ESG RECHERCHE

DOCUMENT DE TRAVAIL / WORKING PAPER

No. 2018-06

Structural change tests for GEL criteria

Alain Guay *et* Jean-François Lamarche

Janvier 2018

ESG UQÀM

Département des sciences économiques
École des sciences de la gestion
Université du Québec à Montréal

Structural change tests for GEL criteria

Alain Guay, Université du Québec à Montréal, Canada; CIPRÉE; et CIREQ
Jean-François Lamarche, Brock University, Canada

Document de travail No. 2018-06

Janvier 2018

Département des Sciences Économiques
Université du Québec à Montréal
Case postale 8888,
Succ. Centre-Ville
Montréal, (Québec), H3C 3P8, Canada
Courriel : brisson.lorraine@uqam.ca
Site web : http://economie.esg.uqam.ca

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

December 2, 2015

First version: February 17, 2010

Abstract

This paper examines structural change tests based on generalized empirical likelihood methods in the time series context, allowing for dependent data. Standard structural change tests for the GMM 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 the GMM, in cases with known and unknown breakpoints. New test statistics specific to GEL methods, and that are robust to weak identification, are also introduced. A simulation study examines the small sample properties of the tests and reveals that GEL-based robust tests performed well, both in terms of the presence and location of a structural change and in terms of the nature of identification.

Keywords: Generalized empirical likelihood, generalized method of moments, parameter instability, weak identification, structural change
JEL codes: C12, C13, C15, 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). An additional document contains lemmas and a longer version of the proofs. It can be found at `http://coffee.econ.brocku.ca/jfl/research`.
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

As the GMM is limited in a number of ways (e.g., the existence of a bias, the requirement to compute a particular weighting matrix), a number of alternative estimators have been proposed: the Continuous Updated Estimator (CUE, Hansen, Heaton and Yaron, 1996), the Empirical Likelihood estimator (EL, Qin and Lawless, 1994) and the Exponential Tilting estimator (ET, Kitamura and Stutzer, 1997). These three estimators can be viewed as special cases of the Generalized Empirical Likelihood (GEL) framework considered by Smith (1997). For example, the CUE shares the same objective function as the GMM but with a weighting matrix that gets continuoulsy updated as opposed to only once under two-steps estimation (Hansen *et al.* (1996)). Newey and Smith (2004) showed (in an i.i.d. setting) that the GEL and GMM estimators have the same asymptotic distribution but have different higher order properties. It was shown that the expression for the second order asymptotic bias of GEL has fewer components than the asymptotic bias for the GMM (with EL having the fewest). Anatolyev (2005) extended these findings to allow for serial correlation and showed that smoothing the moment conditions reduces bias even further. These findings generated interesting avenues of theoretical and empirical reseach.

An important aspect of the validation of an estimation strategy is the stability of the parameters and of the objective function. Both the GMM and the GEL approaches assume that parameters and moment restrictions are stable across time. Detecting structural changes is important and given that the GEL methods have interesting properties, it is important to study structural change tests using these estimators.

We introduce a class of partial-sample GEL (PS-GEL) estimators. As an original contribution to the literature, we establish that the weak convergence of the sequence of PS-GEL estimators is a function of Brownian motions. Also, the sequence of PS-GEL Lagrange multiplier estimators weakly converges to a function of Brownian motions and is asymptotically uncorrelated with the sequence of PS-GEL estimators. These findings allow us to demonstrate that stability tests based on the Lagrange multipliers are stability tests for the overidentifying restrictions that are then orthogonal to the identification conditions. These asymptotic distributions are derived under the null hypothesis of stability and general alternatives of structural change (see Sowell, 1996) for an unknown breakpoint. The distributions are derived for dependent data characterized as near epoch dependent. Amongst temporal dependency concepts, near epoch dependency is the least restrictive and allows for certain nonstationarity as well as for the analysis of local alternatives.

We consider cases of structural change which can occur in the parameters of interest or in the overidentifying restrictions used to estimate these parameters. First, we study standard Wald ($GELW$), Lagrange multiplier ($GELM$) and likelihood ratio ($GELR$) 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 properly smoothed moment conditions, follow the same asymptotic distribution as in the GMM context (Andrews, 1993). Second, we examine tests for the stability of overidentifying restrictions because changes could be attributed to violations of the moment conditions themselves. Equivalent test statistics to Hall and Sen's

(1999) statistics in the GMM context are adapted to the GEL with smoothed moment conditions. Two new tests specific to the GEL framework are also proposed to detect instability of the overidentifying restrictions. We show that these new statistics have the same asymptotic distribution, at first order, as the distribution derived by Hall and Sen (1999) when the moment conditions are properly smoothed. A related paper (Guay and Lamarche, 2012) uses and extend the theoretical results derived in this paper to propose Pearson-type based statistics based on implied probabilities to detect structural change. The Pearson-type tests are easily computed and have a nice intuitive interpretation. Recently, Hall *et al.* (2011) propose Andrews-based tests but derived from an information-metric perspective.

In the last decade, weak identification has received a large amount of attention (Stock and Wright, 2000). When the presence of weak instruments is suspected, structural change tests robust to weak identification need to be implemented so that the correct distributions can be used to perform inference. 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 partial-sample 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 in the 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 allowing for test statistics to be quite flexible.

Our simulation study, based on the widely used design of Stock and Wright (2000), makes use of the exponential tilting estimation method to evaluate the performance of the tests under different types of parameter identification. With standard identification we find that GEL-based tests perform very well (and better than GMM-based tests). Things deterioate quickly with weak parameter identification. The robust GEL-based tests (for example our GEL^R, $GELIPS^R$ and $GELK$ tests presented in section 3.3) performed well in terms of: 1) the presence and location of a structural change and 2) the nature of identification. These test statistics should then be added to the Pearson-type statistics based on implied probabilities to detect structural change presented by Guay and Lamarche (2012) to complement the specification and testing arsenal of the practitioners.

The paper is organized as follows. Section 2 presents the full-sample and partial-sample GMM and GEL estimators. Section 3 presents the test statistics and their respective asymptotic distributions. The simulation results are in Section 4 and 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 a 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 and partial structural change. Pure structural change is when no subvector δ appears and the entire parameter vector is subject to structural change under the alternative while partial structural change corresponds to cases where only a subvector β is subject to structural change under the alternative hypothesis. The generic null of parameter stability can be written as follows:

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

The assumption under the alternative hypothesis is that at some point in the sample there is a single structural break, for instance:

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

where s determines the fraction of the sample before and after the assumed breakpoint and $[\cdot]$ denotes the greatest integer function generated under such conditions. The separation $[Ts]$ represents a possible breakpoint which is governed by an unknown parameter s. Hence, consider 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:

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

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

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

2.1 Definitions

We assume a triangular array of random variables $\{x_{T,t} : 1 \leq t \leq T, T \geq 1\}$ that are supposed to be near epoch dependent which allows temporal dependence but does not impose stationarity.[1] Triangular arrays of random

[1] See Section 5.1 in the Appendix for more details.

variables are required to study local power of the structural change tests, however, to simplify the notation $x_{T,t}$ is denoted as 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 $(\beta', \delta')' \in B \times \Delta \subset R^{r+\nu}$ which are estimated by the population orthogonality conditions

$$\frac{1}{T}\sum_{t=1}^{T} E[g(x_t, \beta_0, \delta_0)] = 0.$$

These moment conditions allow for nonstationary data as in Andrews (1993).

In this study of a single structural change we consider two subsamples, the first is based on observations $t = 1, \ldots, [Ts]$ and the second covers $t = [Ts]+1, \ldots, T$ where $s \in S \subset (0,1)$. 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:

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

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

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

Thus, under the null $H_0 : \lambda_1 = \lambda_2 = \lambda_0 = 0$. 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, 2011 and Guggenberger and Smith, 2008). The adjustment consists of smoothing the original moment conditions $g(x_t, \theta)$. Defining the smoothed moment conditions as

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

for $t = 1, \ldots, T$ and M_T is a bandwidth parameter, $k(\cdot)$ a kernel function and we define $k_j = \int_{-\infty}^{\infty} k(a)^j da$. Let $\rho(\phi)$ be a function of a scalar ϕ that is concave on its domain, an open interval Φ that contains 0, 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, 2011).

The restricted Generalized Empirical Likelihood (GEL) estimator using the entire sample is formally defined as:

Definition 2.1. *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* $\{(\tilde{\beta}_T, \tilde{\delta}_T)\}$ *is a sequence of random vectors such that:*

$$\left(\tilde{\beta}_T', \tilde{\delta}_T'\right)' = \arg \min_{(\beta,\delta)\in B\times\Delta} \sup_{\lambda\in\widetilde{\Lambda}_T(\beta,\delta)} \sum_{t=1}^{T} \frac{\left[\rho\left(k\lambda' g_{tT}(\beta,\delta)\right) - \rho(0)\right]}{T}$$

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

We call $\tilde{\theta}_T = \left(\tilde{\beta}_T', \tilde{\beta}_T', \tilde{\delta}_T'\right)'$ the full-sample GEL estimator of θ. The objective function is normalized so that $\rho_1 = \rho_2 = -1$ (see Smith, 2011). As mentioned earlier, the GEL estimator admits a number of special cases recently proposed in the econometrics literature. The CUE corresponds to the following quadratic function $\rho(\phi) = -(1+\phi)^2/2$. The EL estimator is a GEL estimator with $\rho(\phi) = \ln(1-\phi)$ and finally the ET estimator is obtained with $\rho(\phi) = -\exp(\phi)$.

More precisely, the GEL estimator is obtained as the solution of a saddle point problem. Firstly, the criterion is maximized with respect to the parameter vector (β, δ):

$$\tilde{\lambda}_T(\beta,\delta) = \arg \sup_{\lambda\in\widetilde{\Lambda}_T(\beta,\delta)} \sum_{t=1}^{T} \frac{\left[\rho\left(k\lambda' g_{tT}(\beta,\delta)\right) - \rho(0)\right]}{T}.$$

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

$$\left(\tilde{\beta}_T', \tilde{\delta}_T'\right)' = \arg \min_{(\beta,\delta)\in B\times\Delta} \sum_{t=1}^{T} \frac{\left[\rho\left(k\tilde{\lambda}_T(\beta,\delta)' g_{tT}(\beta,\delta)\right) - \rho(0)\right]}{T}.$$

To characterize the asymptotic distribution we need to define the following gradient matrices:

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

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

and the long run covariance matrix of the moment conditions is defined as:

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

The smoothed moment conditions are obtained the truncated kernel (see Kitamura and Stutzer, 1997 and Guggenberger and Smith, 2008) defined as

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

yielding[2]

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

[2] 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.

where K_T is related to the bandwidth parameter M_T. To handle the endpoints in the smoothing we use the approach of Smith (2011) and Guggenberger and Smith (2008) which sets

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

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

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

The weighting matrix thus obtained using this type of kernel is similar to the matrix obtained with the Bartlett kernel estimator of the long run covariance matrix of the moment conditions (see Smith, 2011). Along the same lines we will define the derivatives of the smoothed moment conditions as:

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

If we consider a possible breakpoint $[Ts]$, we need to 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 is the vector for the second part of the sample ($t = [Ts]+1,\ldots,T$). The sequence of partial-sample GEL estimators for $s \in S$ based on the first and the second subsamples are formally defined as:

Definition 2.2. *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 sequence of partial-sample Generalized Empirical Likelihood (PS-GEL) estimators $\{\hat{\theta}_T(s)\}$ is a sequence of random vectors such that:*

$$\begin{aligned} \hat{\theta}_T(s) &= \arg\min_{\theta\in\Theta} \sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta,s)} \sum_{t=1}^{T} \frac{[\rho(\lambda(s)' g_{tT}(\theta,s)) - \rho(0)]}{T} \\ &= \arg\min_{\theta\in\Theta} \sup_{\lambda(s)\in\widehat{\Lambda}_T(\theta,s)} \left(\sum_{t=1}^{[Ts]} \frac{[\rho(\lambda_1' g_{tT}(\beta_1,\delta)) - \rho(0)]}{T} + \sum_{t=[Ts]+1}^{T} \frac{[\rho(\lambda_2' g_{tT}(\beta_2,\delta)) - \rho(0)]}{T} \right). \end{aligned}$$

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, $\hat{\lambda}_T(\theta,s) = \left(\hat{\lambda}_{1T}(\beta_1,\delta,s)', \hat{\lambda}_{2T}(\beta_2,\delta,s)'\right)'$ where

$$\begin{aligned} \hat{\lambda}_{1T}(\beta_1,\delta,s) &= \arg \sup_{\lambda_1\in\widehat{\Lambda}_{1T}(\beta_1,\delta,s)} \sum_{t=1}^{[Ts]} \frac{[\rho(\lambda_1' g_{tT}(\beta_1,\delta)) - \rho(0)]}{T}, \\ \hat{\lambda}_{2T}(\beta_2,\delta,s) &= \arg \sup_{\lambda_2\in\widehat{\Lambda}_{2T}(\beta_2,\delta,s)} \sum_{t=[Ts]+1}^{T} \frac{[\rho(\lambda_2' g_{tT}(\beta_2,\delta)) - \rho(0)]}{T} \end{aligned}$$

with $\widehat{\Lambda}_{1T}(\beta_1,\delta,s)=\{\lambda_1:\lambda_1' g_{tT}(\beta_1,\delta)\in\Phi, t=1,\ldots,[Ts]\}$ and $\widehat{\Lambda}_{2T}(\beta_2,\delta,s)=\{\lambda_2:\lambda_2' g_{tT}(\beta_2,\delta)\in\Phi, t=[Ts]+1,\ldots,T\}$. 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:

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

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.

Let us define the long run covariance matrix $\Omega(s)$ as

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

which under the null (2) is asymptotically equal to

$$\Omega(s)=\left[\begin{array}{cc} s\Omega & 0\\ 0 & (1-s)\Omega\end{array}\right].$$

Now we define the estimator

$$\widehat{\Omega}_T(s)=\left[\begin{array}{cc} s\widehat{\Omega}_{1T}(s) & 0\\ 0 & (1-s)\widehat{\Omega}_{2T}(s)\end{array}\right],$$

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

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

Finally let

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

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 of GEL estimators $\hat{\theta}_T(\cdot)$ and $\hat{\lambda}_T(\cdot)$ 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 enables us to construct various tests for structural change in the spirit of Sowell (1996a) and Hall and Sen (1999).

2.2 Refining the null hypothesis

The moment conditions for the full sample under the null can be written as: $Eg(x_t,\beta_0,\delta_0) = 0, \forall t = 1,\ldots,T$. 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:

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

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 of stability of moment conditions. For instance, following Hall and Sen (1999) we can consider the null, denoted $H_0^I(s)$, 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:

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

Moreover, the overidentifying restrictions are stable if the restrictions 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:

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

We can then write the general null hypothesis of stability as $H_0 : H_0^I(s)$ & $H_0^O(s)$. 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 separately. 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:*

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

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[3]. 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.

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

[3]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.

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 stability

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

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

where $\hat{V}_T(s) = \left(\hat{V}_{1T}(s)/s + \hat{V}_{2T}(s)/(1-s)\right)$ and $\hat{V}_{iT}(s) = \left(\hat{G}^{\beta}_{i,tT}(s)'\hat{\Omega}^{-1}_{i,T}(s)\hat{G}^{\beta}_{i,tT}(s)\right)^{-1}$ for $i = 1, 2$ corresponding to the first and the second parts of the sample and respectively:

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

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

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

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 $GELM_T(s)$ 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:

$$GELR_T(s) = \frac{2T}{2K+1}\left[\sum_{t=1}^{T}\frac{\left[\rho(\hat{\lambda}_T(\tilde{\theta}_T,s)'g_{tT}(\tilde{\theta}_T,s)) - \rho(0)\right]}{T} - \sum_{t=1}^{T}\frac{\left[\rho(\hat{\lambda}_T(\hat{\theta}_T(s),s)'g_{tT}(\hat{\theta}_T(s),s)) - \rho(0)\right]}{T}\right]$$

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

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

Theorem 3.1. *Under the null hypothesis 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:*

$$GELW_T(s) \Rightarrow Q_r(s), GELM_T(s) \Rightarrow Q_r(s), GELR_T(s) \Rightarrow Q_r(s),$$

with

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

and under local alternatives (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 a 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.

Given Theorem 3.1, we see that the asymptotic distributions under the null of the Wald, LM and LR-type 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 Theorem 3.1 is only valid for moment conditions properly smoothed. Indeed, smoothing the moment conditions is necessary to obtain test statistics whose asymptotic distributions do not depend on nuisance parameters (except s), a finding that 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 (1994) 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:

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

where various choices of c determine power against close or more distant alternatives and $F(\cdot)$ is the weight function over the value of $s \in S$. For instance, Andrews (1993) suggests a possible breakpoint in the interval $S = [.15, .85]$ with an uniform weighting, namely, $F(s) = s$, be considered. Other intervals and weighting can be considered depending on the a priori of the possible breakpoint. In the case of close alternatives ($c = 0$), the optimal test statistic takes the average form, $aveQ_T = \int_S Q_T(s)dF(s)$. For a distant alternative ($c = \infty$), the optimal test statistics take the exponential form, $expQ_T = log\left(\int_S \exp[\frac{1}{2}Q_T(s)]dF(s)\right)$. 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-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:*

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

with

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

This result is directly obtained through the application of the continuous mapping theorem (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 $GELW_T(s)$, $GELM_T(s)$ or $GELR_T(s)$ 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 in 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:

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

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:

$$
GELO_T(s) = GELO1_T(s) + GELO2_T(s)
$$

where

$$
GELO1_T(s) = \frac{2[Ts]}{2K+1}\sum_{t=1}^{[Ts]}\frac{\left[\rho(\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s)' g_{tT}(\hat{\beta}_{1T}(s))) - \rho(0)\right]}{[Ts]}
$$

and

$$
GELO2_T(s) = \frac{2(T-[Ts])}{2K+1}\sum_{t=[Ts]+1}^{T}\frac{\left[\rho(\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s)' g_{tT}(\hat{\beta}_{2T}(s))) - \rho(0)\right]}{T-[Ts]}.
$$

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:

$$
GELMO_T(s) = GELMO1_T(s) + GELMO2_T(s)
$$

where

$$GELMO1_T(s) = \frac{[Ts]}{(2K_T+1)^2}\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s),s)'\widehat{\Omega}_{1T}(s)\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s),s)$$

and

$$GELMO2_T(s) = \frac{[T-Ts]}{(2K_T+1)^2}\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s),s)'\widehat{\Omega}_{2T}(s)\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s),s).$$

The following Theorem provides the asymptotic distribution of $Q_T^O(s)$ which equals $O_T(s)$, $GELO_T(s)$ and $GELMO_T(s)$ under the null and the alternative hypotheses for the supremum, average and exponential mappings.

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:*

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

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 alternatives (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 alternatives 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 the equivalence, 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 the critical values 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[4].

[4]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 as the critical values in Guay (2003) but for $2q-2r$.

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

We propose in this section test statistics robust to the nature of identification as defined by Stock and Wright (2000). Consider the pure structural change case, namely: $\theta = (\beta', \beta')'$. 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 the weakly identified case. In the presence of weak identification, some of the parameters are not consistent so we can't 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 statistics 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 mentioned, 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:

$$GEL_T^R(s) = GEL1_T^R(s) + GEL2_T^R(s) = \frac{2T}{2K_T+1}\widehat{P}(\tilde{\theta}_T(s), \hat{\lambda}_T(\tilde{\theta}_T(s), s), s)$$

where

$$GEL1_T^R(s) = \frac{2[Ts]}{2K+1}\sum_{t=1}^{[Ts]}\frac{\left[\rho(\hat{\lambda}_{1T}(\tilde{\beta}_T, s)' g_{tT}(\tilde{\beta}_T(s))) - \rho(0)\right]}{[Ts]}$$

and

$$GEL2_T^R(s) = \frac{2(T-[Ts])}{2K+1}\sum_{t=[Ts]+1}^{T}\frac{\left[\rho(\hat{\lambda}_{2T}(\tilde{\beta}_T, s)' g_{tT}(\tilde{\beta}_T(s))) - \rho(0)\right]}{T-[Ts]}.$$

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 $\tilde{\theta}_T(s)$[5]. The statistic is defined as:

$$GELM_T^R(s) = GELM1_T^R(s) + GELM2_T^R(s)$$

where

$$GELM1_T^R(s) = \frac{[Ts]}{(2K_T+1)^2}\hat{\lambda}_{1T}(\tilde{\beta}_T(s), s)'\widehat{\Omega}_{1T}(\tilde{\beta}_T(s), s)\hat{\lambda}_{1T}(\tilde{\beta}_T(s), s)$$

and

$$GELM2_T^R(s) = \frac{[T-Ts]}{(2K_T+1)^2}\hat{\lambda}_{2T}(\tilde{\beta}_T(s), s)'\widehat{\Omega}_{2T}(\tilde{\beta}_T(s), s)\hat{\lambda}_{2T}(\tilde{\beta}_T(s), s).$$

We show (see 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 $GEL_T^R(s)$ and $GELM_T^R(s)$ 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:*

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

[5]We can also propose a LR-type test statistic as in Caner (2007) but for the 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 but we do study it the the simulations section.

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

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

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 as those in 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:

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

For the first subsample, we define

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

and for the second

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

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

$$\tilde{\theta}_{K,T}(s) \quad = \quad \arg\min_{\theta\in\Theta} \left(K_{1T}(\beta, s) + K_{2T}(\beta, s)\right)$$

for all $s \in S$ where $K_{1T}(\beta, s) \in R^{r\times 1}$ and $K_{2T}(\beta, s) \in R^{r\times 1}$.

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

$$GELK_T(s) = K_{1T}(\tilde{\beta}_{K,T(s)}, s) + K_{2T}(\tilde{\beta}_{K,T(s)}, s).$$

The GEL framework also allows us to propose an asymptotically equivalent test statistic based on the Lagrange multiplier parameters $\lambda(\cdot)$. By defining the following restricted estimator

$$\tilde{\theta}_{GELMK,T}(s) \quad = \quad \arg\min_{\theta\in\Theta} \left(GELMK1^R_T(\beta, s) + GELMK2^R_T(\beta, s)\right)$$

for all $s \in S$ where

$$GELMK1^R_T(\beta, s) = \frac{[Ts]}{(2K_T+1)^2} \hat{\lambda}_{1T}(\beta, s)'\hat{D}_{1T}(\beta, s)\left(\hat{D}_{1T}(\beta, s)'\hat{\Omega}_{1T}(\beta, s)^{-1}\hat{D}_{1T}(\beta, s)\right)^{-1}\hat{D}_{1T}(\beta, s)'\hat{\lambda}_{1T}(\beta, s)$$

and

$$GELMK2^R_T(\beta, s) = \frac{[T-Ts]}{(2K_T+1)^2} \hat{\lambda}_{2T}(\beta, s)'\hat{D}_{2T}(\beta, s)\left(\hat{D}_{2T}(\beta, s)'\hat{\Omega}_{2T}(\beta, s)^{-1}\hat{D}_{2T}(\beta, s)\right)^{-1}\hat{D}_{2T}(\beta, s)'\hat{\lambda}_{2T}(\beta, s).$$

The statistic is defined as:

$$GELMK^R_T(s) = GELMK1^R_T(\tilde{\beta}_{GELMK,T(s)}, s) + GELMK2^R_T(\tilde{\beta}_{GELMK,T(s)}, s)$$

Theorem 3.5. *The $GELK_T(s)$ and $GELMK^R_T(s)$ 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:*

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

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

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

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 $q \times p$. With weak identification, 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

To examine the finite sample properties of the proposed tests of structural change we use a consumption CAPM environment with CRRA preferences. In this environment, we generate dividend and consumption streams that are used to price stocks as well as a risk-free bond and then to obtain returns for a stock, r_t^s, and for a riskless asset, r_t^f. This environment is very flexible in that we can study different degrees of identification and it has become somewhat of a standard. Using this set-up, Stock and Wright (2000) considered the GMM under weak identification, Wright (2003) proposed a test for detecting a lack of identification (see also the recent contribution by Inoue and Rossi (2008)) and Kleibergen (2005) suggested a framework for testing parameters in the GMM that allows unidentification. In a GEL context, Guggenberger and Smith (2005) studied the properties of GEL estimators with respect to identification in an i.i.d. setting, Otsu (2006) and Guggenberger and Smith (2008) extended their results to the case of dependent data. Finally, and closest to our work, Caner (2007) looked at tests of structural change under the GMM with different degrees of identification.

The design of the experiment follows Tauchen (1986), Kocherlakota (1990), Hansen *et al.* (1996) and the data generating process uses the methods of Tauchen and Hussey (1991)[6]. In particular, their method fits a 16 states Markov chain to the law of motion for consumption growth and dividend growth. This law of motion is calibrated to approximate a $VAR(1)$. Letting δ be the discount factor, γ the coefficient of risk aversion, ι a $G \times 1$ vector of ones, C_t consumption and R_t a $G \times 1$ vector of asset returns we have the following Euler equations

$$h(x_t, \theta) = \delta(C_{t+1}/C_t)^{-\gamma} R_{t+1} - \iota$$

or using the notation of section 2.1

$$g(x_t, \theta) = [\delta(C_{t+1}/C_t)^{-\gamma} R_{t+1} - \iota] \otimes Z_t$$

where Z_t is a K-dimensional vector of instruments and $\theta = (\gamma, \delta)'$. The function $g(x_t, \theta)$ then maps the G Euler equations into $GK \equiv q$ moment conditions. The $VAR(1)$ approximation for the log of consumption growth and of dividend growth is given by the general formulation

$$x_t = A_0 + A_1 x_{t-1} + e_t$$

or

$$\begin{bmatrix} c_t \\ d_t \end{bmatrix} = \begin{bmatrix} 0.021 \\ 0.004 \end{bmatrix} + \begin{bmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{bmatrix} \begin{bmatrix} c_{t-1} \\ d_{t-1} \end{bmatrix} + \begin{bmatrix} e_{ct} \\ e_{dt} \end{bmatrix}$$

with a variance/covariance matrix for the error terms set to

$$\Sigma = \begin{bmatrix} 0.0012 & 0.00177 \\ 0.00177 & 0.014 \end{bmatrix}.$$

For the size properties we study designs that differ in their treatment of $\theta = (\gamma, \delta)'$. We consider the case of one asset, three instruments (designs 3 and 4 in Caner (2007) are not looked at) and one estimated parameter, the coefficient of risk aversion γ. The discount factor δ is set to its true value as it is the most easily identifiable parameter. The designs are presented in Table 1. We considered designs with more instruments/assets or with more estimated parameters but the computational burden increased substantially[7]. The layout of the designs is similar to those found in Stock and Wright (2000) and Caner (2007). For example, designs W1 and W2 focus on weak identification. Designs S1 and S2 consider standard identification (these are the designs 5 and 6 in Caner (2007)). For example, design S1 takes W1 and replaces Σ by 2Σ and design S2 takes W1 and replaces the covariance between e_{ct} and e_{dt} by 0.0039. Samples of sizes 250 (first row in Tables 2 to 4) and 500 observations (second row) are studied while the number of Monte Carlo replications is set to 1000[8]. The results are presented in Tables 2 to 4. We use a nominal size of 5% and a trimming rule with $s = 0.15$.

[6] We thank Mehmet Caner for providing us with his computer code for the data generating process.

[7] A small scale study revealed that results deteriorate as p or q increased. Those results are available from the authors upon request. See for instance Hansen *et al.* (2008) and Cattaneo *et al.* (2015) for estimation and inference problems with many instruments and regressors

[8] We only report results with 250 observations for the cases of mixed identification and different break location under the alternative.

Power properties are studied by considering a partial break in the parameters (only one parameter is affected). The coefficient of risk aversion is allowed to switch at mid sample (we also report in Table 4 results for a break at one fourth of the sample size). Unadjusted power properties for designs S2 and W2 are reported in Table 3. For S2, the value of γ changes from 1.3 to 3 while for W2 it moves from 13.7 to 9. We also consider the case were we move from standard identification (S1) to weak identification as well as allowing for a change in γ (a change from 1.3 to 3). This 'mixed' case is reported in the lower panel of Table 3. As in Stock and Wright (2000), the errors have a martingale difference property so that smoothing is not required ($k = 0$ in all experiments). The effects of smoothing are under consideration in a separate paper. Before discussing our results we provide a justification on our choice of GEL estimator.

Our choice of GEL estimator is based on prior theoretical research and simulation results. The exponential tilting estimator is found by setting $\rho(\phi) = -\exp(\phi)$, the empirical likelihood estimator considers $\rho(\phi) = \ln(1-\phi)$ while the continuously updated estimator is obtained using the quadratic function $\rho(\phi) = -(1+\phi)^2/2$. On the theoretical side, Newey and Smith (2004) showed that the EL estimator removes asymptotically a bias component that the ET and CU estimators do not remove. Indeed, EL uses an efficient weighted average in the estimation of the optimal weighting matrix in contrast to ET and CU. This efficient weighted average removes the asymptotic bias resulting from the third moments of the moment conditions. This result also holds in the time series context for the EL estimator using the truncated kernel (see Anatolyev, 2005). However, the small sample properties of the EL estimator can suffer from misspecification due to the computation of the implied probabilities. Imbens, Spady and Johnson (1998) mentioned that the ET estimators are more robust to misspecification. Schennach (2007) formalized this issue by showing that the EL estimators' asymptotic properties can suffer greatly, even with a small amount of misspecification. The problem originates from the computation of $\rho_1(\cdot)$ appearing in the first order conditions which can become unbounded under misspecification. In contrast, the ET estimator does not suffer from this problem.

On the simulation side, Guggenberger (2008) studied the properties of the ET, EL and CU estimators in linear models and found that their properties are almost identical. Caner (2010) found that tests based on ET estimators performed well in simulation studies involving weak instruments. Further, in a previous version of our paper, we looked at the exponential tilting, empirical likelihood and continuously updated estimators as little was known on their finite sample properties and computational performances. The simulated environment was identical to the one found in Guay and Lamarche (2012) who proposed test statistics to detect structural change that are based on the estimated weights of a GEL problem. This environment, also used by Ghysels *et al.* (1997) and by Hall and Sen (1999), consists of a time series characterized by an autoregressive process of order one. Overall, tests computed with exponential tilting performed best and were much less sensitive to misspecifications than tests obtained with either empirical likelihood or the continuously updated estimators[9]. In a related paper (Guay and Lamarche, 2012), we use the theoretical results derived here to propose alternative structural change tests in the GEL setting but based on implied probabilities.

Two groups of tests are studied in our simulation experiments. The first group contains tests that are valid

[9]The results for these simulations are available upon request.

under standard identification (upper panels in Tables 2 to 4) while the second contains tests robust to weak identification (lower panels in Tables 2 to 4). For the first group of tests, we consider Andrews-type tests aimed at a structural change in the parameters ($GELW$, $GELM$, $GELR$, $GELIPS^I$) as well as Hall and Sen-type tests targeting a violation in the moment conditions themselves ($GELO$, $GELMO$, $GELIPS^O$). The tests with a GMM label are computed using a continuously updated GMM procedure for comparison purposes. All other tests are GEL-based, including the tests based on implied probabilities proposed by Guay and Lamarche (2012), $GELIPS^I$ and $GELIPS^O$. All robust tests are valid for changes in the parameters only and the experiments are such that only parameter changes are allowed. For this reason, the O-type tests should have worst finite sample properties under H_1. GMM-based tests performed consistently worst than all other tests and are not discussed in many details.

For the case of weak identification, the GEL^R test is based on the renormalized objective functions of each subsample while an asymptotically equivalent test to it is the $GELM^R$, which is based on the Lagrange multipliers of each subsample. These tests have asymptotic bounds that depend on the number of moment conditions which can be quite high. These tests are similar to the $supS$ test of Caner (2007), S^{GMM} in our tables, but for the GEL context. Although two versions of the $GELK$ tests were computed in the simulations, we only report the results for the $GELK$ tests computed using the covariance matrices displayed after Theorem 3.4. The results with covariance matrices computed using implied probabilities had more distortions (see Guay and Lamarche (2012) for more on this subject). We also report the K test computed under GMM. We can then evaluate the performance of the widely used LR under weak identification as well as the performance of different estimation techniques. The LR-type test of Caner (2007), computed for the GEL framework, is labeled as $GELRC^R$. As in Caner (2007), this version of the LR test does not perform well at all.

Under standard identification, designs S1 and S2 (upper panel of Table 2), the classical tests often used empirical applications ($GELW$, $GELM$, O, $GELO$) record empirical sizes that are relatively close to the nominal size of 5%. The test based on implied probabilities aimed at instability in the parameters overrejected substantially while the $GELIPS^O$ test performed quite well. Tests that are robust to weak identification have a tendency to underrejects quite badly and we can see that this problem is not solved when the sample size is increased. The GEL-based tests GEL^R and $GELIPS^R$ show, by far, the best results.

Designs W1 and W2, reflecting the case of weak identification (lower panel of Table 2), can yield considerable size distortions for the classical tests. For example, nominal sizes for the $GELM$ test increase (under W2) while those for the O decrease (under W1). Interestingly, the $GELW$ and the O tests are not affected significantly by the identification assumption. The same can be said about the tests robust to this assumption. The GEL^R, $GELK$ and $GELIPS^R$ tests perform very well under weak identification, much better than those suggested by Caner (2007), the S^{GMM} and K^{GMM} tests.

When the coefficient of risk aversion (the parameter that is not well identified) is allowed to change at mid sample, the rejection frequencies increase (as they should) for all tests (Table 3). Because we are presenting size-unadjusted frequencies the reader must be careful when looking at the power of the tests that overrejected greatly under the null hypothesis (the $GELIPS^I$ test for example). One should also note that, under standard identification (upper panel of Table 3), the frequencies for the tests valid for changes in the overidentifying restrictions (all the O-type tests) are much smaller than those of tests that target instability in the parameter. The classical tests (those targeting instability in the parameters) have the most power, followed by the robust tests and the O-tests. For all, except the O-tests, increasing the sample size increases power. Once again, the tests computed under GEL and those robust to identification perform best.

Allowing for weak identification substantially distorts the rejection frequencies, increasing them for the most part and by a very large amount for the O-tests (middle panel of Table 3). A few interesting features come out. First, we see that the robust tests have higher power than the standard tests with the $GELIPS^R$ test performing very well. Second, we see that the robust test $GELK$ is not affected very much by identification with rejection frequencies (for a sample size of 250) going from 0.51 to 0.49. This is a very nice feature of a test statistic as it can be trusted no matter the type of identification suspected. Finally, when both the type of identification and the value of the parameter change (lower panel of Table 3) all tests have some difficulty disentangling the change in the parameter properly. This finding comes as no surprise since the tests considered in this paper are designed to have power against parameter change first. However, the GEL-based tests and those robust to identification did very well.

Our last experiment considers the location of the break. We move the break from mid sample to a break located earlier in the sample (quarter of the sample, Table 4). We see that the robust tests seem to pick up the break earlier in the sample at least as well as a break located mid sample, and so regardless of the type of identification. On the other hand, the standard tests are more sensitive to the location of the break.

5 Conclusion

In this paper we have studied 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 is important to study structural change tests for these methods of estimation.

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 to 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 as 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 as the one derived by Hall and Sen (1999). We also presented stability tests, computed in the GEL context, that are robust to weak identification.

Our simulation study, based on the widely used design of Stock and Wright (2000) and makes use of the exponential tilting estimation procedure. We found that the GEL-based robust tests (for example our GEL^R, $GELIPS^R$ and $GELK$ tests) performed well in terms of: 1) the presence and location of a structural change and 2) the nature of identification. These test statistics should then be added to the Pearson-type statistics based on implied probabilities to detect structural change presented by Guay and Lamarche (2012) to complement the specification and testing arsenal of the practitioners.

6 Appendix

6.1 Assumptions

We[10] consider triangular arrays because they are required to derive asymptotic results under the Pitman drift alternatives. Define $\mathbf{X}$ to be the domain of $g(\cdot,\theta)$ which includes the support of $x_{T,t}, \forall t, \forall T$. 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 $\mu_{T,t}$ denote the distribution of $x_{T,t}$ and let $\bar{\mu}_T = (1/T)\sum_{t=1}^{T}\mu_{T,t}$. 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. *$\{x_{T,t} : t \leq T, T \geq 1\}$ is a triangular array of $\mathbf{X}$-valued rv's that is L^0-near epoch dependent (NED) on a strong mixing base $\{Y_{T,t} : t = \ldots, 0, 1, \ldots; T \geq 1\}$, where $\mathbf{X}$ is a Borel subset of R^k, and $\{\mu_{T,t} : T \geq 1\}$ is tight on $\mathbf{X}$*[11].

Define the smoothed moment conditions as:[12]

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

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):

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

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 $\eta > 1$*[13].

Assumption 6.3. *For some $d > \max\left(2, \frac{2\eta}{\eta-1}\right)$, $\{g(x_{T,t},\beta_0,\delta_0) : t \leq T, T \geq 1\}$ 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 $\{Y_{T,t} : t = \ldots, 0, 1, \ldots; T \geq 1\}$, of size $-d/(d-2)$ and $\sup\|g(x_{T,t},\beta_0,\delta_0)\|^d < \infty$.*

Assumption 6.4. *$Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{Ts} g(x_{T,t},\beta_0,\delta_0)\right) \to s\Omega \;\; \forall s \in [0,1]$ 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 1.1 and 1.2). In the following, x_t is used to denote $x_{T,t}$ for notational simplicity.

[10] An additional document contains lemmas and a longer version of the proofs. It can be found at `http://coffee.econ.brocku.ca/jfl/research`.

[11] 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).

[12] Note here that g_{tT} denotes the smoothed moment conditions and $x_{T,t}$ a triangular array of random variables.

[13] This assumption is slightly different than that in Smith, 2011 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 $(\beta, \delta, x) \in B \times \Delta \times \zeta$ for all compact sets $\zeta \subset \mathbf{X}$.*

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, 2011 and Smith, 2011 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 1.3).

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, 2011.

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. $\sup_{s\in S} \|\widehat{\Omega}_{iT}(s) - \Omega\| \xrightarrow{p} 0$ *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, 2011. To respect these conditions, Assumption 6.3 can be replaced by the following assumption:

Assumption 6.3′. $\{g(x_t, \beta_0, \delta_0) : t \leq T, T \geq 1\}$ 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 $sup_{t\leq T, T\geq 1} E\|g(x_t, \beta_0, \delta_0)\|^d < \infty$ 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

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

and

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

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 (β,δ), $\forall(\beta,\delta)\in B_0\times\Delta_0$ $\forall x\in\mathbf{X_0}\subset\mathbf{X}$ for a Borel measurable set $\mathbf{X_0}$ that satisfies $P\left(x_t\in\mathbf{X_0}\right)=1\forall t\leq T,T\geq1$, $g(x,\beta,\delta)$ 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 $Eg_{tT}(\beta,\delta)=Eg(x_t,\beta,\delta)+o_p(1)$ and $E\partial g_{tT}(\beta_0,\delta_0)/\partial\left(\beta',\delta'\right)=E\partial g(x_t,\beta_0,\delta_0)/\partial\left(\beta',\delta'\right)+o_p(1)$ 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 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, 2011 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, 2011.

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, 2011, 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, 2011, Assumptions 6.5, and 6.2, Lemmas 1.3 and 1.5, we can show $\sup_{s\in S}\|\hat{g}_T(\hat{\theta}_T(s),s)\|=O_p(T^{-1/2})$. By the result that $\sup_{s\in S}\|\hat{g}_T(\hat{\theta}_T(s),s\|=O_p(T^{-1/2})$ we have $\sup_{s\in S}\hat{g}_T(\hat{\theta}_T(s),s)\xrightarrow{p}0$. By Lemma 1.4, $\sup_{s\in S}\sup_{\theta\in\Theta}\|\hat{g}_T(\theta,s)-g(\theta,s)\|\xrightarrow{p}0$ and $\tilde{g}(\beta,\delta)$ 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 $\tilde{g}(\beta,\delta)=0$ 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 $\sup_{s\in S}\|\hat{\theta}_T(s)-\theta_0\|\xrightarrow{p}0$.

The proof to show that $\sup_{s\in S}\|\hat{\lambda}_T(s)\|=O_p\left(\left(T/(2K+1)^2\right)^{-1/2}\right)$ and $\sup_{s\in S}\|\hat{\lambda}_T(s)\|\xrightarrow{p}0$ is similar to the one for Theorem 2.2 in Smith, 2011 but uniformly for $s\in S$.

Proof of Theorem 2.2

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:

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

where

$$\bar{M}(s) = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} 0 & \bar{M}_{12}(s) \\ \bar{M}_{21}(s) & \bar{M}_{22}(s) \end{bmatrix}$$

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

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

By Lemma 1.3; $\sup_{s\in S}\sup_{1\le t\le T} |\hat{\lambda}_T(s)' g_{tT}(\hat{\theta}_T(s), s)| \xrightarrow{p} 0$ 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 $\sup_{s\in S}\max_{1\le t\le T} |\rho_2\left(\hat{\lambda}_T(s)' g_{tT}(\hat{\theta}_T(s), s)\right) - \rho_2(0)| \xrightarrow{p} 0$. To show that

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

and

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

it remains to show that

$$\sup_{s\in S}\left\|\frac{1}{T}\sum_{t=1}^{T} G_{tT}(\bar{\theta}_T(s), s) - G(s)\right\| \xrightarrow{p} 0 \tag{6}$$

and

$$\sup_{s\in S}\left\|\frac{1}{T}\sum_{t=1}^{T}G_{tT}(\hat{\theta}_T(s),s)-G(s)\right\|\xrightarrow{p}0. \tag{7}$$

This can be easily shown using arguments in Andrews (1993).

Moreover, 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

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

By Assumption 6.12, this gives

$$M(s)^{-1}=\begin{bmatrix}-\Sigma(s) & H(s)\\ H(s)' & P(s)\end{bmatrix}$$

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 1.1, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}(\theta_0,s)\Rightarrow J(s)$ 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 1.2 and the proof of Theorem 2.2.

Proof of Theorem 3.1

Using results derived above, we get for terms in the $GELW_T(s)$ statistic:

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

and terms in the LM statistic:

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

The asymptotic distributions for the $GELW_T(s)$ and $GELM_T(s)$ under the null can then be directly derived using the expressions above from similar arguments than in the proof of Theorem 3 in Andrews (1993). The asymptotic distribution under the alternative is a direct implication of Theorem 2.3. For the $GELR_T(s)$ 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 $GELR_T(s)$ 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 $\sum_{t=1}^{[Ts]} g(x_t, \hat{\beta}_{1T}(s))$ and $\sum_{t=[Ts]+1}^{T} g_{tT}(\hat{\beta}_{2T}(s))$ with $\sum_{t=[Ts]+1}^{T} g(x_t, \hat{\beta}_{2T}(s))$ is a direct implication of the Lemmas 1.1 and 1.2 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 proofs for Theorems 2.2 and 2.3 and subsection A.2 in Hall and Sen (1999).

Second, for the statistic $GELO_T(s)$, as in the proof of Theorem 3.2, we can show that:

$$\frac{2[Ts]}{2K_T+1}\sum_{t=1}^{[Ts]}\frac{\left[\rho(\hat{\lambda}_{1T}(\hat{\beta}_{1T}(s), s)'g_{tT}(\hat{\beta}_{1T}(s))) - \rho(0)\right]}{[Ts]} = GELO1_T(s) + o_p(1)$$

and

$$\frac{2(T-[Ts])}{2K_T+1}\sum_{t=[Ts]+1}^{T}\frac{\left[\rho(\hat{\lambda}_{2T}(\hat{\beta}_{2T}(s), s)'g_{tT}(\hat{\beta}_{2T}(s))) - \rho(0)\right]}{T-[Ts]} = GELO2_T(s) + o_p(1).$$

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

Finally, for the statistic $GELMO_T(s)$, 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 1.1, the result under the null follows. Lemma 1.2 yields the asymptotic distribution under the alternative. The equivalence for the statistic $GELM_T^R(\tilde{\theta}_T(s),s)$ 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 $sup_{t\leq T,T\geq 1}E\|G\left(x_{Tt},\beta_0\right)\|^d<$

∞ 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

$$\hat{\Omega}_{Gg,1T}(\beta_0, s) = \frac{2K+1}{T}\sum_{t=1}^{[Ts]} vec\left(G_{tT}(\beta_0)\right) g_{tT}(\beta_0)' \xrightarrow{p} s\Omega_{Gg}(\beta_0), \tag{8}$$

$$\hat{\Omega}_{Gg,2T}(\beta_0, s) = \frac{2K+1}{T}\sum_{t=[Ts]+1}^{T} vec\left(G_{tT}(\beta_0)\right) g_{tT}(\theta_0)' \xrightarrow{p} (1-s)\Omega_{Gg}(\beta_0), \tag{9}$$

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 1.1 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:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} vec\left(G_{tT}(\beta_0) - EG_{tT}(\beta_0)\right) \Rightarrow \Omega_{GG}(\beta_0)^{1/2} B_{qr}(s) \tag{10}$$

where $B_{qr}(s)$ is a qr-dimensional vector of standard Brownian motion. Using Lemma 1.1, this yields for the whole vector $(g_{tT}(\beta_0)', (vec(G_{tT}(\beta_0) - EG_{tT}(\beta_0)))')'$

$$T^{-1/2}\sum_{t=1}^{[Ts]} (g_{tT}(\beta_0)', (vec(G_{tT}(\beta_0) - EG_{tT}(\beta_0)))')' \Rightarrow \Sigma(\beta_0)^{1/2} B_{q+qr}(s) \tag{11}$$

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

We also need the following assumptions:

Assumption 6.13. *Suppose Assumption 6.9 but for $\partial g(x_t,\beta)/\partial\beta_i$ for $i = 1, \ldots, r$.*

Let $\hat{D}_{1T}(\beta_0, s) = \left[\hat{D}_{1,1T}(\beta_0, s), \hat{D}_{2,1T}(\beta_0, s), \ldots, \hat{D}_{r,1T}(\beta_0, s)\right]$ 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 $i = 1, \ldots, p$ 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 $\sup_{s\in S}\max_{1\leq t\leq T}|\rho_2(\hat{\lambda}_{iT}(\beta_0,s)'g_{tT}(\beta_0)) - \rho_2(0)| \xrightarrow{p} 0$ for $i = 1, 2$.

Using (8), (9), (10), (11), Lemma 1.1 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.

Since $\tilde{\theta}_{K,T}(s) = \left(\tilde{\beta}_{K,T}(s)', \tilde{\beta}_{K,T}(s)'\right)'$ minimize the objective function

$$GELK_T(s) = K_{1T}(\tilde{\beta}_{K,T(s)}, s) + K_{2T}(\tilde{\beta}_{K,T(s)}, s) \leq K_{1T}(\beta_0,s) + K_{2T}(\beta_0,s)$$

for all $s\in S$ and all T. The result follows directly under the null. The derivation under the alternative can be easily obtained. The proof for $GELMK_T^R(s)$ is straightforward considering that

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

and

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

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 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

Caner, Mehmet (2010) 'Exponential tilting and weak instruments: estimation and testing.' *Oxford Bulletin of Economics and Statistics* 72, 307–325

Cattaneo, M.D., M. Jansson, and W.K. Newey (2015) 'Alternative asymptotics and the partially linear model with many regressors.' Working Paper, Department of Economics, MIT

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

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

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

Guay, A., and J-F Lamarche (2012) 'Structural change tests based on implied probabilities for gel criteria.' *Econometric Theory* 28, 1186–1228. The tables of critical values can be found at http://coffee.econ.brocku.ca/jfl/iprob_cvalues.pdf.

Guggenberger, P., and R. J. Smith (2005) 'Generalized empirical likelihood estimators and tests under partial, weak and strong identification.' *Econometric Theory* pp. 667–709

_ (2008) 'Generalized empirical likelihood tests in time series models with potential identification failure.' *Journal of Econometrics* 142, 134–161

Guggenberger, Patrick (2008) 'Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator.' *Econometric Reviews* 26, 526–541

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

Hall, A.R., Y. Li, and C.D. Orme (2011) 'Testing for structural instability in moment restriction models: an info-metric approach.' Working Paper

Hansen, C.B., J. A. Hausman, and W.K. Newey (2008) 'Estimation with many instrumental variables.' *Journal of Business and Economic Statistics* 26, 398–422

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

Imbens, G.W., R.H. Spady, and P. Johnson (1998) 'Information theoretic approaches to inference in moment condition models.' *Econometrica* 66, 333–357

Inoue, A., and B. Rossi (2008) 'Testing for weak identification in possibly nonlinear models.' Working Paper

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

Kocherlakota, N. (1990) 'On tests of representative consumer asset pricing models.' *Journal of Monetary Economics* 25, 43–48

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

Schennach, S.M. (2007) 'Point estimation with exponetially tilted empirical likelihood.' *Annals of Statistics* 35, 634–672

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

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

Tauchen, George (1986) 'Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data.' *Journal of Business and Economic Statistics* 4, 397–425

Tauchen, George, and R. Hussey (1991) 'Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models.' *Econometrica* 59, 371–396

Wright, J.H. (2003) 'Detecting lack of identification in GMM.' *Econometric Theory* 19, 322–330

Table 1: Designs for the data generating processes

Design	(γ, δ)	Assets R_t	Instruments Z_t	Moment conditions q
W1	(1.3,0.97)	r_t^s	$1, r_{t-1}^s, c_{t-1}$	3
W2	(13.7,1.139)	r_t^s	$1, r_{t-1}^s, c_{t-1}$	3

Table 2: Rejection frequencies under the null hypothesis

Standard identification

DGP	$GELW$	W^{GMM}	$GELM$	$GELR$	$GELIPS^I$	O	O^{GMM}	$GELO$	$GELMO$	$GELIPS^O$
$S1$	0.0767	0.0800	0.0633	0.1400	0.2267	0.0500	0.0033	0.0900	0.0200	0.0467
	0.0700	0.0367	0.0600	0.0967	0.1933	0.0900	0.0167	0.0933	0.0300	0.0633
$S2$	0.0700	0.0700	0.0967	0.2900	0.3233	0.0733	0.0033	0.0467	0.0267	0.0367
	0.0600	0.0800	0.0933	0.2900	0.3300	0.1100	0.0267	0.0267	0.0333	0.0367

DGP	$GELM^R$	GEL^R	$GELRC^R$	$GELIPS^R$	LR^{GMM}	S^{GMM}	$GELK$	K^{GMM}
$S1$	0.0333	0.0900	0.0133	0.0700	0.0000	0.0133	0.0600	0.0100
	0.0333	0.0833	0.0100	0.0767	0.0000	0.0233	0.0233	0.0033
$S2$	0.0633	0.1500	0.0267	0.1200	0.0000	0.0100	0.0667	0.0067
	0.0533	0.1000	0.0300	0.0900	0.0000	0.0433	0.0567	0.0200

Weak identification

DGP	$GELW$	W^{GMM}	$GELM$	$GELR$	$GELIPS^I$	O	O^{GMM}	$GELO$	$GELMO$	$GELIPS^O$
$W1$	0.0800	0.0800	0.0733	0.1867	0.3200	0.0567	0.0100	0.0833	0.0367	0.0433
	0.1000	0.0833	0.0800	0.1433	0.2367	0.0600	0.0167	0.0567	0.0067	0.0300
$W2$	0.0833	0.0667	0.2033	0.1633	0.2267	0.0833	0.0100	0.1033	0.0467	0.0500
	0.0900	0.0700	0.1767	0.1733	0.2600	0.1167	0.0300	0.0867	0.0333	0.0567

DGP	$GELM^R$	GEL^R	$GELRC^R$	$GELIPS^R$	LR^{GMM}	S^{GMM}	$GELK$	K^{GMM}
$W1$	0.0467	0.1233	0.0133	0.1000	0.0000	0.0100	0.0600	0.0100
	0.0300	0.0800	0.0167	0.0633	0.0000	0.0133	0.0733	0.0133
$W2$	0.0600	0.1033	0.0100	0.1067	0.0100	0.0133	0.0933	0.0167
	0.0567	0.1000	0.0067	0.1033	0.0000	0.0133	0.1100	0.0167

Table 3: Rejection frequencies under the alternative hypothesis (break located at mid point of sample)

Standard identification

DGP	$GELW$	W^{GMM}	$GELM$	$GELR$	$GELIPS^I$	O	O^{GMM}	$GELO$	$GELMO$	$GELIPS^O$
$S2$	0.6100	0.5800	0.4633	0.7667	0.8200	0.3267	0.0100	0.2800	0.1667	0.2667
	0.8733	0.8500	0.6200	0.9167	0.9500	0.4333	0.0333	0.3000	0.2400	0.2567

	$GELM^R$	GEL^R	$GELRC^R$	$GELIPS^R$	LR^{GMM}	S^{GMM}	$GELK$	K^{GMM}
$S2$	0.3833	0.5833	0.3500	0.5367	0.0367	0.1733	0.5100	0.2600
	0.6967	0.7733	0.6367	0.7433	0.1967	0.4200	0.7267	0.6233

Weak identification

DGP	$GELW$	W^{GMM}	$GELM$	$GELR$	$GELIPS^I$	O	O^{GMM}	$GELO$	$GELMO$	$GELIPS^O$
$W2$	0.7500	0.6667	0.6067	0.9933	1.0000	1.0000	0.9867	1.0000	0.9833	1.0000
	0.9467	0.9000	0.8000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	$GELM^R$	GEL^R	$GELRC^R$	$GELIPS^R$	LR^{GMM}	S^{GMM}	$GELK$	K^{GMM}
$W2$	1.0000	1.0000	0.9633	1.0000	0.0367	0.9733	0.4867	0.2267
	1.0000	1.0000	1.0000	1.0000	0.0200	1.0000	0.4900	0.5033

Mixed identification

DGP	$GELW$	W^{GMM}	$GELM$	$GELR$	$GELIPS^I$	O	O^{GMM}	$GELO$	$GELMO$	$GELIPS^O$
$S1/W2$	0.4300	0.3200	0.3033	0.5133	0.6700	0.2267	0.0100	0.3333	0.1600	0.2067

	$GELM^R$	GEL^R	$GELRC^R$	$GELIPS^R$	LR^{GMM}	S^{GMM}	$GELK$	K^{GMM}
$S1/W2$	0.3367	0.4867	0.2567	0.4600	0.0200	0.0767	0.4167	0.1833

Table 4: Rejection frequencies under the alternative hypothesis (break located at quarter of sample)

Standard identification

DGP	$GELW$	W^{GMM}	$GELM$	$GELR$	$GELIPS^I$	O	O^{GMM}	$GELO$	$GELMO$	$GELIPS^O$
$S2$	0.3833	0.2567	0.4500	0.7267	0.8533	0.3867	0.0100	0.3467	0.2667	0.3433
	$GELM^R$	GEL^R	$GELRC^R$	$GELIPS^R$	LR^{GMM}	S^{GMM}	$GELK$	K^{GMM}		
$S2$	0.6067	0.7433	0.5600	0.7200	0.0467	0.1967	0.5133	0.2700		

Weak identification

DGP	$GELW$	W^{GMM}	$GELM$	$GELR$	$GELIPS^I$	O	O^{GMM}	$GELO$	$GELMO$	$GELIPS^O$
$W2$	0.7867	0.6200	0.6800	0.9800	1.0000	1.0000	0.9233	1.0000	1.0000	1.0000
	$GELM^R$	GEL^R	$GELRC^R$	$GELIPS^R$	LR^{GMM}	S^{GMM}	$GELK$	K^{GMM}		
$W2$	1.0000	1.0000	0.9800	1.0000	0.0200	0.9167	0.5033	0.3267		