

The Information Content of Implied Probabilities to Detect Structural Change

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

^a Université du Québec à Montréal, Québec, Canada, CIRPÉE and CIREQ.

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

Preliminary and incomplete version

First Version: July 2005
This Version: May 28, 2007

Abstract

This paper proposes to use implied probabilities resulting from estimation methods based on unconditional moment conditions to detect structural change. The class of GEL estimators (Smith 1997) assigns a set of probabilities to each observation such that moment conditions are satisfied. These restricted probabilities are called implied probabilities. Implied probabilities may also be constructed for the standard GMM (Bach and Brown 1993). The proposed structural change statistic tests are based on the information content in these implied probabilities. We consider cases of structural change which can occur in the parameters of interest or in the overidentifying restrictions used to estimate these parameters.

Keywords: Generalized empirical likelihood, generalized method of moments, parameter instability, structural change.

JEL codes: C12, C32

[†] The first author gratefully acknowledges financial support of Fonds pour la Formation de Chercheurs et l'aide à la Recherche (FCAR). The first and second authors are thankful for the financial support of the Social Sciences and Humanities Research Council of Canada (SSHRC).

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 of interest through moment restrictions. However, Monte Carlo results reveal that the GMM estimators may be seriously biased in small sample.¹ Recently, a number of alternative estimators to GMM have been proposed. Hansen, Heaton and Yaron (1996) suggest the continuous updated estimator (CUE) which shares same objective function that the GMM but with a weighting matrix depending on the parameters of interest. The empirical likelihood (EL) estimator (Qin and Lawless 1994) and the exponential tilting (ET) (Kitamura and Stutzer 1997) have also be proposed. These alternative estimators are special cases of the Generalized Empirical Likelihood (GEL) class considered by Smith (1997). These estimators may be shown to be based on the Cressie and Read (1984) family of power divergence criteria. Recently, Newey and Smith (2004) for an i.i.d. setting and Anatolyev (2005) with temporal dependence show that, in general, the asymptotic bias of GEL estimators is less than the GMM estimators especially with many moment conditions.

GEL estimators assigned a probability to each observation such that the moment conditions are satisfied (see Smith (2004)). This resulting empirical measure is called implied probabilities. Implied probabilities may also be constructed for the standard GMM as shown by Back and Brown (1993). The interpretation of the implied probabilities is the following: less weight is assigned to an observation for which the moment restrictions are not satisfied at the estimated values of the parameters and more weight to an observation for which the moment restrictions are satisfied. As suggested by Back and Brown (1993), implied probabilities may then provide a useful diagnostic device for the model specification. In particular, implied probabilities may content interesting information to detect instability in the sample. Consequently, we propose the use of these weights in detecting an unknown structural change in the model. To our knowledge this procedure was first proposed by Back and Brown (1993) but never pursued. Bertille, Bonnal and Renault (2007) also related the informational content of the moment conditions to the weights and one of their contribution is in the use of the weights in a three step estimation procedure. Schennach (2004) also discusses the use of these weights in the context of model misspecification. Ramalho and Smith (2005) considered Pearson-type test statistics (statistics based on the difference between restricted and unrestricted estimators of the weights) for the validity of moment restrictions and parametric restrictions (see also Otsu (2006)).

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 $\frac{1}{T}$. These statistic tests fall in three categories: 1) structural change tests to detect instability in the identifying restrictions

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

(see for example Andrews (1993)); 2) structural change tests to detect instability in the overidentifying restrictions (see for example Hall and Sen (1999)) and 3) and optimal predictive tests (see Ghysels, Guay and Hall (1997) and Guay (2003)). The asymptotic distribution of the statistic tests is derived for the case of unknown breakpoint under the null and under the alternative hypotheses.

The paper is organized as follows. A discussion on the GMM and GEL estimators are presented in section 2. Section 3 presents formally the full sample and partial sample GMM and 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 GMM and GEL Estimators

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

We consider a random variable $\{x_t : 1 \leq t \leq T, T \geq 1\}$. Suppose we have data that is summarized by a $T \times q$ matrix $g(x, \theta)$ with typical element $g_{ti}(x_t, \theta)$ with $i = 1, \dots, q$ which depends on some unknown p -vector of parameters $\theta \in \Theta$ with $\Theta \subset R^p$ (the parameter space will be refined subsequently) and that in the population their expected value is 0. That is,

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

In this paper we consider the overidentifying case with $q > p$. To put this into context, consider the consumption-based asset pricing model (see (Lucas 1978), Hansen and Singleton (1982) and Hall (2005) for a treatment in the context of GMM) which tries to explain how asset are priced and how consumption spending evolves. In this model a representative agent maximizes her discounted expected utility

$$E\left[\sum_{i=0}^{\infty} \delta_0^i U(c_{t+i}) | I_t\right]$$

subject to a budget constraint

$$c_t + \sum_{j=1}^N p_{j,t} q_{j,t} = \sum_{j=1}^N r_{j,t} q_{j,t-m_j} + w_t \tag{1}$$

In the above c_t denotes consumption, there are N assets with price $p_{j,t}$, quantity purchased $q_{j,t}$ and return $r_{j,t}$, w_t is labor income, δ is a discount factor and I is the information set of the agent.

The optimal consumption and investment paths satisfy the Euler equations which describes the co-movements of consumption and asset prices.

$$p_{j,t}U'(C_t) = \delta_0^{m_j} E[r_{j,t+m_j}U'(C_{t+m_j})|I_t]$$

$$E[\delta_0^{m_j}(r_{j,t+m_j}/p_{j,t})(U'(C_{t+m_j})/U'(C_t))|I_t] - 1 = 0.$$

If we use a CRRA functional form for the utility function $U(c_t) = (c_t^\gamma - 1)/\gamma$, then

$$E[\delta_0^{m_j}(r_{j,t+m_j}/p_{j,t})(c_{t+m_j}/c_t)^{\gamma-1}|I_t] - 1 = 0.$$

This first-order condition contains two parameters of interest that needs to be estimated. We can get around this problem by creating additional conditions upon which estimation will be based. For this, choose a set of instruments $z_t \in I_t$. Now let

$$u_{j,t}(\delta, \gamma) = \delta_0^{m_j}(r_{j,t+m_j}/p_{j,t})(c_{t+m_j}/c_t)^{\gamma-1} - 1$$

and using a law of iterated expectations we obtain

$$E[u_{j,t}(\delta, \gamma)z_t] = E[[u_{j,t}(\delta, \gamma)|I_t]z_t] = 0$$

Typically z_t contains $r_{j,t}/p_{j,t-m_j}$, c_t/c_{t-m_j} and a constant.

Now we can have at least as many population moment conditions than parameters to be estimated. Using the previous notation

$$E[g(x_t, \theta_0)] = E[[u_{j,t}(\delta, \gamma)|I_t]z_t] = 0$$

and θ_0 contains δ and γ , $p = 2$ and if we use one asset $q = 3$. Estimation could be done via maximum likelihood but this requires making a strong distributional assumption. Instead we use a method of moments.

The standard GMM estimators are obtained as the solution of

$$\tilde{\theta}_T = \arg \min_{\theta \in \Theta} g_T(\theta)'W_T g_T(\theta)$$

where W_T is a random positive definite symmetric $q \times q$ matrix and $g_T(\theta) = \frac{1}{T} \sum_{t=1}^T g(x_t, \theta)$. The optimal weighting matrix is defined to be the inverse of the covariance matrix of the moment conditions,

$W_T = \Omega_T^{-1}$ where Ω_T is a consistent estimator of

$$\Omega = \lim_{T \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T g(x_t, \theta) \right)$$

To address the time series properties of the data we use the long run covariance matrix estimator for Ω

$$\Omega_T = \Gamma_0 + \sum_{j=1}^{T-1} k(j/S_T)(\Gamma(j) + \Gamma(j)') = \Gamma_0 + \sum_{j=1}^{S_T} \left(1 - \frac{j}{S_T}\right)(\Gamma(j) + \Gamma(j)'),$$

if the Bartlett kernel, suggested by Newey and West (1987), is used with

$$\Gamma_0 = \frac{1}{T} \sum_{t=1}^T g(x_t, \theta)g(x_t, \theta)'$$

and

$$\Gamma(j) = \frac{1}{T} \sum_{t=j+1}^T g(x_t, \theta)g(x_{t-j}, \theta)'$$

We note that, in the GMM criterion function, the moment conditions receive equal weight ($1/T$) for each observations. Back and Brown (1993) derived, however, a set of *implicit* weights using the GMM estimators given by

$$\pi_t(\theta) = \frac{1}{T} - \frac{1}{T-p} \left[\tilde{J}_{tT}(\theta) - \bar{J}_T(\theta) \right]' \hat{\Omega}_T(\theta)^{-1} \bar{g}_T(\theta) \quad (2)$$

with $\bar{J}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \tilde{J}_{tT}(\theta)$ and

$$\tilde{J}_{tT}(\theta) = \sum_{j=1}^T \kappa(|t-j|)g(x_{t-j}, \theta)$$

where $\kappa(|t-j|)$ is a real valued weighting function and $\bar{g}_T(\theta) = \frac{1}{T} \sum_{t=1}^T g(x_t, \theta)$. The following estimator

$$\hat{\Omega}_T(\theta) = \frac{1}{T-p} \sum_{t=1}^T J_{tT}(\theta)g(x_t, \theta)'$$

is a consistent and positive definite estimator of the variance-covariance matrix and it has the usual form of a heteroskedasticity and autocorrelation consistent (HAC) weight matrix for standard kernels $\kappa(|t-s|)$ considered in the vast literature on consistent estimation of variance-covariance matrix (see e.g. Andrews (1991), Newey and West (1987) and Newey and West (1994)). The implied probabilities such defined above are the empirical measure that ensure that moment conditions are satisfied in the sample.

Now letting the T -vector of *explicit* weights be $\{\pi_t : 1 \leq t \leq T, T \geq 1\}$ 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

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

In practice, the vector π is obtained by solving

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

subject to $\sum_{t=1}^T \pi_t = 1$ and $\sum_{t=1}^T g(x_t, \theta)' \pi_t = 0$. In this case, $I(\pi)$ can be interpreted as the entropy measure of Shannon (1948) and it captures the degree of uncertainty in the distribution of π given the information that we have, the data. Equivalently, we could solve the cross-entropy problem using the Kullback-Leibler (Kullback and Leibler (1951)) information criterion

$$\Xi(\pi, 1/T) = \sum_{t=1}^T \pi_t \ln \pi_t + \ln(T).$$

$\Xi(\pi, 1/T)$ can then be interpreted as a directed sum of discrepancies between π and the empirical distribution $1/T$. Minimizing Ξ is then equivalent to maximizing I .

Maximum uncertainty is reached when $I(\pi) = \ln T$ or when $\Xi(\cdot) = 0$ with $\pi_t = 1/T, \forall t$. In this sense, the frequency distribution of π is chosen such that the missing information is maximized. Maximizing I subject to zero functions will result, unless $q = p$, in a frequency distribution of π that is different from the empirical distribution $1/T$, but as close to it as possible. The probabilities reveal and summarize, therefore, all the necessary information contained in the sample. 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 $I(\pi)$ as large as possible.

Estimates of π , given θ , are obtained by maximizing I subject to the weighted zero functions and the probability constraint. The solution to the Lagrangean yields

$$\pi_t^{ET}(\theta) = \frac{\exp(\gamma' g(x_t, \theta))}{\sum_{t=1}^T \exp(\gamma' g(x_t, \theta))} \quad (3)$$

where the q -vector γ contains the Lagrange multipliers and as such measure the degree of departure from zero of the moment conditions and the superscript stands for exponential tilting (see Kitamura and

Stutzer (1997)). Estimates of θ are obtained by substituting π in $I(\pi)$, maximizing it with respect to λ and then with respect to θ (see for example Kitamura and Stutzer (1997)).

If the sum of discrepancies between π and the empirical distribution $1/T$ is not directed (not efficiently weighted) we obtain a different set of weights given by

$$\pi_t^{EL}(\theta) = \frac{1}{T [1 + \gamma'g(x_t, \theta)]} \quad (4)$$

where the superscript relates to the empirical likelihood approach (see Qin and Lawless (1994) for example). When we evaluate the weights at some estimators we obtain $\pi_t^{ET}(\tilde{\theta}_T)$ and $\pi_t^{EL}(\tilde{\theta}_T)$. Recently, Schennach (2004) combined ET and EL into the ETEL estimator that combines the advantages of each approach.

To provide more intuition on 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, similar to the one used by Imbens, Spady and Johnson (1998) consists of estimating a single parameter, θ , with 2 moment conditions $x_t - \theta = 0$ and $(x_t - \theta)^2 - 4 = 0$ with a sample of 100 observations and $x_t = \theta_t + \epsilon_t$ where $\epsilon_t \sim N(0, 4)$. This data generating process (DGP) will be denoted as DGP1. The pulse, DGP2, is such that we have $\theta_t = 5$ for $t = 50$ and $\theta = 10$ otherwise. The break, DGP3, considers $\theta_t = 10$ for $t \leq 20$ and $t > 80$ while $\theta_t = 15$ for $21 \leq t \leq 80$. Finally, the regime shift, DGP4, has $\theta_t = 10$ for $t \leq 20$ and $\theta_t = 15$ otherwise. In these cases, structural change occurs via the parameters and can be tested using procedures proposed by Andrews (1993) and by Andrews and Ploberger (1994).

In contrast, the next two examples consist of structural change through the moment conditions. Following Hall and Horowitz (1996) and Gregory, Lamarche and Smith (2002) we study a simulated environment with CRRA which was presented above making a distributional assumption on consumption growth, x_t , with i.i.d data and $n = 100$. This is DGP5. 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: $E_t \exp[-\gamma \ln x_{t+1} - 9\sigma^2/2 + (3 - \gamma)z_t] = 1$ and $E_t z_t \exp[-\gamma \ln x_{t+1} - 9\sigma^2/2 + (3 - \gamma)z_t - 1] = 0$ 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 Ghysels et al. (1997), we have the estimation of an autoregressive parameter using two moment conditions when the DGP is an $AR(1)$ ($x_t = 0.5x_{t-1} + \epsilon_t$), for $t \leq 50$, and an $ARMA(1, 2)$ otherwise ($x_t = 0.5x_{t-1} + \epsilon_t + 0.5\epsilon_{t-2}$). There are 100 observations and $\epsilon_t \sim N(0, 1)$. This is DGP6. The two instruments used are the first and second lags of x_t .

Figure 1 shows the average of the vector of implied probabilities π over 10,000 replications. The key feature of these figures is that when there is no break, the weights fluctuate around $1/100$. With a

structural break in the parameter or in the moment conditions, however, more importance is given to observations (and moment conditions) for which there is no break. This paper focuses on the information contained in the estimated implied probabilities to detect these two types of structural change.

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

$$\tilde{\theta}_T = \arg \min_{\theta \in \Theta} \sup_{\gamma \in \tilde{\Gamma}_T(\theta)} \sum_{t=1}^T \frac{[\rho(\gamma'g(x_t, \theta)) - \rho_0]}{T} \quad (5)$$

where $\rho_j(\cdot) = \partial^j \rho(\cdot) / \partial \phi^j$ and $\rho_j = \rho_j(0)$ for $j = 0, 1, 2, \dots$. 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 continuously updated estimator, as opposed to the two-step estimator presented above, of Hansen et al. (1996) can also be obtained from the GEL formulation by using a quadratic function for $\rho(\phi)$.

A similar adjustment for the dynamic structure of $g(x_t, \theta)$ can also be made in the GEL context (see Kitamura and Stutzer (1997), Smith (2000), Smith (2004) and Guggenberger and Smith (2005)). The adjustment consists of smoothing the original moment conditions $g(x_t, \theta)$. Defining the smoothed moment conditions as

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

for $t = 1, \dots, T$ and S_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:

$$\sum_{t=1}^T \frac{[\rho(k\gamma'g_{tT}(\theta)) - \rho_0]}{T} \quad (6)$$

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

3 Full and Partial-Samples GMM and GEL Estimators

To establish the asymptotic distribution theory of tests of parameters based on implied probabilities, for structural change we need to elaborate on the specification of the parameter vector in our generic setup. We will consider parametric models indexed by parameters (β, δ) where $\beta \in B$, with $B \subset R^r$ and $\delta \in \Delta \subset R^v$. Following Andrews (1993) we make a distinction between pure structural change when no subvector δ appears and the entire parameter vector is subject to structural change under the alternative

and partial structural change which corresponds to cases where only a subvector β is subject to structural change under the alternative. The generic null can be written as follows:

$$H_0 : \beta_t = \beta_0 \quad \forall t = 1, \dots, T. \quad (7)$$

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

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

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

$$H_0 : \beta_1 = \beta_2 = \beta_0. \quad (8)$$

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 we have defined the moment conditions we will also translate this into over-identifying restrictions and relate it to structural change tests, following the analysis of Sowell (1996) and Hall and Sen (1999).

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 GMM estimators using the entire data sample or subsamples of observations. We discuss a set of regularity conditions required to obtain weak convergence of partial sample GMM estimators to a function of Brownian motions. To streamline the presentation we only summarize the assumptions and provide a detailed description of them in Appendix A. We now formally define the standard GMM estimator using the full sample.

Definition 3.1. *The full sample General Method of Moments estimator $\{\tilde{\theta}_T\}$ is a sequence of random vectors such that:*

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

where \hat{W}_T is a random positive definite symmetric $q \times q$ matrix.

The optimal weighting matrix W is defined to be the inverse of Ω which was defined in Section 2. The optimal weighting matrix can be estimated consistently using methods developed by Gallant (1987), Andrews and Monahan (1992) and Newey and West (1994), among several others.

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

Definition 3.2. A partial-sample General Method of Moments estimator $\{\hat{\theta}_T(s)\}$ is a sequence of random vectors such that:

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

for all $s \in S$. Define $g_t(\theta, s) = (g(x_t, \beta_1, \delta)', 0)'$ for $t = [Ts] + 1, \dots, T$ and $g_t(\theta, s) = (g(x_t, \beta_2, \delta)', 0)'$ for $t = [Ts] + 1, \dots, T$ such that

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

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

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

$$\Omega(s) = \lim_{T \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{T}} \begin{bmatrix} \sum_{t=1}^{[Ts]} g(x_t, \theta_0) \\ \sum_{t=[Ts]+1}^T g(x_t, \theta_0) \end{bmatrix} \right)$$

which under the null (8) is asymptotically equal to

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

To characterize the asymptotic distribution we define the following matrices:

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

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

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

In the GEL setting, the parameter vector is augmented by a vector of auxiliary parameters 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 function of this vector of auxiliary parameters can be written as follows:

$$H_0 : \gamma_t = \gamma_0 = 0 \quad \forall t = 1, \dots, T. \quad (9)$$

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

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

Thus under the null $H_0 = \gamma_{10} = \gamma_{20} = \gamma_0 = 0$. We will show later that a structural change in γ is associated with instability in over-identifying restrictions.

We now formally define the Generalized Empirical Likelihood (GEL) estimator using the full sample.

Definition 3.3. *Let $\rho(\phi)$ be a function of a scalar ϕ that is concave on its domain, an open interval Φ that contains 0. Also, let $\tilde{\Gamma}_T(\theta) = \{\gamma : k\gamma'g_{tT}(\tilde{\theta}_T) \in \Phi, t = 1, \dots, T\}$ with $k = \frac{k_1}{k_2}$. Then, the full sample GEL estimator $\{\tilde{\theta}_T\}$ is a sequence of random vectors such that:*

$$\tilde{\theta}_T = \arg \min_{\theta \in \Theta} \sup_{\gamma \in \tilde{\Gamma}_T(\theta)} \sum_{t=1}^T \frac{[\rho(k\gamma'g_{tT}(\theta)) - \rho_0]}{T} \quad (10)$$

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

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

More precisely, the GEL estimator is obtained as the solution to a saddle point problem. Firstly, the criterion is maximized for given θ . Thus,

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

Secondly, the GEL estimator $\tilde{\theta}_T$ is given by the following minimization problem:

$$\tilde{\theta}_T = \arg \min_{\theta \in \Theta} \sum_{t=1}^T \frac{[\rho(k\tilde{\gamma}(\theta)'g_{tT}(\theta)) - \rho_0]}{T}.$$

From now on, following Kitamura and Stutzer (1997) and Guggenberger and Smith (2005) we focus on the truncated kernel defined by

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

to obtain

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

To handle the endpoints in the smoothing we use the approach of ? which sets

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

Following Smith (2004), we can easily show for this kernel that $k_1 = 1$ and $k_2 = 1$. A consistent estimator of the long run covariance matrix is given by:

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

The weights thus obtained using this type of kernel are similar to those obtained with the Bartlett kernel estimator of the long run covariance matrix of the moment conditions. Define also the derivative of the smoothed moment conditions as:

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

Now consider a possible breakpoint $[Ts]$. Define the vector of auxiliary parameters $\gamma = (\gamma'_1, \gamma'_2)'$ where γ_1 is the vector of the auxiliary parameters for the first part of the sample e.g. $t = 1, \dots, [Ts]$ and γ_2 for the second part of the sample; $t = [Ts] + 1, \dots, T$. The partial-sample GEL estimators for $s \in S$ based on the first and the second subsamples are formally defined as:

Definition 3.4. A partial-sample General Empirical Likelihood (PS-GEL) estimator $\{\hat{\theta}_T(s)\}$ is a sequence of random vectors such that:

$$\begin{aligned} \hat{\theta}_T(s) &= \arg \min_{\theta \in \Theta} \sup_{\gamma \in \hat{\Gamma}_T(\theta)} \sum_{t=1}^T \frac{[\rho(\gamma(s))' g_{tT}(\theta, s) - \rho_0]}{T} \\ &= \arg \min_{\theta \in \Theta} \sup_{\gamma \in \hat{\Gamma}_T(\theta)} \left(\sum_{t=1}^{[Ts]} \frac{[\rho(\gamma'_1) g_{tT}(\beta_1, \delta) - \rho_0]}{T} + \sum_{t=[Ts]+1}^T \frac{[\rho(\gamma'_2) g_{tT}(\beta_2, \delta) - \rho_0]}{T} \right) \end{aligned}$$

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

Existence of partial-sample GEL estimators can be established under standard conditions. As explained earlier, the GEL estimators are obtained as the solution to a saddle point problem. Let us examine the first orders conditions of this saddle point problem for the partial-sample GEL. Firstly, the

first order conditions corresponding to the Lagrange multiplier γ are obtained the maximization of the partial-sample GEL criterion for a given β, δ . Thus,

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

The first order conditions can then be evaluated at $\hat{\beta}_1, \hat{\delta}$ for the first part of the sample $t = 1, \dots, [Ts]$ to obtain $\hat{\gamma}_1 = \gamma(\hat{\beta}_1, \hat{\delta})$ and at $\hat{\beta}_2, \hat{\delta}$ for the second part $t = [Ts] + 1, \dots, T$ to obtain $\hat{\gamma}_2 = \gamma(\hat{\beta}_2, \hat{\delta})$.

The corresponding first order conditions are given by:

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^{[Ts]} \rho_1(\hat{\gamma}_1(\beta_1, \delta))' g_{tT}(\beta_1, \delta) g_{tT}(\beta_1, \delta) &= 0 \\ \frac{1}{T} \sum_{t=[Ts]+1}^T \rho_1(\hat{\gamma}_2(\beta_2, \delta))' g_{tT}(\beta_2, \delta) g_{tT}(\beta_2, \delta) &= 0.\end{aligned}$$

Secondly, the partial-sample GEL estimators $\hat{\beta}_1, \hat{\beta}_2, \hat{\delta}$ are the minimizer of the partial-sample GEL criterion (3.2). The corresponding first order conditions are:

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^{[Ts]} \rho_1(\hat{\gamma}_1(\hat{\beta}_1, \hat{\delta}))' g_{tT}(\hat{\beta}_1, \hat{\gamma}_1) G_{tT}^\beta(\hat{\beta}_1, \hat{\delta})' \hat{\gamma}_1(\hat{\beta}_1, \hat{\delta}) &= 0, \\ \frac{1}{T} \sum_{t=[Ts]+1}^T \rho_1(\hat{\gamma}_2(\hat{\beta}_2, \hat{\delta}))' g_{tT}(\hat{\beta}_2, \hat{\gamma}_2) G_{tT}^\beta(\hat{\beta}_2, \hat{\delta})' \hat{\gamma}_2(\hat{\beta}_2, \hat{\delta}) &= 0,\end{aligned}$$

and

$$\frac{1}{T} \sum_{t=1}^T \rho_1(\hat{\gamma}(\hat{\theta}(s), s))' g_{tK}(\hat{\theta}(s), s) G_{tK}^\delta(\hat{\theta}(s), s)' \hat{\gamma}(\hat{\theta}(s), s) = 0. \quad (11)$$

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

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

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

Theorem 3.1. *Under Assumptions 7.1 to 7.10, every sequence of PS-GEL estimators $\{\hat{\theta}(\cdot), \hat{\gamma}(\cdot), T \geq 1\}$ satisfies*

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

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

Proof: see Appendix.

We now present the corresponding implied probabilities such defined by Back and Brown (1993) and Smith (2004) for the most commonly used full and partial-sample estimators. Following Back and Brown (1993), the full sample GMM implied probabilities are defined as:

$$\tilde{\pi}_t(\tilde{\theta}_T) = \frac{1}{T} - \frac{1}{T-p} \left[\tilde{J}_{tT}(\tilde{\theta}_T) - \bar{J}_T(\tilde{\theta}_T) \right]' \hat{\Omega}_T(\tilde{\theta}_T)^{-1} \bar{g}_T(\tilde{\theta}_T) \quad (12)$$

with $\bar{J}_T(\tilde{\theta}_T) = \frac{1}{T} \sum_{t=1}^T \tilde{J}_{tT}(\tilde{\theta}_T)$ and

$$\tilde{J}_{tT}(\tilde{\theta}_T) = \sum_{j=1}^T \kappa(|t-j|) g(x_{t-j}, \tilde{\theta}_T)$$

where $\kappa(|t-j|)$ is a real valued weighting function and $\bar{g}_T(\theta) = \frac{1}{T} \sum_{t=1}^T g(x_t, \theta)$. In practice, however, some of the estimated probabilities may be negative in finite sample although these probabilities are asymptotically positive. Antoine, Bonnal, and Renault (2007) proposes a shrinkage procedure defined as weighted average of the standard 2S-GMM's implied probabilities ($1/T$) and the computed implied probabilities to guarantee the non-negativity of these implied probabilities in finite sample.

Implied probabilities for the full sample ET and EL estimators are respectively given by:

$$\tilde{\pi}_t^{ET}(\tilde{\theta}_T) = \frac{\exp[\tilde{\gamma}' g_{tT}(\tilde{\theta}_T)]}{\sum_{t=1}^T \exp[\tilde{\gamma}' g_{tT}(\tilde{\theta}_T)]} \quad (13)$$

and

$$\tilde{\pi}_t^{EL}(\tilde{\theta}_T) = \frac{1}{T[1 + \tilde{\gamma}' g_{tT}(\tilde{\theta}_T)]}. \quad (14)$$

The corresponding unrestricted partial-sample implied probabilities are defined for $s \in S$ as:

$$\hat{\pi}_t(\hat{\theta}_T(s)) = \frac{1}{T} - \frac{1}{T-p} \left[\tilde{J}_{tT}(\hat{\theta}_T(s), s) - \bar{J}_T(\hat{\theta}_T(s), s) \right]' \hat{\Omega}_T(s)^{-1} \bar{g}_T(\hat{\theta}_T(s), s) \quad (15)$$

with $\bar{J}_T(\hat{\theta}_T(s), s) = \frac{1}{T} \sum_{t=1}^T \tilde{J}_{tT}(\hat{\theta}_T(s), s)$ and

$$\tilde{J}_{tT}(\hat{\theta}_T(s), s) = \sum_{j=1}^T \kappa(|t-j|) g_{t-j}(\hat{\theta}_T(s), s)$$

where $\kappa(|t-j|)$ is a real valued weighting function and $\bar{g}_T(\hat{\theta}_T(s), s) = \frac{1}{T} \sum_{t=1}^T g_t(\hat{\theta}_T(s), s)$. Finally,

$$\hat{\pi}_t^{ET}(\hat{\theta}_T(s)) = \frac{\exp[\hat{\gamma}(s)' g_{tT}(\hat{\theta}_T(s), s)]}{\sum_{t=1}^T \exp[\hat{\gamma}(s)' g_{tT}(\hat{\theta}_T(s), s)]} \quad (16)$$

and

$$\widehat{\pi}_t^{EL}(\widehat{\theta}_T(s)) = \frac{1}{T[1 + \widehat{\gamma}(s)'g_{tT}(\widehat{\theta}_T(s), s)]}. \quad (17)$$

The purpose of the next section is to refine the null hypothesis of no structural change. Such a refinement will enable us to construct various tests for structural change in the spirit of Sowell (1993, 1996a) and Hall and Sen (1999). In a first subsection we refine the null hypothesis, a second subsection is devoted to tests for parameter constancy, a final subsection covers tests for stability of over-identifying restrictions.

3.2 Refining the Null Hypothesis

The moment conditions for the full sample under the null can be written as: $Eg_t(\theta_0) = 0, \forall t = 1, \dots, T$. Following Sowell (1993, 1996a), we can project the moment conditions on the subspace identifying the parameters and the subspace of over-identifying restrictions. In particular, considering the (standardized) moment conditions for the full sample GMM estimator, such a decomposition corresponds to:

$$\Omega^{-1/2}Eg_t(\theta_0) = P_G\Omega^{-1/2}Eg_t(\theta_0) + (I_q - g_F)\Omega^{-1/2}Eg_t(\theta_0),$$

where $P_G = \Omega^{-1/2}G[G'^{-1}G]^{-1}G'\Omega^{-1/2}$. The first term is the projection identifying the parameter vector and the second term is the projection for the over-identifying restrictions. The projection argument enables us to refine the null hypothesis (8). For instance, following Hall and Sen (1999) we can consider the null, for the case of a single breakpoint, which separates the identifying restrictions across the two subsamples:

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

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:

$$\begin{aligned} H_0^{O1}(s) : (I_q - P_G)\Omega^{-1/2}E[g_t(\theta_0)] &= 0 & \forall t = 1, \dots, [Ts] \\ H_0^{O2}(s) : (I_q - P_G)\Omega^{-1/2}E[g_t(\theta_0)] &= 0 & \forall t = [Ts] + 1, \dots, T \end{aligned}$$

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 local alternatives based on the moment conditions:

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

$$Eg_t(\theta_0) = h(\eta, \tau, \frac{t}{T})/\sqrt{T} \quad (18)$$

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

3.3 Tests for parameter constancy

In this section we introduce several tests for structural change for parameter stability and establish their asymptotic distribution. The null hypothesis is (8), or more precisely $H_0^I(\pi)$. We present Wald, Lagrange multiplier and likelihood ratio-type statistics. Predictive tests will be discussed in the next section. For brevity, we consider only test statistics based on the optimal weighting matrix. The results obtained are valid under an arbitrary weighting matrix. The first is the usual Wald statistic which is given by:

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

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

$$\begin{aligned} \hat{G}_{1,tT}^\beta &= \frac{1}{[Ts]} \sum_{t=1}^{[Ts]} \frac{\partial g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))}{\partial \beta_1'} \\ \hat{\Omega}_{1T}(s) &= \frac{2K_T + 1}{[Ts]} \sum_{t=1}^{[Ts]} g_{tT}(\beta_1(s), \delta) g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))' \end{aligned}$$

and for the second part of the sample:

$$\begin{aligned} \hat{G}_{2,tT}^\beta &= \frac{1}{T - [Ts]} \sum_{t=[Ts]+1}^T \frac{\partial g_{tT}(\hat{\beta}_2(s), \hat{\delta}(s))}{\partial \beta_2'} \\ \hat{\Omega}_{2T}(s) &= \frac{2K_T + 1}{T - [Ts]} \sum_{t=[Ts]+1}^T g_{tT}(\hat{\beta}_2(s), \hat{\delta}(s)) g_{tT}(\hat{\beta}_2(s), \hat{\delta}(s))'. \end{aligned}$$

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

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

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

where

$$\begin{aligned} g_{1T}(\tilde{\theta}_T, s) &= \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} g_{tT}(\tilde{\theta}_T), \\ \hat{G}_{tT}^\beta &= \frac{1}{T} \sum_{t=1}^T \frac{\partial g_{tT}(\tilde{\beta}, \tilde{\delta})}{\partial \beta'}, \\ \hat{\Omega}_T &= \frac{2K_T + 1}{T} \sum_{t=1}^T g_{tT}(\tilde{\beta}, \tilde{\delta}) g_{tT}(\tilde{\beta}, \tilde{\delta})'. \end{aligned}$$

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

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

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

We state now the main theorem which establishes the asymptotic distribution of the Wald, LM and LR-type test statistics under the local alternative (18).

Theorem 3.2. *Under the null hypothesis H_0 in (8) and Assumptions 7.1 to 7.10, the following processes indexed by s for a given set S whose closure lies in $(0, 1)$ satisfy:*

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

with

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

and under the alternative (18)

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

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

Proof: See Appendix.

The result in Theorem 3.2 tells us that the asymptotic distributions under the null of the Wald, LR-type and LM statistics are the same as those obtained by Andrews (1993) for the GMM estimator.

When s is unknown, i.e. the case of unknown breakpoint, we can use the above result to construct statistics across $s \in S$. In the context of maximum likelihood estimation, Andrews and Ploberger (1994)

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 Sowell (1996a) optimal tests are a generalization of the Andrews and Ploberger approach to the case of two measures that do not admit densities. The most powerful test is given by the Radon-Nikodym derivative of the probability measure implied by the local alternative with respect to the probability measure implied by the null hypothesis.

The optimal average exponential form is the following:

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

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

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

Theorem 3.3. *Under the null hypothesis H_0 in (8) and Assumptions 7.1 to 7.10, the following processes indexed by s for a given set S whose closure lies in $(0,1)$ satisfy:*

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

with

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

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

3.4 Tests for Stability of Overidentifying Restrictions

Tests presented in the preceding section are based on the projection of the moment conditions on the subspace of identifying restrictions. In this section we are interested with testing against violations of $H_0^{O1}(\pi)$ or $H_0^{O2}(\pi)$. The local alternatives are given by the projection of the moment condition on the

subspace orthogonal to the identifying restrictions. For instance, in the case of a single breakpoint, the local alternatives by Assumption 3.1 correspond to:

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

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

We propose here equivalent statistics to the Hall and Sen's statistics specially designed to detect instability before and after the possible breakpoint. The first statistic is based on the same statistic that the one of Hall and Sen (1999) except that is computed with smoothed moment conditions. The $O_T(s)$ statistic is the sum of the GMM criterion function in each subsample

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

where

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

and

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

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

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

where

$$O1_T^{GEL}(s) = \frac{2[Ts]}{2K+1} \sum_{t=1}^{[Ts]} \frac{[\rho(\lambda_1(s)'g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))) - \rho_0]}{[Ts]}$$

and

$$O2_T^{GEL}(s) = \frac{2(T - [Ts])}{2K+1} \sum_{t=[Ts]+1}^T \frac{[\rho(\lambda_1(s)'g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))) - \rho_0]}{T - [Ts]}.$$

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

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

where

$$LM1_T^O(s) = \frac{T}{(2K_T + 1)^2} \hat{\gamma}_{1T}(s)' \hat{\Omega}_{1T}(s) \hat{\gamma}_{1T}(s)$$

and

$$LM2_T^O(s) = \frac{T}{(2K_T + 1)^2} \hat{\gamma}_{2T}(s)' \hat{\Omega}_{2T}(s) \hat{\gamma}_{2T}(s)$$

The equivalence with the overidentifying test of instability results from that $\sqrt{T}/(2K_T + 1) \hat{\gamma}_1(s)$ is asymptotically equivalent at the first order to $\Omega_1(s)^{-1} T^{1/2} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))$ and $\sqrt{T}/(2K_T + 1) \hat{\gamma}_2(s)$ to $\Omega_2(s)^{-1} T^{1/2} \sum_{t=[Ts]+1}^T g_{tT}(\hat{\beta}_2(s), \hat{\delta}(s))$.

The following theorem provides the asymptotic distribution under the null and the alternative of those test statistics for the supremum mapping: $\sup Q_T^O = \sup_{s \in S} Q_T^O(s)$, the average mapping: $ave Q_T^O = \int_S Q_T^O(s) dJ(s)$ and exponential mapping: $exp Q_T^O = \log \left(\int_S \exp \frac{1}{2} Q_T^O(s) dJ(s) \right)$ with $Q_T^O(s)$ equal $O_T(s)$, $O_T^{GEL}(s)$ and $LM_T^O(s)$.

Theorem 3.4. *Under the null of no structural change and Assumptions 7.1 to 7.10, the following processes indexed by s for a given set S whose closure lies in $(0,1)$ satisfy:*

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

with

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

and under the alternative (18)

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

where $B_{q-p}(s)$ is a $q-p$ -dimensional vector of independent Brownian motion and $P_G = \Omega^{-1/2} G (G' \Omega^{-1} G)^{-1} G' \Omega^{-1/2}$.

Proof: See Appendix.

The asymptotic critical values are those test statistics are tabulated in Hall and Sen (1999). The last two terms in the asymptotic distribution under the alternative given in Theorem 3.4 show that the test statistics have non trivial power to detect instability before and after the possible breakpoint point.

4 Tests for a structural change based on implied probabilities

We first consider a statistic test based on the partial sum of the implied probabilities evaluated at the restricted estimator for the GMM and GEL. The test statistic proposed to detect structural change is given by

$$IPS_T^j(s) = \frac{1}{2K_T + 1} \left[\sum_{t=1}^{\lfloor Ts \rfloor} \left(T\pi_t^j(\tilde{\theta}_T) - 1 \right) \right]^2$$

The next Theorem establishes the asymptotic distribution of the partial sample overidentifying moment restriction test.

Theorem 4.1. *Under Assumptions (7.1) to (7.10), the following processes indexed by s for a given $S \in \mathcal{S}$ whose closure lies in $(0, 1)$ satisfy*

$$IPS_T^{ET}(s), IPS_T^{EL}(s), IPS_T^{GMM}(s) \Rightarrow BB_p(s)'BB_p(s) + B_{(q-p)}(s)'B_{(q-p)}(s)$$

and under the alternative (18)

$$\begin{aligned} IPS_T^{ET}(s), IPS_T^{EL}(s), IPS_T^{GMM}(s) \Rightarrow & BB_p(s)'BB_p(s) + (H(s) - sH(1))' \Omega^{-1/2} P_G \Omega^{-1/2} (H(s) - sH(1)) \\ & + B_{q-p}(s)'B_{q-p}(s) + H(s)' \Omega^{-1/2} (I - P_G) \Omega^{-1/2} H(s), \end{aligned}$$

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.

See the appendix for the proof.

This 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 is function of the Brownian bridges corresponds to a parameter stability tests and the second component to a stability of overidentifying restrictions. This statistic test based on implied probabilities can be viewed as a more general of misspecification due to instability that just a test for parameter variation. The predictive tests proposed by Ghysels, Guay and Hall (1997) shares the same properties.

The following Theorem gives the asymptotic distribution for the exponential mapping for Q_T^{IP} where Q_T^{IP} corresponds to the statistic presented above based on the implied probabilities.

Theorem 4.2. *Under the null hypothesis H_0 in (8) and Assumptions 7.1 to 7.10, the following processes indexed by s for a given set S whose closure lies in $(0, 1)$ satisfy:*

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

with

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

This result is obtained through the application of the continuous mapping theorem (see Pollard (1984)).

The simulated critical values for $\pi_T^{ET}(s)$, $\pi_T^{EL}(s)$, $\pi_T^{GMM}(s)$ can be found in XX.

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

4.1 Tests for a structural change in the parameters based on implied probabilities

The statistic proposed to detect parameters instability is based on the difference between the partial sum of unrestricted implied probabilities (appropriately scaled) with the corresponding partial sum of restricted implied probability. More precisely the statistic tests are defined as:

$$IPST_T^{I,j}(s) = \frac{1}{2K+1} \sum_{t=1}^T \left(T\pi_t^j(\hat{\theta}_T(s)) - T\pi_t^j(\tilde{\theta}_T) \right)^2$$

for $j = ET, EL$ or GMM .

Theorem 4.3. *Under the null hypothesis H_0 in (8) and Assumptions Under Assumptions (7.1) to (7.10), the following processes indexed by s for a given set S whose closure lies in $(0, 1)$ satisfy:*

$$IPST_T^{I,ET}(s) \Rightarrow Q_r(s), IPST_T^{I,EL}(s) \Rightarrow Q_r(s), IPST_T^{I,GMM}(s) \Rightarrow Q_r(s),$$

with

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

and under the alternative (18)

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

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

Proof: See Appendix B.

The Theorem shows that the asymptotic distribution of the proposed test based on implied probabilities is asymptotically equivalent under the null and the alternative to the Wald, LM and LR tests for parameter instabilities. However, the small sample properties can differ compared to those more standard tests.

By Lemma 7.4, the two following modified statistics are equivalent the the one defined above:

$$IPSM1_T^{I,j}(s) = \frac{1}{2K+1} \sum_{t=1}^T \frac{\left(T\pi_t^j(\hat{\theta}_T(s)) - T\pi_t^j(\tilde{\theta}_T)\right)^2}{T\pi_t^j(\hat{\theta}_T(s))}$$

and

$$IPSM2_T^{I,j}(s) = \frac{1}{2K+1} \sum_{t=1}^T \frac{\left(T\pi_t^j(\hat{\theta}_T(s)) - T\pi_t^j(\tilde{\theta}_T)\right)^2}{T\pi_t^j(\tilde{\theta}_T)}$$

for $j = ET, EL$ or GMM .

4.2 Tests for a structural change for overidentifying restrictions based on implied probabilities

Ramalho and Smith (2005) introduce in i.i.d. setting a specification tests for moment conditions based on implied probabilities similar in spirit to the classical Pearson Chi-Square goodness-of-fit tests. The test is based on the following statistic:

$$\sum_{t=1}^T \left(T\pi_t^j(\hat{\theta}_{1T}(s)) - 1\right)^2.$$

They show that such statistic is asymptotically equivalent to the overidentifying moment restrictions J-test proposed by Hansen (1982). As shown by Ghysels and Hall, the J-test has no power to detect structural change, a property than shared by the specification test proposed by those authors.

Here, we propose a statistic specially design to detect overidentifying restrictions instability based on implied probabilities. The statistic is asymptotically equivalent to the ones proposed by Hall and Sen (1999) and thus shares its asymptotic properties. More precisely the statistic test is defined as:

$$IPST^{O,j}(s) = \frac{1}{2K_T+1} \sum_{t=1}^{[Ts]} \left([Ts]\pi_t^j(\hat{\theta}_{1T}(s)) - 1\right)^2 + \frac{1}{2K_T+1} \sum_{t=[Ts]+1}^T \left((T-[Ts])\pi_t^j(\hat{\theta}_{2T}(s)) - 1\right)^2 \quad (19)$$

for $j = ET, EL$ or GMM . The statistic is the sum of the overidentifying restrictions statistic proposed by Ramalho and Smith (2005) for the first and the second part of the sample adapted for dependent data.

The next Theorem establishes the asymptotic distribution of the partial sample overidentifying moment restriction test.

Theorem 4.4. *Under Assumptions (7.1) to (7.10), the following processes indexed by s for a given $S \in \mathcal{S}$ whose closure lies in $(0, 1)$ satisfy*

$$IPST^{O,ET}(s) \Rightarrow Q_{q-p}(s), IPS_T^{O,EL}(s) \Rightarrow Q_{q-p}(s), IPS_T^{O,GMM}(s) \Rightarrow Q_{q-p}(s)$$

with under the null of no structural change

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

and under the alternative (18)

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

where $B_{(q-p)}(s)$ is a $(q-p)$ -vector of standard Brownian motion.

See the appendix for the proof.

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

By Lemma 7.4, an asymptotic equivalent statistic to (19) in the spirit of the Neyman-modified chi-square is given by:

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

for $j = ET, EL$ or GMM .

5 Robust structural change tests to weak identification based on implied probabilities

To be completed

6 Simulation Evidence

To evaluate the performance of the test statistics we use the data generating process of (Ghysels et al. 1997) and of (Hall and Sen 1999). We set $p = 1$ and generate x_t as either an AR(1) process

$$x_t = \theta_i x_{t-1} + u_t \tag{20}$$

or as an ARMA(1,2)

$$x_t = \theta_i x_{t-1} + u_t + \alpha \epsilon_{t-2} \tag{21}$$

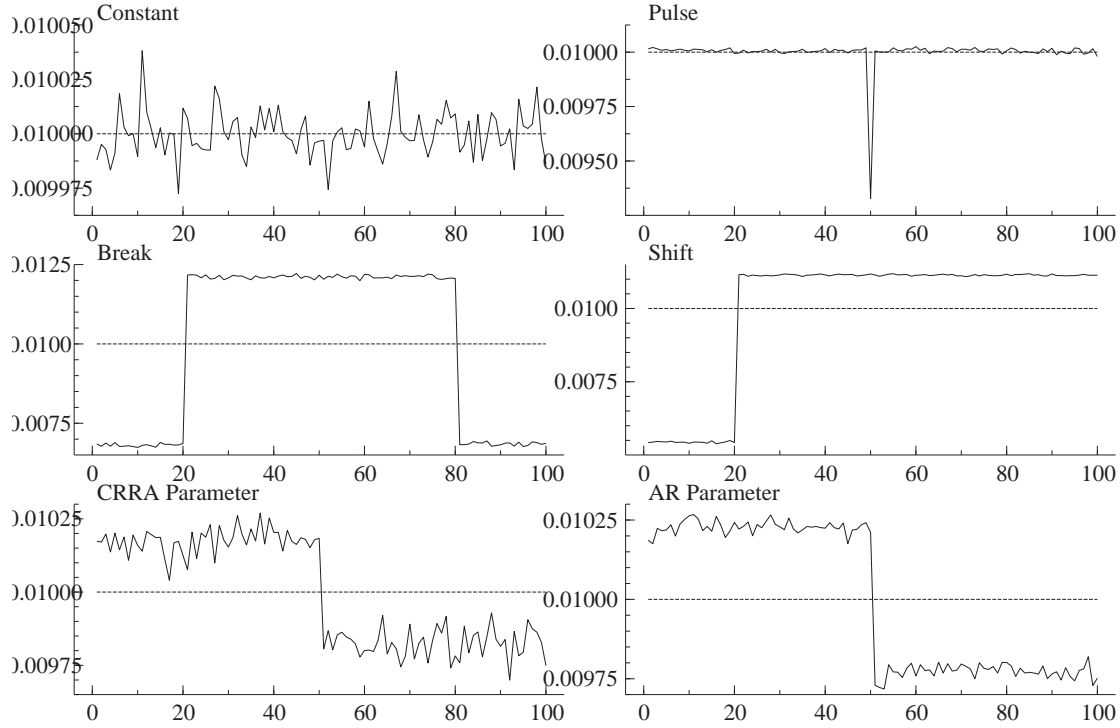
for $t = 1, \dots, T$. Structural change in the identifying restrictions is studied by considering different values of θ_i where $i = 1, 2$ denotes the first or second subsample while structural stability in the overidentifying restrictions is studied by considering nonzero values of α in the second subsample. The change is set at $T/2$. We use two moment conditions, $q = 2$, constructed using the first and second lags of x_t . In the above, $u_t \sim N(0, 1)$. The following table summarizes the different parametrization and is taken from (Hall and Sen 1999):

Table 1: Data Generating Processes

	$H_0^{SS}(r)$	$H_A^I(r)$	$H_A^O(r)$
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		$\theta_1 = 0.4, \theta_2 = 0.8$	$\alpha = 0$
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	$\theta_1 = \theta_2 = 0.4$		$\alpha = -0.9$

TO BE COMPLETED

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 CRR 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 in all cases the horizontal lines correspond to the empirical weights of $1/100$.

7 Appendix

TO BE IMPROVED AND COMPLETED

7.1 Assumptions

Assumption 7.1. The process $\{x_t\}_{t=1}^\infty$ is a finite dimensional stationary and strong mixing coefficients $\sum_{j=1}^\infty \alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$.

Assumption 7.2. (a) $\theta_0 \in \Theta$ is the unique solution to $E[g(x_t, \theta)] = 0$ where Θ is compact, $g(x_t, \theta)$ is continuous at each $\theta \in \Theta$ with probability one; (b) $E[\sup_{\theta \in \Theta} \|g(x_t, \theta)\|^\alpha] < \infty$ for some $\alpha > \max\left(4\nu, \frac{1}{\eta}\right)$; (c) $\Omega(\theta)$ is finite and positive definite for all $\theta \in \Theta$.

Define the smoothed moment conditions as:

$$g_{tT}(\theta) = \frac{1}{S} \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) g_{t-s}(x_t, \theta)$$

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

$$g_{tT}(\theta) = \frac{1}{2K_T + 1} \sum_{s=-K_T}^{K_T} g_{t-s}(x_t, \theta)$$

Assumption 7.3. $K_T^2/T \rightarrow 0$ and $K_T \rightarrow \infty$ as $T \rightarrow \infty$.

Assumption 7.4. (a) $\rho(\cdot)$ is twice continuously differential and concave on its domain, an open interval V containing θ , $\rho_1 = \rho_2 = -1$; (b) $\gamma \in \Gamma$ where $\Gamma = [\gamma : \|\gamma\| \leq D (T/K_T^2)^{-\zeta}]$ for some $D > 0$ with $\frac{1}{2} > \zeta > \frac{1}{2\alpha\eta}$.

Under Assumptions 7.1 to 7.4, Smith (2004) shows that for the full sample estimators $\tilde{\theta}_T \xrightarrow{p} \theta_0$, $\tilde{\gamma}_T \xrightarrow{p} 0$, $\|\tilde{\gamma}_T\| = O_p[(T/2K_T + 1)^2]^{-1/2}$ and $\|\frac{1}{T} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T)\| = O_p(T^{1/2})$. Now, Let $G(\theta) = E[\partial g(x_t, \theta)/\partial \theta]$ and $G = G(\theta_0)$.

Assumption 7.5. $g(\cdot, \theta)$ is differentiable in $\theta_0 \in \Theta_0$ where Θ_0 is a neighborhood of θ_0 , $E[\sup_{\theta \in \Theta_0} \|\partial g(x_t, \theta)/\partial \theta'\|^\alpha / (\alpha-1)] < \infty$ and $\text{rank}(G) = p$.

Assumptions 7.1 to 7.5 yield the asymptotic distribution of $T^{1/2}(\tilde{\theta}_T)$ and $(T/(2K_T + 1))^2 \tilde{\gamma}_T$ (see Smith 2004). Also, under these assumptions:

$$\frac{1}{T} \sum_{t=1}^T G_{tT}(\tilde{\theta}_T) \xrightarrow{p} G$$

and

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

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

Assumption 7.6. $\text{Var}\left(\frac{1}{T} \sum_{t=1}^{Ts} g(x_t, \theta)\right) \rightarrow s\Omega \forall s \in [0, 1]$.

Assumption 7.7. $\sup_{s \in S} \|\hat{\Omega}_T(s) - \Omega(s)\| \xrightarrow{p} 0$ where $\Omega(s)$ is defined in Section 3.1.

Assumption 7.8. $\sup_{s \in S} \|\hat{\theta}_T(s) - \theta_0\| \xrightarrow{p} 0$ for some θ_0 in the interior of Θ and $\sup_{s \in S} \|\hat{\gamma}_T(s) - 0\| \xrightarrow{p} 0$.

Assumption 7.9. $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{Ts} E \partial g(x_t, \theta_0) / \partial \theta'$ exists uniformly over $s \in S$ and equals $sG \forall s \in S$.

Assumption 7.10. $G(s)' \Omega(s)^{-1} G(s)$ is nonsingular $\forall s \in S$ and has eigenvalues bounded away from zero.

7.2 Lemmas

Lemma 7.1. *Under Assumptions 7.1 to 7.1, the asymptotic distribution of the smoothed moment conditions under the null is given by:*

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) \Rightarrow B(s),$$

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

Proof of Lemma 7.1 First, by a well established result:

$$\Omega^{-1/2} \sum_{t=1}^{[Ts]} g_t(\theta_0) \Rightarrow B(s)$$

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

Second, the smoothed moment condition are defined as:

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \frac{1}{2K_T + 1} \sum_{j=-K_T}^{K_T} g_{t-j}(\theta_0).$$

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

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

which implies that

$$\sum_{t=1}^{[Ts]} g_t(\theta_0) = \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) + O_p\left(\frac{K_T^2}{2K_T+1}\right),$$

Under the Assumptions that $\max_{1 \leq t \leq T} \|g_t(\theta_0)\| = o_p(T^{1/2})$ and $K_T^2/T \rightarrow 0$, we get

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_t(\theta_0) = \Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) + o_p(1).$$

which yields the asymptotic equivalence.

The following Lemma provides the asymptotic distribution of the smoothed moment condition under the general sequence of local alternatives appearing in Assumption 18.

Lemma 7.2. *Let Assumptions 7.1 to 7.10 and Assumption 18 hold, then*

$$\frac{1}{\sqrt{T}} \Omega^{-1/2} \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) \Rightarrow B(s) + \Omega^{-1/2} H(s)$$

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

Proof of Lemma 7.2

Under the alternative (18), the sample smoothed moments satisfy:

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \frac{1}{2K_T + 1} \sum_{j=-K_T}^{K_T} \left(g_{t-j}(\theta_0) - \frac{h((t-j)/T)}{\sqrt{T}} \right) \Rightarrow B(s),$$

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

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \sum_{j=-K_T}^{K_T} g_{t-j}(\theta_0) - \Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \frac{1}{2K_T + 1} \sum_{j=-K_T}^{K_T} \frac{h((t-j)/T)}{\sqrt{T}}.$$

Let us now examine the last term,

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

which equals

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

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

The next Lemma gives the asymptotic distribution of the partial sum of the full sample estimator for the general local alternatives based on smoothed moment conditions. With this result, we can obtain the asymptotic distribution for optimal tests for the stability of overidentifying restrictions with GEL. The asymptotic distributions are the same than derived by Sowell and Hall and Sen

Lemma 7.3. *Under Assumptions 7.1 to 7.10 and under the alternative (18), then*

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T) &\Rightarrow \Omega^{1/2} B(s) + H(s) - \\ & sG(G'\Omega^{-1}G)^{-1}G'\Omega^{-1} \left[\Omega^{1/2} B(1) + H(1) \right]. \end{aligned}$$

where $B(\pi)$ is a q -dimensional vectors of mutually independent Brownian motion.

Lemma 7.4. *Under Assumptions 7.1 to 7.10,*

$$\pi_t^j(\tilde{\theta}_T) = \frac{1}{T} - \frac{1}{T} g_{tT}(\tilde{\theta}_T)' \Omega^{-1} \frac{2K_T+1}{T} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T) + O_p(T^{-2})$$

and $\pi_t^j(\tilde{\theta}_T) = \frac{1}{T} + o_p(1)$ uniformly in $t = 1, \dots, T$ and $j = ET, EL$ or GMM .

Proof of Lemma 7.4

We need to derive the asymptotic distributions of the implied probabilities for the three following cases evaluated at the full sample estimator:³

³By (3.1) of Smith (2004)

- Exponential Tilting:

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

- Empirical Likelihood:

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

- Standard Generalized Method of Moments⁴

$$\pi_t^{GMM}(\tilde{\theta}_T) = \frac{1}{T} - \frac{1}{T} \left[g(x_t, \tilde{\theta}_T) - \frac{1}{T} \sum_{t=1}^T g(x_t, \tilde{\theta}_T) \right]' \left[\text{Var}g(x_t, \tilde{\theta}_T) \right]^{-1} \frac{1}{T} \sum_{t=1}^T g(x_t, \tilde{\theta}_T)$$

First, consider the implied probabilities for the exponential tilting. A mean value expansion around $(\tilde{\theta}_T, \tilde{\gamma}_T) = (\tilde{\theta}_T, 0)$ yields:

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

where $\tilde{\gamma}$ lies on the line segment joining $\tilde{\theta}$ and 0 and may differ from row to row.

Since $\tilde{\gamma}_T$ converges in probability to 0, thus:

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

As $\sum_{t=1}^T g_{tT}(\tilde{\theta}_T)/T = O_p(T^{-1/2})$, $g_{tT}(\tilde{\theta}_T) = O_p((2K_T + 1)^{-1/2})$ ⁵ and $\tilde{\gamma}_T = O_p(2K_T + 1/\sqrt{T})$ yields:

$$\pi_t^{ET}(\tilde{\theta}_T) = \frac{1}{T} + \frac{1}{T} g_{tT}(\tilde{\theta}_T)' \tilde{\gamma}_T + O_p(T^{-2})$$

and

$$\pi_t^{ET}(\tilde{\theta}_T) = \frac{1}{T} (1 + o_p(1)).$$

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

$$T\pi_t^{ET}(\tilde{\theta}_T) - 1 = g_{tT}(\tilde{\theta}_T)' \tilde{\gamma}_T + o_p(1). \quad (22)$$

uniformly in $t = 1, \dots, T$.

By similar arguments, we can show for the Empirical likelihood:

$$T\pi_t^{EL}(\tilde{\theta}_T) - 1 = g_{tT}(\tilde{\theta}_T)' \tilde{\gamma}_T + o_p(1). \quad (23)$$

uniformly in $t = 1, \dots, T$. And considering that $\tilde{\gamma}_T = -\Omega^{-1} \frac{2K_T + 1}{T} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T) + o_p(1)$ and results derive in Lemma 7.1, the asymptotic equivalence with $T\pi_t^{GMM}(\tilde{\theta}_T) - 1$ follows.

⁴See Back and Brown (1993)

⁵See Kitamura and Stutzer (1997), p. 871

7.3 Proofs of Theorems

Proof of Theorem 3.1

The FOC (11) can be rewritten as

$$\frac{1}{T} \sum_{t=1}^T \rho_1(\hat{\gamma}(\hat{\theta}, s)' g_{tT}(\hat{\theta}, s)) G_{tT}^\delta(\hat{\theta}, s)' \frac{1}{2K_T+1} \hat{\gamma}(\hat{\theta}, s) = 0.$$

By a mean-value expansion of the first order conditions for the partial-sample GEL where $\Xi_T = (\hat{\beta}'_1, \hat{\beta}'_2, \hat{\delta}', \frac{\hat{\gamma}'_1}{2K_T+1}, \frac{\hat{\gamma}'_2}{2K_T+1})'$ and $\Xi_0 = (\beta'_0, \beta'_0, \delta'_0, 0, 0)'$:

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

where

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

and

$$\bar{M}(s) = \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} 0 & \rho_1(\hat{\gamma}_t(s)' g_{tT}(\hat{\theta}(s), s)) G_{tT}(\hat{\theta}(s), s)' \\ \rho_1(\bar{\gamma}_t(s)' g_{tT}(\bar{\theta}(s), s)) G_{tT}(\hat{\theta}(s), s)' & (2K_T+1) \rho_2(\hat{\gamma}_t(s)' g_{tT}(\hat{\theta}(s), s)) g_{tT}(\bar{\theta}(s), s) g_{tT}(\hat{\theta}(s), s)' \end{bmatrix}$$

where $\bar{\theta}(s)$ is a random vector on the line segment joining $\hat{\theta}(s)$ and θ_0 and $\bar{\gamma}(s)$ is a random vector joining $\hat{\gamma}(s)$ to $(0', 0)'$.

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

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

First, under Assumption 7.8,

$$\max_{1 \leq t \leq T} |\rho_1(\bar{\gamma}_t(s)' g_{tT}(\bar{\theta}(s), s)) - \rho_1(0)| \xrightarrow{p} 0$$

and

$$\max_{1 \leq t \leq T} |\rho_2(\bar{\gamma}_t(s)' g_{tT}(\bar{\theta}(s), s)) - \rho_2(0)| \xrightarrow{p} 0.$$

By a proof similar to the one in Lemma 1, we can show that $\sum_{t=1}^{[Ts]} G_t(\theta_0, s) = \sum_{t=1}^{[Ts]} G_{tT}(\theta_0, s) + o_p(1)$ which directly implies by Assumption 7.9 that

$$\frac{1}{T} \sum_{t=1}^T \rho_1(\bar{\gamma}_t(s)' g_{tT}(\bar{\theta}(s), s)) G_{tT}(\hat{\theta}(s), s) \xrightarrow{p} -G(s).$$

By Assumptions 7.6 and 7.8:

$$\frac{2K_T+1}{T} \sum_{t=1}^{[Ts]} g_{tT}(\bar{\beta}_1, \bar{\delta}) g_{tT}(\hat{\beta}_1, \hat{\delta})' \xrightarrow{p} s\Omega$$

and

$$\frac{2K_T+1}{T} \sum_{t=[Ts]+1}^T g_{tT}(\hat{\beta}_2, \hat{\delta}) g_{tT}(\bar{\beta}_2, \bar{\delta})' \xrightarrow{p} (1-s)\Omega$$

which results imply

$$\frac{2K_T + 1}{T} \sum_{t=1}^T \rho_2(\tilde{\gamma}_t(s)' g_{tT}(\tilde{\theta}(s), s)) g_{tT}(\tilde{\theta}(s), s) g_{tT}(\hat{\theta}(s), s)' \xrightarrow{p} -\Omega(s).$$

Moreover under Assumption 7.10:

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

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

As $\bar{M}(s)$ is positive definite with probability approaching one, we obtain:

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

We have also by Lemma 7.1,

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

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

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

and

$$\begin{aligned} \sqrt{T}(\hat{\gamma}(s) - \gamma_0) &= -\left(\Omega^{-1}(s) - \Omega^{-1}(s) G(s) (G(s)' \Omega(s)^{-1} G(s))^{-1} G(s)' \Omega(s)^{-1} \right) \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\theta_0, s) + o_p(1) \\ &\Rightarrow -\left(\Omega^{-1}(s) - \Omega^{-1}(s) G(s) (G(s)' \Omega(s)^{-1} G(s))^{-1} G(s)' \Omega(s)^{-1} \right) J(s). \end{aligned}$$

Proof of Theorem 3.3

The results for the $Wald_T(s)$ and $LM_T(s)$ under the null can be directly derived from similar arguments than in the proof of Theorem 3 in Andrews (1993). The asymptotic distribution under the alternative is a direct implication of the Theorem 3.2. For the $LR_T(s)$ statistic, expanding the partial-sample GEL objective function evaluated at the unrestricted estimator about $\gamma = 0$ yields,

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

By the fact that $\frac{2K_T+1}{T} \sum_{t=1}^T g_{tT}(\hat{\theta}_T(s), s)g_{tT}(\hat{\theta}_T(s), s)$ is a consistent estimator of $\Omega(s)$ and by Theorem (3.1), $\sqrt{T}/(2K_T+1)\hat{\gamma}(s) = -\Omega(s)^{-1}\sqrt{T} \sum_{t=1}^T g_{tT}(\hat{\theta}_T(s), s) + o_p(1)$, we get

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

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

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

The $LR_T(s)$ is then asymptotically equivalent to the LR statistic defined in Andrews (1993) for the standard GMM.

Proof of Theorem 3.4

First, for the statistic $O_T(s)$, the asymptotic equivalence between $\sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))$ with $\sum_{t=1}^{[Ts]} g(x_t \hat{\beta}_1(s), \hat{\delta}(s))$ and $\sum_{t=[Ts]+1}^T g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))$ with $\sum_{t=[Ts]+1}^T g(x_t \hat{\beta}_1(s), \hat{\delta}(s))$ is direct implication of the Lemma 7.1 and Lemma 7.2 and by the asymptotic consistency of the estimator $\hat{\Omega}_{1T}(s)$ and $\hat{\Omega}_{2T}(s)$ for Ω , the result under the null and alternative follow directly from the proofs for Theorem 2.1 and subsection A.2 in Hall and Sen (1999).

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

$$\frac{2[Ts]}{2K_T+1} \sum_{t=1}^{[Ts]} \frac{[\rho(\lambda_1(s)' g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))) - \rho_0]}{[Ts]} = O1_T(s) + o_p(1)$$

and

$$\frac{2(T-[Ts])}{2K_T+1} \sum_{t=[Ts]+1}^T \frac{[\rho(\lambda_1(s)' g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))) - \rho_0]}{T-[Ts]} = O2_T(s) + o_p(1)$$

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

Finally, for the statistic $LM_T^Q(s)$, the asymptotic equivalence between $(\sqrt{T}/(2K_T+1))\hat{\gamma}_1(s)$ to $\Omega_1(s)^{-1}T^{1/2} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\beta}_1(s), \hat{\delta}(s))$ and $(\sqrt{T}/(2K_T+1))\hat{\gamma}_2(s)$ to $\Omega_2(s)^{-1}T^{1/2} \sum_{t=[Ts]+1}^T g_{tT}(\hat{\beta}_2(s), \hat{\delta}(s))$ implies directly the asymptotic distribution of this statistic under the null and the alternative.

Proof of Theorem 4.1

First, consider, the Exponential Tilting estimator. A development similar to Kitamura and Stutzer (1997), the Lagrange multiplier has the following expression:

$$\frac{\sqrt{T}}{2K_T+1} \tilde{\gamma}_T = -\tilde{\Omega}_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\theta_0) - \tilde{\Omega}_T^{-1} G \sqrt{T}(\tilde{\theta}_T - \theta_0) + o_p(1) \quad (24)$$

where $\tilde{\Omega}_T$ is a consistent estimator of Ω , $G = E \left[\frac{\partial g_{tT}(\theta_0)}{\partial \theta'} \right] = E \left[\frac{\partial g(\theta_0)}{\partial \theta'} \right]$ and by the result in Kitamura and Stutzer (1997, p. 872):

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) = - (G' \Omega^{-1} G)^{-1} G' \Omega^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\theta_0) + o_p(1). \quad (25)$$

Combining the two preceding results, we obtain

$$\begin{aligned} \frac{\sqrt{T}}{2K_T + 1} \tilde{\gamma}_T &= -\Omega^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\theta_0) + \Omega^{-1} G (G' \Omega^{-1} G)^{-1} G' \Omega^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\theta_0) + o_p(1) \\ &= -\Omega^{-1} [I - G (G' \Omega^{-1} G)^{-1} G' \Omega^{-1}] \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\theta_0) + o_p(1). \end{aligned}$$

By Lemma 1, we have

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) \Rightarrow B(s) \quad (26)$$

where $B(s)$ is a q -vector of standard Brownian motions. Note also that:

$$\frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) = \frac{1}{T} \sum_{t=1}^{[Ts]} g_t(\theta_0) + o_p(1).$$

Consider now the following partial sums and by eq. (23);

$$\sum_{t=1}^{[Ts]} \left[T \pi_t^{ET}(\tilde{\theta}_T) - 1 \right] = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T)' \sqrt{T} \tilde{\gamma}_T + o_p(1). \quad (27)$$

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

$$\Omega^{-1/2} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T) \right] \Rightarrow B(s) - s \Omega^{-1/2} G [G' \Omega^{-1} G]^{-1} G' \Omega^{-1/2} B(1) \quad (28)$$

To see this, by a mean value expansion:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) + \frac{1}{T} \sum_{t=1}^{[Ts]} \frac{\partial g_{tT}(\bar{\theta})}{\partial \theta'} (\tilde{\theta}_T - \theta_0) + o_p(1),$$

where $\bar{\theta}$ lies on the line segment joining $\tilde{\theta}$ and θ_0 and may differ from row to row. By applying eq. (25), this gives

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\theta_0) - s G [G' \Omega^{-1} G]^{-1} G' \Omega^{-1} \frac{1}{T} \sum_{t=1}^T g_{tT}(\theta_0) + o_p(1). \quad (29)$$

It follows by (26),

$$\Omega^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} g_{tT}(\tilde{\theta}_T) \Rightarrow B(s) - s \Omega^{-1/2} G [G' \Omega^{-1/2} G]^{-1} G' \Omega^{-1/2} B(1). \quad (30)$$

Now, let us examine the Lagrange multiplier estimator $\tilde{\gamma}_T$. This parameter is directly function of the smoothed moment conditions by the following relation⁶

$$\sqrt{T}\tilde{\gamma}_T = -(2K_T + 1)\Omega^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T) + o_p(1) \quad (31)$$

Then the partial sum function of the implied probabilities can be rewritten as:

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

The following quadratic form:

$$\left[\frac{\sqrt{2K_T + 1}}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T) \right] \left[\frac{\sqrt{2K_T + 1}}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T) \right]'$$

is a consistent estimator of Ω . We then obtain for the following statistic:

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

By the result above, the quadratic form of the partial sum of interest (28) yields by equations (30) and (??):

$$\frac{1}{2K + 1} \left[\sum_{t=1}^{[Ts]} [T \pi_t^{ET}(\tilde{\theta}_T) - 1] \right]^2 \Rightarrow \left[B(s) - s\Omega^{-1/2}G[G'\Omega^{-1}G]^{-1}G'\Omega^{-1/2}B(1) \right]' \times \quad (32)$$

$$\left[B(s) - s\Omega^{-1/2}G[G'\Omega^{-1}G]^{-1}G'\Omega^{-1/2}B(1) \right]. \quad (33)$$

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

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

We can now decompose:

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

where $CC' = I$ and

$$\Lambda = \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix}$$

and I_p is an identity matrix with dimension $p \times p$. Note also that $CB(s)$ has the same asymptotic distribution as $B(s)$. The RHS of expression (32) 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

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

⁶See Kitamura and Stutzer (1997, p. 871) and Smith (2004, p.12).

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.

By the same line of arguments, we can show that

$$IPST^{EL}(s), IPST^{GMM}(s) \Rightarrow B_{(q-p)}(s)'B_{(q-p)}(s) + BB_p(s)'BB_p(s)$$

In particular, the result for the GMM can be obtained by directly using the expression (24) of the Lagrange multiplier estimator $\hat{\gamma}_T$ in function of the moment conditions.

Proof of Theorem 4.3

By the result given in (23) applied to difference between the partial-sample implied probabilities evaluated at the unrestricted and at the restricted estimators yields:

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

Replacing the unrestricted and the restricted estimators of γ by the corresponding expression (31), we get:

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

which equals

$$T \left[\pi_t(\hat{\theta}_T(s)) - \pi_t(\tilde{\theta}_T) \right] = - \left[\frac{1}{T} \sum_{t=1}^T g_{tT}(\hat{\theta}_T(s), s) - \frac{1}{T} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T, s) \right]' \quad (34)$$

$$\times \Omega^{-1}(s) (2K_T + 1) g_{tT}(\hat{\theta}_T(s), s) + o_p(1) \quad (35)$$

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

Now, consider the following mean value expansion:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\tilde{\theta}_T, s) = \frac{1}{\sqrt{T}} \sum_{t=1}^T g_{tT}(\hat{\theta}_T(s), s) + \frac{1}{T} \sum_{t=1}^T \frac{\partial g_{tT}(\hat{\theta}_T(s), s)}{\partial \theta'} \sqrt{T} (\tilde{\theta}_T - \hat{\theta}_T(s)) + o_p(1). \quad (36)$$

A premultiplication by $\left[\frac{1}{T} \sum_{t=1}^T \partial g_{tT}(\hat{\theta}_T(s), s) / \partial \theta' \right]' \hat{\Omega}(s)^{-1}$ and by the fact that the F.O.C of the partial sample criteria are equal to zero yields

$$\sqrt{T} (\tilde{\theta}_T - \hat{\theta}_T(s)) = - [G(\theta, s)' \Omega(s)^{-1} G(\theta, s)]^{-1} \left(\frac{1}{T} \sum_{t=1}^T \frac{\partial g_{tT}(\hat{\theta}_T(s), s)}{\partial \theta'} \right)' \hat{\Omega}(s)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{g}_{tT}(\tilde{\theta}_T) + o_p(1). \quad (37)$$

By using (37), the RHS expression (34) and summing over $t = 1, \dots, T$,

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

as $(2K_T + 1) \frac{1}{T} \sum_{t=1}^T g_{tT}(\hat{\theta}_T(s), s) g_{tT}(\hat{\theta}_T(s), s)'$ is a consistent estimator of $\Omega(s)$. Andrews (1993, p. 851-852) shows that this expression is asymptotically equivalent to the $LM_T(s)$ for parameters instability. The result follows.

Proof of Theorem 4.4

For the partial-sample GMM and the partial-sample GEL, we get

$$\Omega^{-1/2} \left[\frac{1}{T} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\theta}_{1T}(s)) \right] \Rightarrow B(s) - s\Omega^{-1/2} G[G'\Omega^{-1}G]^{-1} G'\Omega^{-1/2} B(s) \quad (38)$$

and

$$\Omega^{-1/2} \left[\frac{1}{\sqrt{T}} \sum_{t=[Ts]+1}^T g_{tT}(\hat{\theta}_{2T}(s)) \right] \Rightarrow (B(1) - B(s)) - (1-s)\Omega^{-1/2} G[G'\Omega^{-1}G]^{-1} G'\Omega^{-1/2} (B(1) - B(s)) \quad (39)$$

Now consider the first term of the statistic. For the first part of the sample by a direct application of (23), this gives

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

uniformly in $t = 1, \dots, [Ts]$ or equivalently

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

By summing the expression above to $t = \dots, [Ts]$, and by (31) applied to the sample $t = 1, \dots, [Ts]$, this yields

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

As $(2K_T + 1) \frac{1}{[Ts]} \sum_{t=1}^{[Ts]} g_{tT}(\hat{\theta}_{1T}(s)) g_{tT}(\hat{\theta}_{1T}(s))'$ is a consistent estimator of Ω and by (38), this gives the following asymptotic distribution for this statistic based implied probabilities under the null:

$$\frac{1}{2K+1} \sum_{t=1}^{[Ts]} \left([Ts] \pi_t^{ET}(\hat{\theta}_{1T}(s)) - 1 \right)^2 \Rightarrow \frac{B_{q-p}(s)' B_{q-p}(s)}{s}.$$

The asymptotic distribution under the alternative can also be show to be given by:

$$\frac{1}{2K+1} \sum_{t=1}^{[Ts]} \left([Ts] \pi_t^{ET}(\hat{\theta}_{1T}(s)) - 1 \right)^2 \Rightarrow \frac{B_{q-p}(s)' B_{q-p}(s)}{s} + \frac{H(s)' \Omega^{-1/2} (I - P_G) \Omega^{-1/2} H(s)}{s}.$$

The proof is similar for the second term based on the second part of the sample of the statistic test, then the result follows.

References

- Anatolyev, Stanislav, 2005, GMM, GEL, serial correlation, and asymptotic bias, *Econometrica* 73(3), 983–1002.
- Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.
<http://links.jstor.org/sici?sici=0012-9682%28199105%2959%3A3%3C817%3AHAACM%3E2.0.CO%3B2-P>.
- 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821–856. <http://links.jstor.org/sici?sici=0012-9682%28199307%2961%3A4%3C821%3ATFPIAS%3E2.0.CO%3B2-I>.
- Andrews, D.W.K., and J.C. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* 60, 953–966.
<http://links.jstor.org/sici?sici=0012-9682%28199207%2960%3A4%3C953%3AAIHAAC%3E2.0.CO%3B2-Q>.
- Andrews, D.W.K., and W. Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383–1414.
<http://links.jstor.org/sici?sici=0012-9682%28199411%2962%3A6%3C1383%3AOTWANP%3E2.0.CO%3B2-7>.
- Back, K., and D.P. Brown, 1993, Implied probabilities in GMM estimators, *Econometrica* 61, 971–975.
<http://links.jstor.org/sici?sici=0012-9682%28199307%2961%3A4%3C971%3AIPIGE%3E2.0.CO%3B2-E>.
- Bertille, Antoine, Hélène Bonnal, and Eric Renault, 2007, On the efficient use of the informational content of estimating equations: implied probabilities and euclidean empirical likelihood, *Journal of Econometrics* 138, 461–487.
- Caner, Mehmet, 2004, Nearly singular design in GMM and generalized empirical likelihood estimators, University of Pittsburg.
- 2005, Exponential tilting and weak instruments: estimation and testing, University of Pittsburg.
- Ghysels, E., A. Guay, and A. Hall, 1997, Predictive tests for structural change with unknown breakpoint, *Journal of Econometrics* 82, 209–233.
- Gregory, A.W., Jean-Francois Lamarche, and Gregor W. Smith, 2002, Information-theoretic estimation of preference parameters: macroeconomic applications and simulation evidence, *Journal of Econometrics* 107, 213–233.
- Guay, A., 2003, Optimal predictive tests, *Econometric Reviews* 22(4), 379–410.
- Guggenberger, Patrick, and Richard J. Smith, 2005, Generalized empirical likelihood tests in time series models with potential identification failure.

- Hall, Alastair R., 2005, Generalized method of moments (Oxford University Press, New York).
- Hall, A.R., and A. Sen, 1999, Structural stability testing in models estimated by generalized methods of moments, *Journal of Business and Economic Statistics* 17, 335–348.
- Hall, P., and J.L. Horowitz, 1996, Bootstrap critical values for tests based on generalized-method-of-moments estimators, *Econometrica* 64, 891–916.
<http://links.jstor.org/sici?sici=0012-9682/28199607%2964%3A4%3C943%3AEWSM%3E2.O.CO%3B2-3>.
- Hansen, L.P., and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1296–1286.
<http://links.jstor.org/sici?sici=0012-9682/28198209%2950%3A5%3C1269%3AGIVEON%3E2.O.CO%3B2-G>.
- Hansen, L.P., J. Heaton, and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262–280.
- Imbens, G.W., R.H. Spady, and P. Johnson, 1998, Information theoretic approaches to inference in moment condition models, *Econometrica* 66, 333–357.
<http://links.jstor.org/sici?sici=0012-9682/28199803%2966%3A2%3C333%3AITATII%3E2.O.CO%3B2-M>.
- Kitamura, Y., and M. Stutzer, 1997, An information-theoretic alternative to generalized method of moments estimation, *Econometrica* 65, 861–874.
<http://links.jstor.org/sici?sici=0012-9682/28199707%2965%3A4%3C861%3AAIATGM%3E2.O.CO%3B2-V>.
- Kullback, S., and R.A. Leibler, 1951, On information and sufficiency, *Annals of Mathematical Statistics* 22, 79–86. <http://links.jstor.org/sici?sici=0003-4851/28195103%2922%3A1%3C79%3A0IAS%3E2.O.CO%3B2-3>.
- Newey, W.K., and K.D. West, 1987, A simple positive semidefinite, heteroskedastic and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
<http://links.jstor.org/sici?sici=0012-9682/28198705%2955%3A3%3C703%3AASPSHA%3E2.O.CO%3B2-F>.
- 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631–654. <http://links.jstor.org/sici?sici=0034-6527/28199410%2961%3A4%3C631%3AALSICM%3E2.O.CO%3B2-Q>.
- Newey, W.K., and R.J. Smith, 2004, Higher order proprieties of GMM and generalized empirical likelihood estimators, *Econometrica* 72, 219–256.
- Otsu, Taisuke, 2006, Generalized empirical likelihood inference for nonlinear and time series models under weak identification, *Econometric Theory* 22, 513–527.
- Qin, J., and J. Lawless, 1994, Empirical likelihood and general estimating equations, *Annals of Statistics* 22, 300–325. <http://links.jstor.org/sici?sici=0090-5364/28199403%2922%3A1%3C300%3AELAGEE%3E2.O.CO%3B2-F>.
- Ramalho, Joaquim J.S., and Richard J. Smith, 2005, Goodness of fit tests for moment condition models.

Schennach, S.M., 2004, Point estimation with exponentially tilted empirical likelihood, University of Chicago.

Shannon, C.E., 1948, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Transactions on information theory* 26, 26–37.

Smith, Richard J., 2004, GEL criteria for moment condition models.

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

Sowell, F., 1996, Optimal tests for parameter instability in the generalized method of moments framework, *Econometrica* 64, 1085–1107.

<http://links.jstor.org/sici?sici=0012-9682/28199609%2964%3A5%3C1085%3AOTFPII%3E2.O.CO%3B2-X>.