Econometric Theory, **28**, 2012, 1186–1228.
doi:10.1017/S0266466612000114

STRUCTURAL CHANGE TESTS BASED ON IMPLIED PROBABILITIES FOR GEL CRITERIA

ALAIN GUAY
Université du Québec à Montréal

JEAN-FRANÇOIS LAMARCHE
Brock University

This paper proposes Pearson-type statistics based on implied probabilities to detect structural change. The class of generalized empirical likelihood estimators (see Smith, 1997, *The Economic Journal* 107, 503–519) assigns a set of implied probabilities to each observation such that moment conditions are satisfied. The proposed test statistics for structural change are based on the information content in these implied probabilities. We consider cases of structural change with unknown breakpoint that can occur in the parameters of interest or in the overidentifying restrictions used to estimate these parameters. We also propose a structural change test based on implied probabilities that is robust to weak identification or cases in which parameters are completely unidentified. The test statistics considered here have competitive size and power properties. Moreover, they are computed in a single step, which eliminates the need to compute the weighting matrix required for generalized method of moments estimation.

1. INTRODUCTION

This paper proposes structural change tests based on implied probabilities resulting from estimation methods based on unconditional moment restrictions. The generalized method of moments (GMM) is the standard method to estimate parameters through moment restrictions. Monte Carlo results reveal, however, that GMM estimators may be seriously biased in small samples.[1] Recently, a number of alternative estimators to GMM have been proposed. Hansen, Heaton, and Yaron (1996) suggested the continuous updating estimator (CUE), which shares the same objective function as GMM but with a weighting matrix that depends on

A previous version, October 2008, of this paper was originally circulated under the title "The Information Content of Implied Probabilities to Detect Structural Change." Guay gratefully acknowledges financial support of le Fonds québécois de la recherche sur la société et la culture (FQRSC). Both authors are thankful for the financial support of the Social Sciences and Humanities Research Council of Canada (SSHRC). We also are grateful to Chuan Goh for his comments and also for the comments of participants at the CIREQ Conference on Generalized Method of Moments, Montréal, 2007; ESEM Budapest 2007; and CEA meeting 2007. Finally, we thank the editor, co-editor, and two referees for their time and very helpful suggestions that helped improved the paper. Address correspondence to Alain Guay, Département de sciences économiques, Université du Québec à Montréal, Québec, Canada; e-mail: guay.alain@uqam.ca.

 © Cambridge University Press 2012

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

the parameters of interest. The empirical likelihood (EL) (see Qin and Lawless, 1994) and the exponential tilting (ET) estimators (see Kitamura and Stutzer, 1997) have also been proposed. Kitamura (2001) showed that tests for the validity of moment restrictions based on EL have optimal properties in terms of large deviations. In particular, EL-based tests are shown to be more powerful than other standard tests. These alternative estimators are special cases of the generalized empirical likelihood (GEL) class considered by Smith (1997) and may be shown to be based on the Cressie and Read (1984) family of power divergence criteria. Newey and Smith (2004) showed (in an i.i.d. setting) that, although estimators based on GMM, EL, ET, or continuous updating have the same first-order asymptotic distribution, they have different higher-order asymptotic properties. Among their findings, it is shown that the expression for the second-order asymptotic bias of GEL has fewer components than that of GMM (with EL having the fewest). Anatolyev (2005) extended the Newey and Smith setting to allow for weakly dependent data correlation and show that the asymptotic bias of smoothed GEL estimators is less than that of GMM estimators, especially with many moment conditions.

GEL estimators assign a probability to each observation such that the moment conditions are satisfied (see Smith, 2004). This resulting empirical measure is called implied probabilities. An interpretation of the implied probabilities is the following: Less weight is assigned to an observation for which the moment restrictions are not satisfied at the estimated values of the parameters and more weight to an observation for which the moment restrictions are satisfied. As suggested by Back and Brown (1993), implied probabilities may then provide a useful diagnostic device for model specification. In particular, implied probabilities may contain interesting information to detect instability in the sample. Consequently, we propose to use these weights in detecting an unknown structural change in the model. Antoine, Bonnal, and Renault (2007) use the weights given by the implied probabilities to propose a three-step estimator with higher-order asymptotic properties equivalent to those of EL. Schennach (2007) also discusses the use of these weights in the context of model misspecification. Ramalho and Smith (2005) considered Pearson-type test statistics (statistics based on the difference between restricted and unrestricted estimators of the weights) for the validity of moment restrictions and parametric restrictions using implied probabilities.

The proposed test statistics to detect structural change are based on different measures of the discrepancy between these implied probabilities and the unconstrained empirical probabilities $\frac{1}{T}$ (T is the sample size). The test statistics fall into three categories: (1) general structural change tests to detect instability in the identifying restrictions and in the overidentifying restrictions; (2) structural change tests specifically designed to detect instability in the identifying restrictions (see, for example, Andrews, 1993; and (3) structural change tests to detect instability in the overidentifying restrictions (see, for example, Hall and Sen, 1999). As an additional contribution of this paper, we propose a boundedly pivotal structural change test based on implied probabilities robust to cases of weak identification or cases in which parameters are completely unidentified. In particular,

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

our derivation allows us to examine the power of the test under different assumptions about identification, and so for general alternatives about parameters instability. The asymptotic distributions of the test statistics are derived under the null hypothesis and under the alternative of an unknown breakpoint. In a simulation study, we find that targeted tests based on implied probabilities performed very well if the structural change corresponds to the target. That is, if instability is present in the identifying restrictions or in the overidentifying restrictions, then the targeted tests have competitive size and power properties compared to more standard structural change tests for GEL criteria (see Guay and Lamarche, 2009). Overall, the test statistics based on implied probabilities considered in this paper have a nice intuitive appeal, are based on an estimation method that has been shown to have better higher-order asymptotic properties, and are computed in a single step that eliminates the need to compute the weighting matrix required to perform alternative structural change tests in GEL or GMM settings. This weighting matrix is often linked to the poor finite sample properties of GMM-based estimators and test statistics.

The paper is organized as follows. A discussion on GEL estimators is presented in Section 2. Section 3 presents formally the full-sample and partial-sample GEL estimators. Section 4 presents the test statistics proposed based on the implied probabilities. The simulation results are in Section 5, while the proofs are in the Appendix.

2. DISCUSSION OF GEL ESTIMATORS

In this section we present the estimators used in this paper. We start with an entropy-based formulation of the problem, which puts emphasis on the informational content of the estimated weights. We then move to the more recent GEL formulation (see Newey and Smith, 2004; Smith, 2004).

We consider a triangular array of random variables $\{x_{Tt} : 1 \leq t \leq T, T \geq 1\}$. Triangular arrays are considered because they are required to examine the results under local alternatives. For notational simplicity x_t refers to x_{Tt} hereafter. Suppose we have a $q \times 1$ vector function of the data, $g(x_t, \theta)$, which depends on some unknown p-vector of parameters $\theta \in \Theta$ with $\Theta \subset R^p$, and that in the population their expected value is 0, namely

$$\mathbb{E}[g(x_t, \theta_0)] = 0.$$

We study the overidentifying case with $q > p$. Now letting the T-vector of *explicit* weights (explicit because the weights are used directly in estimation as opposed to the *implicit* weights obtained via GMM estimation, as explained by Back and Brown, 1993) be $\{\pi_t : 1 \leq t \leq T, T \geq 1\}$, we can recast the population moment conditions as

$$\mathbb{E}_\pi [g(x_t, \theta_0)] = 0.$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

The vector π is determined by finding the most probable data distribution of the outcomes given the data. We can think of π as containing information on the content of the moment conditions. Therefore, $g(x,\theta)$ is viewed as a message. That is, when π is small, the message is informative and vice versa. This relation is summarized by the function $f(\pi) = -\ln \pi$. The average information is then

$$S(\pi) \equiv \mathbb{E}_\pi f(\pi) = -\sum_{t=1}^{T} \pi_t \ln \pi_t.$$

In this case, $S(\pi)$ can be interpreted as the entropy measure of Shannon (1948), and it captures the degree of uncertainty in the distribution π with respect to whether or not the distribution is concentrated or dispersed (see Mittelhammer, Judge, and Miller, 2000, for a fuller discussion). The vector π is obtained by maximizing entropy,

$$\max S(\pi) = -\sum_{t=1}^{T} \pi_t \ln \pi_t,$$

subject to $\sum_{t=1}^{T} \pi_t = 1$ and $\sum_{t=1}^{T} \pi_t g(x_t,\theta) = 0$.

With no second constraint, we get $\pi_t = 1/T, \forall t$, the maximally uninformative uniform distribution, while with both constraints, we want to choose π_t to be as maximally uninformative as the moment conditions will allow. We do not want to assert more about the distribution than is known via the moment conditions. In this sense, the probabilities make use of all the information that is available, and nothing more. In particular, we focus on detecting a structural change in the moment conditions. With no structural change, the weights will fluctuate around $1/T$; otherwise, the entropy formulation will attempt to reduce the weight on the observation characterized by the change, and at the same time put more weights on the remaining observations so as to make $S(\pi)$ as large as possible.

We can transpose this formulation using the Kullback and Leibler (1951) information criterion (KLIC), which measures the discrepancy between two distributions, p_t and π_t. If the subject distribution is π_t and the reference distribution is $p_t = 1/T, \forall t$, we have

$$KLIC(\pi, p = 1/T) = \sum_{t=1}^{T} \pi_t [\ln(\pi_t) - \ln(1/T)] = \sum_{t=1}^{T} \pi_t \ln(\pi_t) + \ln(T).$$

Therefore, maximizing entropy is equivalent to minimizing the $KLIC(\pi, p = 1/T)$. Estimates of π, given θ, are obtained by maximizing $S(\pi)$ (or by minimizing $KLIC(\pi, p = 1/T)$) subject to the weighted zero functions and the probability constraint. Note that the averaging is done with respect to the weights on the observations in the sample, the vector π. The solution to the Lagrangian yields

$$\pi_t^{ET}(\theta) = \frac{\exp(\gamma' g(x_t,\theta))}{\sum_{t=1}^{T} \exp(\gamma' g(x_t,\theta))}, \tag{1}$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

where the q-vector γ contains the Lagrange multipliers and as such measures the degree of departure from zero of the moment conditions, and ET stands for exponential tilting. Estimates of θ are obtained by substituting π in $S(\pi)$, maximizing it with respect to γ, and then with respect to θ (see, for example, Kitamura and Stutzer, 1997).

If we interchange the subject/reference distributions we get

$$KLIC(p = 1/T, \pi) = \sum_{t=1}^{T} (1/T)[\ln(1/T) - \ln(\pi_t)],$$

and the solution to the optimization problem yields a different set of weights given by

$$\pi_t^{EL}(\theta) = \frac{1}{T\,[1 + \gamma' g(x_t, \theta)]},$$

where EL stands for empirical likelihood. When we evaluate the weights at some estimators, $\tilde{\theta}_T$, we obtain $\pi_t^{ET}(\tilde{\theta}_T)$ and $\pi_t^{EL}(\tilde{\theta}_T)$. In the EL context, averaging is done with respect to the empirical weights $1/T$. We therefore have two different interpretations for the mean information. In the latter, EL sense, the average information is such that it puts more importance on the empirical weights $1/T$ relative to π. In the former, ET sense, the average information instead puts more importance on probabilities, π, that are computed using observed data. Recently, Schennach (2007) combined ET and EL into the ETEL estimator that combines the advantages of each approach.

We mentioned in the Introduction that less weight is assigned to an observation for which the moment conditions are not satisfied. In this section we have seen that the vector π contains all the relevant information with respect to the moment conditions. We now provide a graphical illustration of the use of the weights in the detection of a structural change. We consider a small simulation study that contains three examples that have been studied in the structural change and entropy literature. The first example, which encompasses three cases, is similar to the one used by Imbens, Spady, and Johnson (1998) and consists of estimating a single parameter, θ, with two moment conditions:

$$\mathbb{E}[x_t - \theta] = 0 \quad \text{and} \quad \mathbb{E}\left[(x_t - \theta)^2 - 4\right] = 0,$$

with a sample of 100 observations and $x_t = \theta_t + \epsilon_t$ where $\epsilon_t \sim N(0, 4)$. We consider a pulse (a one-time short break), a break (a longer but temporary break), and a regime shift (a permanent break). In the pulse case we have $\theta_t = 5$ for $t = 50$ and $\theta = 10$ otherwise. For the break case we have $\theta_t = 10$ for $t \leq 20$ and $t > 80$, while $\theta_t = 15$ for $21 \leq t \leq 80$. Finally, the regime shift case has $\theta_t = 10$ for $t \leq 20$ and $\theta_t = 15$ otherwise. In these cases, structural change occurs via the parameters and can be tested using procedures proposed by Andrews (1993) and by Andrews and Ploberger (1994).

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

In contrast, the next two examples consist of structural change through the moment conditions. Following Hall and Horowitz (1996) and Gregory, Lamarche, and Smith (2002) we study a simulated environment with constant relative risk aversion (CRRA) preferences and making a distributional assumption on consumption growth, x_t, with i.i.d data and $T = 100$. In particular, we assume that the natural logarithm of consumption growth follows a $N(0, \sigma^2 = 0.16)$. There is a single parameter to be estimated, θ, the coefficient of CRRA and two moment conditions are used:

$$\begin{aligned} \mathbb{E}\exp[-\theta \ln x_{t+1} - 9\sigma^2/2 + (3-\theta)z_t] &= 1 \\ \mathbb{E}z_t \exp[-\theta \ln x_{t+1} - 9\sigma^2/2 + (3-\theta)z_t - 1] &= 0, \end{aligned}$$

with $z_t \sim N(0, \sigma^2)$. The moment conditions are satisfied when $\theta = 3$. The structural break occurs in period 51 and is summarized by a shift in θ from 3 to 4.

Lastly, as in Ghysels, Guay, and Hall (1997), we have the estimation of an autoregressive parameter using two moment conditions when the data generating process is an $AR(1)$ ($x_t = \beta x_{t-1} + \epsilon_t$), for $t \leq 50$, and an $ARMA(1,2)$ otherwise ($x_t = \beta x_{t-1} + \epsilon_t + 0.5\epsilon_{t-2}$). There are 100 observations and $\epsilon_t \sim N(0,1)$. The two instruments used are the first and second lags of x_t. The two moment conditions are then

$$\mathbb{E}\left[x_{t-1}(x_t - \beta x_{t-1})\right] = 0 \quad \text{and} \quad \mathbb{E}\left[x_{t-2}(x_t - \beta x_{t-1})\right] = 0.$$

The instability occurs because the second moment condition is violated for $t > 50$ (and β was set to 0.5).

Figure 1 shows the average of the vector of implied probabilities π (calculated using (1)) over 10,000 replications. The key feature of these panels is that when there is no break, the weights fluctuate around $1/T = 1/100$ (upper left panel). With a structural break in the parameter or in the moment conditions, however, more weight is given to observations (and moment conditions) for which there is no break, while less weight is assigned to an observation that violates the moment conditions. This simple simulation study clearly shows that implied probabilities contain interesting information to detect structural change. We can try to understand analytically the results from the simulation experiments. By a Taylor expansion of the implied probabilities (1), we get

$$\pi_t^{ET}(\theta) \approx \frac{1}{T} + \frac{1}{T}\left[g(x_t,\theta) - \frac{1}{T}\sum_{t=1}^{T} g(x_t,\theta)\right]' \gamma.$$

Under the null of no structural change $\mathbb{E}[g(x_t,\theta_0)] = 0$, $\forall t = 1,\ldots,T$, the second part of the right-hand side (RHS) expression evaluated at θ_0 will be zero and $\pi_t = \frac{1}{T}$. Now suppose that a structural change occurs such that $\mathbb{E}[g(x_t,\theta_1)] = 0$ for the first part of the sample and $\mathbb{E}[g(x_t,\theta_2)] = 0$ for the second part of the sample but $\theta_1 \neq \theta_2$. The implied probabilities evaluated at a restricted, stable estimator $\tilde{\theta}_T$ will differ from $\frac{1}{T}$ by the second term of the RHS expression. Indeed, the

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

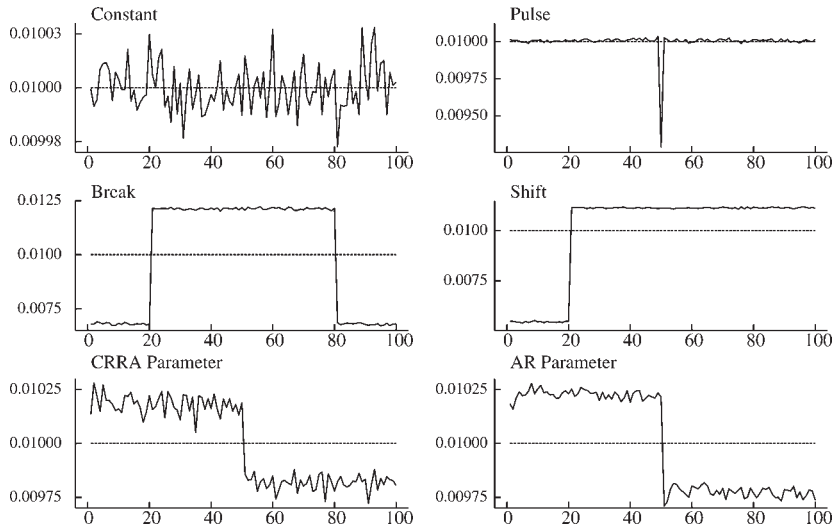


FIGURE 1. Simulated implied probabilities. *Note:* The *constant* case refers to no break, the *pulse* case to a one-time temporary jump at observation 50, the *break* case to a change in the mean for 60 time periods, and the *shift* to a permanent change in the mean at observation 20. For the CRRA case, a preference parameter is estimated using two moments. The moment conditions are violated at observation 51. For the AR case an autoregressive parameter is estimated. The data generating process is represented by an $AR(1)$ process for $t \leq 50$ and by an $ARMA(1,2)$ otherwise. The sample is of size 100, and in all cases the horizontal lines correspond to the empirical weights of $1/100$.

moment conditions $g(x_t, \tilde{\theta}_T)$ for each t will differ with a high probability from $\frac{1}{T}\sum_{t=1}^{T} g(x_t, \tilde{\theta}_T)$, and the estimator $\tilde{\gamma}_T$ will be different from zero with a high probability. The aim of this paper is then to examine the information contained in the estimated implied probabilities to detect structural change and propose test statistics based on some function of these implied probabilities.

Now following the recent econometric literature on GEL (see Newey and Smith, 2004; Smith, 2004; Ramalho and Smith, 2005; Caner, 2008a, 2008b; Guggenberger and Smith, 2008), we 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 $\tilde{\Gamma}_T(\theta) = \{\gamma : \gamma' g(x_t, \theta) \in \Phi, t = 1, \ldots, T\}$. Then, the GEL estimator is a solution to the problem

$$\tilde{\theta}_T = \arg\min_{\theta \in \Theta} \sup_{\gamma \in \tilde{\Gamma}_T(\theta)} \sum_{t=1}^{T} \frac{\left[\rho(\gamma' g(x_t, \theta)) - \rho_0\right]}{T},$$

where $\rho_j(\cdot) = \partial^j \rho(\cdot)/\partial \phi^j$ and $\rho_j = \rho_j(0)$ for $j = 0, 1, 2, \ldots$. Under this formulation, a number of estimators can be obtained. First, the ET estimator of θ is found by setting $\rho(\phi) = -\exp(\phi)$. Second, the EL estimator of θ is found by setting $\rho(\phi) = \ln(1 - \phi)$. Third, the continuous updating estimator of Hansen et al.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

(1996) can also be obtained from the GEL formulation by using a quadratic function for $\rho(\phi)$, $-(1+\phi)^2/2$.

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

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

for $t=1,\ldots,T$ and M_T is a bandwidth parameter, $k(\cdot)$ a kernel function, and we define $k_j = \int_{-\infty}^{\infty} k(a)^j da$. In the time series context, the criteria is then given by $\sum_{t=1}^{T} \frac{[\rho(k\gamma' g_{tT}(\theta))-\rho_0]}{T}$, where $k=\frac{k_1}{k_2}$ (see Smith, 2004).

3. FULL- AND PARTIAL-SAMPLE GEL ESTIMATORS

To establish the asymptotic distribution theory of tests for structural change in the parameters based on implied probabilities we need to elaborate on the specification of the parameter vector in our generic setup. We consider parametric models indexed by a vector of parameters $(\beta,\delta)\subset B\times\Delta\in R^p$ with $p=r+\nu$. Following Andrews (1993), we make a distinction between pure structural change in parameters when no subvector δ appears and the entire parameter vector is subject to structural change under the alternative and partial structural change, which corresponds to cases where only a subvector β is subject to structural change, under the alternative. The moment conditions for the full sample under the null of no structural change are then written as

$$\mathbb{E}g(x_t,\beta_0,\delta_0)=0, \qquad \forall t=1,\ldots,T. \tag{2}$$

3.1. Definitions

We need to impose restrictions on the admissible class of functions and processes involved in estimation to guarantee well-behaved asymptotic properties of GEL estimators using the entire data sample or subsamples of observations. A set of regularity conditions is also required to obtain weak convergence of partial-sample GEL estimators to a function of Brownian motions. To streamline the presentation we provide a detailed description of them in the Appendix. We now formally define the GEL estimator using the full sample.

DEFINITION 3.1. *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 $\tilde{\Gamma}_T(\beta,\delta)=\{\gamma : k\gamma' g_{tT}(\beta,\delta)\in\Phi, t=1,\ldots,T\}$ with $k=\frac{k_1}{k_2}$ (see the definition of k_j just before Section 3). Then, the full-sample GEL estimator $\{(\tilde{\beta}_T,\tilde{\delta}_T)\}$ is a sequence of random vectors*

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

such that

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

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

The criteria are normalized so that $\rho_1 = \rho_2 = -1$ (see Smith, 2004). This GEL estimator encompasses the ET, EL, and continuous updating estimators, as mentioned earlier. Note that the vector γ represents the Lagrange multipliers associated with the second constraint in the entropy problem of Section 2. Hence, in the GEL setting, the parameter vector $\left(\beta', \delta'\right)'$ is augmented by a vector of auxiliary parameters γ where each element of this vector is associated with an element of the smoothed moment conditions $g_{tT}(\beta, \delta)$.

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

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

Second, the GEL estimator $(\tilde{\beta}_T', \tilde{\delta}_T')'$ is given by the minimization problem

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

To characterize the asymptotic distributions in our estimators and test statistics we define the following matrices. First, the long-run covariance matrix of the moment conditions is defined as

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

Second, the gradient vectors are defined as

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

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

and $G = \left[\begin{array}{cc} G_\beta & G_\delta \end{array}\right] \in R^{q\times p}$ where $p = r + v$.

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

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

to obtain the smoothed moment conditions

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

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

where K_T is related to the bandwidth parameter used in the GMM (see Section 5 and the Appendix for more discussion of K_T). To handle the endpoints, we use the approach of Smith (2004) and Guggenberger and Smith (2008), which sets

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

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

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

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

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

Several tests for structural change involving partial-sample GMM estimators were defined by Andrews (1993). We focus here on the GEL setting. We consider two subsamples: The first covers observations $t=1,\ldots,[Ts]$ and the second covers $t=[Ts]+1,\ldots,T$ where $s\in S\subset(0,1)$. As mentioned earlier, the subvector β is allowed to vary across subsamples such that β_1 denotes the corresponding subvector for the first subsample and β_2 for the second subsample. We also define the vector of auxiliary parameters $\gamma(s)=\left(\gamma_1',\gamma_2'\right)'$ where γ_1 is the vector of the auxiliary parameters for the first part of the sample and γ_2 for the second part of the sample. The partial-sample GEL estimators for $s\in S$ based on the first and second subsamples are formally defined next.

DEFINITION 3.2. *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 $\hat{\Gamma}_T(\theta,s)=\{\gamma(s)=\left(\gamma_1',\gamma_2'\right)':\gamma(s)'g_{tT}(\theta,s)\in\Phi,t=1,\ldots,T\}$. A partial-sample general empirical likelihood (PS-GEL) estimator $\{\hat{\theta}_T(s)\}$ is a sequence of random vectors such that*

$$\begin{aligned}\hat{\theta}_T(s) &= \arg\min_{\theta\in\Theta}\sup_{\gamma(s)\in\hat{\Gamma}_T(\theta,s)}\sum_{t=1}^{T}\frac{\left[\rho(\gamma(s)'g_{tT}(\theta,s))-\rho_0\right]}{T}\\ &= \arg\min_{\theta\in\Theta}\sup_{\gamma(s)\in\hat{\Gamma}_T(\theta,s)}\left(\sum_{t=1}^{[Ts]}\frac{\left[\rho(\gamma_1'g_{tT}(\beta_1,\delta))-\rho_0\right]}{T}\right.\\ &\qquad\left.+\sum_{t=[Ts]+1}^{T}\frac{\left[\rho(\gamma_2'g_{tT}(\beta_2,\delta))-\rho_0\right]}{T}\right)\end{aligned}$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

for all $s \in S$, *where* $g_{tT}(\theta, s) = \left(g_{tT}(\beta_1, \delta)', 0'\right)' \in R^{2q\times 1}$ *for* $t = 1, \ldots, [Ts]$ *and* $g_{tT}(\theta, s) = \left(0', g_{tT}(\beta_2, \delta)'\right)' \in R^{2q\times 1}$ *for* $t = [Ts]+1, \ldots, T$ *with* $\gamma(s) = \left(\gamma_1', \gamma_2'\right)' \in R^{2q\times 1}$.

According to the definition above, $\hat{\theta}_T(s) = \left(\hat{\beta}_{1T}(s)', \hat{\beta}_{2T}(s)', \hat{\delta}_T(s)'\right)'$ is a $2r+\nu$-vector with an estimator $\hat{\beta}_{1T}(s)$ that uses the first subsample $t = 1, \ldots, [Ts]$, an estimator $\hat{\beta}_{2T}(s)$ that uses the second subsample $t = [Ts]+1, \ldots, T$, and an estimator $\hat{\delta}_T(s)$ that uses all the sample. The vector $\hat{\theta}_T(s)$ is called the partial-sample GEL estimator of $\theta = \left(\beta_1', \beta_2', \delta'\right)'$. Similarly, the vector $\tilde{\theta}_T = \left(\tilde{\beta}_T', \tilde{\beta}_T', \tilde{\delta}_T'\right)'$ is called the full-sample GEL estimator of θ (see Definition 3.1).

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

$$\hat{\gamma}_{1T}(\beta_1, \delta, s) = \arg \sup_{\gamma_1 \in \hat{\Gamma}_{1T}(\beta_1, \delta, s)} \sum_{t=1}^{[Ts]} \frac{\left[\rho(\gamma_1(\beta_1, \delta)' g_{tT}(\beta_1, \delta)) - \rho_0\right]}{T},$$

$$\hat{\gamma}_{2T}(\beta_2, \delta, s) = \arg \sup_{\gamma_2 \in \hat{\Gamma}_{2T}(\beta_2, \delta, s)} \sum_{t=[Ts]+1}^{T} \frac{\left[\rho(\gamma_2(\beta_2, \delta)' g_{tT}(\beta_2, \delta)) - \rho_0\right]}{T}$$

with $\hat{\Gamma}_{1T}(\beta_1, \delta, s) = \{\gamma_1 : \gamma_1' g_{tT}(\beta_1, \delta) \in \Phi, t = 1, \ldots, [Ts]\}$ and $\hat{\Gamma}_{2T}(\beta_2, \delta, s) = \{\gamma_2 : \gamma_2' g_{tT}(\beta_2, \delta) \in \Phi, t = [Ts]+1, \ldots, T\}$. The estimator $\hat{\gamma}_T(\hat{\theta}_T(s), s) = (\hat{\gamma}_{1T}(\hat{\beta}_{1T}, \hat{\delta}_T, s)', \hat{\gamma}_{2T}(\hat{\beta}_{2T}, \hat{\delta}_T, s)')'$ is hereafter denoted $\hat{\gamma}_T(s)$ to simplify the notation.

We need to define $\Omega(s)$ as

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

which under the null of no structural change is equal to

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

We also define

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

A consistent estimator of estimator $\Omega(s)$ is given by

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

with

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

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

and

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

We now present the corresponding implied probabilities. The general formula of the implied probabilities for the full-sample GEL estimator is defined by the following ratio (see Smith, 2004):

$$\pi_t^{GEL}\left(\tilde{\beta}_T,\tilde{\delta}_T\right) = \frac{\rho_1\left(\tilde{\gamma}_T' g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right)}{\sum_{t=1}^{T}\rho_1\left(\tilde{\gamma}_T' g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right)}$$

with $\tilde{\gamma}_T = \tilde{\gamma}_T\left(\tilde{\beta}_T,\tilde{\delta}_T\right)$.

Implied probabilities for the full-sample ET, EL, and CU estimators with the smoothed moment conditions are respectively given by

$$\pi_t^{ET}\left(\tilde{\beta}_T,\tilde{\delta}_T\right) = \frac{\exp\left[\tilde{\gamma_T}' g_{tT}\left(\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right)\right]}{\sum_{t=1}^{T}\exp\left[\tilde{\gamma}_T' g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right]},$$

$$\pi_t^{EL}\left(\tilde{\beta}_T,\tilde{\delta}_T\right) = \frac{1}{T\left[1+\tilde{\gamma}_T' g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right]},$$

and

$$\pi_t^{CUE}\left(\tilde{\beta}_T,\tilde{\delta}_T\right) = \frac{1}{T} - \frac{1}{T}g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\left[\frac{1}{T}\sum_{t-1}^{T} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)'\right]^{-1} \times\frac{1}{T}\sum_{t=1}^{T} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right).$$

Note that $\frac{2K+1}{T}\sum_{t-1}^{T} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)'$ is a consistent estimator of Ω.

The general formula for the unrestricted partial-sample implied probabilities for the GEL is defined for $s \in S$ as

$$\pi_t^{GEL}\left(\hat{\theta}_T(s),s\right) = \frac{\rho_1\left(\hat{\gamma}_T(s)' g_{tT}\left(\hat{\theta}_T(s),s\right)\right)}{\sum_{t=1}^{T}\rho_1\left(\hat{\gamma}_T(s)' g_{tT}\left(\hat{\theta}_T(s),s\right)\right)}. \quad \textbf{(3)}$$

For example, in the case of t between observations 1 and $[Ts]$, we get for the unrestricted implied probabilities at t,

$$\pi_t^{GEL}\left(\hat{\theta}_T(s),s\right) = \frac{\rho_1\left(\hat{\gamma}_{1T}(s)' g_{tT}\left(\hat{\beta}_{1T}(s),\hat{\delta}_T(s)\right)\right)}{\sum_{t=1}^{T}\rho_1\left(\hat{\gamma}_T(s)' g_{tT}\left(\hat{\theta}_T(s),s\right)\right)}$$

with $\hat{\gamma}_{1T}(s) = \hat{\gamma}_{1T}\left(\hat{\beta}_{1T}(s),\hat{\delta}_T(s),s\right)$.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

For the commonly used GEL partial-sample estimators, we get

$$\pi_t^{ET}\left(\hat{\theta}_T(s),s\right) = \frac{\exp\left[\hat{\gamma}_T(s)'g_{tT}\left(\hat{\theta}_T(s),s\right)\right]}{\sum_{t=1}^{T}\exp\left[\hat{\gamma}_T(s)'g_{tT}\left(\hat{\theta}_T(s),s\right)\right]},$$

$$\pi_t^{EL}\left(\hat{\theta}_T(s),s\right) = \frac{1}{T\left[1+\hat{\gamma}_T(s)'g_{tT}\left(\hat{\theta}_T(s),s\right)\right]},$$

$$\pi_t^{CUE}\left(\hat{\theta}_T(s),s\right) = \frac{1}{T}-\frac{1}{T}g_{tT}\left(\hat{\theta}_T(s),s\right)\left[\frac{1}{T}\sum_{t-1}^{T}g_{tT}\left(\hat{\theta}_T(s),s\right)g_{tT}\left(\hat{\theta}_T(s),s\right)'\right]^{-1}$$
$$\times\frac{1}{T}\sum_{t=1}^{T}g_{tT}\left(\hat{\theta}_T(s),s\right).$$

The purpose of the next subsection is to refine the null hypothesis (2) of no structural change. Such a refinement will enable us to construct various tests for structural change in the spirit of Sowell (1996a) and Hall and Sen (1999).

3.2. The Null and the Alternative Hypotheses

Following Sowell (1996b), 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

$$\Omega^{-1/2}\mathbb{E}g(x_t,\beta_0,\delta_0) = P_G\Omega^{-1/2}\mathbb{E}g(x_t,\beta_0,\delta_0)+(I_q-P_G)\Omega^{-1/2}$$
$$\times\mathbb{E}g(x_t,\beta_0,\delta_0), \tag{4}$$

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 no structural change. For instance, following Hall and Sen (1999) we can consider the null, $H_0^I(s)$, for the case of a single possible breakpoint s, which separates the identifying restrictions across the two subsamples:

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

The null $H_0^I(s)$ is equivalent to the null hypothesis that

$$H_0(s)\ :\ \beta_1(s)=\beta_2(s)=\beta_0, \tag{5}$$

since only the subvector β is allowed to vary across subsamples.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

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

$$H_0^{O1}(s) : (I_q - P_G)\Omega^{-1/2}\mathbb{E}[g(x_t, \beta_0, \delta_0)] = 0 \qquad \forall t = 1, \ldots, [Ts]$$

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

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 the overidentifying restrictions can be violated for both subsamples under the alternative of a structural change, while we can always find parameter values that satisfy the identifying restrictions for one subsample under the alternative of structural change. This is the reason why $H_0^O(s)$ is divided into each subsample, namely $H_0^{O1}(s)$ and $H_0^{O2}(s)$ but not $H_0^I(s)$. We are more explicit hereafter about the violation of $H_0^I(s)$ and $H_0^O(s)$ under the alternative of structural change. Various tests can be constructed with local power properties against any one of these three null hypotheses. To elaborate further on this, we consider a sequence of Pitman local alternatives based on the moment conditions.

Assumption 3.1. A sequence of local alternatives is specified as

$$\mathbb{E}g(x_{Tt}, \beta_0, \delta_0) = \frac{h\left(\eta, \tau, \frac{t}{T}\right)}{\sqrt{T}} \tag{6}$$

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

Following the decomposition in equation (4) , the sequence of alternatives can be rewritten as

$$\Omega^{-1/2}\mathbb{E}g(x_{Tt}, \beta_0, \delta_0) = P_G\Omega^{-1/2}\frac{h\left(\eta, \tau, \frac{t}{T}\right)}{\sqrt{T}} + \left(I_q - P_G\right)\Omega^{-1/2}\frac{h\left(\eta, \tau, \frac{t}{T}\right)}{\sqrt{T}}, \tag{7}$$

where the first component is the local alternative on the identifying moments and the second is the local alternative on the overidentifying restrictions.

For instability in the parameter vector β, consider a general local alternative of the form (see Sowell, 1996a)

$$\beta_{Tt} = \beta_0 + \frac{f\left(\eta, \tau, \frac{t}{T}\right)}{\sqrt{T}}$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

for $t = 1, \ldots, T$. Under this local alternative, $\mathbb{E}g(x_{Tt}, \beta_{Tt}, \delta_0) = 0$. A Taylor expansion of $g(x_{Tt}, \beta_{Tt}, \delta_0)$ around β_0 yields

$$\mathbb{E}g(x_{Tt}, \beta_0, \delta_0) = -G_\beta \frac{f\left(\eta, \tau, \frac{t}{T}\right)}{\sqrt{T}} + o_p\left(T^{-1/2}\right). \tag{8}$$

Thus, $\frac{h(\eta,\tau,\frac{t}{T})}{\sqrt{T}} = -G_\beta \frac{f(\eta,\tau,\frac{t}{T})}{\sqrt{T}} + o_p(T^{-1/2})$. If we substitute this expression into (7), the expression (8) is orthogonal to the last component of (7) (up to the $o_p(T^{-1/2})$ term) and therefore puts restrictions on the first component (the identifying restrictions). In the case of pure structural change, $P_G = P_{G_\beta} = \Omega^{-1/2} G_\beta [G_\beta' \Omega^{-1} G_\beta]^{-1} G_\beta' \Omega^{-1/2}$. The alternative that at some point there is a single structural break at τ, $H_A^I(\tau)$, is represented as

$$\beta_{Tt} = \begin{cases} \beta_0 & t = 1, ..., [T\tau] \\ \beta_0 + \frac{\eta}{\sqrt{T}} & t = [T\tau] + 1, ..., T, \end{cases}$$

which corresponds to a specific form for $f(\eta, \tau, \frac{t}{T})$. Thus, the null hypothesis H_0^I is respected for the first subsample but not for the second at β_0.

For instability of the overidentifying restrictions at a single breakpoint τ occurring before and/or after the breakpoint, this is formally stated as $H_A^O(\tau) = H_A^{O1}(\tau) \cap H_A^{O2}(\tau)$ with

$$H_A^{O1}(\tau) : (I_q - P_G)\Omega^{-1/2}\mathbb{E}[g(x_t, \beta_0, \delta_0)] = \frac{\eta_1}{\sqrt{T}} \qquad \forall t = 1, \ldots, [T\tau]$$

$$H_A^{O2}(\tau) : (I_q - P_G)\Omega^{-1/2}\mathbb{E}[g(x_t, \beta_0, \delta_0)] = \frac{\eta_2}{\sqrt{T}} \qquad \forall t = [T\tau] + 1, \ldots, T,$$

and $\eta_1 \neq \eta_2$.[3] This formulation of the alternative for a single breakpoint corresponds to a specific form of $h(\eta, \tau, \frac{t}{T})$ in (7).

4. TESTS FOR A STRUCTURAL CHANGE BASED ON IMPLIED PROBABILITIES

Ramalho and Smith (2005) introduced in an i.i.d. setting a specification test for moment conditions based on implied probabilities similar in spirit to the classical Pearson chi-square goodness-of-fit test. The test is based on the statistic

$$\sum_{t=1}^{T} \left(T\pi_t^{GEL}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) - 1\right)^2.$$

They showed that such a statistic is asymptotically equivalent to the overidentifying moment restrictions J-test proposed by Hansen (1982). Guay and Pelgrin (2007) and Guggenberger, Ramalho, and Smith (2007) also used this statistic but adapted for weakly dependent data and showed that

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

$$\frac{1}{2K_T+1}\sum_{t=1}^{T}\left(T\pi_t^{GEL}(\tilde{\beta}_T,\tilde{\delta}_T)-1\right)^2 \tag{9}$$

is asymptotically first-order equivalent to the overidentifying moment restrictions J-test. However, as shown by Ghysels and Hall (1990), the J-test has no power to detect structural change in parameter values, a property that is shared by the specification tests, (10), and the one just above, as we demonstrate below.

In the same spirit, we first consider a test statistic based on the partial-sample implied probabilities evaluated at the restricted estimator for GEL. The implied probabilities structural change (IPSC) test statistic is given by the following partial sum:

$$IPSC_T^{GEL}(s)=\frac{s}{2K_T+1}\sum_{t=1}^{[Ts]}\left(T\pi_t^{GEL}(\tilde{\theta}_T,s)-1\right)^2 \tag{10}$$

with

$$\pi_t^{GEL}(\tilde{\theta}_T,s)=\frac{\rho_1\left(\tilde{\gamma}_{1T}(s)'g_{tT}(\tilde{\theta}_T,s)\right)}{\sum_{t=1}^{T}\rho_1\left(\tilde{\gamma}_T(s)'g_{tT}(\tilde{\theta}_T,s)\right)},$$

where $\tilde{\gamma}_{1T}(s)=\tilde{\gamma}_{1T}(\tilde{\beta}_T,\tilde{\delta}_T,s)$ is the solution of the following maximization problem:

$$\gamma_{1T}(\beta,\delta,s) \;=\; \arg\sup_{\gamma_1\in\hat{\Gamma}_{1T}(\beta,\delta,s)}\sum_{t=1}^{[Ts]}\frac{\left[\rho(\gamma_1(\beta,\delta)'g_{tT}(\beta,\delta))-\rho_0\right]}{T} \tag{11}$$

evaluated at the restricted estimator $\tilde{\theta}_T$ and for the denominator $\tilde{\gamma}_T(s)=\left(\tilde{\gamma}_{1T}(s)',\ \tilde{\gamma}_{2T}(s)'\right)'$ with $\tilde{\gamma}_{1T}(s)$ defined as above, and $\tilde{\gamma}_{2T}(s)=\tilde{\gamma}_{2T}(\tilde{\beta}_T,\tilde{\delta}_T,s)$ is the solution of

$$\gamma_{2T}(\beta,\delta,s) \;=\; \arg\sup_{\gamma_2\in\hat{\Gamma}_{2T}(\beta,\delta,s)}\sum_{t=[Ts]+1}^{T}\frac{\left[\rho(\gamma_2(\beta,\delta)'g_{tT}(\beta,\delta))-\rho_0\right]}{T}. \tag{12}$$

It is crucial to note that this statistic is based on the **unrestricted** implied probabilities evaluated at the **restricted** estimator of θ.

The next theorem establishes the asymptotic distribution for this test of a structural change under the null and the sequence of alternatives defined in (6).

THEOREM 4.1. *Under Assumptions (A.1) to (A.12), the following process indexed by s for* $s\in[0,1]$ *satisfies, under the null (2),*

$$IPSC_T^{GEL}(s)\Rightarrow BB_p(s)'BB_p(s)+B'_{(q-p)}(s)B_{(q-p)}(s),$$

and under the alternative (6),

$$\begin{aligned}IPSC_T^{GEL}(s)\;\Rightarrow\;& BB_p(s)'BB_p(s)+(H(s)-sH(1))'\Omega^{-1/2}P_G\Omega^{-1/2}\\&\times(H(s)-sH(1))+B_{q-p}(s)'B_{q-p}(s)+H(s)'\Omega^{-1/2}\\&\times(I-P_G)\Omega^{-1/2}H(s),\end{aligned}$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

where $B_{(q-p)}(s)$ *is a* $(q-p)$*-vector of standard Brownian motion,* $BB_p(s) = B_p(s) - sB_p(1)$ *is a* p*-vector of Brownian bridge with* $p = r + v$*, and* $H(s) = \int_0^s h(\eta, \tau, r)dr$.

Proof: See the Appendix.

The theorem shows that the structural change test based on this quadratic form of the partial-sample sum of the implied probabilities evaluated at the full-sample estimator combines two components. The first component of the limiting distribution, which is a function of Brownian bridges, corresponds to a stability test for the whole set of parameters (β and δ), and the second component corresponds to a stability test for the overidentifying restrictions. The predictive tests proposed by Ghysels et al. (1997) shares the same properties. In the Appendix we show that the IPSC test is asymptotically equivalent to the test statistic proposed by Sowell (1996b). He showed that his test statistic is optimal for a one-time jump in all moment conditions where the location of the jump is unknown and consistent for arbitrary alternatives. These properties are then shared by our test. Note that the limiting distribution exists for $s = 0$, which is trivially equal to 0. For $s = 1$, the test statistic corresponds to the specification test for moment conditions developed by Guay and Pelgrin (2007) and Guggenberger et al. (2007), and by the above theorem the limiting distribution is given by

$$B_{q-p}(1)' B_{q-p}(1) + H(1)' \Omega^{-1/2} (I - P_G) \Omega^{-1/2} H(1).$$

This limiting distribution shows that this test statistic has a chi-square distribution with $q - p$ degrees of freedom under the null and has local power equal to the size to detect instability in parameter values, as the J-test proposed by Hansen (1982). Moreover, the test statistic (9) can not asymptotically detect instability in the overidentifying restrictions for which $(I - P_G)\Omega^{-1/2} H(1) = 0$.[4]

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

The optimal average exponential form applied to a statistic $Q_T(s)$ for $s \in S$ has the form

$$\exp - Q_T = (1+c)^{-q/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c} Q_T(s)\right) dJ(s),$$

where various choices of c determine power against close or more distant alternatives and $J(\cdot)$ is the weight function over the value of $s \in S$. In the case of close alternatives ($c = 0$), the optimal test statistic takes the average form,

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

$\text{ave}Q_T = \int_S Q_T(s)dJ(s)$. For a distant alternative ($c = \infty$), the optimal test statistic takes the exponential form, $\exp Q_T = \log(\int_S \exp[\frac{1}{2}Q_T(s)]dJ(s))$. The supremum form often used in the literature corresponds to the case where $\frac{c}{(1+c)} \to \infty$. The sup test is given by $\sup Q_T = \sup_{s\in S} Q_T(s)$.

Theorem 4.2 gives the asymptotic distribution for the average exponential mapping for $Q_T^{IPSC}(s)$, where $Q_T^{IPSC}(s)$ corresponds to the statistic presented above based on the implied probabilities.

THEOREM 4.2. *Under the null hypothesis H_0 in (2) and Assumptions A.1 to A.12, the following processes indexed by s for a given set S whose closure lies in [0,1] satisfy*

$$\sup Q_T^{IPSC} \Rightarrow \sup_{s\in S} Q_{p,q-p}(s), \qquad \text{ave}Q_T^{IPSC} \Rightarrow \int_S Q_{p,q-p}(s)dJ(s),$$

$$\exp Q_T^{IPSC} \Rightarrow \log\left(\int_S \exp\left[\frac{1}{2}Q_{p,q-p}(s)\right]dJ(s)\right),$$

with

$$Q_{p,q-p}(s) = BB_p(s)'BB_p(s) + B_{(q-p)}(s)'B_{(q-p)}(s),$$

and $J(s)$ is the weighting distribution function for the location of the instability s.

This result is obtained through the application of the continuous mapping theorem (see Pollard, 1984). The asymptotic critical values were obtained using simulated Brownian motions and Brownian bridges over 10,000 replications for maximum values of p and $q-p$ set at 10. The critical values are tabulated for symmetric intervals $S = [s_0, 1-s_0]$.[5]

An asymptotically equivalent modified statistic to (10) in the spirit of the Neyman-modified chi-square test is given by

$$IPSCM_T^{GEL}(s) = \frac{s}{2K_T+1}\sum_{t=1}^{[Ts]} \frac{\left(T\pi_t^{GEL}(\tilde{\theta}_T, s) - 1\right)^2}{T\pi_t^{GEL}(\tilde{\theta}_T, s)}, \tag{13}$$

since $T\pi_t^{GEL}(\tilde{\theta}_T, s) = 1 + o_p(1)$ under the null.

In the sequel, we propose structural change tests based on implied probabilities specially designed to detect instability in the parameters of interest or in the overidentifying restrictions.

4.1. Tests for a Structural Change in the Parameters Based on Implied Probabilities

The test statistics proposed to specifically detect parameter instability are based on the difference between the partial sum of unrestricted implied probabilities evaluated at the unrestricted estimator $\hat{\theta}_T(s)$ and the partial sum of unrestricted implied probability evaluated at the restricted estimator $\tilde{\theta}_T$. More precisely, the test statistic is defined as

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

$$IPSC_T^{I,GEL}(s) = \frac{1}{2K_T+1}\sum_{t=1}^{T}\left(T\pi_t^{GEL}(\hat{\theta}_T(s),s) - T\pi_t^{GEL}(\tilde{\theta}_T,s)\right)^2, \tag{14}$$

with

$$\pi_t^{GEL}(\tilde{\theta}_T,s) = \frac{\rho_1\left(\tilde{\gamma}_T(s)'g_{tT}(\tilde{\theta}_T,s)\right)}{\sum_{t=1}^{T}\rho_1\left(\tilde{\gamma}_T(s)'g_{tT}(\tilde{\theta}_T,s)\right)},$$

where $\tilde{\gamma}_T(s) = \left(\tilde{\gamma}_{1T}(s)'\tilde{\gamma}_2(s)'\right)'$ is the solution of the respective maximization problems defined in equations (11) and (12) and $\pi_t^{GEL}(\hat{\theta}_T(s),s)$ is defined in equation (3).

The next theorem establishes the asymptotic distribution for this test of structural change in the parameter values under the null that the vector β is constant throughout the sample.

THEOREM 4.3. *Under the null hypothesis in (5) and Assumptions (A.1) to (A.12), the following process indexed by s for a given set S whose closure lies in* $(0,1)$ *satisfies*

$$IPSC_T^{I,GEL}(s) \Rightarrow Q_r(s) = \frac{BB_r(s)'BB_r(s)}{s(1-s)},$$

and under the alternative (6),

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

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

$$\sup IPSC_T^{I,GEL} \Rightarrow \sup_{s\in S} Q_r(s), \qquad \text{ave}\, IPSC_T^{I,GEL} \Rightarrow \int_S Q_r(s)dJ(s),$$

$$\exp IPSC_T^{I,GEL} \Rightarrow \log\left(\int_S \exp\left[\frac{1}{2}Q_r(s)\right]dJ(s)\right).$$

Proof: See the Appendix.

The theorem shows that the asymptotic distribution of the test based on implied probabilities for smoothed moment conditions is asymptotically equivalent under the null and the alternative to the Wald, LM, and LR tests for parameter instability (see Andrews, 1993). More precisely, the limiting distribution is a function of an r-vector of Brownian bridge with the same dimension as the parameter vector β. However, the small sample properties of tests based on implied probabilities can differ from those of more standard tests. Note that, in contrast to Theorem 4.2, the limiting distribution in Theorem 4.3 is valid only for S in the open interval

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

$(0, 1)$. The asymptotic distribution derived above is valid under the assumption that the parameter vector δ is stable across time, as discussed in Li and Müller (2009), a property shared by the Wald, LM, and LR tests for parameter instability. These authors proposed a modified test statistic that is robust to possible local instability in the parameter vector δ. However, if it is known for sure that the parameter vector δ is stable, we can easily show that the test above based on implied probabilities is more powerful than the test proposed by Li and Müller.

The two following modified statistics are asymptotically equivalent to the one defined above:

$$IPSCM1_T^{I,GEL}(s) = \frac{1}{2K_T+1}\sum_{t=1}^{T}\frac{\left(T\pi_t^{GEL}\left(\hat{\theta}_T(s),s\right)-T\pi_t^{GEL}\left(\tilde{\theta}_T,s\right)\right)^2}{T\pi_t^{GEL}\left(\hat{\theta}_T(s),s\right)} \quad \textbf{(15)}$$

and

$$IPSCM2_T^{I,GEL}(s) = \frac{1}{2K_T+1}\sum_{t=1}^{T}\frac{\left(T\pi_t^{GEL}\left(\hat{\theta}_T(s),s\right)-T\pi_t^{GEL}\left(\tilde{\theta}_T,s\right)\right)^2}{T\pi_t^{GEL}\left(\tilde{\theta}_T,s\right)}. \quad \textbf{(16)}$$

4.2. Tests for a Structural Change in the Overidentifying Restrictions Based on Implied Probabilities

Now, we propose a test statistic based on implied probabilities designed specially to detect instability in the overidentifying restrictions. The statistic is powerful against violation of $H_0^{O1}(s)$ and $H_0^{O2}(s)$ and is asymptotically equivalent to the ones proposed by Hall and Sen (1999) and thus shares its asymptotic properties, but perhaps not its small sample properties. In particular, the small sample properties of Hall and Sen's statistic can be affected by the estimation of the covariance matrix needed to compute the overidentifying restrictions test as shown by Hall, Inoue, and Peixe 2003.

Again, the sample is split in two subamples with a single breakpoint at $[Ts]$. An estimator of the parameter vector is obtained with the first subsample (for $t = 1, \ldots, [Ts]$) and with the second subsample (for $t = [Ts]+1, \ldots, T$). The entire parameter vector is allowed to vary for both subsamples. The proposed statistic designed to detect instability in the overidentifying restrictions is based on the specification test for moment conditions given in equation (10) for the first and second subsamples evaluated at the unrestricted estimator.

More precisely, the statistic is defined as

$$\begin{aligned} IPSC_T^{O,GEL}(s) \;=\; & \frac{1}{2K_T+1}\sum_{t=1}^{[Ts]}\left([Ts]\pi_t^{GEL}(\hat{\theta}_{1T}(s),s)-1\right)^2 \\ & +\frac{1}{2K_T+1}\sum_{t=[Ts]+1}^{T}\left((T-[Ts])\pi_t^{GEL}(\hat{\theta}_{2T}(s),s)-1\right)^2, \quad \textbf{(17)} \end{aligned}$$

where $\hat{\theta}_{1T}(s) = \hat{\beta}_{1T}(s)$ and $\hat{\theta}_{2T}(s) = \hat{\beta}_{2T}(s)$.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

The next theorem establishes the asymptotic distribution of this statistic and the corresponding mappings.

THEOREM 4.4. *Under Assumptions (A.1) to (A.12), the following process indexed by s for a given set S whose closure lies in* $(0,1)$ *satisfies* $IPSC_T^{O,GEL}(s) \Rightarrow Q_{q-r}(s)$ *with, under the null of no structural change,*

$$Q_{q-r}(s) \Rightarrow \frac{B_{q-r}(s)' B_{q-r}(s)}{s} + \frac{\left[B_{q-r}(1) - B_{q-r}(s)\right]' \left[B_{q-r}(1) - B_{q-r}(s)\right]}{1-s},$$

and, under the alternative (6),

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

where $B_{q-r}(s)$ *is a* $q-r$*-vector of standard Brownian motion and* $P_G = P_{G_\beta}$ *with* $P_{G_\beta} = \Omega^{-1/2} G_\beta (G_\beta' \Omega^{-1} G_\beta)^{-1} G' \Omega^{-1/2}$. *Moreover,*

$$\sup IPSC_T^{O,GEL} \Rightarrow \sup_{s \in S} Q_{q-r}(s), \qquad \text{ave} IPSC_T^{O,GEL} \Rightarrow \int_S Q_{q-r}(s) dJ(s),$$

$$\exp IPSC_T^{O,GEL} \Rightarrow \log \left(\int_S \exp \left[\frac{1}{2} Q_{q-r}(s) \right] dJ(s) \right).$$

Proof: See the Appendix.

The theorem shows that the proposed test statistics can detect instability occurring in overidentifying restrictions before and after the breakpoint. Indeed, the term $\frac{H(s)' \Omega^{-1/2} (I - P_G) \Omega^{-1/2} H(s)}{s}$ corresponds to a structural change in the moment conditions before the breakpoint s, while the term $\frac{[H(1) - H(s)]' \Omega^{-1/2} (I - P_G) \Omega^{-1/2} [H(1) - H(s)]}{(1-s)}$ corresponds to a structural change after the breakpoint s.

The asymptotic critical values for the interval $S = [.15, .85]$ can be found in Hall and Sen (1999). For another symmetric interval $[s_0, 1 - s_0]$, critical values can be obtained in Guay (2003, Tables 1–3) for a number of overidentifying restrictions divided by 2 (in those tables). To see this, note that the critical values for the supremum, the average, and the log exponential mappings applied to $\frac{B_{2q-2r}(s)' B_{2q-2r}(s)}{s}$ are equivalent to ones corresponding to $\frac{B_{q-r}(s)' B_{q-r}(s)}{s} + \frac{(B_{q-r}(1) - B_{q-r}(s))' (B_{q-r}(1) - B_{q-r}(s))}{1-s}$ for a symmetric interval S.[6]

An asymptotically equivalent statistic to $IPSC_T^{O,GEL}(s)$ in the spirit of the Neyman-modified chi-square test is given by

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

$$IPSCM_T^{O,GEL}(s) = \frac{1}{2K_T+1}\sum_{t=1}^{[Ts]}\frac{\left([Ts]\pi_t^{GEL}(\hat{\theta}_{1T}(s),s)-1\right)^2}{[Ts]\pi_t^{GEL}(\hat{\theta}_{1T}(s),s)} + \frac{1}{2K_T+1}\sum_{t=[Ts]+1}^{T}\frac{\left((T-[Ts])\pi_t^{GEL}(\hat{\theta}_{2T}(s),s)-1\right)^2}{(T-[Ts])\pi_t^{GEL}(\hat{\theta}_{2T}(s),s)}. \quad (18)$$

4.3. A Structural Change Test Robust to Weak Identification or Cases in Which Parameters Are Completely Unidentified

We propose here a test statistic robust to the context of weak identification, as defined by Stock and Wright (2000), or robust to cases in which parameters are completely unidentified. Consider the pure structural change case with $\theta = (\beta', \beta')'$. We need to introduce a restricted estimator $\tilde{\theta}_T(s)$ 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

$$\begin{aligned}\tilde{\theta}_T(s) &= \arg\min_{\theta\in\Theta}\sup_{\gamma(s)\in\tilde{\Gamma}_T(\theta,s)}\sum_{t=1}^{T}\frac{\left[\rho(k\gamma(s)'g_{tT}(\theta,s))-\rho_0\right]}{T}\\ &= \arg\min_{\beta\in B}\sup_{\gamma(s)\in\tilde{\Gamma}_T(\theta,s)}\left(\sum_{t=1}^{[Ts]}\frac{\left[\rho(k\gamma_1'g_{tT}(\beta))-\rho_0\right]}{T} + \sum_{t=[Ts]+1}^{T}\frac{\left[\rho(k\gamma_2'g_{tT}(\beta))-\rho_0\right]}{T}\right)\end{aligned}$$

for all $s \in S$ where $g_{tT}(\theta,s) = \left(g_{tT}(\beta)', 0'\right)' \in R^{2q\times 1}$ for $t = 1,\ldots,[Ts]$ and $g_{tT}(\theta,s) = \left(0', g_{tT}(\beta)'\right)' \in R^{2q\times 1}$ for $t = [Ts]+1,\ldots,T$ with $\gamma(s) = \left(\gamma_1', \gamma_2'\right)' \in R^{2q\times 1}$ and $\tilde{\Gamma}_T(\theta,s) = \{\gamma(s) = \left(\gamma_1', \gamma_2'\right)' : k\gamma(s)'g_{tT}(\theta,s)\}$.

The statistic is defined as

$$IPSC_T^{R,GEL}(s) = \frac{1}{2K_T+1}\sum_{t=1}^{[Ts]}\left([Ts]\pi_t^{GEL}(\tilde{\theta}_T(s),s)-1\right)^2 + \frac{1}{2K_T+1}\sum_{t=[Ts]+1}^{T}\left((T-[Ts])\pi_t^{GEL}(\tilde{\theta}_T(s),s)-1\right)^2.$$

The statistic depends on **unrestricted** implied probabilities for the first and the second parts of the sample evaluated at the **restricted** partial-sample GEL estimators.

We show in the Appendix that this test statistic is asymptotically equivalent at first-order to the S-based test in Caner (2007). The test statistic is not asymptotically pivotal but asymptotically boundedly pivotal. The bound is then free of

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

nuisance parameters 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 (6).

THEOREM 4.5. *Under Assumptions (A.1) to (A.5) and (A.7) to (A.12), the process* $IPSC_T^{R,GEL}(s)$ *indexed by* s *for a given set* S *whose closure lies in* $(0,1)$ *is asymptotically boundedly pivotal and the asymptotic bound distribution is given by*

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

under the null of no structural change and under the alternative (6) by

$$\begin{aligned} Q_q(s) \;\Rightarrow\; & \frac{B_q(s)'B_q(s)}{s} + \frac{H(s)'\Omega^{-1}H(s)}{s} \\ & \times \frac{\left[B_q(1)-B_q(s)\right]'\left[B_q(1)-B_q(s)\right]}{1-s} \\ & + \frac{[H(1)-H(s)]'\,\Omega^{-1}\,[H(1)-H(s)]}{(1-s)}, \end{aligned}$$

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

Proof: See the Appendix.

The asymptotic bounds are valid for strong, weakly identified, and completely unidentified cases. Critical values under the null for the different mappings are given in the same tables as those of Section 4.2. The asymptotic bound under the local alternative allows us to examine the power of the test under different assumptions with respect to identification. Consider the decomposition of the alternative according to (7), namely

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

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_β (see equation (8)). Under weak identification, G_β has a weak value, which means that $G_{\beta,T}(\theta_0) = \frac{C}{\sqrt{T}}$ for a C matrix of dimension $q \times p$. In this case, the test statistic has trivial power equal to the size. Obviously, it is also the case when the parameters are unidentified because $G = 0$. 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 (again see Stock and Wright, 2000). For instance, the test statistic will detect structural change in the parameter values with no trivial power for the following fixed alternative:

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

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

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

5. SIMULATION EVIDENCE

To evaluate the performance of the test statistics, and to facilitate comparison with previous studies, we use the data generating process found in Ghysels et al. (1997) and in Hall and Sen (1999). The time series model used is an AR(1) process for the variable x_t (see also the Introduction). One parameter is estimated, the autoregressive parameter (denoted by β in the expression below), using two moment conditions formed with the lagged values of x_t. In this context, the parameters are well identified. A simulation study for cases of weak identification is left for future research.

The data generating process (DGP) is given by

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

for $t = 1, \ldots, T$. Structural change in the identifying restrictions (in the parameter) is studied by considering different values of β_i where the index $i = 1, 2$ denotes the first or second subsamples and $\alpha = 0$ for the whole sample. Structural change in the overidentifying restrictions is studied by considering nonzero values of α in the second subsample. The break is set at $T/2$. In the above, $u_t \sim N(0, 1)$. The sample size was set to 200 observations, and the number of Monte Carlo replications was 2,000 (for each replication one constrained and 242 unconstrained optimization problems were solved for a total of 486,000 numerical optimization). A trimming rule of 0.15 was used, namely $S = [.15, .85]$.

We consider 10 parameter configurations. DGP 1 to DGP 3 correspond to the null hypothesis of structural stability denoted by H_0. Experiments with DGP 1 to DGP 3 cover cases with $\beta = 0, .4, .8$, respectively, for the whole sample and $\alpha = 0$. For those DGPs, we then vary the magnitude of the autoregressive parameter β. The alternatives of instability in the parameters denoted by H_A^I are investigated under DGP 4 to DGP 6, which covers the following cases: DGP 4 has $\beta_1 = 0$, $\beta_2 = .4$; DGP 5 has $\beta_1 = 0$, $\beta_2 = .6$; while DGP 6 has $\beta_1 = .4$, $\beta_2 = .8$, with $\alpha = 0$ for all three DGPs. To investigate the alternatives of instability in the overidentifying restrictions denoted by H_A^O, we consider various values of the moving average parameter. For DGP 7 to DGP 10, α is fixed to zero for the first part of the sample and to $.5, .9, -.5, -.9$, respectively, for the second part of the sample. Under H_0, where $\alpha = 0$, the instruments are appropriate. Under the first class of alternative hypothesis (H_A^I) the two instruments are also valid, while they are no longer for the second part of the sample with the second class of alternative hypothesis (H_A^O). This version of the paper reports results for DGP 1 (the null hypothesis), 5 (instability in the parameter), and 8 (instability in the overidentifying restrictions). The supremum mapping of the tests is reported.[7]

Smoothing the moment conditions can be done via an appropriate choice of smoothing parameter K_T. In a GMM setting this is equivalent to using some

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

form of estimate of the long-run covariance matrix of the moment conditions (for example, using the Newey–West estimator as the weight matrix in the GMM quadratic form). Most of the previous simulation work considered a fixed degree of smoothing (see, for example, Gregory et al., 2002; Guggenberger and Smith, 2008). Otsu (2006) did look at fixed smoothing but also looked at applying the automatic bandwidth selection rule of Newey and West (1994). We follow Otsu's approach and consider fixed values of $K_T = \{0, 1, 2\}$ as well as a value of K_T selected using the data-driven procedure. For the former, the highest fixed K_T is set to 2 due to small samples issues. Indeed, with a 15% trimming and 200 observations, we have samples of 30 observations with only 6 nonoverlapping blocks of data such that $\frac{2K_T+1}{[T\pi]} = \frac{5}{30}$. For the latter, we select a bandwidth parameter, m_T, via an automatic, data-driven procedure, which is then transformed as $K_T^{m_T} = [(m_T - 1)/2]$. In our simulations, K_T ranges from 1.4 to 2.3. The smoothing parameter was increasing in the autoregressive parameter β and in α. In general, as the smoothing parameter increases the rejection frequencies deteriorates. For example, the LM test, (20), and the O_T test, (22), become more conservative, while the rejection frequencies for the other tests increase. Guay (2003) obtained similar results in the GMM context with the data driven procedure but with more size distortions in the majority of the cases.

The rejection frequencies, associated with a nominal size of 5%, are presented in Table 1. Power figures (for DGPs 5 and 8) are size-corrected as rejection frequencies under the null were distorted for some values of K_T. For DGP 5 we used corrected critical values from the null DGP 1, while for DGP 8 critical values were obtained from DGP 2.[8] In these tables, the second column contains the

TABLE 1. Rejection frequencies for tests of structural change

DGP	K_T	supIPSC	supLM	supW	supLR	supIPSC^I	supIPSC^O	sup$O$
DGP1	0	0.0655	0.0365	0.0855	0.0825	0.0760	0.0715	0.0210
	1	0.0835	0.0230	0.1390	0.0985	0.0945	0.1190	0.0075
	2	0.1045	0.0190	0.2010	0.1245	0.1235	0.1920	0.005
	$K_T^{m_T}$	0.0930	0.0220	0.1605	0.1060	0.1055	0.1515	0.0075
DGP5	0	0.8335	0.9610	0.9510	0.9595	0.9645	0.1455	0.1670
	1	0.8595	0.9205	0.9405	0.9595	0.9690	0.0995	0.1310
	2	0.8350	0.8875	0.9180	0.9555	0.9655	0.1065	0.1225
	$K_T^{m_T}$	0.8460	0.8635	0.9235	0.9560	0.9650	0.1600	0.0800
DGP8	0	0.7360	0.2380	0.2505	0.2700	0.3210	0.8880	0.8935
	1	0.6980	0.1390	0.3525	0.3855	0.5385	0.7625	0.8140
	2	0.6840	0.1280	0.3805	0.4405	0.6365	0.5965	0.7475
	$K_T^{m_T}$	0.6880	0.1275	0.3675	0.4065	0.5895	0.6360	0.7120

Note: DGP 1 corresponds to the null hypothesis of structural stability. The alternative hypotheses of instability in the parameters and in the overidentifying restrictions are investigated under DGPs 5 and 8, respectively. Results for the other DGPs are available in the online version of the paper.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

smoothing parameter labeled as either 0, 1, 2 for the fixed case or $K_T^{m_T}$ for the data-driven case. We present the the supremum, exponential, and average version of the test statistics. Results are presented for the general specification tests IPSC, (10), for the test statistics designed to have power against a structural change in the parameters (test IPSC^I, (14), including the comparison tests Wald, LM, and LR appearing below) as well as for test statistics that are designed to have power against a structural change in the overidentifying restrictions (test IPSC^O, (17), and the comparison O test, (22)). All test statistics were computed in the GEL setting using exponential tilting. The tests used for comparison appear in Guay and Lamarche (2009). For completeness, we report the Wald, LM, LR, and O tests (also computed using exponential tilting). The Wald test is

$$Wald_T(s) = T\left(\hat{\beta}_{1T}(s) - \hat{\beta}_{2T}(s)\right)'\left(\hat{V}_\Omega(s)\right)^{-1}\left(\hat{\beta}_{1T}(s) - \hat{\beta}_{2T}(s)\right), \tag{19}$$

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

$$\hat{G}^{\beta}_{1,tT}(s) = \frac{1}{[Ts]}\sum_{t=1}^{[Ts]}\frac{\partial g_{tT}\left(\hat{\beta}_1(s), \hat{\delta}(s)\right)}{\partial \beta_1'},$$

$$\widehat{\Omega}_{1T}(s) = \frac{2K_T+1}{[Ts]}\sum_{t=1}^{[Ts]} g_{tT}(\beta_1(s), \delta)\, g_{tT}\left(\hat{\beta}_1(s), \hat{\delta}(s)\right)',$$

and for the second part of the sample,

$$\widehat{G}^{\beta}_{2,tT} = \frac{1}{T-[Ts]}\sum_{t=[Ts]+1}^{T}\frac{\partial g_{tT}\left(\hat{\beta}_2(s), \hat{\delta}(s)\right)}{\partial \beta_2'},$$

$$\widehat{\Omega}_{2T}(s) = \frac{2K_T+1}{T-[Ts]}\sum_{t=[Ts]+1}^{T} g_{tT}\left(\hat{\beta}_2(s), \hat{\delta}(s)\right) g_{tT}\left(\hat{\beta}_2(s), \hat{\delta}(s)\right)'.$$

The Lagrange multiplier statistic is given by

$$LM_T(s) = \frac{T}{s(1-s)} g_{1T}\left(\tilde{\theta}_T, s\right)'\widetilde{\Omega}_T^{-1}\tilde{G}^{\beta}_{tT}\left[\left(\tilde{G}^{\beta}_{tT}\right)'\widetilde{\Omega}_T^{-1}\tilde{G}^{\beta}_{tT}\right]^{-1} \times \left(\tilde{G}^{\beta}_{tT}\right)'\widetilde{\Omega}_T^{-1} g_{1T}\left(\tilde{\theta}_T, s\right), \tag{20}$$

where $g_{1T}(\tilde{\theta}_T, s) = \frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T)$, $\tilde{G}^{\beta}_{tT} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_{tT}(\tilde{\beta}, \tilde{\delta})}{\partial \beta'}$, $\widetilde{\Omega}_T = \frac{2K_T+1}{T}$ $\sum_{t=1}^{T} g_{tT}(\tilde{\beta}, \tilde{\delta}) g_{tT}(\tilde{\beta}, \tilde{\delta})'$.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

The LR-like statistic is defined as

$$LR_T(s) = \frac{2T}{2K+1}\left[\sum_{t=1}^{T}\frac{\left[\rho\left(\tilde{\gamma}_T(s)'g_{tT}\left(\tilde{\theta},s\right)\right)-\rho_0\right]}{T} - \sum_{t=1}^{T}\frac{\left[\rho\left(\hat{\gamma}_T(s)'g_{tT}\left(\hat{\theta}(s),s\right)\right)-\rho_0\right]}{T}\right], \tag{21}$$

with $\tilde{\gamma}_T(s) = \left(\tilde{\gamma}_{2T}(\tilde{\beta}_T,\tilde{\gamma}_T,s)', \tilde{\gamma}_{1T}(\tilde{\beta}_T,\tilde{\gamma}_T,s)'\right)'$ as defined in equations (11) and (12). And finally, the test statistic proposed by Hall and Sen (1999) is

$$O_T(s) = O1_T(s) + O2_T(s), \tag{22}$$

where

$$O1_T(s) = \left[\frac{1}{\sqrt{[Ts]}}\sum_{t=1}^{[Ts]} g_{tT}\left(\hat{\beta}_1(s),\hat{\delta}(s)\right)\right]' \hat{\Omega}_{1T}^{-1}(s)\left[\frac{1}{\sqrt{[Ts]}}\sum_{t=1}^{[Ts]} g_{tT}\left(\hat{\beta}_1(s),\hat{\delta}(s)\right)\right]$$

and

$$O2_T(s) = \left[\frac{1}{\sqrt{(T-[Ts])}}\sum_{t=[Ts]+1}^{[T]} g_{tT}\left(\hat{\beta}_2(s),\hat{\delta}(s)\right)\right]' \times \hat{\Omega}_{2T}^{-1}(s)\left[\frac{1}{\sqrt{(T-[Ts])}}\sum_{t=[Ts]+1}^{[T]} g_{tT}\left(\hat{\beta}_2(s),\hat{\delta}(s)\right)\right].$$

We begin the discussion of the results with some general comments. First, even if all test statistics showed some overrejection or underrejection, the new test statistics based on implied probabilities, IPSC^I, (14), and IPSC^O, (17), achieve size distortions comparable to those of the traditional tests and better power under the alternatives, while the general specification test IPSC, (10) has some power against both a change in the parameters and in the overidentifying restrictions. Second, the modified tests IPSCM, IPSCM1^I, IPSCM2^I, and IPSCM^O (given by (13), (15), (16), and (18)) have rejection frequencies that are much too large and for this reason are not reported in the tables. The intuition for this is that the modified test statistics contain a more volatile term in the denominator that can inflate the value of the tests and hence increase the rejection frequencies. Finally, in terms of smoothing, we note that the data-driven choice of K_T was on average somewhere between 1 and 2, depending on the DGP. This explains why the results for the data-driven K_T are similar to those obtained with a fixed smoothing parameter of 1 or 2. In general, a higher K_T increases size except for LM and O tests, for which the size decreases and reduces power. We now turn to the results in more detail.

Focusing first on size (DGP 1), we observe some overrejection for all tests with the exception of the LM, (20), and O, (22), tests that are conservative

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

(as mentioned above, a similar results was found in the GMM context). The degree of overrejection can be as small as about 1% (usually for the average mapping) to as large as 15% (for the supremum mapping). Typically, the size of all tests worsens with an increasing K_T. Under the null of no structural change and $K_T = 0$, all tests deliver an empirical size close to the nominal size. This is not surprising, since for those cases the error terms are uncorrelated. The data-driven procedure, which is specifically designed to minimize the mean squared error of the GMM estimator, seems not to be well adapted for structural change testing procedure. In fact, there is no theoretical reason that such a data-driven procedure tailored for estimation purposes delivers good performance in a testing context. Among the tests considered in this paper, the IPSC test, (10), obtained the best rejection frequencies under the null hypothesis. Further, the performance of the IPSC^I test, (14) is, in general, comparable to the LR test and better than the Wald test. Based on the small sample size properties, no test clearly dominates the others, and the empirical size is close to the nominal size for $K_T = 0$ for all tests.

The study of power is divided into two cases. In case 1, structural change occurs in the parameter β, while in case 2, structural change occurs in the overidentifying restrictions. Under the alternative of instability in the parameter, H_A^I (DGP 5), we see that the newly proposed test statistics based on implied probabilities have good rejection frequencies. The power of these test statistics is greater than the power of the standard Wald, LM, and LR tests. The more general specification test IPSC, (10), has some reasonable power.

Under the alternative of instability in the overidentifying restrictions, e.g., H_A^O, (DGP 8), we see that the test statistics specially designed to detect a change in the parameter IPSC^I, (14), and the LM, Wald, and LR tests have some power, while the targeted IPSC^O, (17), and O, (22), tests have very good power.[9] Importantly the average tests based on implied probabilities are seen to have higher power than the O tests in all cases except one. As expected, the general specification tests based on implied probabilities IPSC, (10), have rejection frequencies that fall in between those of IPSC^O, (17), and IPSC^I, (14).

6. CONCLUSION

In this paper, we suggest different types of testing procedures for the detection of a structural change. Each procedure is based on different measures of the discrepancy between the estimated weights and the unconstrained weights $\frac{1}{T}$ computed in a generalized empirical likelihood (GEL) setting. Specifically, we propose general structural change tests to detect instability both in the identifying restrictions and in the overidentifying restrictions, instability in the identifying restrictions, and instability in the overidentifying restrictions. We found that tests based on these implied probabilities in a well-identified context have comparable finite sample size and better power properties than standard tests. We also propose a boundedly pivotal structural change test based on implied probabilities robust to cases of weak identification and cases in which parameters are completely unidentified.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

The relative small sample performance of this test compared to alternative tests proposed by Caner (2007) is left for future work. Moreover, additional research is needed to improve the small sample size properties of existing structural change tests relative to the choice of the smoothing parameter in GMM and GEL settings.

Another interesting extension is to develop optimal tests like the ones proposed by Rossi (2005), but based on implied GEL probabilities, to jointly test for parameter instability and for a null hypothesis on a subset of the parameters. Such tests (much like those used in this paper) would eliminate the need to compute the weighting matrix as required to perform optimal tests proposed by Rossi and thereby could improve the small sample properties.

NOTES

1. See in particular the special issue of *Journal of Business and Economic Statistics*, 1996, vol. 14.
2. The function $h(\cdot)$ allows for a wide range of alternative hypotheses (see Sowell, 1996b). 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 $0 < \tau_1 < \tau_2 < \cdots < \tau_j < 1$ and θ^* is in the interior of Θ. Therefore it can accommodate multiple breaks.
3. This specification allows for the overidentifying restrictions to be violated just after the breakpoint ($\eta_1 = 0$ and $\eta_2 \neq 0$), just before the breakpoint ($\eta_1 \neq 0$ and $\eta_2 = 0$), or both ($\eta_1 \neq 0$, $\eta_2 \neq 0$ and $\eta_1 \neq \eta 2$).
4. We can also show that the asymptotic distribution of the statistic $IPSC_T^{GEL}(s)$ evaluated at the restricted implied probabilities is equal to $s^2\left[B_{q-p}(1)'B_{q-p}(1) + H(1)'\Omega^{-1/2}(I - P_G)\Omega^{-1/2}H(1)\right]$ such that the same comments hold.
5. Tables of critical values can be obtained at http://coffee.econ.brocku.ca/jfl/research.
6. 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 $2q - 2r$.
7. The full set of results is available in a working paper at http://ideas.repec.org/e/pla154.html.
8. The results are not qualitatively different depending on the use of size-corrected critical values from DGP 1 or DGP 2.
9. The result that the tests specifically designed to detect structural change in parameter values have some power is not surprising. Indeed, the violation of the overidentifying restrictions results in a biased estimator of the parameter β in the second part of the sample.

REFERENCES

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

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

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

Antoine, B., H. Bonnal, & E. Renault (2007) On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics* 138, 461–487.

Back, K. & D.P. Brown (1993) Implied probabilities in GMM estimators. *Econometrica* 61, 971–975.

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.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

Caner, M. (2008a) Exponential Tilting and Weak Instruments: Estimation and Testing. Working paper, North Carolina State University.
Caner, M. (2008b) Nearly-singular design in GMM and generalized empirical likelihood estimators. *Journal of Econometrics* 144, 511–523.
Cressie, N. & T. Read (1984) Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society B* 46, 440–464.
Gallant, A.R. & H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell.
Ghysels, E., A. Guay, & A. Hall (1997) Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics* 82, 209–233.
Ghysels, E. & A. Hall (1990) Are consumption-based intertemporal capital asset pricing models structural. *Journal of Econometrics* 45, 121–139.
Gregory, A.W., J.-F. Lamarche, & G.W. Smith (2002) Information-theoretic estimation of preference parameters: Macroeconomic applications and simulation evidence. *Journal of Econometrics* 107, 213–233.
Guay, A. (2003) Optimal predictive tests. *Econometric Reviews* 22(4), 379–410.
Guay, A. & J.-F. Lamarche (2009) Structural Change Tests for GEL Criteria. Working paper 0904, Brock University.
Guay, A. & F. Pelgrin (2007) Using Implied Probabilities to Improve Estimation in Unconditional Moment Models: Re-assessing the New Keynesian Phillips Curve, Manuscript.
Guggenberger, P., J.J.S. Ramalho, & R.J. Smith (2007) GEL Pearson-type Statistics under Weak Identification. Manuscript.
Guggenberger, P. & R.J. Smith (2008) Generalized empirical likelihood tests in time series models with potential identification failure. *Journal of Econometrics* 142, 134–161.
Hall, A.R., A. Inoue, & F.P.M. Peixe (2003) Covariance matrix estimation and the limiting behavior of the overidentifying restrictions test in the presence of neglected structural instability. *Econometric Theory* 19, 962–983.
Hall, A.R. & A. Sen (1999) Structural stability testing in models estimated by generalized methods of moments. *Journal of Business and Economic Statistics* 17, 335–348.
Hall, P. & J.L. Horowitz (1996) Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 64, 891–916.
Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
Hansen, L.P., J. Heaton, & 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, & P. Johnson (1998) Information theoretic approaches to inference in moment condition models. *Econometrica* 66, 333–357.
Kitamura, Y. (2001) Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 69, 1661–1672.
Kitamura, Y. & M. Stutzer (1997) An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 65, 861–874.
Kullback, S. & R.A. Leibler (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.
Li, H. & U.K. Müller (2009) Valid inference in partially unstable generalized method of moments models. *Review of Economic Studies* 76, 343–365.
Mittelhammer, R.C., G.G. Judge, & D.J. Miller (2000) *Econometric Foundations*, 1st ed. Cambridge University Press.
Newey, W.K. & R.J. Smith (2004) Higher order properites of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–256.
Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–654.
Otsu, T. (2006) Generalized empirical likelihood inference for nonlinear and time series models under weak identification. *Econometric Theory* 22, 513–527.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag.
Qin, J. & J. Lawless (1994) Empirical likelihood and general estimating equations. *Annals of Statistics* 22, 300–325.
Ramalho, J.J.S. & R.J. Smith (2005) Goodness of Fit Tests for Moment Condition Models. Working paper 5_2005, University of Évora.
Rossi, B. (2005) Optimal tests for nested model selection with underlying parameter instability. *Econometric Theory* 21, 962–990.
Schennach, S.M. (2007) Point estimation with exponetially tilted empirical likelihood. *Annals of Statistics* 35, 634–672.
Shannon, C.E. (1948) Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* 26, 26–37.
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 M. Salmon and P. Marriott (eds.), *Applications of Differential Geometry to Econometrics*. pp. 119–150. Cambridge University Press.
Smith, R.J. (2004) GEL Criteria for Moment Condition Model. CEMMAP Working Paper CWP19/04.
Sowell, F. (1996a) Optimal tests of parameter variation in the generalized method of moments framework. *Econometrica* 64, 1085–1108.
Sowell, F. (1996b) Test for Violations of Moment Conditions. Manuscript, Carnegie Mellon University.
Stock, James H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.

APPENDIX

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

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

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

Define the smoothed moment conditions as

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

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

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

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

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

Assumption A.2. $K_T\to\infty$ as $T\to\infty$ and $K_T=O_p\left(T^{1/2\eta}\right)$ for some $\eta>1$.

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

Assumption A.3. For some $d>\max\left(2,\frac{2\eta}{\eta-1}\right)$, $\{g(x_{Tt},\beta,\delta):t\le T,T\ge 1\}$ is a triangular array of mean zero R^q-valued random variables that are L^2-near epoch dependent of size $-\frac{1}{2}$ on a strong mixing base $\{Y_{Tt}:t=\ldots,0,1,\ldots;T\ge 1\}$, of size $-d/(d-2)$ and $\sup\|g(x_{Tt},\beta,\delta)\|^d<\infty$.

Assumption A.4. $\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g(x_{Tt},\beta,\delta)\right)\to s\Omega\ \ \forall s\in[0,1]$ for some positive definite $q\times q$ matrix Ω.

Guay and Lamarche (2009) show that 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. In the following, x_t is used to denote x_{Tt} for notational simplicity.

Assumption A.5. $\tilde{g}(\beta_0,\delta_0)=0$ with $(\beta_0,\delta_0)\in B\times\Delta$ where $\tilde{g}(\beta,\delta)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}g(x_t,\beta,\delta)$ and B and Δ are bounded subsets of R^r and R^v, $g(x_t,\beta,\delta)$ is continuous in x for all $(\beta,\delta)\in B\times\Delta$ and is continuous in (β,δ) uniformly over $(\beta,\delta,x)\in B\times\Delta\times C$ for all compact sets $C\subset\mathbf{X}$.

Assumption A.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 $g(\theta,s)=\left(s\tilde{g}(\beta_1,\delta)',(1-s)\tilde{g}(\beta_2,\delta)'\right)'$.

Assumption A.7. (a) $\rho(\cdot)$ is twice continuously differential and concave on its domain, an open interval Φ containing 0, $\rho_1=\rho_2=-1$; (b) $\gamma(s)\in\Gamma(s)$ where $\Gamma(s)=\{\gamma(s):\|\gamma(s)\|\le D(T/(K_T^2)^{-\zeta}\}$ for some $D>0$ with $\frac{1}{2}>\zeta>\frac{1}{d(1-1/\eta)}$.

Under Assumptions A.1, A.2, A.3, A.5, A.6, and A.7, Guay and Lamarche (2009) show for the partial-sample GEL that $\sup_{s\in S}\|\hat{\theta}_T(s)-\theta_0\|\xrightarrow{p}0$, $\sup_{s\in S}\|\hat{\gamma}_T(s)\|\xrightarrow{p}0$, $\|\hat{\gamma}_T(s)\|=O_p(T/K_T^2)^{-1/2}$ 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 A.6 and A.7 (b) (see Guay and Lamarche, 2009). The following high-level assumptions are sufficient to derive the weak convergence under the null and the alternative of the PS-GEL estimators $\hat{\theta}_T(s)$ and $\hat{\gamma}_T(s)$ (see Guay and Lamarche). These assumptions are similar to the ones in Andrews (1993).

Assumption A.8. $\sup_{s\in S}\|\hat{\Omega}_{iT}(s)-\Omega\|\xrightarrow{p}0$ where Ω is defined in Section 3.1 and S whose closure lies in $(0,1)$ for $i=1,2$.

Assumption A.8 holds under conditions given in Andrews (1991) and Lemma A.3 in Smith (2004). To respect these conditions, Assumption A.3 can be replaced by the following assumption.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

Assumption A.3′. $\{g(x_{Tt},\theta) : t \leq T, T \geq 1\}$ 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 $\sup_{t\leq T, T\geq 1}\mathbb{E}\|g(x_{Tt},\theta)\|^d < \infty$ for some $d > \max(4\nu, \frac{2\eta}{\eta-1})$.

Assumptions A.3′ and A.8 guarantee for the full-sample and partial-sample GEL that

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

and

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

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

Assumption A.9. $g(x,\beta,\delta)$ is differentiable in $(\beta,\delta), \forall(\beta,\delta) \in B_0\times\Delta_0\ \forall x \in \mathbf{X_0} \subset \mathbf{X}$ for a Borel measurable set $\mathbf{X_0}$ that satisfies $P(x_t \in \mathbf{X_0}) = 1 \forall t \leq T, T \geq 1$, $g(x,\beta,\delta)$ is Borel measurable in $x\ \forall(\beta,\delta) \in B_0\times\Delta_0$, $\partial g(x_t,\beta,\delta)/\partial\left(\beta',\delta'\right)$ is continuous in (x,β,δ) on $\mathbf{X}\times B_0\times\Delta_0$,

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

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

Assumption A.10. $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{[Ts]}\mathbb{E}g_{tT}(\beta,\delta)$ exists uniformly over $(\beta,\delta,s) \in B\times\Delta\times S$ and equals $s\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}g(x_t,\beta,\delta) = s\tilde{g}(\beta,\delta)$.

Assumption A.11. $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{[Ts]}\mathbb{E}\partial g_{tT}(\beta_0,\delta_0)/\partial\left(\beta',\delta'\right)$ exists uniformly over $s \in S$ and equals $sG\ \forall s \in S$ and S whose closure lies in $(0,1)$.

Assumption A.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 $(0,1)$.

Lemmas. The following lemmas are necessary to establish the proofs of the theorems.

LEMMA A.1. *We denote* $\{B(s) : s \in [0,1]\}$ *as* q*-dimensional vectors of mutually independent Brownian motion 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 motion. Under Assumptions A.1 to A.12, every sequence of PS-GEL estimators* $\{\hat{\theta}_T(\cdot),\hat{\gamma}_T(\cdot), T \geq 1\}$ *satisfies*

$$\sqrt{T}\left(\hat{\theta}_T(\cdot)-\theta_0\right) \Rightarrow -\left(G(\cdot)'\Omega(\cdot)^{-1}G(\cdot)\right)^{-1}G(\cdot)'\Omega(\cdot)^{-1}J(\cdot),$$

$$\frac{\sqrt{T}}{2K_T+1}\hat{\gamma}_T(\cdot) \Rightarrow -\left(\Omega(\cdot)^{-1}-\Omega(\cdot)^{-1}\left(G(\cdot)'\Omega(\cdot)^{-1}G(\cdot)\right)^{-1}G(\cdot)'\Omega(\cdot)^{-1}\right)J(\cdot)$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

as a process indexed by $s \in S$, where S has closure in (0,1). Further, the sequence of GEL estimators $\hat{\theta}_T(\cdot)$ and the sequence of auxiliary estimators $\hat{\gamma}_T(\cdot)$ are asymptotically uncorrelated. Under the alternative (6), the same results hold except that

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

where $H(s) = \int_0^s h(r)dr$ with $h(r) = h(\eta, \tau, r)$ to simplify the notation.

Proof of Lemma A.1. See Guay and Lamarche (2009).

LEMMA A.2. *Under Assumptions A.1 to A.12, for the unrestricted implied probabilities evaluated at the restricted estimator, we get*

$$\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) = \frac{1}{T} + \frac{1}{T} g_{tT}\left(\tilde{\theta}_T, s\right)' \tilde{\gamma}_T(s)(1 + o_p(1)) + O_p\left(K_T / T^2\right)$$

and $\pi_t^{GEL}\left(\tilde{\theta}_T\right) = \frac{1}{T} + o_p(1)$ uniformly in $t = 1, \ldots, T$. According to the notation in Definition 3.2, for $t = 1, \ldots, [Ts]$,

$$T\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) - 1 = g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)' \tilde{\gamma}_{1T}(s)(1 + o_p(1)) + O_p\left(K_T / T\right),$$

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

$$T\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) - 1 = g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)' \tilde{\gamma}_{2T}(s)(1 + o_p(1)) + O_p\left(K_T / T\right).$$

Proof of Lemma A.2. We need to derive the asymptotic distribution of the partial-sample implied probabilities evaluated at the full-sample estimator:

$$\pi_t^{GEL}(\tilde{\theta}_T, s) = \frac{\rho_1\left(\tilde{\gamma}_T(s)' g_{tT}\left(\tilde{\theta}_T, s\right)\right)}{\sum_{t=1}^T \rho_1\left(\tilde{\gamma}_T(s)' g_{tT}\left(\tilde{\theta}_T, s\right)\right)}.$$

with $\tilde{\gamma}_T(s) = (\tilde{\gamma}_{1T}(s)', \tilde{\gamma}_{2T}(s)')'$, $\tilde{\gamma}_{1T}(s) = \tilde{\gamma}_{1T}(\tilde{\beta}_T, \tilde{\gamma}_T, s)$, and $\tilde{\gamma}_{1T}(s) = \tilde{\gamma}_{1T}(\tilde{\beta}_T, \tilde{\gamma}_T, s)$. A mean value expansion for $\pi_t^{GEL}(\tilde{\theta}_T, s)$ around $(\tilde{\theta}_T, \tilde{\gamma}_T(s)) = (\tilde{\theta}_T, 0)$ yields

$$\begin{aligned} \pi_t^{GEL}(\tilde{\theta}_T, s) \;=\; & \frac{1}{T} + \frac{\rho_2\left(\bar{\gamma}_T(s)' g_{tT}\left(\widetilde{\theta}_T, s\right)\right)}{\sum_{t=1}^T \rho_1\left(\bar{\gamma}_T(s)' g_{tT}\left(\widetilde{\theta}_T, s\right)\right)} g_{tT}\left(\widetilde{\theta}_T, s\right)' \widetilde{\gamma}_T(s) \\ & - \frac{\rho_1\left(\bar{\gamma}_T(s)' g_{tT}\left(\widetilde{\theta}_T, s\right)\right)}{\left[\sum_{t=1}^T \rho_1\left(\bar{\gamma}_T(s)' g_{tT}\left(\widetilde{\theta}_T, s\right)\right)\right]^2} \\ & \times \sum_{t=1}^T \rho_2\left(\bar{\gamma}_T(s)' g_{tT}\left(\widetilde{\theta}_T, s\right)\right) g_{tT}\left(\widetilde{\theta}_T, s\right)' \widetilde{\gamma}_T(s), \end{aligned}$$

where $\bar{\gamma}_T(s)$ lies on the line segment joining $\widetilde{\gamma}_T(s)$ and 0 and may differ from row to row.

Since $\bar{\gamma}_T(s)$ converges in probability to 0 and using Lemma 4.3 in Guay and Lamarche (2009), this yields that $\rho_j(\bar{\gamma}_T(s)' g_{tT}(\widetilde{\theta}_T, s)) = \rho_j(0) + o_p(1) = -1 + o_p(1)$, for $j = 1, 2$, $\forall t$. Thus, we get

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

$$\pi_t^{GEL}(\tilde{\theta}_T, s) = \frac{1}{T} + \frac{1}{T}\left[g_{tT}\left(\tilde{\theta}_T, s\right)'\tilde{\gamma}_T(s)\left(1+o_p(1)\right)\right] - \frac{1}{T^2}\left[\sum_{t=1}^{T} g_{tT}\left(\tilde{\theta}_T, s\right)'\tilde{\gamma}_T(s)\left(1+o_p(1)\right)\right].$$

As $\sum_{t=1}^{T} g(x_t, \tilde{\beta}_T, \tilde{\delta}_T) = \sum_{t=1}^{T} g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) + O_p\left(K^2/(2K+1)\right) = O_p(T^{1/2})$, which implies $\sum_{t=1}^{T} g_{tT}(\tilde{\theta}_T, s) = O_p(T^{1/2})$, $g_{tT}(\tilde{\theta}_T, s) = O_p((2K_T+1)^{-1/2})$, and $\tilde{\gamma}_T(s) = O_p(2K_T + 1/\sqrt{T})$ yields (see Kitamura and Stutzer, 1997; Guay and Lamarche, 2009)

$$\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) = \frac{1}{T} + \frac{1}{T}\left[g_{tT}\left(\tilde{\theta}_T, s\right)'\tilde{\gamma}_T(s)\left(1+o_p(1)\right)\right] + O_p\left(K_T/T^2\right)$$

and

$$\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) = \frac{1}{T}\left(1+o_p(1)\right)$$

uniformly in $t = 1, \ldots, T$. Thus, we get

$$T\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) - 1 = g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)'\tilde{\gamma}_{1T}(s)(1+o_p(1)) + O_p\left(K_T/T\right)$$

uniformly in $t = 1, \ldots, [Ts]$ and

$$T\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) - 1 = g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)'\tilde{\gamma}_{2T}(s)(1+o_p(1)) + O_p\left(K_T/T\right)$$

uniformly in $t = [Ts]+1, \ldots, T$.

Proofs of Theorems.

Proof of Theorem 4.1. We expand the FOC of the Lagrange multiplier vector for the partial-sample GEL evaluated at the restricted estimator in a Taylor series about 0 for the first part of the sample $t = 1, \ldots, [Ts]$. Thus

$$\frac{1}{T}\sum_{t=1}^{[Ts]} \rho_1\left(\tilde{\gamma}_{1T}(s)'g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)\right) g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)$$

$$= -\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) - \frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)'\tilde{\gamma}_{1T}(s)$$

$$+ \frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)\sum_{j=2}^{\infty}\frac{1}{j!}\rho_{j+1}(0)\left(g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)'\tilde{\gamma}_{1T}(s)\right)^j,$$

where $\tilde{\gamma}_{1T}$ is the partial-sample Lagrange multiplier evaluated for the restricted estimator for $t = 1, \ldots, [Ts]$ and $\rho_1(0) = \rho_2(0) = -1$. By the fact that $g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) = O_p((2K_T+1)^{-1/2})$, this yields

$$\frac{1}{T}\sum_{t=1}^{[Ts]} \rho_1\left(\tilde{\gamma}_{1T}(s)'g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)\right) g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)$$

$$= -\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) - \frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)'\tilde{\gamma}_{1T}(s)$$

$$+ O_p\left((2K_T+1)^{-3/2}\|\tilde{\gamma}_{1T}(s)\|^2\right).$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

By using a consistent estimator $\tilde{\Omega}_T = \frac{2K+1}{[Ts]} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T)'$ and $\tilde{\gamma}_{1T}(s) = O_p\left((2K_T+1)/\sqrt{T}\right)$, we get under the null,

$$0 = \frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) + \frac{s}{2K+1} \tilde{\Omega}_T \tilde{\gamma}_{1T}(s) + O_p\left((2K_T+1)^{-3/2} \|\tilde{\gamma}_{1T}(s)\|^2\right),$$

which yields

$$\frac{s}{2K_T+1} \tilde{\gamma}_{1T}(s) = -\tilde{\Omega}_T^{-1} \frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) + O_p\left(K_T^{1/2}/T\right). \tag{A.1}$$

We also can easily show for the restricted estimators under the null that (see also Smith, 2004)

$$\sqrt{T}\begin{pmatrix} \tilde{\beta}_T - \beta_0 \\ \tilde{\delta}_T - \delta_0 \end{pmatrix} = -\left(G'\Omega^{-1}G\right)^{-1} G'\Omega^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{tT}(\beta_0, \delta_0) + o_p(1). \tag{A.2}$$

By Lemma 6.1 in Guay and Lamarche (2009), under the null of no structural change, we have

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\beta_0, \delta_0) \Rightarrow B(s), \tag{A.3}$$

and under the generic alternative (6),

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\beta_0, \delta_0) \Rightarrow B(s) + \Omega^{-1/2} H(s), \tag{A.4}$$

where $B(s)$ is a q-vector of standard Brownian motions and $H(s) = \int_0^s h(r)dr$. Note also that

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

Using Lemma A.2 and eq. (A.1),

$$\begin{aligned} T\pi_t^{GEL}\left(\tilde{\theta}_T, s\right) - 1 \;=\; & -\left(\frac{2K+1}{s} \tilde{\Omega}_T^{-1} \frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) + O_p\left(K_T^{3/2}/T\right)\right)' \\ & \times g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)\left(1 + o_p(1)\right) + O_p(K_T/T). \end{aligned} \tag{A.5}$$

Let us now examine the asymptotic distribution of the expression on the right-hand side. First, we show the following asymptotic result for the partial sums of the moments conditions evaluated at $\tilde{\theta}_T$:

$$\Omega^{-1/2}\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right)\right] \Rightarrow B(s) - s\Omega^{-1/2} G[G'\Omega^{-1}G]^{-1} G'\Omega^{-1/2} B(1).$$

By a mean value expansion,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T, \tilde{\delta}_T\right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\beta_0, \delta_0) + \frac{1}{T} \sum_{t=1}^{[Ts]} \frac{\partial g_{tT}\left(\bar{\beta}_T, \bar{\delta}_T\right)}{\partial\left(\beta', \delta'\right)} \begin{pmatrix} \tilde{\beta}_T - \beta_0 \\ \tilde{\delta}_T - \delta_0 \end{pmatrix},$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

where $\bar{\beta}_T$ lies on the line segment joining $\tilde{\beta}_T$ and β_0 and may differ from row to row and respectively for $\bar{\delta}_T$. By applying equation (A.2), we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}(\beta_0,\delta_0) - sG[G'\Omega^{-1}G]^{-1}G' \times \Omega^{-1}\frac{1}{T}\sum_{t=1}^{T} g_{tT}(\beta_0,\delta_0)+o_p(1).$$

It follows by (A.3) that

$$\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right) \Rightarrow B(s) - s\Omega^{-1/2}G[G'\Omega^{-1}G]^{-1}G'\Omega^{-1/2}B(1). \qquad \textbf{(A.6)}$$

By (A.5), the partial sum can be shown to be

$$\sum_{t=1}^{[Ts]}[T\ \pi_t^{GEL}\left(\tilde{\theta}_T,s\right)-1]^2 = \left(\frac{2K+1}{s}\tilde{\Omega}_T^{-1}\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right)' \sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right) g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)' \times \left(\frac{2K+1}{s}\tilde{\Omega}_T^{-1}\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right)+o_p(1).$$

Considering that $\tilde{\Omega}_T = \frac{2K+1}{[Ts]}\sum_{t=1}^{[Ts]} g_{tT}(\tilde{\beta}_T,\tilde{\delta}_T)g_{tT}(\tilde{\beta}_T,\tilde{\delta}_T)'$, the expression above gives

$$\sum_{t=1}^{[Ts]}[T\ \pi_t^{GEL}(\tilde{\theta}_T,s)-1]^2 = \frac{2K+1}{s}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right)'\tilde{\Omega}_T^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}\left(\tilde{\beta}_T,\tilde{\delta}_T\right)\right)+o_p(1).$$

The partial sum of interest yields (by equation (A.6) and the consistency of $\tilde{\Omega}_T$)

$$\frac{s}{2K+1}\sum_{t=1}^{[Ts]}\left[T\pi_t^{GEL}\left(\tilde{\theta}_T,s\right)-1\right]^2 \Rightarrow \left[B(s)-sP_GB(1)\right]'\left[B(s)-sP_GB(1)\right]. \qquad \textbf{(A.7)}$$

The expression $\left[B(s)-sP_GB(1)\right]$ can be rewritten as $\left[I-P_G\right]B(s)+P_G\left[B(s)-sB(1)\right]$. We can now decompose: $\Omega^{-1/2}G(G'\Omega^{-1}G)^{-1}G'\Omega^{-1/2} = C'\Lambda C$, where $CC' = I$, $\Lambda = \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix}$, and I_p is an identity matrix with dimension $p\times p$, where $p = r+\nu$ with r and ν the respective dimensions of the vectors β and δ. Note also that $CB(s)$ has the same asymptotic distribution as $B(s)$. The RHS of expression (A.7) can then be rewritten as $B(s)'\left[I-C'\Lambda C\right]B(s)+[B(s)-sB(1)]'C'\Lambda C\,[B(s)-sB(1)]$, which is equal in distribution to $B_{(q-p)}(s)'B_{(q-p)}(s)+BB_p(s)'BB_p(s)$, where $B_{(q-p)}(s)$ is a $(q-p)$-vector of standard Brownian motion and $BB_p(s) = B_p(s)-sB_p(1)$ is a p-vector of Brownian

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

bridge. The result follows. The asymptotic distribution under the alternative (6) can be easily obtained similarly using (A.4).

Proof of Theorem 4.3. By Lemma A.2 applied to the difference between the partial-sample implied probabilities evaluated at the unrestricted and at the restricted estimators, we get

$$T\left[\hat{\pi}_t(\hat{\theta}_T(s),s)-\tilde{\pi}_t(\tilde{\theta}_T,s)\right] = g_{tT}(\hat{\theta}_T(s),s)'\hat{\gamma}_T(s)(1+o_p(1)) - g_{tT}(\tilde{\theta}_T,s)'\tilde{\gamma}_T(s)(1+o_p(1))+O_p(K_T/T).$$

Let us define the following selection matrices:

$$H_1=\begin{bmatrix} I_{r\times r} & 0_{r\times r} & 0_{r\times v} \\ 0_{v\times r} & 0_{v\times r} & I_{v\times v}\end{bmatrix} \quad \text{and} \quad H_2=\begin{bmatrix} 0_{r\times r} & I_{r\times r} & 0_{r\times v} \\ 0_{v\times r} & 0_{v\times r} & I_{v\times v}\end{bmatrix},$$

and the corresponding estimator: $\hat{\theta}_{1T}(s)=H_1\hat{\theta}_T(s)=(\hat{\beta}_{1T}(s)',\hat{\delta}_T(s)')'$ and $\hat{\theta}_{2T}(s)=H_2\hat{\theta}_T(s)=(\hat{\beta}_{2T}(s)',\hat{\delta}_T(s)')'$. Similarly, for the restricted estimator $\tilde{\theta}_{1T}(s)=H_1\tilde{\theta}_T(s)=(\tilde{\beta}_T(s)',\tilde{\delta}_T(s)')'$ and $\tilde{\theta}_{2T}(s)=H_2\tilde{\theta}_T(s)=(\tilde{\beta}_T(s)',\hat{\delta}_T(s)')'$. Accordingly, we define $\theta_{1,0}=H_1\theta_0=(\beta_0',\delta_0')'$ and $\theta_{2,0}=H_2\theta_0=(\beta_0',\delta_0')'$.

Let us examine the expression for the first part of the sample, namely $t=1,\ldots,[Ts]$. Replacing the unrestricted and the restricted estimators of γ by the corresponding expression (A.1), we get

$$\begin{aligned} &T\left[\pi_t\left(\hat{\theta}_T(s),s\right)-\pi_t\left(\tilde{\theta}_T,s\right)\right] \\ &\quad=-\left[(2K_T+1)\Omega(s)^{-1}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}\left(\hat{\theta}_T(s),s\right)+o_p(1)\right]'g_{tT}\left(\hat{\theta}_T(s),s\right)(1+o_p(1)) \\ &\qquad+\left[(2K_T+1)\Omega(s)^{-1}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}\left(\tilde{\theta}_T,s\right)+o_p(1)\right]'g_{tT}(\tilde{\theta}_T,s)(1+o_p(1)) \\ &\qquad+O_p(K_T/T), \end{aligned}$$

which equals

$$T\left[\pi_t\left(\hat{\theta}_T(s),s\right)-\pi_t\left(\tilde{\theta}_T,s\right)\right]=-\left[\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}\left(\hat{\theta}_{1T}(s)\right)-\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}\left(\tilde{\theta}_{1T}\right)+o_p(1)\right]' \times\Omega^{-1}(2K_T+1)g_{tT}\left(\hat{\theta}_{1T}(s)\right)(1+o_p(1))+O_p(K_T/T),$$

since $G_{tT}(\cdot)=O_p((2K_T+1)^{-1/2})$ and $\hat{\theta}_{1T}(s)-\tilde{\theta}_{1T}=O_p(T^{-1/2})$, $\forall s\in S$.

Now, consider the mean value expansion

$$\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\tilde{\theta}_{1T})=\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\theta}_{1T}(s))+\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}\frac{\partial g_{tT}(\bar{\theta}_{1T}(s))}{\partial\theta_1'}\left(\tilde{\theta}_{1T}-\hat{\theta}_{1T}(s)\right),$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

where $\bar{\theta}_{1T}(s)$ lies on the line segment joining $\hat{\theta}_{1T}(s)$ and $\tilde{\theta}_{1T}$ and may differ from row to row. Thus,

$$T\left[\pi_t(\hat{\theta}_T(s),s)-\pi_t(\tilde{\theta}_T,s)\right]=\left[\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}\frac{\partial g_{tT}(\bar{\theta}_{1T}(s))}{\partial\theta_1'}\left(\tilde{\theta}_{1T}-\hat{\theta}_{1T}(s)\right)+o_p(1)\right]'$$
$$\times\,\Omega^{-1}(2K_T+1)g_{tT}\left(\hat{\theta}_{1T}(s)\right)(1+o_p(1))+O_p(K_T/T).$$

The partial sum over the first subsample for the square of the LHS expression above yields

$$\frac{1}{2K+1}\sum_{t=1}^{[Ts]}\left[T\pi_t\left(\hat{\theta}_T(s),s\right)-T\pi_t\left(\tilde{\theta}_T,s\right)\right]^2$$
$$=[Ts]\left(\tilde{\theta}_{1T}-\hat{\theta}_{1T}(s)\right)'G'\Omega^{-1}G\left(\tilde{\theta}_{1T}-\hat{\theta}_{1T}(s)\right)+o_p(1)$$

as $(2K_T+1)\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\theta}_{1T}(s),s)g_{tT}(\hat{\theta}_{1T}(s),s)'$ is a consistent estimator of Ω.

Similarly for the second subsample $t=[Ts]+1,\ldots,T$, we obtain

$$\frac{1}{2K+1}\sum_{t=[Ts]+1}^{T}\left[T\pi_t\left(\hat{\theta}_T(s),s\right)-T\pi_t\left(\tilde{\theta}_T,s\right)\right]^2$$
$$=(T-[Ts])\left(\tilde{\theta}_{2T}-\hat{\theta}_{2T}(s)\right)'G'\Omega^{-1}G\left(\tilde{\theta}_{2T}-\hat{\theta}_{2T}(s)\right)+o_p(1).$$

By results derived above and some calculations, we obtain

$$\frac{1}{2K+1}\sum_{t=1}^{T}\left[T\pi_t\left(\hat{\theta}_T(s),s\right)-T\pi_t\left(\tilde{\theta}_T,s\right)\right]^2$$
$$=T\left(\tilde{\theta}_T-\hat{\theta}_T(s)\right)'G(\theta_0,s)'\Omega^{-1}(s)G(\theta_0,s)\left(\tilde{\theta}_T-\hat{\theta}_T(s)\right)+o_p(1).$$

Andrews (1993, pp. 851–852) shows that this expression is asymptotically equivalent to the $LM_T(s)$ statistic for parameter instability. This gives the result under the null. Under the alternative, Guay and Lamarche (2009) show that the $LM_T(s)$ statistic for the restricted GEL estimator has the same asymptotic distribution as in Theorem 4.3. Given the asymptotic equivalence between the expression above and the $LM_T(s)$ statistic, the result follows.

Proof of Theorem 4.4. The two subsamples are now evaluated separately. Using Lemma A.2 but for the estimation of the subsample with the first $[Ts]$ observations and that $g_{tT}(\hat{\theta}_{1T}(s))=O_p((2K_T+1)^{-1/2})$ and $\hat{\gamma}_{1T}(s)=O_p(2K_T+1/\sqrt{[Ts]}$ gives

$$[Ts]\pi_t^{GEL}\left(\hat{\theta}_{1T}(s),s\right)-1=g_{tT}\left(\hat{\theta}_{1T}(s)\right)'\hat{\gamma}_{1T}(s)+o_p\left(K_T^{1/2}/[Ts]^{1/2}\right)$$

uniformly in $t=1,\ldots,[Ts]$. This is obtained with a proof similar to the one of Lemma A.2 but only for the first part of the sample. This yields

$$\left([Ts]\pi_t^{GEL}(\hat{\theta}_{1T}(s),s)-1\right)^2=\hat{\gamma}_{1T}(s)'g_{tT}(\hat{\theta}_{1T}(s))g_{tT}(\hat{\theta}_{1T}(s)'\hat{\gamma}_{1T}(s))+o_p(K_T/[Ts]).$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

By summing the expression above to $t = \ldots, [Ts]$, and by a result similar to (A.1) but for the unrestricted estimator applied to the sample $t = 1, \ldots, [Ts]$, this yields

$$\begin{aligned}
&\frac{1}{2K_T+1}\sum_{t=1}^{[Ts]}\left([Ts]\,\pi_t^{GEL}(\hat{\theta}_{1T}(s),s)-1\right)^2 \\
&\quad=(2K_T+1)\frac{1}{\sqrt{[Ts]}}\sum_{t=1}^{[Ts]}g_{tT}\left(\hat{\theta}_{1T}(s)\right)'\Omega^{-1}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}\left(\hat{\theta}_{1T}(s)\right) \\
&\qquad\times\left(g_{tT}\left(\hat{\theta}_{1T}(s)\right)\right)'\Omega^{-1}\frac{1}{\sqrt{[Ts]}}\sum_{t=1}^{[Ts]}g_{tT}\left(\hat{\theta}_{1T}(s)\right)+o_p(1).
\end{aligned}$$

As $(2K_T+1)\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\theta}_{1T}(s))g_{tT}(\hat{\theta}_{1T}(s))'$ is a consistent estimator of Ω, this gives

$$\begin{aligned}
&\frac{1}{2K_T+1}\sum_{t=1}^{[Ts]}\left([Ts]\,\pi_t^{GEL}(\hat{\theta}_{1T}(s),s)-1\right)^2 \\
&\quad=\frac{1}{\sqrt{[Ts]}}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\theta}_{1T}(s))'\Omega^{-1}\frac{1}{\sqrt{[Ts]}}\sum_{t=1}^{[Ts]}g_{tT}\left(\hat{\theta}_{1T}(s)\right)+o_p(1). \qquad \textbf{(A.8)}
\end{aligned}$$

Similarly, we can easily obtain that

$$\begin{aligned}
&\frac{1}{2K_T+1}\sum_{t=[Ts]+1}^{T}\left((T-[Ts])\,\pi_t^{GEL}\left(\hat{\theta}_{2T}(s),s\right)-1\right)^2 \\
&\quad=\frac{1}{\sqrt{T-[Ts]}}\sum_{t=[Ts]+1}^{T}g_{tT}\left(\hat{\theta}_{2T}(s)\right)' \\
&\qquad\times\Omega^{-1}\frac{1}{\sqrt{T-[Ts]}}\sum_{t=[Ts]+1}^{T}g_{tT}\left(\hat{\theta}_{2T}(s)\right)+o_p(1). \qquad \textbf{(A.9)}
\end{aligned}$$

Now, by a mean value expansion and the consistency of $\hat{\theta}_T(s)$,

$$\begin{aligned}
\Omega(s)^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}(\hat{\theta}_T(s),s) \;=\; &\Omega(s)^{-1/2}\begin{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0)\\ \frac{1}{\sqrt{T}}\sum_{t=[Ts]+1}^{T}g_{tT}(\beta_0,\delta_0)\end{bmatrix} \\
&+\Omega(s)^{-1/2}G(s)\sqrt{T}\left(\hat{\theta}_T(s)-\theta_0\right)+o_p(1).
\end{aligned}$$

By Lemma A.1 and (A.3),

$$\begin{aligned}
\Omega(s)^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}\left(\hat{\theta}_T(s),s\right) \;\Rightarrow\; &\begin{bmatrix}\frac{1}{\sqrt{s}}B(s)\\ \frac{1}{\sqrt{(1-s)}}(B(1)-B(s))\end{bmatrix} \\
&-\Omega(s)^{-1/2}G(s)\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1} \\
&\times G(s)'\Omega(s)^{-1/2}\begin{bmatrix}B(s)\\ (B(1)-B(s))\end{bmatrix}.
\end{aligned}$$

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

The RHS can be rewritten as

$$\left(I-\Omega(s)^{-1/2}G(s)\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1}G(s)'\Omega(s)^{-1/2}\right)\begin{bmatrix}\frac{1}{\sqrt{s}}B(s)\\ \frac{1}{\sqrt{(1-s)}}(B(1)-B(s))\end{bmatrix}.$$

Since the entire parameter vector is estimated for both subsample, e.g., $\hat{\theta}_{1T}(s)=\hat{\beta}_{1T}(s)$ and $\hat{\theta}_{2T}(s)=\hat{\beta}_{2T}(s)$, the matrices $\Omega(s)$ and $G(s)$ are block-diagonal and by the definition of $g_{tT}(\hat{\theta}_T(s),s)$, we get

$$\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\beta}_{1T}(s))$$
$$\Rightarrow\left(I_q-\Omega^{-1/2}G_\beta\left(G_\beta'\Omega^{-1}G_\beta\right)^{-1}G_\beta'\Omega^{-1/2}\right)B(s) \tag{A.10}$$

and

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

We can now decompose: $\left(I-\Omega^{-1/2}G_\beta\left(G_\beta'\Omega^{-1}G_\beta\right)^{-1}G'\Omega^{-1/2}\right)=\left(I-C'\Lambda C\right)$, where $CC'=I$ with $\Lambda=\begin{bmatrix}I_r & 0\\ 0 & 0\end{bmatrix}$ and I_r is an identity matrix with dimension $r\times r$ with r is the dimension of the vectors β. Noting that $C\left(I-C'\Lambda C\right)B(s)=(I-\Lambda)CB(s)$ and $CB(s)$ is also an r-vector of Brownian motion. Now consider the multiplication of the RHS term of equation (A.10) by the matrix C; this yields

$$C\Omega^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\beta}_{1T}(s))\Rightarrow\left(I_q-\Lambda\right)B(s).$$

Using this result, we obtain directly the asymptotic distribution of the statistic for the first subsample by equation (A.8) considering that $C'C=I_q$. The proof is similar for the second subsample (equation A.9), hence obtaining the asymptotic distribution under the null. Under the alternative (6), the asymptotic distribution under the alternative is derived similarly using Lemma A.1.

Proof of Theorem 4.5. Define $P(\theta(s),\gamma(s),s)=\sum_{t=1}^{T}\frac{[\rho(\gamma(s)'g_{tT}(\theta,s))-\rho_0]}{T}$ and $\hat{g}_T(\theta,s)=\frac{1}{T}\sum_{t=1}^{T}g_{tT}(\theta,s)$. Since $\tilde{\theta}_T(s)=\arg\min_{\theta\in\Theta}P(\theta,\gamma(s),s)$, this implies that $\sup_{\gamma(s)\in\tilde{\Gamma}_T(s)}P(\tilde{\theta}_T(s),\gamma(s),s)\leq\sup_{\gamma(s)\in\tilde{\Gamma}_T(s)}P(\theta_0,\gamma(s),s)$, and this holds for all $s\in S$. Let $\bar{\gamma}_T(s)=\arg\max_{\gamma_s\in\Gamma_T(s)}P(\theta_0,\gamma(s),s)$ and $\dot{\gamma}(s)=\tau\bar{\gamma}(s)$, $0\leq\tau\leq1$ and we have $(2K_T+1)\sum_{t=1}^{T}\rho_2(\dot{\gamma}(s)'g_{tT}(\theta_0,s))g_{tT}(\theta_0,s)g_{tT}\theta_0,s)'/T\xrightarrow{p}-\Omega(s)$. By a second-order Taylor expansion with Lagrange remainder,

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

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

w.p.a.1, where the second equality holds by $\frac{1}{2K_T+1}\bar{\gamma}(s)=-\tilde{\Omega}(s)^{-1}\hat{g}_T(\theta_0,s)+o_p(1)$. Similarly, we can obtain

$$\frac{1}{2K_T+1}P(\tilde{\theta}_T(s),\tilde{\gamma}(s),s)=\hat{g}_T\left(\tilde{\theta}_T(s),s\right)'\tilde{\Omega}_T(s)^{-1}\hat{g}_T\left(\tilde{\theta}_T(s),s\right)/2+o_p(1)$$

with $\tilde{\gamma}_T(s)=\arg\max_{\gamma_s\in\Gamma_T(s)}P(\tilde{\theta}_T(s),\gamma(s),s)$ and

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

with

$$\tilde{\Omega}_{1T}(s) \;=\; \frac{2K_T+1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}\left(\tilde{\beta}_T(s)\right)g_{tT}\left(\tilde{\beta}_T(s)\right)'$$

and

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

Next, we show that the statistic $IPSC_T^{R,GEL}(s)$ is asymptotically equivalent at first-order to $\frac{2}{2K_T+1}P(\tilde{\theta}_T(s),\tilde{\gamma}(s),s)$. The following partial sum evaluated at the restricted partial-sample GEL estimator yields

$$\begin{aligned}
&\sum_{t=1}^{[Ts]}\left[T\,\pi_t^{GEL}(\tilde{\theta}_T(s),s)-1\right]^2\\
&\quad=\left(\frac{2K+1}{s}\tilde{\Omega}_{1T}(s)^{-1}\frac{1}{T}\sum_{t=1}^{[Ts]}g_{tT}\left(\tilde{\beta}_T\right)\right)'\sum_{t=1}^{[Ts]}g_{tT}\left(\tilde{\beta}_T(s)\right)g_{tT}\left(\tilde{\beta}_T(s)\right)'\\
&\qquad\times\left(\frac{2K+1}{s}\tilde{\Omega}_{1T}(s)^{-1}\frac{1}{T}\sum_{t=1}^{[Ts]}g_{tT}\left(\tilde{\beta}_T(s)\right)\right)+o_p(1).
\end{aligned}$$

Considering that $\tilde{\Omega}_{1T}(s)=\frac{2K+1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\tilde{\beta}_T(s))g_{tT}(\tilde{\beta}_T(s))'$, the expression above gives

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press

$$\sum_{t=1}^{[Ts]} [T\ \pi_t^{GEL}(\tilde{\theta}_T(s), s) - 1]^2$$

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

Similarly, for the second subsample,

$$\sum_{t=[Ts]+1}^{T} \left[T\ \pi_t^{GEL}\left(\tilde{\theta}_T(s), s\right) - 1 \right]^2 = \frac{2K+1}{1-s} \left(\frac{1}{\sqrt{T}} \sum_{t=[Ts]+1}^{T} g_{tT}\left(\tilde{\beta}_T(s)\right) \right)' \tilde{\Omega}_{2T}(s)^{-1}$$

$$\times \left(\frac{1}{\sqrt{T}} \sum_{t=[Ts]+1}^{T} g_{tT}\left(\tilde{\beta}_T(s)\right) \right) + o_p(1).$$

Then, we get for the statistic

$$IPSC_T^{R,GEL}(s)$$

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

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

$$= T\, \hat{g}_T(\tilde{\theta}_T(s), s)' \tilde{\Omega}_T(s)^{-1} \hat{g}_T(\tilde{\theta}_T(s), s) + o_p(1),$$

with $\tilde{\theta}_T(s) = \left(\tilde{\beta}_T(s)', \tilde{\beta}_T(s)'\right)'$. Since $\frac{2T}{2K_T+1} P(\tilde{\theta}_T(s), \tilde{\gamma}(s), s)$ is asymptotically bounded by $\frac{2T}{2K_T+1} P(\theta_0, \bar{\gamma}(s), s)$ for all $s \in S$ and that $IPSC_T^{R,GEL}(s)$ is asymptotically equivalent at first-order to $\frac{2T}{2K_T+1} P(\tilde{\theta}_T(s), \tilde{\gamma}(s), s)$ for all $s \in S$, the statistic $IPSC_T^{R,GEL}(s)$ is asymptotically bounded for all $s \in S$ by the asymptotic distribution of $T\, \hat{g}_T(\theta_0, s)' \Omega(s)^{-1} \hat{g}_T(\theta_0, s)$, which is derived in Caner (2007). The result under the null follows, and the asymptotic bound under the alternative can be easily derived.

https://doi.org/10.1017/S0266466612000114 Published online by Cambridge University Press