Structural Change tests based on Implied Probabilities for GEL Criteria

Alain Guay^a, Jean-François Lamarche^{b,\dagger}

^a Département de sciences économiques, Université du Québec à Montréal, Québec, Canada. CIRPÉE and CIREQ. E-mail: guay.alain@uqam.ca

^b Department of Economics, Brock University, St. Catharines, Ontario, Canada. E-mail: jfl@brocku.ca

February 7, 2013

Abstract

This paper proposes Pearson-type statistics based on implied probabilities to detect structural change. The class of generalized empirical likelihood estimators (see ?) assigns a set of probabilities to each observation such that moment conditions are satisfied. These probabilities are called implied probabilities. 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 which 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 good size and competitive power properties. Moreover, they are computed in a single step which eliminates the need to compute the weighting matrix required for GMM estimation.

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

[†] A previous version, October 2008, of this paper was originally circulated under the title: "The Information Content of Implied Probabilities to Detect Structural Change". The first author gratefully acknowledges financial support of le Fonds québécois de la recherche sur

la société et la culture (FQRSC). The first and second authors are thankful for the financial support of the Social Sciences and Humanities Research Council of Canada (SSHRC). We 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.

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. ? suggested the continuous updating estimator (CUE) which shares the same objective function as GMM but with a weighting matrix that depends on the parameters of interest. The empirical likelihood (EL) (see ?) and the exponential tilting (ET) estimators (see ?) have also been proposed. ? 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 ? and may be shown to be based on the ? family of power divergence criteria. ? showed (in an i.i.d. setting) that, although estimators based on GMM, EL, ET or CU have the same first-order asymptotic distribution, they have different higher-order asymptotic properties. Amongst their findings, it is shown that the expression for the second-order asymptotic bias of GEL has fewer components than the one of GMM (with EL having the fewest). ? 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 GMM estimators, especially with many moment conditions.

GEL estimators assign a probability to each observation such that the moment conditions are satisfied (see ?). 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 ?, 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. ? use the weights given by the implied probabilities to propose a three-step estimator with higher-order asymptotic properties equivalent to those of EL. ? also discusses the use of these weights in the context of model misspecification. ? considered Pearson-type test statistics (statistics based on the difference between restricted and unrestricted estimators of the weights) for the validity of moment

¹See in particular the special issue of *Journal of Business and Economic Statistics*, 1996, volume 14.

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 in 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 ?) and 3) structural change tests to detect instability in the overidentifying restrictions (see for example ?). 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, 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 good size and competitive power properties compared to more standard structural change tests for GEL criteria (see ?). 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 nice higher-order asymptotic properties and are computed in a single step which 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 are 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 entropybased 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 ? and ?).

We consider a triangular array of random variables $\{x_{Tt} : 1 \le t \le T, T \ge 1\}$. 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 $\Theta \subset \mathbb{R}^p$ and that in the population their expected value is 0, namely

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

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 ?) be $\{\pi_t : 1 \le t \le T, T \ge 1\}$ we can recast the population moment conditions as

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

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

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

In this case, $S(\pi)$ can be interpreted as the entropy measure of ? and it captures the degree of uncertainty in the distribution π with respect to whether or not the distribution is concentrated or dispersed. 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 constraint, we get $\pi_t = 1/T \quad \forall t$, the maximally uninformative uniform distribution, while with 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 ? information criterion which measures the discrepancy between two distributions p and π . If the subject distribution is π and the reference distribution is $p_t = 1/T, \forall t$ we have

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

So that 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. The solution to the Lagrangian yields

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

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 ?).

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

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

where EL stands for empirical likelihood. When we evaluate the weights at some estimators we obtain $\pi_t^{ET}(\tilde{\theta}_T)$ and $\pi_t^{EL}(\tilde{\theta}_T)$. Recently, ? 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 ? and consists of estimating a single parameter, θ , with 2 moment conditions:

$$E[x_t - \theta] = 0$$
 and $E[(x_t - \theta)^2 - 4] = 0$

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 break and a regime shift cases. 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 ? and by ?.

In contrast, the next two examples consist of structural change through the moment conditions. Following ? and ? we study a simulated environment with 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 consumption growth follows a $N(0, \sigma^2 = 0.16)$. There is a single parameter to be estimated, γ , the coefficient of CRRA and two moment conditions are used:

$$E_t \exp[-\gamma \ln x_{t+1} - 9\sigma^2/2 + (3-\gamma)z_t] = 1$$
$$E_t z_t \exp[-\gamma \ln x_{t+1} - 9\sigma^2/2 + (3-\gamma)z_t - 1] = 0$$

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

Lastly, as in ?, we have the estimation of an autoregressive parameter using two moment conditions when the data generating process is an AR(1) ($x_t = \rho x_{t-1} + \epsilon_t$), for $t \leq 50$, and an ARMA(1,2) otherwise ($x_t = \rho x_{t-1} + \epsilon_t + 0.5\epsilon_{t-2}$). 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:

$$E [x_{t-1} (x_t - \rho x_{t-1})] = 0$$
$$E [x_{t-2} (x_t - \rho x_{t-1})] = 0$$

The instability occurs because the second moment condition is violated for t > 50.

Figure 1 shows the average of the vector of implied probabilities π 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 right 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 which violates the moment conditions. This simple simulation study clearly show that implied probabilities contain interesting information to detect structural change. In this paper, we 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 ?, ?, ?, ?, ?, and ?), 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 $\tilde{\Gamma}_T(\theta) = \{\gamma : \gamma' g(x_t, \theta) \in \Phi, t = 1, ..., T\}$. Then, the GEL estimator is a solution to the problem

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

where $\rho_j() = \partial^j \rho()/\partial \phi^j$ and $\rho_j = \rho_j(0)$ for j = 0, 1, 2, ... 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 by setting $\rho(\phi) = \ln(1 - \phi)$. Third, the continuous updating estimator of ? 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 ?, ?, ? and ?). 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, ..., T and M_T is a bandwidth parameter, $k(\cdot)$ a kernel function and we define where $k_j = \int_{-\infty}^{\infty} k(a)^j da$. In the time series context, the criteria is then given by:

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

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

3 Full and partial-samples 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 will consider parametric models indexed by parameters (β, δ) . With no structural change we define a vector of parameters $(\beta, \delta) \subset B \times \Delta \in R^p$ with $p = r + \nu$. Following ?, we make a distinction between pure structural change when no subvector δ appears and the entire parameter vector is subject to structural change under the alternative and partial structural change which corresponds to cases where only a subvector β is subject to structural change under the alternative. The generic null can be written as follows:

$$H_0: \beta_t = \beta_0 \qquad \forall t = 1, \dots, T.$$
(1)

The tests considered assume as alternative that at some point in the sample there is a single structural break. For example,

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

where s determines the fraction of the sample before and after the assumed breakpoint and [.] denotes the greatest integer function. The separation [Ts] represents a possible breakpoint which is governed by an unknown parameter $s \in S \subset (0, 1)$. Hence, we will consider a setup with a parameter vector which encompasses any kind of partial or pure structural change involving a single breakpoint. In particular, we consider a $2r + \nu$ dimensional parameter vector $\theta = (\beta'_1, \beta'_2, \delta')'$ where β_1 and $\beta_2 \in B \subset R^r$ and $\theta \in \Theta = B \times B \times \Delta \subset R^{2r+\nu}$. The parameters β_1 and β_2 apply to the samples before and after the presumed breakpoint and the null implies that:

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

We will formulate all our models in terms of θ . Special cases could be considered whenever restrictions are imposed in the general parametric formulation. One such restriction would be that $\theta_0 = (\beta'_0, \beta'_0)'$, which would correspond to the null of a pure structural change hypothesis. Once moment conditions are defined, we will translate them into a subspace identifying the parameters and a subspace for the overidentifying restrictions and relate these two translations to structural change tests, following the analysis of ? and ?.

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 Appendix 7.1.

In the GEL setting, the parameter vector θ is augmented by a vector of auxiliary parameters γ where each element of this vector is associated to an element of the smoothed moment conditions $g_{tT}(\theta)$. Under the null of no structural change relative to the specification of the model, the generic null hypothesis for this vector of auxiliary parameters can be written as follows:

$$H_0: \gamma_t = \gamma_0 = 0 \qquad \forall t = 1, \dots, T.$$
(3)

As for the parameter vector β , the tests that we will consider assume as alternative that

at some point in the sample there is a single structural break, namely:

$$\gamma_t = \begin{cases} \gamma_1 & t = 1, ..., [Ts] \\ \gamma_2 & t = [Ts] + 1, ..., T. \end{cases}$$

Thus, under the null of no structural change H_0 : $\gamma_1 = \gamma_2 = \gamma_0 = 0$. ? show that a structural change in γ is associated with instability in the overidentifying restrictions.

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 $\tilde{\Gamma}_T(\beta, \delta) = \{\gamma : k\gamma' g_{tT}(\beta, \delta) \in \Phi, t = 1, ..., T\}$ with $k = \frac{k_1}{k_2}$. Then, the full-sample GEL estimator $\{(\tilde{\beta}_T, \tilde{\delta}_T)\}$ is a sequence of random vectors such that:

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

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

The criteria is normalized so that $\rho_1 = \rho_2 = -1$ (see ?). This GEL estimator encompasses the ET, EL and CU estimators, as mentioned earlier. Let $\tilde{\theta}_T = \left(\tilde{\beta}'_T, \tilde{\beta}'_T, \tilde{\delta}'_T\right)'$ be a $2r + \nu$ -vector. Hereafter, the vector $\tilde{\theta}_T$ is called the full-sample estimator of θ . This restricted estimator is consistent only under the null that $\beta_1 = \beta_2$.

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

$$\tilde{\gamma}_{T}\left(\beta,\delta\right) = \arg \sup_{\gamma \in \tilde{\Gamma}(\beta,\delta)} \sum_{t=1}^{T} \frac{\left[\rho\left(k\gamma' g_{tT}\left(\beta,\delta\right)\right) - \rho_{0}\right]}{T}$$

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

$$\left(\tilde{\beta}_{T}^{\prime},\tilde{\delta}_{T}^{\prime}\right)^{\prime} = \arg\min_{(\beta,\delta)\in B\times\Delta}\sum_{t=1}^{T}\frac{\left[\rho\left(k\tilde{\gamma}_{T}\left(\beta,\delta\right)^{\prime}g_{tT}\left(\beta,\delta\right)\right)-\rho_{0}\right]}{T}.$$

To characterize the asymptotic distribution we define the following matrices:

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

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

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

$$G = [\begin{array}{cc} G_{\beta} & G_{\delta} \end{array}] \in R^{q \times p},$$

where $p = r + \nu$. From now on, following ? and ? we focus on the truncated kernel defined by

$$k(x) = 1$$
 if $|x| \le 1$ and $k(x) = 0$ otherwise

to obtain the following smoothed moment conditions

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

To handle the endpoints, we use the approach of ? and ? which sets

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

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

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

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 ?). Define also the derivatives of the smoothed moment conditions as:

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

Now consider a possible breakpoint [Ts]. Define the vector of auxiliary parameters $\gamma(s) = (\gamma_1(s)', \gamma_2(s)')'$ where γ_1 is the vector of the auxiliary parameters for the first part of the sample e.g. $t = 1, \ldots, [Ts]$ and γ_2 for the second part of the sample; $t = [Ts] + 1, \ldots, T$. The partial-sample GEL estimators for $s \in S$ based on the first and the second subsamples are formally defined as:

Definition 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 $\hat{\Gamma}_T(\theta, s) = \{\gamma(s) = (\gamma'_1, \gamma'_2)' : k\gamma(s)'g_{tT}(\theta, s)\}$. A partial-sample General Empirical Likelihood (PS-GEL) estimator $\{\hat{\theta}_T(s)\}$ is a sequence of random vectors such that:

$$\hat{\theta}_{T}(s) = \arg\min_{\theta \in \Theta} \sup_{\gamma(s) \in \hat{\Gamma}_{T}(\theta,s)} \sum_{t=1}^{T} \frac{\left[\rho(k\gamma(s)'g_{tT}(\theta,s)) - \rho_{0}\right]}{T}$$

$$= \arg\min_{\theta \in \Theta} \sup_{\gamma(s) \in \hat{\Gamma}_{T}(\theta,s)} \left(\sum_{t=1}^{[Ts]} \frac{\left[\rho(k\gamma'_{1}g_{tT}(\beta_{1},\delta)) - \rho_{0}\right]}{T} + \sum_{t=[Ts]+1}^{T} \frac{\left[\rho(k\gamma'_{2}g_{tT}(\beta_{2},\delta)) - \rho_{0}\right]}{T}\right)$$

for all $s \in S$, where $g_{tT}(\theta, s) = (g_{tT}(\beta_1, \delta)', 0')' \in R^{2q \times 1}$ for $t = 1, \dots, [Ts]$ and $g_{tT}(\theta, s) = (0', g_{tT}(\beta_2, \delta)')' \in R^{2q \times 1}$ for $t = [Ts] + 1, \dots, T$ with $\gamma(s) = (\gamma'_1, \gamma'_2)' \in R^{2q \times 1}$.

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.

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,

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

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

with $\hat{\Gamma}_{1T}(\beta_1, \delta, s) = \{\gamma_1 : k\gamma'_1 g_{tT}(\beta_1, \delta) \in \Phi, t = 1, \dots, [Ts]\}$ and $\hat{\Gamma}_{2T}(\beta_2, \delta, s) = \{\gamma_2 : k\gamma'_2 g_{tT}(\beta_2, \delta) \in \Phi, t = [Ts] + 1, \dots, T\}.$

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

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

which under the null (2) is asymptotically equal to

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

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

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

with

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

and

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

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):

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

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

$$\pi_t^{ET}(\tilde{\theta}_T) = \frac{\exp[\tilde{\gamma}_T'g_{tT}(\theta_T)]}{\sum_{t=1}^T \exp[\tilde{\gamma}_T'g_{tT}(\tilde{\theta}_T)]},$$
$$\pi_t^{EL}(\tilde{\theta}_T) = \frac{1}{T[1 + \tilde{\gamma}_T'g_{tT}(\tilde{\theta}_T)]},$$

and

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

Note that $\frac{2K+1}{T} \sum_{t=1}^{T} g_{tT}(\tilde{\theta}_T) g_{tT}(\tilde{\theta}_T)'$ is a consistent estimator of Ω .

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

$$\pi_t^{GEL}(\hat{\theta}_T(s), s) = \frac{\rho_1\left(\hat{\gamma}_T(s)'g_{tT}(\hat{\theta}_T(s), s)\right)}{\sum_{t=1}^T \rho_1\left(\hat{\gamma}_T(s)'g_{tT}(\hat{\theta}_T(s), s)\right)}.$$
(4)

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

$$\pi_t^{GEL}(\hat{\beta}_{1T}(s), \hat{\delta}_T(s), s) = \frac{\rho_1\left(\hat{\gamma}_{1T}(s)'g_{tT}(\hat{\beta}_{1T}(s), \hat{\delta}_T(s))\right)}{\sum_{t=1}^T \rho_1\left(\hat{\gamma}_{1T}(s)'g_{tT}(\hat{\beta}_{1T}(s), \hat{\beta}_{2T}(s), \hat{\delta}_T(s), s)\right)}.$$

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

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

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

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

3.2 Refining the null hypothesis

The moment conditions for the full sample under the null can be written as:

$$Eg(x_t, \theta_0) = 0, \quad \forall t = 1, \dots, T.$$
(5)

Following ?, we can project the moment conditions on the subspace identifying the parameters and the subspace of overidentifying restrictions. In particular, considering the (standardized) moment conditions for the full-sample GMM estimator, such a decomposition corresponds to:

$$\Omega^{-1/2} Eg(x_t, \theta_0) = P_G \Omega^{-1/2} Eg(x_t, \theta_0) + (I_q - P_G) \Omega^{-1/2} Eg(x_t, \theta_0),$$
(6)

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 the null hypothesis (2). For instance, following Hall and Sen (1999) we can consider the null, for the case of a single possible breakpoint s, which separates the identifying restrictions across the two subsamples:

$$H_0^I(s) = \begin{cases} P_G \Omega^{-1/2} E[g(x_t, \theta_0)] = 0 & \forall t = 1, \dots, [Ts] \\ P_G \Omega^{-1/2} E[g(x_t, \theta_0)] = 0 & \forall t = [Ts] + 1, \dots, T \end{cases}$$

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

$$H_0^{O1}(s) : (I_q - P_G)\Omega^{-1/2} E[g(x_t, \theta_0)] = 0 \quad \forall t = 1, \dots, [Ts]$$

$$H_0^{O2}(s) : (I_q - P_G)\Omega^{-1/2} E[g(x_t, \theta_0)] = 0 \quad \forall t = [Ts] + 1, \dots, T.$$

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 $H_0^{O2}(s)$. Various tests can be constructed with local power properties against any particular one of these three null hypotheses (and typically no power against the others). To elaborate further on this we consider a sequence of Pitman local alternatives based on the moment conditions:

Assumption 3.1. A sequence of local alternatives is specified as:

$$Eg(x_{Tt},\theta_0) = \frac{h(\eta,\tau,\frac{t}{T})}{\sqrt{T}}$$
(7)

where $h(\eta, \tau, s)$, for $r \in [0, 1]$, is a q-dimensional function. The parameter τ locates structural changes as a fraction of the sample size and the vector η defines the local alternatives.²

²The 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, $\eta \in \mathbb{R}^i$, $\tau \in \mathbb{R}^j$ such that $0 < \tau_1 < \tau_2 < \ldots < \tau_j < 1$ and θ^* is in the interior of Θ . Therefore it can accommodate multiple breaks.

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 non trivial 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 (6), the sequence of alternatives can be rewritten as:

$$\Omega^{-1/2} Eg(x_{Tt}, \theta_0) = P_G \Omega^{-1/2} \frac{h(\eta, \tau, \frac{t}{T})}{\sqrt{T}} + (I_q - P_G) \,\Omega^{-1/2} \frac{h(\eta, \tau, \frac{t}{T})}{\sqrt{T}} \tag{8}$$

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, consider a general local alternative of the form (see ?)

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

for $t = 1, \ldots, T$. A Taylor expansion of $g(x_{Tt}, \theta_{Tt})$ yields

$$Eg(x_{Tt},\beta_0) = -G_\beta \frac{f(\eta,\tau,\frac{t}{T})}{\sqrt{T}} + o_p(1).$$
(9)

If we substitute this expression into (8), the expression above is orthogonal to the second component of (8) and therefore puts restrictions on the first component (the identifying restrictions). In the case of pure structural $P_G = P_{G_\beta} = \Omega^{-1/2} G_\beta \left[G'_\beta \Omega^{-1} G_\beta \right]^{-1} G'_\beta \Omega^{-1/2}$. The alternative that at some point there is a single structural break at τ , $H^I_A(\tau)$, 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 $f(\eta, \tau, \frac{t}{T})$.

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:

$$H_A^{O1}(\tau) : (I_q - P_G)\Omega^{-1/2} E[g(x_t, \theta_0)] = \frac{\eta_1}{\sqrt{T}} \quad \forall t = 1, \dots, [T\tau]$$

$$H_A^{O2}(\tau) : (I_q - P_G)\Omega^{-1/2} E[g(x_t, \theta_0)] = \frac{\eta_2}{\sqrt{T}} \quad \forall t = [T\tau] + 1, \dots, T.$$

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 (8).

³This 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 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 following statistic:

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

They showed that such statistic is asymptotically equivalent to the overidentifying moment restrictions J-test proposed by ?. ? and ? also used this statistic in the time series context and showed that:

$$\frac{1}{2K+1} \sum_{t=1}^{T} \left(T\pi_t(\tilde{\theta}_T) - 1 \right)^2$$
(10)

is asymptotically first-order equivalent to the overidentifying moment restrictions J-test. However, as shown by ?, the J-test has no power to detect structural change in parameter values, a property that is shared by the specification tests, (11), 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 proposed to detect instability is given by the following partial sum:

$$IPSC_{T}^{GEL}(s) = \frac{s}{2K_{T}+1} \sum_{t=1}^{[Ts]} \left(T\pi_{t}^{GEL}(\tilde{\theta}_{T}, s) - 1 \right)^{2}$$
(11)

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

$$\gamma_{1T}(\beta,\delta,s) = \arg \sup_{\gamma_1 \in \hat{\Gamma}_T(\beta,\delta,s)} \sum_{t=1}^{[Ts]} \frac{[\rho(\gamma_1(\beta,\delta)'g_{tT}(\beta,\delta)) - \rho_0]}{T}$$
(12)

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

$$\gamma_{2T}(\beta,\delta,s) = \arg \sup_{\gamma_2 \in \hat{\Gamma}_T(\beta,\delta,s)} \sum_{t=[Ts]+1}^T \frac{\left[\rho(\gamma_2(\beta,\delta)'g_{tT}(\beta,\delta)) - \rho_0\right]}{T}.$$
 (13)

at $\tilde{\theta}_T$. 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 general test of a structural change under the null and the sequence of alternatives defined in (7).

Theorem 4.1. Under Assumptions (7.1) to (7.12), the following processes indexed by s for $s \in [0, 1]$ satisfy, under the null (5),

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

and under the alternative (7)

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

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 $p = r + \nu$ 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 to a stability test for the overidentifying restrictions. This test statistic, based on implied probabilities, can be viewed as a more general form of misspecification due to instability than just a test for parameter variation. The predictive tests proposed by Ghysels, Guay and Hall (1997) shares the same properties. In the Appendix we show that the *IPSC* test is asymptotically equivalent to the test statistic proposed by ?. 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 ? and ? 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 (10) can not detect asymptotically instability in the overidentifying restrictions for which $(I - P_G) \Omega^{-1/2} H(1) = 0$.

When the breakpoint is unknown, one can construct statistics across $s \in S$. In the context of maximum likelihood estimation, ? derive asymptotic optimal tests for a gaussian a priori of the amplitude of the structural change based on the Neyman-Pearson approach which are characterized by an average exponential form. The ? 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 following 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, $aveQ_T = \int_S Q_T(s)dJ(s)$. For a distant alternative $(c = \infty)$, the optimal test statistics takes the exponential form, $expQ_T = log\left(\int_S \exp[\frac{1}{2}Q_T(s)]dJ(s)\right)$. The supremum form often used in the literature corresponds to the case where $\frac{c}{(1+c)} \to \infty$. The sup test is given by $\sup Q_T = \sup_{s \in S} Q_T(s)$.

The following Theorem gives the asymptotic distribution for the 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 (5) and Assumptions 7.1 to 7.12, the following processes indexed by s for a given set S whose closure lies in [0,1] satisfy:

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

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 ?). 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 appear in Tables 1 to 3 for symmetric intervals $S = [s_0, 1 - s_0]$. Critical values for the entire sample appear at $s_0 = 0^4$.

An asymptotically equivalent modified statistic to (11) in the spirit of the Neymanmodified chi-square test is given by:

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

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:

$$IPSC_{T}^{I,GEL}(s) = \frac{1}{2K_{T}+1} \sum_{t=1}^{T} \left(T\pi_{t}^{GEL}(\hat{\theta}_{T}(s), s) - T\pi_{t}^{GEL}(\tilde{\theta}_{T}, s) \right)^{2}$$

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 $\tilde{\gamma}_T(s) = (\tilde{\gamma}_{1T}(s)'\tilde{\gamma}_{2T}(s)')'$ is the solution of the respective maximization problems defined in equations (12) and (13) and $\pi_t^{GEL}(\hat{\theta}_T(s), s)$ is defined in equation (4).

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 H_0 in (2) and Assumptions (7.1) to (7.12), the following processes indexed by s for a given set S whose closure lies in (0, 1) satisfy:

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

⁴Tables 1 to 3 can be obtained at http://coffee.econ.brocku.ca/jfl/research

and under the alternative (7)

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

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,

$$\sup IPSC_T^{I,GEL} \Rightarrow \sup_{s \in S} Q_r(s), \quad aveIPSC_T^{I,GEL} \Rightarrow \int_S Q_r(s)dJ(s),$$
$$expIPSC_T^{I,GEL} \Rightarrow \log\left(\int_S \exp[\frac{1}{2}Q_r(s)]dJ(s)\right).$$

Proof: See the Appendix.

The Theorem shows that the asymptotic distribution of the test based on implied probabilities 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 function of a *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 to 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 (0, 1).

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

$$IPSCM1_{T}^{I,GEL}(s) = \frac{1}{2K_{T}+1} \sum_{t=1}^{T} \frac{\left(T\pi_{t}^{GEL}(\hat{\theta}_{T}(s),s) - T\pi_{t}^{GEL}(\tilde{\theta}_{T},s)\right)^{2}}{T\pi_{t}^{GEL}(\hat{\theta}_{T}(s),s)}$$

and

$$IPSCM2_{T}^{I,GEL}(s) = \frac{1}{2K_{T}+1} \sum_{t=1}^{T} \frac{\left(T\pi_{t}^{GEL}(\hat{\theta}_{T}(s),s) - T\pi_{t}^{GEL}(\tilde{\theta}_{T},s)\right)^{2}}{T\pi_{t}^{GEL}(\tilde{\theta}_{T},s)}.$$

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. As previously, 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, ..., [Ts]) and with the second subsample (for t = [Ts] + 1, ..., 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.

More precisely the statistic is defined as:

$$IPSC_{T}^{O,GEL}(s) = \frac{1}{2K_{T}+1} \sum_{t=1}^{[Ts]} \left([Ts]\pi_{t}^{GEL}(\hat{\theta}_{1T}(s),s) - 1 \right)^{2} + \frac{1}{2K_{T}+1} \sum_{t=[Ts]+1}^{T} \left((T-[Ts])\pi_{t}^{GEL}(\hat{\theta}_{2T}(s),s) - 1 \right)^{2}$$

where $\hat{\theta}_{1T}(s) = \hat{\beta}_{1T}(s)$ and $\hat{\theta}_{2T}(s) = \hat{\beta}_{2T}(s)$. The statistic is the sum of the overidentifying restrictions statistics (10) but for the first and the second parts of the sample evaluated at the unrestricted estimator.

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

Theorem 4.4. Under Assumptions (7.1) to (7.12), the following processes indexed by s for a given set S whose closure lies in (0, 1) satisfy

$$IPSC_T^{O,GEL}(s) \Rightarrow Q_{q-r}(s)$$

with under the null of no structural change

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

and under the alternative (7)

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

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

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

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

 $\frac{[H(1)-H(s)]'\Omega^{-1/2}(I-P_G)\Omega^{-1/2}[H(1)-H(s)]}{(1-s)} \text{ to a structural change after the breakpoint } s.$

The asymptotic critical values for the interval S = [.15, .85] can be found in Hall and Sen (1999). For other symmetric interval $[s_0, 1 - s_0]$, critical values can be obtained in Guay (2003), Tables 1 to 3 for a number of overidentifying restrictions divided by 2 (in those tables). To see this, note that the critical values for the supremum, the average and the log exponential mappings applied to $\frac{B_{2q-2r}(s)'B_{2q-2r}(s)}{s}$ are equivalent to ones corresponding to $\frac{B_{q-r}(s)'B_{q-r}(s)}{s} + \frac{(B_{q-r}(1)-B_{q-r}(s))'(B_{q-r}(1)-B_{q-r}(s))}{1-s}$ for a symmetric interval $S.^5$

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

0

$$IPSCM_{T}^{O,GEL}(s) = \frac{1}{2K_{T}+1} \sum_{t=1}^{[Ts]} \frac{\left([Ts]\pi_{t}^{GEL}(\hat{\theta}_{1T}(s),s)-1\right)^{2}}{[Ts]\pi_{t}^{GEL}(\hat{\theta}_{1T}(s),s)} + \frac{1}{2K_{T}+1} \sum_{t=[Ts]+1}^{T} \frac{\left((T-[Ts])\pi_{t}^{GEL}(\hat{\theta}_{2T}(s),s)-1\right)^{2}}{(T-[Ts])\pi_{t}^{GEL}(\hat{\theta}_{2T}(s),s)}$$

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 ?, 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 General Empirical Likelihood estimator $\{\tilde{\theta}_T(s)\}$ is a sequence of random vectors such that:

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

⁵This is verified by comparing the critical values in Hall and Sen (1999) and Guay (2003). The critical values in Table 1 in Hall and Sen for q - r in our notation are the same than the critical values in Guay (2003) but for 2q - 2r.

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

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 ?. The test statistic is not asymptotically pivotal but asymptotically boundedly pivotal. The bound is then free of 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 (7).

Theorem 4.5. Under Assumptions (7.1) to (7.12), the following processes indexed by s for a given set S whose closure lies in (0,1) satisfy

$$IPSC_T^{R,GEL}(s) \Rightarrow Q_q(s)$$

are asymptotically boundedly pivotal and the asymptotic bound distribution is given by:

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

under the null of no structural change and under the alternative (7) by

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

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

Proof: See the Appendix.

The asymptotic 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 subsection 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 (8), namely

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

This decomposition and the asymptotic bound under the alternative show that the ability of the test statistic to detect a structural change in the parameter values depends on the Jacobian matrix G_{β} (see equation 9). 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 equals 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 ?). For instance, the test statistic will detect structural change in the parameter values with no trivial power for the following fixed alternative:

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

The discussion above also holds for the S-based test statistic proposed by ? 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 . One parameter is estimated, the autoregressive parameter (denoted by θ in the expression below), using two moment conditions formed with the lagged values of x_t .

The data generating process is given by

$$x_t = \theta_i x_{t-1} + u_t \tag{14}$$

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

$$x_t = \theta_i x_{t-1} + u_t + \alpha u_{t-2} \tag{15}$$

and 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 2000 (for each replication one constrained and 242 unconstrained optimization problems were solved for a total of 486,000 numerical optimization).

Table 4 summarizes the different parameterizations. The null hypothesis of structural stability is denoted by H_0^{SS} (DGP 1 to 3). For those DGPs we vary the magnitude of the autoregressive parameter θ . The alternatives of instability in the parameters or in the overidentifying restrictions are denoted by H_A^I (DGP 4 to 6), where we vary the magnitude of the change in the autoregressive parameter, and H_A^O (DGP 7 to 10) where we consider various values of the moving average parameter, respectively. In this situation only one parameter is estimated using two moment conditions created with the first two lags of x_t . Under H_0^{SS} , where $\alpha = 0$, the instruments are appropriate. Under the first class of alternative hypothesis (H_A^I) the two instruments are also valid while they no longer are for the second part of the sample with the second class of alternative hypothesis (H_A^O) .

Smoothing the moment conditions can be done via an appropriate choice of K_T . In a GMM setting this is equivalent to using some 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 quadratic form). Most of the previous simulation work considered a fixed degree of smoothing (see for example ? and ?). ? did look at fixed smoothing but also looked at applying the automatic bandwidth selection rule of Newey and West (1994).

Our Monte Carlo study is not particularly designed to investigate smoothing because under the null hypothesis the optimal value of K_T is 0 while it is not 0 under the alternative hypothesis H_A^O . For this reason we set $K_T = 0$ for all DGPs except DGPs 7, 8, 9 and 10. For these DGPs we select a bandwidth parameter chosen via the automatic, data-driven, procedure which chooses a bandwidth m_T which is then transformed as $K_T = [(m_T - 1)/2]$. The average, taken over Monte Carlo replications, K_T was found to vary between 1.6 and 2.3, increasing with the moving average component. A complete analysis of the effects of smoothing is left for future work. Lastly, a trimming rule of 0.15 was used, namely S =[.15, .85].

Table 5 contains the results for the general specification tests *IPSC* and *IPSCM* (the supremum, exponential and average version of the test statistics are presented). Table 6 contains the rejection frequencies for the test statistics designed to have power against a structural change in the parameters while Table 7 presents the results for test statistics which are designed to have power against a structural change in the overidentifying restrictions. All test statistics were computed in the GEL setting using exponential tilting. The tests used for comparison appear in ?. For completeness we report the Wald, LM, LR and O tests (all

computed in a GEL environment) below:

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

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

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

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

and for the second part of the sample:

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

The Lagrange Multiplier statistic is given by:

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

where

$$g_{1T}(\tilde{\theta}_T, s) = \frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T),$$

$$\hat{G}_{tT}^{\beta} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{tT}(\tilde{\beta}, \tilde{\delta})}{\partial \beta'},$$

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

The LR-like statistic is defined as:

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

And finally, the test statistic proposed by Hall and Sen (1999) is

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

where

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

and

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

Focusing first on size, all tests, except the modified tests, have rejection frequencies that are quite good, with some overrejection except for the O test statistics which slightly underrejects. The supremum and exponential versions of the test statistics show some overrejection while the average version exhibits rejection frequencies closer to the desired sizes. The LMtest statistics (supremum, average and exponential mappings) significantly underreject for all DGPs (1 to 3) under the null and have poor power for other DGPs. For these reasons, the rejection frequencies for the LM test statistics are not reported in the tables.

The modified tests $(IPSCM, IPSCM1^{I}, IPSCM2^{I} \text{ and } IPSCM^{O})$ based on implied probabilities have rejection frequencies that are much too large. The intuition for this is that the modified test statistics contain a more volatile term in the denominator which can inflate the value of the tests and hence increase the rejection frequencies.⁶

The study of power is divided into two cases. In case 1, structural change occurs in the parameter θ while in case 2, structural change occurs in the overidentifying restrictions. Under the alternative of instability in the parameter, H_A^I (DGP 4 to 6), we see that the 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* and *LR* tests. The *IPSC^O* and *O* test statistics, geared toward instability in the overidentifying restrictions, have no useful power while the more general specification test, *IPSC*, has some reasonable power.

Under the alternative of instability in the overidentifying restrictions, e.g. H_A^O , (DGP 7 to 10), we see that the test statistics specially designed to detect a change in the parameter $IPSC^I$, and the standard Wald and LR tests have little power while the targeted $IPSC^O$ and O tests have very good power. Importantly the tests based on implied probabilities are seen to have significantly higher power than O tests for all cases. In some cases, the gain in power can be twice as important. As expected, the general specification tests based on implied probabilities, IPSC, have rejection frequencies that fall in between those of $IPSC^O$ and $IPSC^I$.

The increase in the autoregressive coefficient from 0 to 0.8 does not impact greatly on the rejection frequencies under the null hypothesis but under the alternative hypotheses the

⁶Only the results for the test IPSCM are included in the tables. The results for the other modified tests can be obtained upon request.

magnitude of the change is important. Under H_A^I , for example, we see that power is close to unity when the change in the autoregressive parameter is quite extreme (0 to 0.8). Under H_A^O , which captures a change in the overidentifying restrictions, an increase (in absolute terms) in the moving average coefficient increases power.

Overall, structural change tests based on impled probabilities have very good size and power properties.

6 Conclusion

As noted by Back and Brown (1993), implied probabilities obtained from estimation of models using estimating equations could be used as an additional tool for model specification in the researcher's tool kit. An important specification test that has received considerable attention in the econometric literature has been a test for structural stability of either the underlying key parameters composing the estimation equations and, or, the stability of the additional equations (the overidentifying restrictions) that are often used in estimation. In this paper we have focused on the class of estimators based on the Generalized Empirical Likelihood approach. This approach is appealing intuitively because it requires the search for a vector of weights, one for each observation, that yields the most probable data distribution. We view these weights as potentially containing information on the content of the moment conditions (the estimating equations). In a pure entropy setting with no estimating equations the vector of weights is found to be maximally uninformative. That is, we obtain weights fluctuating around 1/T where T is the sample size. With constraints, the weights are chosen to be as maximally uninformative as the constraints will allow.

In this sense, the weights make use of the available information contained in the sample. In particular we use the weights to detect an unknown structural change. We suggest different types of testing procedures for the detection of a structural change each based on different measures of the discrepancy between the estimated weights and the unconstrained weights $\frac{1}{T}$. 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 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. We found that tests based on these implied probabilities have good finite sample size and competitive power properties. An important attractive feature is that these tests are computed in a single step which eliminates the need to compute the weighting matrix required to perform for alternative structural

change tests in GEL and for GMM estimation and testing. This weighting matrix is often linked to the poor finite sample properties of GMM-based estimators and test statistics. An issue that was not investigated in length in this paper was the impact of smoothing (to take into account serial dependence) on the performance of the tests. This interesting avenue of research is left for future work.

7 Appendix

7.1 Assumptions

We consider triangular arrays because they are required to derived asymptotic results under the Pitman drift alternatives. Define **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^{ν} 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 7.1. $\{x_{Tt} : t \leq T, T \geq 1\}$ is a triangular array of X-valued rv's that is L^0 near epoch dependent (NED) on a strong mixing base $\{Y_{Tt} : t = ..., 0, 1, ...; T \geq 1\}$, where X is a Borel subset of R^k , and $\{\mu_{Tt} : T \geq 1\}$ is tight on X.⁷

Define the smoothed moment conditions as:⁸

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

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

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

Assumption 7.2. $K_T/T^2 \to 0$ and $K_T \to \infty$ as $T \to \infty$ and $K_T = O_p\left(T^{\frac{1}{2\eta}}\right)$ for some $\eta > 1^9$.

Assumption 7.3. For some $d > \max\left(2, \frac{2\eta}{\eta-1}\right)$, $\{g(x_{Tt}, \beta, \delta) : t \leq T, T \geq 1\}$ 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 $\{Y_{Tt} : t = \ldots, 0, 1, \ldots; T \geq 1\}$, of size -d/(d-2) and $\sup \|g(x_{Tt}, \beta\delta)\|^d < \infty$.

Assumption 7.4. $Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{Ts} g\left(x_{Tt},\beta,\delta\right)\right) \to s\Omega \ \forall s \in [0,1]$ for some positive definite $q \times q$ matrix Ω .

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

⁷For a definition of L^p -near epoch dependence and tightness, see Andrews (1993, p. 829-830). For a presentation of the concept of near epoch dependence, we refer the reader to ?.

⁸Note here that g_{tT} denotes the smoothed moment conditions and x_{Tt} a triangular array of random variables. ⁹This assumption is slightly different than that in Smith (2004) but facilitates the proofs at no real cost.

Assumption 7.5. $\tilde{g}(\beta_0, \delta_0) = 0$ with $(\beta_0, \delta_0) \in B \times \Delta$ where $\tilde{g}(\beta_0, \delta_0) = \lim_{T \to \infty} \sum_{t=1}^{T} Eg(x_t, \beta, \delta)$ and B and Δ are bounded subsets of R^r and R^{ν} , $g(x_t, \beta, \delta)$ is continuous in x for all $(\beta, \delta) \in B \times \Delta$ and is continuous in (β, δ) uniformly over $(\beta, \delta, x) \in B \times \Delta \times C$ for all compact sets $C \subset \mathbf{X}$.

Assumption 7.6. For every neighborhood $\Theta_0 \subset \Theta$ of θ_0 , $\inf_{s \in S} (\inf_{\theta \in \Theta/\Theta_0} ||g(\theta, s)||) > 0$ where $g(\theta, s) = (s\tilde{g}(\beta_1, \delta)', (1 - s)\tilde{g}(\beta_2, \delta)')'$.

Assumption 7.7. (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 $\frac{1}{2} > \zeta > \frac{1}{d(1-1/\eta)}$.

Assumption 7.7 (b) is similar to the assumption in Newey and Smith (2004) and Smith (2004) but for $\gamma(s) = (\gamma_1(s)', \gamma_2(s)')'$. It specifies bounds on $\gamma(s)$ and with the existence of higher than second moments in Assumption 7.3, it leads to the arguments $\gamma(s)'g_{tT}(\theta, s)$ being in the domain Φ of $\rho(\cdot)$ w.p.a. 1 in the first subsample for all β_1, δ and $1 \le t \le [Ts]$ and in the second subsample for all β_2, δ and $[Ts] + 1 \le t \le T$.

Under Assumptions 7.1, 7.2, 7.3, 7.5, 7.6 and 7.7, Guay and Lamarche (2009) show for the partial-sample GEL that $\sup_{s \in S} \|\hat{\theta}_T(s) - \theta_0\| \xrightarrow{p} 0$, $\sup_{s \in S} \|\hat{\gamma}_T(s)\| \xrightarrow{p} 0$, $\|\hat{\gamma}_T(s)\| = O_p \left(T/K_T^2\right)^{-1/2}$ and $\sup_{s \in S} \|\frac{1}{T} \sum_{t=1}^T g_{tT}(\hat{\theta}_T(s), s)\| = O_p(T^{-1/2}).$

The consistency of the full-sample GEL estimator is obtained by slight modifications of Assumptions 7.6 and 7.7 (b). Assumption 7.6 must be modified by a simplified version with $\tilde{g}(\beta, \delta)$ instead of $g(\theta, s)$. Assumption 7.7 (b) holds but for the full sample Lagrange multiplier γ . The consistency result that $\tilde{\theta}_T \xrightarrow{p} \theta_0$ is then derived under weaker conditions than in Smith (2004).

The following high level assumptions are sufficient to derive the weak convergence under the null and the alternative of the PS-GEL estimators $\hat{\theta}_T(s)$ and $\hat{\gamma}_T(s)$ (see Guay and Lamarche, 2009). These assumptions are similar to the ones in Andrews (1993).

Assumption 7.8. $\sup_{s \in S} \|\hat{\Omega}_{iT}(s) - \Omega\| \xrightarrow{p} 0$ where Ω is defined in Section 3.1 and S whose closure lies in (0, 1) for i = 1, 2.

Assumption 7.8 holds under conditions given in Andrews (1991) and Lemma A.3 in Smith (2004). To respect these conditions, Assumption 7.3 can be replaced by the following assumption:

Assumption 7.3'. $\{g(x_{Tt}, \theta) : t \leq T, T \geq 1\}$ 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 $\sup_{t \leq T, T \geq 1} E \|g(x_{Tt}, \theta)\|^d < \infty$ for some $d > \max\left(4\nu, \frac{2\eta}{\eta-1}\right)$.

Assumptions 7.3' and 7.8 guarantee for the full-sample and partial-sample GEL that

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

and

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

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

Assumption 7.9. $g(x, \beta, \delta)$ is differentiable in (β, δ) , $\forall (\beta, \delta) \in B_0 \times \Delta_0 \ \forall x \in \mathbf{X_0} \subset \mathbf{X}$ for a Borel measurable set $\mathbf{X_0}$ that satisfies $P(x_t \in \mathbf{X_0}) = 1 \forall t \leq T, T \geq 1$, $g(x, \beta, \delta)$ is Borel measurable in $x \ \forall (\beta, \delta) \in B_0 \times \Delta_0$, $\partial g(x_t, \beta, \delta) / \partial (\beta', \delta')$ is continuous in (x, β, δ) on $\mathbf{X} \times B_0 \times \Delta_0$,

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

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

Assumption 7.10. $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T_s} Eg_{tT}(\beta, \delta)$ exists uniformly over $(\beta, \delta, s) \in B \times \Delta \times S$ and equals $s \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} Eg(x_t, \beta, \delta) = s\tilde{g}(\beta, \delta).$

Assumption 7.11. $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T_s} E \partial g_{tT}(\beta_0, \delta_0) / \partial(\beta', \delta')$ exists uniformly over $s \in S$ and equals $sG \ \forall s \in S$ and S whose closure lies in (0, 1).

Assumption 7.12. $G(s)'\Omega(s)^{-1}G(s)$ 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).

7.2 Lemmas

The following Lemmas are necessary to establish the proofs of the Theorems:

Lemma 7.1. We denote $\{B(s) : s \in [0,1]\}$ as q-dimensional vectors of mutually independent Brownian motion on [0,1] and define

$$J(s) = \left[\begin{array}{c} \Omega^{1/2}B(s)\\ \Omega^{1/2}(B(1) - B(s)) \end{array}\right]$$

where $B(\pi)$ is a q-dimensional vector of standard Brownian motion. Under Assumptions 7.1 to 7.12, every sequence of PS-GEL estimators $\{\hat{\theta}(\cdot), \hat{\gamma}(\cdot), T \geq 1\}$ satisfies

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

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 (7), the same results hold except that:

$$J(s) = \left[\begin{array}{c} \Omega^{1/2}B(s) + H(s) \\ \Omega^{1/2}(B(1) - B(s)) + H(1) - H(s) \end{array}\right]$$

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

Proof of Lemma 7.1: see Guay and Lamarche (2009).

Lemma 7.2. Under Assumptions 7.1 to 7.12, for the unrestricted implied probabilities evaluated at the restricted estimator, we get

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

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

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

and for t = [Ts] + 1, ..., T,

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

Proof of Lemma 7.2

We need to derive the asymptotic distribution of the partial-sample implied probabilities evaluated at the full-sample estimator:

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

A mean value expansion for $\pi_t^{GEL}(\tilde{\theta}_T, s)$ around $(\tilde{\theta}_T, \tilde{\gamma}_T) = (\tilde{\theta}_T, 0)$ yields:

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

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

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

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

As $\sum_{t=1}^{T} g(\tilde{\theta}_T) = \sum_{t=1}^{T} g_{tT}(\tilde{\theta}_T) + O_p \left(K^2 / (2K+1) \right) = O_p(T^{1/2})$ 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 $\tilde{\gamma}_T(s) = O_p(2K_T+1/\sqrt{T})^{-10}$ yields:

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

and

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

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

$$T\pi_t^{GEL}(\tilde{\theta}_T, s) - 1 = g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T)' \tilde{\gamma}_{1T}(1 + o_p(1)) + O_p(K_T/T)$$
(16)

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

$$T\pi_t^{GEL}(\tilde{\theta}_T, s) - 1 = g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T)' \tilde{\gamma}_{2T}(1 + o_p(1)) + O_p\left(K_T/T\right)$$
(17)

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

¹⁰See Kitamura and Stutzer (1997) and ?.

7.3 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

$$\frac{1}{T} \sum_{t=1}^{[Ts]} \rho_1 \left(\tilde{\gamma}'_{1T} g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) \right) g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) = -\frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) - \frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T)' \tilde{\gamma}_{1T}
+ \frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) \sum_{j=2}^{\infty} \frac{1}{j!} \rho_{j+1}(0) \left(g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T)' \tilde{\gamma}_{1T} \right)^j$$

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 $g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) = O_p\left((2K_T+1)^{-1/2}\right)$, this yields

$$\frac{1}{T} \sum_{t=1}^{[Ts]} \rho_1 \left(\tilde{\gamma}_{1T}' g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) \right) g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) = -\frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) - \frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T) g_{tT} (\tilde{\beta}_T, \tilde{\delta}_T)' \tilde{\gamma}_{1T} + O_p \left((2K_T + 1)^{-3/2} \| \tilde{\gamma}_{1T} \|^2 \right).$$

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 $\tilde{\gamma}_{1T} = O_p\left((2K_T+1)/\sqrt{T}\right)$, we get under the null:

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

which yields

$$\frac{s}{2K_T + 1}\tilde{\gamma}_{1T} = -\tilde{\Omega}_T^{-1}\frac{1}{T}\sum_{t=1}^{[Ts]} g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) + O_p\left(K_T^{1/2}/T\right).$$
(18)

Now, expanding this expression around β_0 and δ_0 gives:

$$\frac{s\sqrt{T}}{2K_T+1}\tilde{\gamma}_{1T} = -\tilde{\Omega}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0) - s\tilde{\Omega}_T^{-1}G\sqrt{T}\left(\begin{array}{c}\tilde{\beta}_T-\beta_0\\\tilde{\delta}_T-\delta_0\end{array}\right) + o_p(1).$$

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

$$\sqrt{T} \left(\begin{array}{c} \tilde{\beta}_T - \beta_0 \\ \tilde{\delta}_T - \delta_0 \end{array} \right) = -\left(G' \Omega^{-1} G \right)^{-1} G' \Omega^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\beta_0, \delta_0) + o_p(1).$$
(19)

Combining the two preceding results, we obtain

$$\frac{s\sqrt{T}}{2K_T+1}\tilde{\gamma}_{1T} = -\Omega^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0) + s\Omega^{-1}G(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{tT}(\beta_0,\delta_0) + o_p(1).$$

By Lemma 5.1 in ?, under the null of no structural change, we have

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\beta_0, \delta_0) \Rightarrow B(s)$$
(20)

and under the generic alternative (7),

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\beta_0, \delta_0) \Rightarrow B(s) + \Omega^{-1/2} H(s)$$
(21)

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

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

Using Lemma 7.2 and the derivation above,

$$T\pi_t^{GEL}(\tilde{\theta}_T, s) - 1 = -\left(\frac{2K+1}{s}\tilde{\Omega}_T^{-1}\frac{1}{T}\sum_{t=1}^{[Ts]}g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) + O_p\left(K_T^{3/2}/T\right)\right)' \times (22)$$
$$g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T)(1 + o_p(1)) + O_p(K_T/T). \tag{23}$$

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

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

By a mean value expansion:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\tilde{\beta}_{T},\tilde{\delta}_{T}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_{0},\delta_{0}) + \frac{1}{T}\sum_{t=1}^{[Ts]}\frac{\partial g_{tT}(\bar{\beta}_{T},\bar{\delta}_{T})}{\partial\left(\beta',\delta'\right)} \left(\begin{array}{c}\tilde{\beta}_{T}-\beta_{0}\\\tilde{\delta}_{T}-\delta_{0}\end{array}\right)$$

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 eq. (19), we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\tilde{\beta}_T,\tilde{\delta}_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}g_{tT}(\beta_0,\delta_0) - sG[G'\Omega^{-1}G]^{-1}G'\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}g_{tT}(\beta_0,\delta_0) + o_p(1).$$
(24)

It follows by (20) that

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\beta}_T, \tilde{\delta}_T) \Rightarrow B(s) - s\Omega^{-1/2} G[G'\Omega^{-1}G]^{-1} G'\Omega^{-1/2} B(1).$$
(25)

By (22), the following partial sum can be shown to be:

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

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:

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

By the result above, the partial sum of interest yields by equations (25) and the consistency of $\tilde{\Omega}_T$:

$$\frac{s}{2K+1} \sum_{t=1}^{[Ts]} [T\pi_t^{GEL}(\tilde{\theta}_T, s) - 1]^2 \quad \Rightarrow \quad \left[B(s) - s\Omega^{-1/2}G[G'\Omega^{-1}G]^{-1}G'\Omega^{-1/2}B(1) \right] (28) \\ \left[B(s) - s\Omega^{-1/2}G[G'\Omega^{-1}G]^{-1}G'\Omega^{-1/2}B(1) \right] (27)$$

The expression $\left[B(s) - s\Omega^{-1/2}G[G'\Omega^{-1}G]^{-1}G'\Omega^{-1/2}B(1)\right]$ can be rewritten as:

 $[I - P_G] B(s) + P_G [B(s) - sB(1)]$

with $P_G = \Omega^{-1/2} G[G' \Omega^{-1} G]^{-1} G' \Omega^{-1/2}$. We can now decompose:

$$\Omega^{-1/2} G(G' \Omega^{-1} G)^{-1} G' \Omega^{-1/2} = C' \Lambda C$$
(28)

where CC' = I and

$$\Lambda = \left[\begin{array}{cc} I_p & 0\\ 0 & 0 \end{array} \right]$$

and I_p is an identity matrix with dimension $p \times p$ where $p = r + \nu$ with r and ν the respective dimension of the vectors β and δ . Note also that CB(s) has the same

asymptotic distribution as B(s). The RHS of expression (27) can then be rewritten as:

$$B(s)' [I - C'\Lambda C] B(s) + [B(s) - sB(1)]' C'\Lambda C [B(s) - sB(1)]$$

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 bridge. The result follows. The asymptotic distribution under the alternative (7) can be easily obtained similarly using (21).

Proof of Theorem 4.3

By Lemma 7.2 applied to the difference between the partial-sample implied probabilities evaluated at the unrestricted and at the restricted estimators we get:

$$T\left[\hat{\pi}_t(\hat{\theta}_T(s),s) - \tilde{\pi}_t(\tilde{\theta}_T,s)\right] = g_{tT}(\hat{\theta}_T(s),s)'\hat{\gamma}_T(s) - g_{tT}(\tilde{\theta}_T,s)'\tilde{\gamma}_T(s) + o_p(T^{-1/2}).$$

Let us define the following selection matrices:

$$H_1 = \begin{bmatrix} I_{r \times r} & 0_{r \times r} & 0_{r \times \nu} \\ 0_{\nu \times r} & 0_{\nu \times r} & I_{\nu \times \nu} \end{bmatrix}$$

and

$$H_2 = \left[\begin{array}{ccc} 0_{r \times r} & I_{r \times r} & 0_{r \times \nu} \\ 0_{\nu \times r} & 0_{\nu \times r} & I_{\nu \times \nu} \end{array} \right].$$

and the corresponding estimator: $\hat{\theta}_{1T}(s) = H_1 \hat{\theta}_T(s) = \left(\hat{\beta}_{1T}(s)', \hat{\delta}_T(s)'\right)'$ and $\hat{\theta}_{2T}(s) = H_2 \hat{\theta}_T(s) = \left(\hat{\beta}_{2T}(s)', \hat{\delta}_T(s)'\right)'$. Similarly, for the restricted estimator $\tilde{\theta}_{1T}(s) = H_1 \tilde{\theta}_T(s) = \left(\tilde{\beta}_T(s)', \tilde{\delta}_T(s)'\right)'$ and $\tilde{\theta}_{2T}(s) = H_2 \tilde{\theta}_T(s) = \left(\tilde{\beta}_T(s)', \hat{\delta}_T(s)'\right)'$ where $\tilde{\theta}_T(s) = \left(\tilde{\beta}_T(s)', \tilde{\beta}_T(s)', \tilde{\delta}_T(s)'\right)'$. 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 (18), we get:

$$T\left[\pi_{t}(\hat{\theta}_{T}(s),s) - \pi_{t}(\tilde{\theta}_{T},s)\right] = -\left[(2K_{T}+1)\Omega(s)^{-1}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\theta}_{T}(s),s) + o_{p}(1)\right]'g_{tT}(\hat{\theta}_{T}(s),s)(1+o_{p}(1))$$
$$+ \left[(2K_{T}+1)\Omega(s)^{-1}\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\tilde{\theta}_{T},s) + o_{p}(1)\right]'g_{tT}(\tilde{\theta}_{T},s)(1+o_{p}(1))$$
$$+ O_{p}(K_{T}/T)$$

which equals

$$T\left[\pi_{t}(\hat{\theta}_{T}(s),s) - \pi_{t}(\tilde{\theta}_{T},s)\right] = -\left[\frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\hat{\theta}_{1T}(s)) - \frac{1}{[Ts]}\sum_{t=1}^{[Ts]}g_{tT}(\tilde{\theta}_{1T}) + o_{p}(1)\right] \times \Omega^{-1}(2K_{T}+1)g_{tT}(\hat{\theta}_{1T}(s))(1+o_{p}(1)) + O_{p}(K_{T}/T30)$$

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

Now, consider the following mean value expansion:

$$\frac{1}{[Ts]} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_{1T}) = \frac{1}{[Ts]} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\theta}_{1T}(s)) + \frac{1}{[Ts]} \sum_{t=1}^{[Ts]} \frac{\partial g_{tT}(\bar{\theta}_{1T}(s))}{\partial \theta'_1} \left(\tilde{\theta}_{1T} - \hat{\theta}_{1T}(s)\right). (31)$$

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,

$$T\left[\pi_{t}(\hat{\theta}_{T}(s),s) - \pi_{t}(\tilde{\theta}_{T},s)\right] = \left[\frac{1}{[Ts]}\sum_{t=1}^{[Ts]} \frac{\partial g_{tT}(\bar{\theta}_{1T}(s))}{\partial \theta_{1}'} \left(\tilde{\theta}_{1T} - \hat{\theta}_{1T}(s)\right) + o_{p}(1)\right]' \times \Omega^{-1}(2K_{T}+1)g_{tT}(\hat{\theta}_{1T}(s))(1+o_{p}(1)) + O_{p}(K_{T}/T).$$

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

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

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

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

By results derived above and some calculations, this yields

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

Andrews (1993, p. 851-852) shows that this expression is asymptotically equivalent to the $LM_T(s)$ statistic for parameters instability. This gives the result under the null. Under the alternative, ? show that the $LM_T(s)$ statistic for the restricted GEL estimator has the same asymptotic distribution as in Theorem 4.3. Since 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. A direct application of Lemma 7.2 but for the estimation of the subsample with the first [Ts] observations gives

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

uniformly in t = 1, ..., [Ts]. This is obtained with a proof similar to the one of Lemma 7.2 but only for the first part of the sample. This yields

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

By summing the expression above to $t = \dots, [Ts]$, and by (18) for the unrestricted estimator applied to the sample $t = 1, \dots, [Ts]$, this yields

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

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

$$\frac{1}{2K_T+1} \sum_{t=1}^{[Ts]} \left([Ts] \ \pi_t^{GEL}(\hat{\theta}_{1T}(s), s) - 1 \right)^2 = \frac{1}{\sqrt{[Ts]}} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\theta}_{1T}(s))' \Omega^{-1} \frac{1}{\sqrt{[Ts]}} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\theta}_{1T}(s)) + o_p(1).$$
(32)

Similarly, we can easily obtain that

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

$$\frac{1}{\sqrt{T-[Ts]}} \sum_{t=[Ts]+1}^{T} g_{tT}(\hat{\theta}_{2T}(s))' \Omega^{-1} \frac{1}{\sqrt{T-[Ts]}} \sum_{t=[Ts]+1}^{T} g_{tT}(\hat{\theta}_{2T}(s)) + o_p(1).$$
(33)

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

$$\Omega(s)^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{tT}(\hat{\theta}_T(s), s) = \Omega(s)^{-1/2} \left[\begin{array}{c} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\beta_0, \delta_0) \\ \frac{1}{\sqrt{T}} \sum_{t=[Ts]+1}^{T} g_{tT}(\beta_0, \delta_0) \end{array} \right] + \Omega(s)^{-1/2} G(s) \sqrt{T} \left(\hat{\theta}_T(s) - \theta_0 \right) + o_T \left(\hat{\theta}_T(s) - \theta_T \right) + o_T$$

By Lemma 7.1 and (20),

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

The RHS can be rewritten as

$$\left(I - \Omega(s)^{-1/2}G(s)\left(G(s)'\Omega(s)^{-1}G(s)\right)^{-1}G(s)'\Omega(s)^{-1/2}\right) \left[\begin{array}{c}\frac{1}{\sqrt{s}}B(s)\\\frac{1}{\sqrt{(1-s)}}(B(1) - B(s))\end{array}\right]$$

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:

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_{1T}(s)) \quad \Rightarrow \left(I_q - \Omega^{-1/2} G_\beta \left(G'_\beta \Omega^{-1} G_\beta \right)^{-1} G'_\beta \Omega^{-1/2} \right) B(s) (34)$$

and

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

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 and

$$\Lambda = \left[\begin{array}{cc} I_r & 0\\ 0 & 0 \end{array} \right]$$

and I_r is an identity matrix with dimension $r \times r +$ with r is the dimension of the vectors β . Noting that $C(I - C'\Lambda C)B(s) = (I - \Lambda)CB(s)$ and CB(s) is also a r-vector of Brownian motion. Now consider the multiplication of the RHS term of equation (34) by the matrix C, this yields

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

Using this result, we obtain directly the asymptotic distribution of the statistic for the first subsample by equation (32) considering that $C'C = I_q$. The proof is similar for the second subsample (eq. 33), hence obtaining the asymptotic distribution under the null.

Under the alternative (7), the asymptotic distribution under the alternative is derived similarly using Lemma 7.1.

Proof of Theorem 4.5

Define

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

and $\hat{g}_T(\theta, s) = \frac{1}{T} \sum_{t=1}^T g_{tT}(\theta, s).$ Since $\tilde{\theta}_T(s) = \arg \min_{\theta \in \Theta} P(\theta, \gamma(s)), s)$

$$\sup_{\gamma(s)\in\tilde{\Gamma}_{T}(s)} P(\theta_{T}(s),\gamma(s),s) \leq \sup_{\gamma(s)\in\tilde{\Gamma}_{T}(s)} P(\theta_{0},\gamma(s),s)$$

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 $\dot{\gamma}(s) = \tau \bar{\gamma}(s), \ 0 \leq \tau \leq 1$ 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,

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

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

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

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

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

and

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

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:

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

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:

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

Similarly for the second subsample

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

Then, we get for the statistic

$$IPSC_{T}^{R,GEL}(s) = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}(\tilde{\beta}_{T}(s))\right)' \frac{\tilde{\Omega}_{1T}(s)^{-1}}{s} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} g_{tT}(\tilde{\beta}_{T}(s))\right) + \left(\frac{1}{\sqrt{T}}\sum_{t=[Ts]+1}^{T} g_{tT}(\tilde{\beta}_{T}(s))\right)' \frac{\tilde{\Omega}_{2T}(s)^{-1}}{1-s} \left(\frac{1}{\sqrt{T}}\sum_{t=[Ts]+1}^{T} g_{tT}(\tilde{\beta}_{T}(s))\right) + o_{p}(1) = T\hat{g}_{T}(\tilde{\theta}_{T}(s), s)'\tilde{\Omega}_{T}(s)^{-1}\hat{g}_{T}(\tilde{\theta}_{T}(s), s) + o_{p}(1)$$

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 ?. The result under the null follows and the asymptotic bound under the alternative can be easily derived.

Table 4: Data Generating Processes							
	H_0^{SS}	H^I_A	H_A^O				
DGP1	$\theta_1 = \theta_2 = 0$		$\alpha = 0$				
DGP2	$\theta_1 = \theta_2 = 0.4$		$\alpha = 0$				
DGP3	$\theta_1 = \theta_2 = 0.8$		$\alpha = 0$				
DGP4		$\theta_1 = 0, \theta_2 = 0.4$	$\alpha = 0$				
DGP5		$\theta_1 = 0, \theta_2 = 0.8$	$\alpha = 0$				
DGP6		$\theta_1 = 0.4, \theta_2 = 0.8$	$\alpha = 0$				
$\mathrm{DGP7}$	$\theta_1 = \theta_2 = 0.4$		$\alpha = 0.5$				
DGP8	$\theta_1 = \theta_2 = 0.4$		$\alpha = 0.9$				
DGP9	$\theta_1 = \theta_2 = 0.4$		$\alpha = -0.5$				
DGP10	$\theta_1 = \theta_2 = 0.4$		$\alpha = -0.9$				

Figure 1: Simulated Implied Probabilities

Notes: 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 a 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.

DGP	Size $(\%)$	supIPSC	aveIPSC	expIPSC	supIPSCM	aveIPSCM	expIPSCM
DGP1	1	0.0155	0.0145	0.0160	0.0770	0.0435	0.0590
-	5	0.0655	0.0615	0.0655	0.1575	0.1035	0.1275
	10	0.1255	0.1255	0.1270	0.2295	0.1720	0.1870
DGP2	1	0.0145	0.0155	0.0145	0.0750	0.0385	0.0595
	5	0.0660	0.0545	0.0610	0.1520	0.1005	0.1225
	10	0.1260	0.1215	0.1245	0.2330	0.1735	0.1925
DGP3	1	0.0225	0.0140	0.0165	0.0835	0.0465	0.0575
	5	0.0685	0.0670	0.0690	0.1525	0.1050	0.1215
	10	0.1195	0.1225	0.1210	0.2300	0.1680	0.1875
DGP4	1	0.1465	0.1590	0.1515	0.4070	0.3190	0.3560
	5	0.3955	0.3980	0.4010	0.5985	0.5365	0.5635
	10	0.5590	0.5980	0.5885	0.7050	0.6840	0.6925
DGP5	1	0.9780	0.9710	0.9765	0.9990	0.9955	0.9980
	5	0.9985	0.9980	0.9985	1.0000	0.9995	0.9995
	10	1.0000	0.9995	1.0000	1.0000	0.9995	1.0000
DGP6	1	0.4095	0.3825	0.3995	0.7020	0.6130	0.6675
	5	0.7040	0.6685	0.6885	0.8450	0.8020	0.8225
	10	0.8170	0.8120	0.8220	0.8970	0.8810	0.8870
$\mathrm{DGP7}$	1	0.2945	0.1835	0.2365	0.5535	0.3680	0.4780
	5	0.5320	0.4005	0.4675	0.7090	0.5360	0.6270
	10	0.6690	0.5660	0.6050	0.7805	0.6515	0.7145
DGP8	1	0.6080	0.4565	0.5340	0.8205	0.6555	0.7645
	5	0.8120	0.7000	0.7620	0.9025	0.7955	0.8580
	10	0.8910	0.8330	0.8645	0.9370	0.8720	0.9040
DGP9	1	0.3970	0.2385	0.3265	0.5580	0.3235	0.4630
	5	0.6200	0.4825	0.5590	0.7005	0.4990	0.6150
	10	0.7390	0.6415	0.6890	0.7810	0.6575	0.7180
DGP10	1	0.7975	0.6695	0.7535	0.8515	0.6685	0.7930
	5	0.9420	0.8790	0.9175	0.9290	0.8545	0.9055
	10	0.9730	0.9525	0.9650	0.9625	0.9435	0.9580

 Table 5: Rejection Frequencies for General Tests of Structural Change

	DGP	Size (%)	$supIPSC^{I}$	$aveIPSC^{I}$	$expIPSC^{I}$	supW	aveW	expW	supLR	aveLR	expL
-		()	1		1	1		1	1		1
	DGP1	1	0.0190	0.0195	0.0220	0.0275	0.0190	0.0275	0.0220	0.0195	0.024
		5	0.0760	0.0765	0.0805	0.0855	0.0725	0.0815	0.0825	0.0730	0.083
		10	0.1455	0.1315	0.1475	0.1450	0.1225	0.1465	0.1435	0.1290	0.147
	DGP2	1	0.0200	0.0150	0.0195	0.0285	0.0150	0.0275	0.0235	0.0150	0.022
		5	0.0760	0.0740	0.0870	0.0880	0.0735	0.0900	0.0845	0.0720	0.087
		10	0.1450	0.1415	0.1555	0.1510	0.1295	0.1500	0.1525	0.1375	0.155
	DGP3	1	0.0160	0.0125	0.0155	0.0320	0.0120	0.0195	0.0170	0.0120	0.014
		5	0.0705	0.0670	0.0750	0.0980	0.0635	0.0825	0.0715	0.0675	0.075
		10	0.1465	0.1320	0.1420	0.1550	0.1300	0.1440	0.1425	0.1290	0.139
	DGP4	1	0.4880	0.5345	0.5450	0.4840	0.5205	0.5420	0.4840	0.5295	0.547
		5	0.6960	0.7540	0.7555	0.6915	0.7430	0.7500	0.6930	0.7510	0.751
		10	0.7915	0.8390	0.8350	0.7860	0.8280	0.8300	0.7875	0.8365	0.832
	DGP5	1	1.0000	1.0000	1.0000	1.0000	0.9995	1.0000	1.0000	1.0000	1.000
		5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
		10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.000
	DGP6	1	0.6955	0.7495	0.7655	0.7060	0.7475	0.7655	0.6950	0.7435	0.762
		5	0.8620	0.8975	0.9005	0.8590	0.8910	0.8955	0.8590	0.8975	0.900
		10	0.9160	0.9420	0.9395	0.9075	0.9345	0.9365	0.9140	0.9385	0.937
	$\mathrm{DGP7}$	1	0.3225	0.2880	0.3395	0.2895	0.2215	0.3045	0.1845	0.1905	0.220
		5	0.5000	0.4965	0.5295	0.4605	0.4170	0.4795	0.3630	0.3855	0.406
		10	0.6190	0.6100	0.6350	0.5705	0.5275	0.5805	0.4915	0.5015	0.519
	DGP8	1	0.5460	0.5020	0.5710	0.4425	0.3800	0.4720	0.3575	0.3545	0.400
		5	0.7075	0.6920	0.7185	0.6160	0.5960	0.6460	0.5530	0.5760	0.599
		10	0.7875	0.7795	0.8085	0.7140	0.6900	0.7305	0.6595	0.6855	0.695
	DGP9	1	0.2910	0.1630	0.2665	0.2255	0.1275	0.2110	0.1120	0.0955	0.124
		5	0.4550	0.3515	0.4300	0.3660	0.2805	0.3640	0.2650	0.2415	0.279
		10	0.5705	0.4725	0.5440	0.4640	0.4015	0.4650	0.3770	0.3665	0.396
	DGP10	1	0.5665	0.3130	0.5175	0.4530	0.2825	0.4410	0.2150	0.1425	0.213
		5	0.7525	0.5545	0.6965	0.5960	0.4890	0.5920	0.4205	0.3340	0.420
		10	0.8295	0.6815	0.7800	0.6790	0.6050	0.6790	0.5540	0.4755	0.538

 Table 6: Rejection Frequencies for Tests of Structural Change in the Parameters

DGP	Size $(\%)$	$supIPSC^O$	$aveIPSC^O$	$expIPSC^O$	supO	aveO	expO
DGP1	1	0.0195	0.0195	0.0175	0.0020	0.0075	0.0050
	5	0.0715	0.0630	0.0675	0.0210	0.0420	0.0390
	10	0.1290	0.1240	0.1320	0.0595	0.1005	0.0865
DGP2	1	0.0280	0.0145	0.0220	0.0040	0.0085	0.0085
	5	0.0825	0.0630	0.0810	0.0270	0.0450	0.0440
	10	0.1300	0.1225	0.1300	0.0595	0.0955	0.0865
DGP3	1	0.0270	0.0170	0.0240	0.0055	0.0090	0.0075
	5	0.0790	0.0650	0.0780	0.0295	0.0490	0.0440
	10	0.1345	0.1230	0.1420	0.0630	0.1035	0.0890
DGP4	1	0.0300	0.0170	0.0250	0.0050	0.0070	0.0080
	5	0.0970	0.0725	0.0925	0.0360	0.0460	0.0445
	10	0.1650	0.1370	0.1640	0.0810	0.0985	0.1075
DGP5	1	0.2930	0.1420	0.2885	0.1605	0.0805	0.1725
	5	0.5165	0.3420	0.4975	0.3805	0.2630	0.4040
	10	0.6380	0.5000	0.6200	0.5305	0.4300	0.5420
$\mathrm{DGP6}$	1	0.0415	0.0220	0.0390	0.0085	0.0110	0.0145
	5	0.1100	0.0740	0.0960	0.0505	0.0515	0.0615
	10	0.1820	0.1430	0.1695	0.0925	0.1055	0.1140
$\mathrm{DGP7}$	1	0.6080	0.5790	0.6340	0.0425	0.1725	0.1290
	5	0.7755	0.7800	0.7965	0.2380	0.5415	0.4740
	10	0.8370	0.8580	0.8695	0.4380	0.7255	0.6790
DGP8	1	0.7955	0.8080	0.8230	0.0850	0.3275	0.2375
	5	0.8935	0.9135	0.9100	0.3620	0.7540	0.6755
	10	0.9275	0.9525	0.9510	0.5875	0.8820	0.8410
DGP9	1	0.8780	0.8900	0.9065	0.0960	0.4185	0.2975
	5	0.9565	0.9625	0.9660	0.4575	0.8230	0.7810
	10	0.9760	0.9845	0.9840	0.7150	0.9445	0.9245
DGP10	1	0.9950	0.9955	0.9960	0.1140	0.6610	0.4790
	5	0.9980	1.0000	0.9995	0.6080	0.9780	0.9470
	10	0.9995	1.0000	1.0000	0.8645	0.9970	0.9925

 $\underline{ {\rm Table 7: Rejection \ Frequencies \ for \ Tests \ of \ Structural \ Change \ in \ the \ Overidentifying \ Restrictions} }$