right)^{-\zeta}\}$ for some $D > 0$ with $\frac{1}{2} > \zeta > \frac{1}{d(1-1/\eta)}$.*

Assumption 6.14. *Suppose Assumption 6.9 but for $\partial g(x_t,\beta)/\partial\beta_i$ for $i = 1,\ldots,r$.*

The Assumption 6.13 guarantees that Lemma 6.3 holds for the objective function defined with $K_{1t}(\beta,s)$ and $K_{1t}(\beta,s)$.

The proof is based on the following bound for all $s \in S$,

$$P_K(\tilde{\theta}_{K,T}(s), \hat{\upsilon}(\tilde{\theta}_{K,T}(s)s), s) \leq P_K(\theta_0, \hat{\upsilon}(\tilde{\theta}_{K,T}(s)s), s) \leq \sup_{\upsilon(s)\in\widehat{\Upsilon}_T(\theta_0,s)} P_K(\theta_0, \upsilon(s), s). \tag{10}$$

Let $\hat{\upsilon}_T(\theta_0, s) = \arg\max_{\upsilon(s)\in\Upsilon_T(\theta_0,s)} \hat{P}_K(\theta_0, \upsilon(s), s)$. The corresponding FOC for the first subsample are

$$\frac{1}{T}\sum_{t=1}^{[Ts]} \rho_1\left(\hat{\upsilon}_{1T}(\beta_0, s)' K_{1tT}(\beta_0, s)\right) K_{1tT}(\beta_0, s) = 0$$

have to hold at the true value $(\beta_0, 0)$ w.p.a.1. Let $\dot{\upsilon}_{1T}(\beta_0, s) = \tau\hat{\upsilon}_{1T}(\beta_0, s)$, $0 \leq \tau \leq 1$ and expanding the FOC in $\upsilon_{1T}(\beta_0, s)$ around 0 yields

$$\begin{aligned} 0 &= -\sum_{t=1}^{[Ts]} K_{1tT}(\beta_0, s) + \left[(2K_T+1)\sum_{t=1}^{[Ts]} \rho_2(\dot{\upsilon}_{1T}(\beta_0, s)' K_{1tT}(\beta_0, s) K_{1tT}(\beta_0, s)'/T\right](2K+1)^{-1}\hat{\upsilon}_{1T}(\beta_0, s) \\ &= -\sum_{t=1}^{[Ts]} K_{1tT}(\beta_0, s) - \left(\hat{D}_{1T}(\beta_0, s)'\hat{\Omega}_{1T}(\beta_0, s)^{-1}\hat{D}_{1T}(\beta_0, s)\right)(2K+1)^{-1}\hat{\upsilon}_{1T}(\beta_0, s) + o_p(1) \end{aligned}$$

by $\sup_{s\in S}\max_{1\leq t\leq T}\left|\rho_2\left(\dot{\upsilon}_{1T}(\theta_0, s)' K_{1tT}(\theta_0, s)\right) - \rho_2(0)\right| \xrightarrow{p} 0$ which holds by imposing Assumptions 6.3$''$ and 6.13.

Therefore,

$$\frac{1}{(2K+1)}\hat{\upsilon}_{1T}(\beta_0, s) = -\left(\hat{D}_{1T}(\beta_0, s)'\hat{\Omega}_{1T}(\beta_0, s)^{-1}\hat{D}_{1T}(\beta_0, s)\right)^{-1}\hat{D}_{1T}(\beta_0, s)'\hat{\Omega}_{1T}(\beta_0, s)^{-1}\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}(\beta_0, s) + o_p(1)$$

w.p.a.1. The derivation for the FOC for the second subsample $\frac{1}{T}\sum_{t=[Ts]+1}^{T}\rho_1\left(\upsilon_{2T}(\beta_0, s)' K_{2tT}(\beta_0, s)\right) K_{2tT}(\beta_0, s) = 0$ is the same and we obtain:

$$\frac{1}{(2K+1)}\hat{\upsilon}_{2T}(\beta_0, s) = -\left(\hat{D}_{2T}(\beta_0, s)'\hat{\Omega}_{2T}(\beta_0, s)^{-1}\hat{D}_{2T}(\beta_0, s)\right)^{-1}\hat{D}_{2T}(\beta_0, s)'\hat{\Omega}_{2T}(\beta_0, s)^{-1}\frac{1}{T}\sum_{t=[Ts]+1}^{T} g_{tT}(\beta_0) + o_p(1).$$

Now define for the vector $\theta = (\beta', \beta')'$:

$$\hat{D}_T(\theta, s) = \begin{bmatrix} \hat{D}_{1T}(\beta, s) & 0 \\ 0 & \hat{D}_{2T}(\beta, s) \end{bmatrix} \in R^{2q\times 2r},$$

$$K_T(\theta, s) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{[Ts]} K_{1tT}(\beta, s) \\ \frac{1}{T}\sum_{t=[Ts]+1}^{T} K_{2tT}(\beta, s) \end{bmatrix} \in R^{2r\times 1}$$

and

$$\Omega(\theta, s) = \begin{bmatrix} \Omega_1(\beta, s) & 0 \\ 0 & \Omega_1(\beta, s) \end{bmatrix} \in R^{2q\times 2q}.$$

By a second-order Taylor expansion of $P_K(\theta_0,\hat{\upsilon}_T(\theta_0,s),s)$ around 0 with $\dot{\upsilon}_T(\theta_0,s)=\tau\hat{\upsilon}_T(\theta_0,s)$, $0\le\tau\le1$,

$$\begin{aligned}\frac{1}{2K_T+1}P_K(\theta_0,\hat{\upsilon}(\theta_0,s),s) &= -\left(\frac{\hat{\upsilon}_T(s)}{2K_T+1}\right)' K_T(\theta_0,s)\\ &+ \left(\frac{\hat{\upsilon}_T(\theta_0,s)}{2K_T+1}\right)'\left(\sum_{t=1}^{T}\rho_2(\dot{\upsilon}_T(s)'K_{tT}(\theta_0,s)K_{tT}(\theta_0,s)K_{tT}(\theta_0,s)'/T\right)\hat{\upsilon}_T(\theta_0,s)/2\\ &= K_T(\theta_0,s)'\left(\hat{D}_T(\theta_0,s)'\hat{\Omega}(\theta_0,s)^{-1}\hat{D}_T(\theta_0,s)\right)^{-1}K_T((\theta_0,s)-\\ &\quad K_T(\theta_0,s)'\left(\hat{D}_T(\theta_0,s)'\hat{\Omega}(\theta_0,s)^{-1}\hat{D}_T(\theta_0,s)\right)^{-1}K_T((\theta_0,s)/2+o_p(1)\\ &= K_T(\theta_0,s)'\left(\hat{D}_T(\theta_0,s)'\hat{\Omega}(\theta_0,s)^{-1}\hat{D}_T(\theta_0,s)\right)^{-1}K_T((\theta_0,s)/2+o_p(1) \qquad (11)\end{aligned}$$

w.p.a.1 where the second equality holds by

$$\frac{1}{2K_T+1}\hat{\upsilon}_T(\theta_0,s)=-\left(\hat{D}_T(\theta_0,s)'\Omega(\theta_0,s)^{-1}\hat{D}_T(\theta_0,s)\right)^{-1}\hat{D}_T(\theta_0,s)'\Omega(\theta_0,s)^{-1}\hat{g}_T(\theta_0,s)+o_p(1)$$

and $K_T(\theta_0,s)=\hat{D}_T(\theta_0,s)'\hat{\Omega}(\theta_0,s)^{-1}\hat{g}_T(\theta_0,s)$.

Now, let $\hat{D}_{1T}(\beta_0,s)=\left[\hat{D}_{1,1T}(\beta_0,s),\hat{D}_{2,1T}(\beta_0,s),\dots,\hat{D}_{r,1T}(\beta_0,s)\right]$ with $\hat{D}_{i,1T}(\beta,s)=\frac{1}{T}\sum_{t=1}^{[Ts]}\rho_1(\hat{\lambda}_{1T}(\beta,s)'g_{tT}(\beta))G_{i,tT}(\beta,s)$ for $i=1,\dots,p$ and respectively for $\hat{D}_{2T}(\beta_0,s)$. By a Taylor expansion of $\hat{D}_{i,1T}(\beta_0,s)$ and $\hat{D}_{i,2T}(\beta_0,s)$ around $\hat{\lambda}_{1T}(\beta_0,s)=0$ and $\hat{\lambda}_{2T}(\beta_0,s)=0$ respectively yields

$$\hat{D}_{i,1T}(\beta_0,s)=-\frac{1}{T}\sum_{t=1}^{[Ts]}G_{i,tT}(\beta_0)+\frac{2K+1}{T}\sum_{t=1}^{[Ts]}G_{i,tT}(\beta_0)g_{tT}(\beta_0)'\hat{\Omega}_{1T}(\beta_0,s)^{-1}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0)+o_p(1)$$

$$\hat{D}_{i,2T}(\beta_0,s)=-\frac{1}{T}\sum_{t=[Ts]+1}^{T}G_{i,tT}(\beta_0)+\frac{2K+1}{T}\sum_{t=[Ts]+1}^{T}G_{i,tT}(\beta_0)g_{tT}(\beta_0)'\hat{\Omega}_{2T}(\beta_0,s)^{-1}\frac{1}{T-[Ts]}\sum_{t=[Ts]+1}^{T}g_{tT}(\beta_0)+o_p(1)$$

using $\frac{1}{2K_T+1}\hat{\lambda}_{1T}(\beta_0,s)=-\hat{\Omega}_{1T}(\beta_0,s)^{-1}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0)+o_p(1)$ and $\frac{1}{2K_T+1}\hat{\lambda}_{2T}(\beta_0,s)=-\hat{\Omega}_{2T}(\beta_0,s)\frac{1}{T-[Ts]}\sum_{t=[Ts]+1}^{T}g_{tT}(\beta_0)+o_p(1)$ with $\sup_{s\in S}\max_{1\le t\le T}|\rho_2(\hat{\lambda}_{iT}(\beta_0,s)'g_{tT}(\beta_0))-\rho_2(0)|\xrightarrow{p}0$ for $i=1,2$.

Using (6), (7), (8), (9), Lemma 6.1 and with $G(\beta_0)=\lim_{T\to\infty}\left[T^{-1}\sum_{t=1}^{T}G_{tT}(\beta_0)\right]$, we obtain that

$$\begin{bmatrix} I_q & 0\\ -\frac{2K+1}{T}\sum_{t=1}^{[Ts]}vec\left(G_{tT}(\beta_0)\right)g_{tT}(\beta_0)'\hat{\Omega}_{1T}(\beta_0)^{-1} & I_{qr}\end{bmatrix}\times\begin{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0)\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}vec\left(G_{tT}(\beta_0)-EG_{tT}(\beta_0)\right)\end{bmatrix}$$

$$=\begin{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0)\\ -\sqrt{T}\left(\hat{D}_{1T}(\beta_0,s)-sG(\beta_0)\right)\end{bmatrix}\Rightarrow\begin{bmatrix}\Omega(\beta_0)^{1/2}B_q(s)\\ \Omega_D(\beta_0)^{1/2}B_{2.1}(s)\end{bmatrix}$$

with $\Omega_D(\beta_0)^{1/2}B_{2.1}(s)=\Omega_{GG}(\beta_0)^{1/2}B_{qr}(s)-\Omega_{Gg}(\beta_0)\Omega(\beta_0)^{-1}\Omega(\beta_0)^{1/2}B_q(s)$, $\Omega_D(\beta_0)=\Omega_{GG}(\beta_0)-\Omega_{Gg}(\beta_0)\Omega(\beta_0)^{-1}\Omega_{Gg}(\beta_0)$ and $B_{2.1}(s)$ is independent of $B_q(s)$. This result is true for any value of $G(\beta_0)$. Thus, $G(\beta_0)$ can be of full rank value, weak value such that $G_T(\beta_0) = \frac{C_1}{T^{1/2}}$ for $q \times r$ matrix C_1 or $G(\beta_0) = 0$ in the case of no identification.

This implies that

$$\sqrt{T}\left(\hat{D}_{1T}(\beta_0, s) - sG(\beta_0)\right) \Rightarrow -\Omega_D(\beta_0)^{1/2}B_{2.1}(s)$$

and

$$\sqrt{T}\left(\hat{D}_{2T}(\beta_0, s) - (1-s)G(\beta_0)\right) \Rightarrow -\Omega_D(\beta_0)^{1/2}(B_{2.1}(1) - B_{2.1}(s)).$$

Since $\hat{D}_{1T}(\beta_0, s)$ and $\hat{D}_{2T}(\beta_0, s)$ are respectively independent of $\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}(\beta_0)$ and $\frac{1}{T}\sum_{t=[Ts]+1}^{T} g_{tT}(\beta_0)$ this yields

$$\left(\hat{D}_{1T}(\beta_0)'\hat{\Omega}_{1T}(\beta_0)^{-1}\hat{D}_{1T}(\beta_0)\right)^{-1/2}\hat{D}_{1T}(\beta_0)'\hat{\Omega}_{1T}(\beta_0)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}(\beta_0) \Rightarrow B_r(s) \tag{12}$$

and

$$\left(\hat{D}_{2T}(\beta_0)'\hat{\Omega}_{2T}(\beta_0)^{-1/2}\hat{D}_{2T}(\beta_0)\right)^{-1}\hat{D}_{2T}(\beta_0)'\hat{\Omega}_{2T}(\beta_0)^{-1}\frac{1}{\sqrt{T}}\sum_{t=[Ts]+1}^{T} g_{tT}(\beta_0) \Rightarrow B_r(1) - B_r(s) \tag{13}$$

where $B_r(s)$ is a r-vector of standard Brownian motion.

By the inequality (10),

$$\frac{2T}{2K+1}P_K(\tilde{\theta}_{K,T}(s), \hat{v}(\tilde{\theta}_{K,T}(s)s), s) \leq \frac{2T}{2K+1}P_K(\theta_0, \hat{v}(\theta_0, s), s)$$

and using (11), (12) and (13)

$$\begin{aligned}
\frac{2T}{2K+1}P_K(\theta_0, \hat{v}(\theta_0, s), s) &= [Ts]\frac{1}{[Ts]}\sum_{t=1}^{[Ts]} g_{tT}(\beta_0)\hat{\Omega}_{1T}(\beta_0, s)^{-1/2}P_{1T,\Omega^{-1/2}\hat{D}}(\beta_0, s)\hat{\Omega}_{1T}(\beta_0, s)^{-1/2}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]} g_{tT}(\beta_0) \\
&+ [T-Ts]\frac{1}{[T-Ts]}\sum_{t=[Ts]+1}^{T} g_{tT}(\beta_0)\hat{\Omega}_{2T}(\beta_0, s)^{-1/2}P_{2T,\Omega^{-1/2}\hat{D}}(\beta_0, s)\hat{\Omega}_{2T}(\beta_0, s)^{-1/2} \\
&\times \frac{1}{[T-Ts]}\sum_{t=[Ts]+1}^{T} g_{tT}(\beta_0) + o_p(1) \\
&\Rightarrow \frac{B_r(s)'B_r(s)}{s} + \frac{[B_r(1) - B_r(s)]'[B_r(1) - B_r(s)]}{1-s}
\end{aligned}$$

with $P_{1T,\Omega^{1/2}\hat{D}}(\beta_0, s) = \hat{\Omega}_{1T}(\beta_0, s)^{-1/2}\hat{D}_{1T}(\beta_0, s)\left(\hat{D}_{1T}(\beta_0, s)'\hat{\Omega}_{1T}(\beta_0, s)^{-1/2}\hat{D}_{1T}(\beta_0, s)\right)^{-1}\hat{D}_{1T}(\beta_0, s)'\hat{\Omega}_{1T}(\beta_0, s)^{-1/2}$ and $P_{2T,\Omega^{1/2}\hat{D}}(\beta_0, s) = \hat{\Omega}_{2T}(\beta_0, s)^{-1/2}\hat{D}_{2T}(\beta_0, s)\left(\hat{D}_{2T}(\beta_0, s)'\hat{\Omega}_{2T}(\beta_0, s)^{-1/2}\hat{D}_{2T}(\beta_0, s)\right)^{-1}\hat{D}_{2T}(\beta_0, s)'\hat{\Omega}_{2T}(\beta_0, s)^{-1/2}$. The result follows directly under the null. The derivation under the alternative can be easily obtained.

References

Anatolyev, S. (2005) 'GMM, GEL, serial correlation, and asymptotic bias.' *Econometrica* 73(3), 983–1002

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

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

Andrews, D.W.K., and J.C. Monahan (1992) 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator.' *Econometrica* 60, 953–966

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

Caner, M. (2007) 'Bounded pivotal structural change tests in continuous updating GMM with strong, weak identification and completely unidentified cases.' *Journal of Econometrics* 137, 28–67

Gallant, A.R. (1987) *Nonlinear statistical models* (New York: Wiley)

Gallant, A.R., and H. White (1988) *A Unified theory of estimation and inference for nonlinear dynamic models* (Basil Blackwell)

Guay, A. (2003) 'Optimal predictive tests.' *Econometric Reviews* 22(4), 379–410

Guay, A., and J-F Lamarche (2009) 'Structural change tests based on implied probabilities for gel criteria.' Working Paper

Guggenberger, P., and R. J. Smith (2008) 'Generalized empirical likelihood tests in time series models with potential identification failure.' *Journal of Econometrics* 142, 134–161

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

Hansen, L.P., J. Heaton, and A. Yaron (1996) 'Finite-sample properties of some alternative GMM estimators.' *Journal of Business and Economic Statistics* 14, 262–280

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

Kleibergen, F. (2005) 'Testing parameters in GMM without assuming that they are identified.' *Econometrica* 73, 1103–1123

Newey, W.K., and K.D. West (1994) 'Automatic lag selection in covariance matrix estimation.' *Review of Economic Studies* 61, 631–654

Newey, W.K., and R.J. Smith (2004) 'Higher order properites of GMM and generalized empirical likelihood estimators.' *Econometrica* 72, 219–256

Otsu, T. (2006) 'Generalized empirical likelihood inference for nonlinear and time series models under weak identification.' *Econometric Theory* 22, 513–527

Pollard, D. (1984) *Convergence of Stochastic Processes* (New York: Springer-Verlag)

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

Smith, R. J. (2004) 'GEL criteria for moment condition models.' CEMMAP Working Paper CWP19/04

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

Smith, R.J. (2000) 'Empirical likelihood estimation and inference.' In *Applications of Differential Geometry to Econometrics,* ed. M. Salmon and P. Marriott (Cambridge: Cambridge University Press) pp. 119–150

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

—— (1996b) 'Test for violations of moment conditions.' Manuscript, Carnegie Mellon University

Stock, James H., and J.H. Wright (2000) 'GMM with weak identification.' *Econometrica* 68, 1055–1096

Wooldridge, J.M., and H. White (1988) 'Some invariance principles and central limit theorems for dependent heterogeneous processes.' *Econometric Theory* 4, 210–230

Table 1: Data Generating Processes

	H_0^S	H_A^I	H_A^O
DGP1	$\beta_1 = \beta_2 = 0$		$\alpha = 0$
DGP2	$\beta_1 = \beta_2 = 0.4$		$\alpha = 0$
DGP3	$\beta_1 = \beta_2 = 0.8$		$\alpha = 0$
DGP4		$\beta_1 = 0, \beta_2 = 0.4$	$\alpha = 0$
DGP5		$\beta_1 = 0, \beta_2 = 0.8$	$\alpha = 0$
DGP6		$\beta_1 = 0.4, \beta_2 = 0.8$	$\alpha = 0$
DGP7	$\beta_1 = \beta_2 = 0.4$		$\alpha = 0.5$
DGP8	$\beta_1 = \beta_2 = 0.4$		$\alpha = 0.9$
DGP9	$\beta_1 = \beta_2 = 0.4$		$\alpha = -0.5$
DGP10	$\beta_1 = \beta_2 = 0.4$		$\alpha = -0.9$

Table 2: Rejection Frequencies for Tests of Structural Change in the Parameters

DGP	Size (%)	*supW*	*aveW*	*expW*	*supLR*	*aveLR*	*expLR*
DGP1	1	0.0275	0.0190	0.0275	0.0220	0.0195	0.0240
	5	0.0855	0.0725	0.0815	0.0825	0.0730	0.0835
	10	0.1450	0.1225	0.1465	0.1435	0.1290	0.1470
DGP2	1	0.0285	0.0150	0.0275	0.0235	0.0150	0.0225
	5	0.0880	0.0735	0.0900	0.0845	0.0720	0.0875
	10	0.1510	0.1295	0.1500	0.1525	0.1375	0.1550
DGP3	1	0.0320	0.0120	0.0195	0.0170	0.0120	0.0145
	5	0.0980	0.0635	0.0825	0.0715	0.0675	0.0750
	10	0.1550	0.1300	0.1440	0.1425	0.1290	0.1390
DGP4	1	0.4840	0.5205	0.5420	0.4840	0.5295	0.5475
	5	0.6915	0.7430	0.7500	0.6930	0.7510	0.7510
	10	0.7860	0.8280	0.8300	0.7875	0.8365	0.8325
DGP5	1	1.0000	0.9995	1.0000	1.0000	1.0000	1.0000
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
DGP6	1	0.7060	0.7475	0.7655	0.6950	0.7435	0.7625
	5	0.8590	0.8910	0.8955	0.8590	0.8975	0.9000
	10	0.9075	0.9345	0.9365	0.9140	0.9385	0.9370
DGP7	1	0.2895	0.2215	0.3045	0.1845	0.1905	0.2205
	5	0.4605	0.4170	0.4795	0.3630	0.3855	0.4065
	10	0.5705	0.5275	0.5805	0.4915	0.5015	0.5190
DGP8	1	0.4425	0.3800	0.4720	0.3575	0.3545	0.4000
	5	0.6160	0.5960	0.6460	0.5530	0.5760	0.5990
	10	0.7140	0.6900	0.7305	0.6595	0.6855	0.6950
DGP9	1	0.2255	0.1275	0.2110	0.1120	0.0955	0.1240
	5	0.3660	0.2805	0.3640	0.2650	0.2415	0.2790
	10	0.4640	0.4015	0.4650	0.3770	0.3665	0.3960
DGP10	1	0.4530	0.2825	0.4410	0.2150	0.1425	0.2130
	5	0.5960	0.4890	0.5920	0.4205	0.3340	0.4200
	10	0.6790	0.6050	0.6790	0.5540	0.4755	0.5385

Table 3: Rejection Frequencies for Tests of Structural Change in the Overidentifying Restrictions

DGP	Size (%)	*supO*	*aveO*	*expO*	*supOg*	*aveOg*	*expOg*	*supLMO*	*aveLMO*	*expLMO*
DGP1	1	0.0020	0.0075	0.0050	0.0205	0.0195	0.0215	0.0055	0.0080	0.0050
	5	0.0210	0.0420	0.0390	0.0735	0.0665	0.0710	0.0305	0.0455	0.0390
	10	0.0595	0.1005	0.0865	0.1300	0.1260	0.1390	0.0730	0.1020	0.0925
DGP2	1	0.0040	0.0085	0.0085	0.0245	0.0155	0.0240	0.0090	0.0080	0.0090
	5	0.0270	0.0450	0.0440	0.0870	0.0645	0.0785	0.0345	0.0455	0.0495
	10	0.0595	0.0955	0.0865	0.1370	0.1240	0.1365	0.0835	0.1015	0.0975
DGP3	1	0.0055	0.0090	0.0075	0.0260	0.0175	0.0260	0.0080	0.0095	0.0095
	5	0.0295	0.0490	0.0440	0.0815	0.0690	0.0785	0.0405	0.0455	0.0500
	10	0.0630	0.1035	0.0890	0.1445	0.1280	0.1445	0.0810	0.1045	0.1020
DGP4	1	0.0050	0.0070	0.0080	0.0295	0.0175	0.0285	0.0090	0.0090	0.0120
	5	0.0360	0.0460	0.0445	0.0965	0.0720	0.0935	0.0480	0.0475	0.0545
	10	0.0810	0.0985	0.1075	0.1700	0.1370	0.1660	0.1005	0.1105	0.1160
DGP5	1	0.1605	0.0805	0.1725	0.2965	0.1435	0.2915	0.1620	0.0835	0.1805
	5	0.3805	0.2630	0.4040	0.5195	0.3475	0.5040	0.3935	0.2670	0.4155
	10	0.5305	0.4300	0.5420	0.6415	0.4990	0.6195	0.5370	0.4345	0.5470
DGP6	1	0.0085	0.0110	0.0145	0.0425	0.0230	0.0410	0.0150	0.0150	0.0210
	5	0.0505	0.0515	0.0615	0.1150	0.0750	0.1030	0.0650	0.0520	0.0675
	10	0.0925	0.1055	0.1140	0.1865	0.1455	0.1760	0.1165	0.1140	0.1280
DGP7	1	0.0425	0.1725	0.1290	0.5640	0.5610	0.6075	0.0685	0.1540	0.1330
	5	0.2380	0.5415	0.4740	0.7450	0.7705	0.7860	0.2525	0.5310	0.4780
	10	0.4380	0.7255	0.6790	0.8305	0.8560	0.8635	0.4645	0.7240	0.6860
DGP8	1	0.0850	0.3275	0.2375	0.7665	0.7975	0.8070	0.1635	0.3150	0.2845
	5	0.3620	0.7540	0.6755	0.8855	0.9185	0.9070	0.4325	0.7540	0.6865
	10	0.5875	0.8820	0.8410	0.9220	0.9525	0.9525	0.6390	0.8960	0.8605
DGP9	1	0.0960	0.4185	0.2975	0.8620	0.8810	0.8980	0.1895	0.3550	0.3220
	5	0.4575	0.8230	0.7810	0.9495	0.9640	0.9645	0.4855	0.8040	0.7525
	10	0.7150	0.9445	0.9245	0.9740	0.9830	0.9825	0.7105	0.9280	0.9060
DGP10	1	0.1140	0.6610	0.4790	0.9935	0.9955	0.9960	0.3950	0.5205	0.5165
	5	0.6080	0.9780	0.9470	0.9985	0.9995	0.9995	0.7060	0.9205	0.8955
	10	0.8645	0.9970	0.9925	0.9995	1.0000	1.0000	0.8905	0.9755	0.9705