ELSEVIER Journal of Econometrics 82 (1997) 209–233

JOURNAL OF
Econometrics

Predictive tests for structural change with unknown breakpoint

Eric Ghysels[a,b,*], Alain Guay[c], Alastair Hall[d,e]

[a]*Pennsylvania State University, Department of Economics, University Park, PA 16802-3305, USA*
[b]*Center for Interuniversity Research and Analysis on Organizations (CIRANO), Montreal H3A 2A5, Canada*
[c]*University de Québec à Montréal*
[d]*University of Birmingham*
[e]*North Carolina State University*

Received 1 July 1994; received in revised form 1 July 1996

Abstract

This paper considers predictive tests for structural change in models estimated via Generalized Method of Moments. Our analysis extends earlier work by Ghysels and Hall (1990a) by allowing for the instability to occur at an unknown point in the sample. We analyse various statistics based on continuous mappings of the sequence of predictive test calculated for a set of possible breakpoints in the sample. The limiting distribution of these statistics is derived under both the null hypothesis and local alternatives. Percentiles are reported for the distribution under the null. A sideproduct of our analysis is that we can illuminate the power properties of the predictive test and also compare its properties to those of the Wald, LR and LM tests for parameter variation. We study those power properties both via local asymptotic analysis and Monte Carlo. © 1997 Elsevier Science S.A.

* Corresponding author.

We would like to thank P. Burridge, J. Galbraith, D. Pappell and Amit Sen for comments on the paper. This paper has also benefitted from the comments of two anonymous referees and participants at the 1995 Allied Social Science Meetings in Washington, D.C., the 1995 Research Triangle Econometrics Conference in the Research Triangle Park, North Carolina, and seminars at North Carolina State University, Tilburg University, the European University Institute, the University of Amsterdam and the University of Birmingham. The first author would like to acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and the Fonds FCAR.

PII S0304-4076(97)00057-2

Key words: Generalized method of moments; Nonlinear models; Structural stability testing; Parameter variation; Overidentifying restrictions
JEL classification: C1, C12, C22

1. Introduction

There is a perennial interest in testing whether parametric econometric models are invariant through time. The vast literature on testing for structural change has paid most attention to the linear regression model while only a handful of tests are available for nonlinear dynamic single and multiple equation models. Advances in econometric theory over the last decade have cleared the horizon to address the more challenging task of testing for structural change in dynamic nonlinear models. Andrews and Fair (1988) considered the problem of testing parameter constancy when the sample can be split at some known breakpoint into two subsamples governed by parameter values equal under the null but different under the alternative. They proposed Wald, likelihood ratio-type (LR) and Lagrange multiplier-type (LM) tests and showed that under some weak regularity conditions such tests have standard asymptotic distributions. These developments were made in the context of the Generalized Method of Moments (GMM) estimator which in its generic form covers a large class of estimators for a wide variety of nonlinear dynamic models.[1] Ghysels and Hall (1990a) proposed a predictive test for structural stability. In this approach, parameter estimates from a first subsample are used to evaluate moment conditions in the second subsample. The essential idea behind such tests is that the predicted moment conditions should be statistically insignificantly different from zero when there is no structural change. The null and alternative hypotheses of these tests are formulated in terms of the structural stability of the moment conditions, rather than the parameter variation employed by Andrews and Fair (1988), and so have different power properties to Wald, LR and LM tests. Intuition suggests that neither type of the test dominates in null situations and so it is of interest to apply both in applications.

One drawback with all these tests is that they assume the breakpoint is known. While in some cases this may be reasonable, such as exploring the impact of specific economic events like the 1973 oil shock, in many cases one

[1] See Hansen (1982), Gallant and White (1988) and the recent surveys by Hall (1993), Newey (1993) and Ogaki (1993) for detailed discussion of GMM estimation. It should, parenthetically, be noted that the LR-type test is only appropriate under more restrictive conditions which are not satisfied in many GMM applications.

may wish to test for structural stability over all points in the sample. Andrews (1993) proposed a procedure for testing parameter stability when the breakpoint is unknown. His strategy is to consider the Wald tests, say, for a set of possible breakpoints and base inference on the supremum of these tests. Andrews shows that this 'Sup-Wald' statistic converges to the supremum square of a standardized tied down Bessel process under the null hypothesis of parameter constancy.[2] This extension takes the statistical theory outside the conventional framework in which the Wald test is asymptotically optimal because a nuisance parameter, the breakpoint, is not present under the null hypothesis. In this setting, the design of optimal tests requires assumptions about the nature of the alternative of interest. Therefore, rather than one optimal test of structural stability, there are, in fact, a whole family of such tests, each of whose structure depends on the nature of its alternative. Andrews and Ploberger (1995) delineate the alternative against which the Sup test is optimal. In the context of maximum likelihood, Andrews and Ploberger (1994) consider two different alternatives and show how these imply that the optimal tests are the averages of the Wald statistics over the set of possible breakpoints or the average of an exponential transformation of these Wald statistics. Sowell (1996a) considers the construction of optimal tests for parameter variation more generally and thereby extends Andrews and Ploberger's (1994) analysis to GMM estimators.

In Section 2, we adopt a similar approach to developing predictive tests for structural change at an unknown breakpoint. Using results from Andrews (1993) and Sowell (1996a) it is shown that under the null hypothesis the predictive test converges to the sum of the square of a standardized tied down Bessel process and the square of the standardized Bessel process. This structure reflects a decomposition of the statistic into a test for parameter variation and a test of the stability of the overidentifying restrictions. This enables us to clarify the relationship between the predictive and Wald tests. In the special case where the number of moment conditions equals the number of parameters, the various predictive tests proposed in this paper are asymptotically equivalent to the analogous Wald tests. In this case, the percentiles of the limiting distributions can be obtained from Andrews (1993) and Andrews and Ploberger (1994). We present percentiles of the limiting distributions for situations in which the number of moment conditions exceeds the number of parameters. In Section 3, we derive the distribution of the predictive test under local alternatives which helps to further illuminate its properties. These results are used to characterize the asymptotic behavior of a sup-predictive test and versions of the statistic based on the average-exponential form analyzed by Andrews and Ploberger

[2] A similar result applies for the Sup-LR and Sup-LM tests. In the remainder of this introduction all discussion of the Wald test, or functions thereof, similarly applies to the LR and LM tests.

(1994) and Sowell (1996a). Section 4 reports the results from a small simulation study. Section 5 concludes the paper with a discussion of some interesting extensions of our procedures. All proofs are relegated to a mathematical appendix.

2. Tests statistics and their asymptotic distributions

In this section, we propose predictive tests with unknown breakpoint and discuss their asymptotic distribution. The details of the proof and the required regularity conditions appear in the Appendix. We consider the class of GMM estimators which subsumes many standard estimators such as quasi-maximum likelihood, certain semi-parametric procedures, as well as least squares and IV procedures. In a general context, the GMM estimator is based on a set of moment conditions:

$$\mathrm{E}[f(x_t, \theta_0)] = 0, \tag{2.1}$$

where $f(\cdot)$ is a $(q \times 1)$ vector of continuous differentiable functions of (x_t, θ_0) with $f(\cdot) \in \mathbb{R}^q$; x_t is a $(s \times 1)$ vector of random variables; θ_0 is a $(p \times 1)$ parameter vector contained in $\Theta \subset \mathbb{R}^p$. This specification follows the usual practice of assuming that the moment conditions are valid throughout the whole sample. If this assumption is invalid for some observations, then the model is said to be structurally unstable. There are various ways in which one could characterize such structural instability. Andrews (1993) considers the situation in which the parameter vector at which the moment conditions are satisfied is indexed by t, θ_t say. This approach allows a wide variety of models against which it is difficult to design a single test. Consequently, he focuses attention on two homogeneous subsamples, i.e. $\theta_t = \theta_1$ for $t = 1, 2, \dots [\pi T]$ and $\theta_t = \theta_2$ for $t = [\pi T] + 1, \dots, T$ where $\pi \in \Pi \subset (0, 1)$ and $[\pi T]$ denotes the integer part of πT. The resulting tests are designed to have power against the explicit alternative of a single breakpoint although, as shown by Andrews (1993), the tests have power against a much wider class of alternatives. The tests we consider are similarly designed to detect situations in which there is a single breakpoint in the sample. However our characterization of structural instability is different. The predictive tests proposed by Ghysels and Hall (1990a) are formulated in terms of changes in the moment conditions without necessarily attributing such changes to the parameter vector. To present the null and alternative hypotheses of the predictive tests we need the following notation. Let $T_1(\pi) = \{1, 2, \dots, [\pi T]\}$ and $T_2(\pi) = \{[\pi T] + 1, \dots, T\}$. In this case where π is fixed a priori, the null and alternative hypotheses are:

$$\mathrm{H}_0\colon\ \mathrm{E}[f(x_t, \theta_0)] = 0 \quad \text{for } t \in T_1(\pi) \cup T_2(\pi), \tag{2.2a}$$

$$\mathrm{H}_1\colon\ \mathrm{E}[f(x_t, \theta_0)] = 0 \quad \text{for } t \in T_1(\pi), \quad \text{but } \mathrm{E}[f(x_t, \theta_0)] \neq 0 \quad \text{for } t \in T_2(\pi). \tag{2.2b}$$

The idea behind the predictive test is based on evaluating the moment conditions for the observations in the second subsample, $T_2(\pi)$, at the parameter estimators based on only the first subsample, $T_1(\pi)$. If the null hypothesis is correct then these estimated moment conditions should be approximately zero. When the breakpoint π is known and so H_0 and H_1 are given by (2.2), one can use the test proposed by Ghysels and Hall (1990a). In the remainder of this section, we consider the generalization of this test to the case where the breakpoint is unknown. In this case, the null hypothesis is that (2.2a) holds for all $\pi \in \Pi$, and the alternative is that (2.2b) holds for some $\pi \in \Pi$.

To proceed with the presentation of the tests, let us first present the required GMM estimators:

Definition 2.1. The set of GMM estimators $\{\theta_T(\pi)\}$ is a sequence of random vectors such that

$$\hat{\theta}_T(\pi) = \operatorname{argmin} \, ([\pi T])^{-1} \sum_{t=1}^{[\pi T]} f(x_t, \theta)' \, \hat{W}_T([\pi T])^{-1} \sum_{t=1}^{[\pi T]} f(x_t, \theta),$$

where $\hat{W}_T$ is a random symmetric matrix which may depend on π. Following Hansen (1982), the optimal weighting matrix W_T is defined to be the inverse of

$$\Omega_0 = \lim_{T \to \infty} \operatorname{Var}\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(x_t, \theta_0)\right]. \tag{2.3}$$

This matrix can be consistently estimated by a variety of procedures, see inter alia Gallant (1987), Newey and West (1987), Andrews nd Monahan (1992). Whenever the covariance estimator only involves data from the first subsample, we denote it by $\hat{S}_1(\pi)$. Likewise, when data starting with observation $[\pi T] + 1$ are used, we shall denote the estimator $\hat{S}_2(\pi)$. Eqs. (A.3) and (A.4) in the Appendix provide generic formula for both estimators. We now proceed with the definition of the predictive test statistics as a function of the unknown breakpoint π:

$$\begin{aligned} \mathrm{PR}_T(\pi) = & \left[(T - [\pi T])^{-1/2} \sum_{t=[\pi T]+1}^{T} f(x_t, \hat{\theta}_T(\pi))\right]' \\ & \times \hat{V}_2^{-1}(\pi) \left[(T - [\pi T])^{-1/2} \sum_{t=[\pi T]+1}^{T} f(x_t, \hat{\theta}_T(\pi))\right], \end{aligned} \tag{2.4}$$

where $\hat{V}_2(\pi)$ is a consistent estimator of

$$V_2(\pi) = S_2(\pi) + dF_2(\pi) \, [F_1(\pi)' S_1^{-1}(\pi) F_1(\pi)]^{-1} F_2(\pi)',$$

and $d = (1 = \pi)/\pi$ while the matrices $F_i(\pi)$ appear in the Appendix as Eq. (A.2). A first theorem establishes weak convergence of the $\mathrm{PR}_T(\pi)$ process indexed by π.

Theorem 2.1. Under the null hypothesis H_0 *in* (2.2a) *and Assumption* A.1–A.13 *of the Appendix, the process* PR *indexed by* π *for a given set* Π *whose closure lies in* (0, 1) *satisfies*:

$$\mathrm{PR}_T(\pi) \Rightarrow \mathrm{BBH}(\pi) + \mathrm{BMH}(\pi) \tag{2.5}$$

where

$$\mathrm{BBH}(\pi) = \frac{[B_p(\pi) - \pi B_p(1)]'[B_p(\pi) - \pi B_p(1)]}{\pi(1-\pi)},$$

$$\mathrm{BMH}(\pi) = \frac{[B_{q-p}(1) - B_{q-p}(\pi)]'[B_{q-p}(1) - B_{q-p}(\pi)]}{(1-\pi)}$$

and $B(\pi)$ *is a vector of q standard Brownian motions whose first p element are* $B_p(\pi)$ *and whose remaining* $q-p$ *elements are* $B_{q-p}(\pi)$.

Hence, the asymptotic distribution is a squared p-dimensional standardized tied-down Bessel process plus the square of a $q-p$ dimensional standardized Bessel process. Following Sowell (1996a), it is possible to establish the role played by the two components of this asymptotic distribution. Asymptotically, the Predictive test is the sum of the Wald test for parameter variation, whose distribution is BBH(π), and a test of the overidentifying restrictions in $T_2(\pi)$, whose distribution in BMH(π). Therefore, the Predictive test can be viewed as a more general test of the model specification than just a test for parameter variation. In Section 3, this structure will play an important role in contrasting the power properties of the Wald and predictive tests. This structure also implies that if $q = p$, then the Wald and Predictive tests are asymptotically equivalent.

The result in Theorem 2.1 enables us to proceed with the formulation of three test statistics. The first statistic is the sup-predictive test

$$\mathrm{Sup\,PR}_T = \sup_{\pi \in \Pi} \mathrm{PR}_T(\pi). \tag{2.6}$$

This approach amounts to basing inference on the breakpoint which maximizes the evidence against structural stability. The other two statistics are motivated by recent work on optimal testing by Andrews and Ploberger (1994) and Sowell (1996a). These statistics are:

$$\mathrm{AvPR}_T = \int_\Pi \mathrm{PR}_T(\pi)\,\mathrm{d}J(\pi) \tag{2.7}$$

$$\mathrm{ExpPR}_T = \log\left\{\int_\Pi \exp[0.5\mathrm{PR}_T(\pi)]\,\mathrm{d}J(\pi)\right\} \tag{2.8}$$

where $J(\pi)$ is the probability density function specified for π. In these statistics, $J(\pi)$ acts as a weighting mechanism which indicates the relative importance of instability at points in the sample. Below, we follow Andrews and Ploberger (1994) and Sowell (1996a) in setting $J(\pi)$ to be the uniform distribution over Π. Notice that with this choice, AvPR_T is just the average of the Predictive tests over Π. Following Andrews and Ploberger (1994), the AvPR_T statistic is anticipated to be powerful for alternatives close to the null; whereas ExpPR_T is anticipated to be powerful against distant alternatives. The distributions of these statistics are presented in Theorem 2.2.

Theorem 2.2. Under the conditions of Theorem 2.1, *we have*

$$\mathrm{SupPR}_T \Rightarrow \sup_{\pi \in \Pi} \{\mathrm{BBH}(\pi) + \mathrm{BMH}(\pi)\},$$

$$\mathrm{AvPR}_T \Rightarrow \int_{\Pi} [\mathrm{BBH}(\pi) + \mathrm{BMH}(\pi)]\, \mathrm{d}J(\pi),$$

$$\mathrm{ExpPR}_T \Rightarrow \log \left\{ \int_{\Pi} \exp[(\mathrm{BBH}(\pi) + \mathrm{BMH}(\pi))/2]\, \mathrm{d}J(\pi) \right\}.$$

In the case where $p = q$ these distributions are the same as those derived by Andrews (1993) and Andrews and Ploberger (1994) for the analogous Wald, LM and LR based tests. Therefore percentiles for our tests when $p = q$ can be obtained from the appropriate tables in these earlier papers. We tabulate the distributions in Theorem 2.2 for the case where $q > p$; and $J(\pi)$ is equal to the uniform distribution on Π. The tables with these critical values are relegated to the Appendix. All calculations were performed using GAUSS 3.0 with 10,000 replications.

3. Asymptotical local power

The predictive tests discussed in the previous section are designed to test for a single breakpoint at which the value of the moment condition changes. However, they have power against a variety of other alternatives as well. In this section, we develop the formal asymptotic arguments to support this claim. The analysis parallels the development in the previous section. We start by characterizing the distribution of the Predictive test under local alternatives when π is fixed a priori. This analysis allows us to further illuminate the differences between the Wald and Predictive tests. This framework can also be used to deduce consistency properties for the Sup-, Av- and Exp- versions of the Predictive test.

Following Andrews (1993) we adopt a very general specification for the sequence of local alternatives to our H_0. In this section we assume

Assumption 3.1. The moment conditions satisfy

$$\sup_{\pi \in \Pi} \left\| T^{-1/2} \sum_{t=1}^{[\pi T]} f(x_t, \theta_0) - \mu_1(\pi) \right\| = o_p(1),$$

$$\sup_{\pi \in \Pi} \left\| T^{-1/2} \sum_{t=[\pi T]+1}^{T} f(x_t, \theta_0) - \mu_2(\pi) \right\| = o_p(1)$$

for some nonrandom bounded $\mathbb{R}^{2q}$-valued function $\mu' = (\mu_1', \mu_2')$ on Π. Notice that this sequence of alternatives allows for violation of the moment conditions in both subsamples. If we put more structure on the problem and assume that $[f(x_t, \theta_0)] = \eta(t/T)/T^{1/2}$ where $\eta(\cdot)$ is some bounded $\mathbb{R}^q$-valued function which is Riemann integrable on $[0, \Pi]$ uniformly over $\pi \in \Pi \cup \{1\}$, then

$$\mu_1(\pi) = \int_0^\pi \eta(s)\,\mathrm{d}s, \qquad \mu_2(\pi) = \int_\pi^1 \eta(s)\,\mathrm{d}s,$$

as discussed in Andrews (1993). It is of interest to specialize these results further to the case of a single breakpoint at unknown time π_0. Suppose the value of the moment conditions in $\eta \neq 0$ for $t \geqslant \pi_0 T$. This can be captured within our framework by putting $\eta(s) = \eta 1(s \geqslant \pi_0)$, where $1(A)$ is an indicator function which equals 1 if A occurs and zero otherwise. From this definition, it follows that $\mu_1(\pi) = \max(\pi - \pi_0, 0)\eta$ and $\mu_2(\pi) = [1 - \max(\pi, \pi_0)]\eta$.

We now present the limiting distribution of $\mathrm{PR}_T(\pi)$ under the class of alternatives in Assumption 3.1.

Theorem 3.1. Under Assumption 3.1 *and* A.1–A.13 *given in the Appendix, we have*

$$\mathrm{PR}_T(\pi) \Rightarrow J_p^*(\pi)' J_p^*(\pi) + K_{q-p}^*(\pi)' K_{q-p}^*(\pi)$$

where

$$J_p^*(\pi) = \frac{[\mathrm{BH}_1(\pi) - \pi \mathrm{BH}_1(1)]}{[\pi(1-\pi)]^{-1/2}} - \left[\frac{(1-\pi)}{\pi}\right]^{1/2} H_1 S^{-1/2} \mu_1(\pi)$$

$$+ \left[\frac{\pi}{(1-\pi)}\right]^{1/2} H_1 S^{-1/2} \mu_2(\pi)$$

$$K_{q-p}^* = \frac{\mathrm{BH}_2(1) - \mathrm{BH}_2(\pi) + H_2 S^{-1/2} \mu_2(\pi)}{(1-\pi)^{1/2}}$$

and $H' = [H'_1 \ \ H'_2]$ *is a matrix whose columns form a set of orthonormal vectors with* H_1, H_2 *of dimensions* $p \times q$, $(q-p) \times q$ *respectively which is defined in the Appendix.*

The two components of this limiting distribution can be interpreted using the asymptotic decomposition of the Predictive test described in the previous section. The component involving $J_p^*(\pi)$ is the distribution of the Wald test for parameter variation under this local alternative; see Andrews (1993, Theorem 4). The remaining component involving $K_{q-p}^*(\pi)$ is the distribution of the overidentifying restrictions test in $T_2(\pi)$.

Before examining the properties of the tests proposed in this paper, it is interesting to use Theorem 3.1 to learn about the power of the predictive test when the breakpoint is known. If π is fixed then $\mathrm{PR}_T(\pi)$ has a χ_q^2 with noncentrality parameter equal to

$$\mathrm{constant}[J_p^*(\pi)]^2 + \mathrm{constant}[K_{q-p}^*(\pi)]^2 \tag{3.1}$$

where constant $[\cdot]$ denotes the nonrandom part of the random vector in the brackets. We use this result to examine the power of the test in two situations. First consider the case where the instability is driven by parameter variation. Following Andrews (1993), we assume $\mathrm{E}[f(x, \theta + \eta(t/T)/T^{1/2})] = 0$ and so

$$\mu_1(\pi) = -F\int_0^\pi \eta(s)\,\mathrm{d}s, \qquad \mu_2(\pi) = -F\int_\pi^1 \eta(s)\,\mathrm{d}s.$$

Substituting these representations into (3.1) and noting that $H_2 S^{-1/2} F = 0$,[3] it can be shown that the predictive test has exactly the same noncentrality parameter as the Wald, LR and LM tests proposed by Andrews and Fair (1988). However the predictive test has $q-p$ more degrees of freedom and so is less powerful if $q > p$. This is intuitively reasonable. The Wald, LR and LM tests are designed to have power against parameter variation and under this alternative there is no additional information in the overidentifying restrictions. We now turn to the situation where there is a single breakpoint π_0 and examine the power of the test against structural instability either before or after the break. For this discussion assume the correct breakpoint is chosen. If the moment condition (2.1) is only invalid before the break, i.e. prior to $\pi_0 T$, then the noncentrality parameter of the test is

$$\mathrm{ncp}_1 = \left[\frac{1-\pi_0}{\pi_0}\right] \eta' S^{-1/2} H'_1 H_1 S^{-1/2} \eta.$$

[3] This follows from Eq. (A.7) in the appendix because the column of H form an orthonormal set.

Whereas if the moment condition (2.1) is only invalid after the break the noncentrality parameter is

$$\mathrm{ncp}_2 = \left[\frac{\pi_0}{1-\pi_0}\right] \eta' S^{-1/2} H_1' H_1 S^{-1/2} \eta + \eta S^{-1/2} H_2' H_2 S^{-1/2} \eta.$$

The relative magnitudes of these two noncentrality parameters depends on π_0 and the moments of various functions of the data. However, note that if $\pi_0 = 0.5$, then $\mathrm{ncp}_2 \geqslant \mathrm{ncp}_1$. In order words, the predictive test has more power against structural instability after the break under these conditions.[4] We also observe that ncp_1 equals the noncentrality parameter of the Wald, LR and LM tests. Therefore if $q > p$ the predictive test is less powerful than the other tests when the moment conditions are only invalid prior to the breakpoint. This follows because instability in the first subsample only affects the test via the parameter estimator $\hat{\theta}_T(\pi_0)$. However if the moment conditions are invalid in the second subsample alone then the predictive test can be more powerful. This is illustrated in Table 1. For simplicity, we consider the case where $H = I_q$, $S^{-1/2}\eta = \varepsilon\, 1_q$ where 1_q is a $q \times 1$ vector of ones. In this case, the predictive test converges to a $\chi_q^2(q\varepsilon^2)$ distribution and the Wald, LR and LM test converge to a $\chi_p^2(p\varepsilon^2)$ distribution, where $\chi_a^2(b)$ is a χ^2 distribution with a degrees of freedom and noncentrality parameter b. From Table 1, it is clear that the predictive test can be much more powerful asymptotically. For example, if $q = 10$, $p = 1$ and $\varepsilon = 1.5$, then the predictive test has power equal to 0.93, while the other three tests only have power equal to 0.32. The Monte Carlo results reported in the

Table 1
Probability a $\chi_k^2(k\varepsilon^2)$ random variables exceeds the 95th percentile of the χ_k^2 distribution

ε	k									
	1	2	3	4	5	6	7	8	9	10
0.5	0.08	0.09	0.10	0.11	0.11	0.12	0.13	0.13	0.14	0.14
1.0	0.17	0.22	0.27	0.32	0.36	0.40	0.44	0.48	0.51	0.53
1.5	0.32	0.46	0.57	0.66	0.74	0.80	0.84	0.88	0.91	0.92
2.0	0.51	0.71	0.84	0.91	0.95	0.97	0.99	0.99	1.00	1.00
2.5	0.71	0.90	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00

[4] Ghysels and Hall (1990a) show that the predictive test has the same power against structural instability either before or after the break. This can be reconciled with the results in this paper because Ghysels and Hall (1990a) concentrate on instability caused by parameter variation alone in which case the second term in ncp_2 is 0.

next section will reinforce this finding. Taken together, these properties of the predictive test suggest that it is desirable to perform the tests in two ways: using the parameter estimators of the first subsample to evaluate the moment conditions in the second subsample and using the parameter estimators of the second subsample to evaluate the moment conditions of the first subsample. The construction of the latter test is analogous to that of the former and its distribution is easily deduced from Theorem 3.1; in particular note this second predictive test has the same distribution under the null hypothesis.

One can derive the limiting distributions of the SupPR_T, AvPR_T and ExpPR_T by applying the approximate continuous mapping to the distribution in Theorem 3.1; for brevity exact formula for these limiting distributions are omitted. However, we do explicitly consider the power properties of these predictive based tests. For this discussion we restrict attention to the class of alternatives

$$\mathrm{E}[f(x_t, \theta_0)] = \xi\, \eta(t/T)/T^{1/2}. \tag{3.2}$$

Corollary 3.1. Under the conditions of Theorem 3.1, *Eq.* (3.1) *holds with* $\eta(\cdot)$ *not equal to a constant vector almost everywhere on* Π *then*

$$\lim_{\xi\to\infty}\lim_{T\to\infty} P[\mathrm{SupPR}_T > c_{\mathrm{sup}}(\alpha)] = 1,$$

$$\lim_{\xi\to\infty}\lim_{T\to\infty} P[\mathrm{AvPR}_T > c_{\mathrm{av}}(\alpha)] = 1,$$

$$\lim_{\xi\to\infty}\lim_{T\to\infty} P[\mathrm{ExpPR}_T > c_{\mathrm{exp}}(\alpha)] = 1,$$

where $c_{\mathrm{sup}}(\alpha)$, $c_{\mathrm{av}}(\alpha)$ *and* $c_{\mathrm{exp}}(\alpha)$ *are the* $100(1-\alpha)$ *percentiles of the limiting distributions of the* SupPR_T, AvPR_T *and* ExpPR_T *tests given in Theorem* 2.2.

Therefore all three tests have nontrivial power against alternatives for which the expectation of the moment condition is not constant over the sample. This result follows from Theorem 3.1 and Corollary 2 of Andrews (1993).

4. Finite sample properties – a simulation study

We now turn our attention to the finite sample properties of the exponential and supremum predictive tests. The design of the simulation study will emphasize the difference between testing for structural change through moment conditions versus through parameters as in Andrews (1993). We present the simulation design first and discuss the results thereafter.

Consider a data series x_t with a sample of size T available. The data are generated by the following equation:

$$x_t = \theta x_{t-1} + \varepsilon_t + \alpha_{2t}\varepsilon_{t-2}, \tag{4.1}$$

where ε_t is i.i.d N(0, 1). We will be interested in comparing two different scenarios: (A) the data generating process is AR(1), i.e., $\alpha_{2t} = 0\ \forall t$ and (B) the data generating process is AR(1) for half the sample and, for the remainder of the sample, it is fixed parameter ARMA(1, 2), with a zero restriction on the first lag of the MA polynomial and

$$\alpha_{2t} = \begin{cases} 0 & \text{if } t \leqslant T/2, \\ \bar{\alpha}_2 \neq 0 & \text{if } t > T/2. \end{cases} \tag{4.2}$$

On first appearance, we are in a typical situation of structural changes, in this case involving the MA parameter α_{2t}. However, the next element will emphasize the differences which may occur between testing for structural change through moment conditions and parameter estimation. Namely, the econometrician estimates only the parameter θ has being the "parameter of interest". Such a situation is indeed not uncommon. For instance, may applications of GMM involving Euler equations entail estimation only of a small set of parameters which usually have an economic interpretation but do not fully describe the DGP. Our setup of an ARMA(1, 2) process with only the estimation of the AR parameter is a simplified example of this commonly encountered situation. The estimator for the parameter θ is based on the following moment function:

$$f(x_t, \theta) = (x_t - \theta x_{t-1})(x_{t-1}, x_{t-2})'. \tag{4.3}$$

Under the null hypothesis, which is assumed to be scenario (A), the lagged dependent variable x_{t-1} and x_{t-2} are valid instruments. It should also be noted that one has a situation of one overidentifying moment condition in (4.3). From the discussion in Section 2 and 3, we known that the Wald, LM and LR tests on the one hand and predictive tests on the other will have different power properties. Under scenario (B), which is chosen here as a specific class of alternatives, neither x_{t-1} nor x_{t-2} are valid instruments for half of the sample. The LR, LM and Wald tests for structural change discussed in Andrews (1993) and Andrews and Ploberger (1994) will be based on statistics involving parameter estimates of θ over the entire sample or subsamples. In our design, θ will be estimated consistently during part of the sample only. One should observe though that the parameter θ actually never changes. Instead, the validity of the orthogonality conditions are affected through the design of the DGP. In particular, using the notation of the previous section:

$$\mathrm{E}[f(x_t, \theta)] = 0 \quad \text{for } t = 1, \ldots, T/2,$$

$$\mathrm{E}[f(x_t, \theta)] \neq 0 \quad \text{for } t = T/2 + 1, \ldots, T.$$

This design stresses in a simple way the differences between structural change tests proposed here and those considered by Andrews (1993) and Andrews and Ploberger (1994). Obviously, the latter tests will have power because of the inconsistent estimation of θ during part of the sample which will be viewed as a structural change.

In Table 2, we report results from a Monte Carlo study involving a total of eight test statistics for twelve parameter settings in Eqs. (4.1) and (4.2) and two

Table 2
Size and power properties of supremum and exponential tests for structural change with unknown breakpoint (5% critical value)

$x_t = \theta x_{t-1} + \varepsilon_1 \quad t \leqslant T/2$

$x_t = \theta x_{t-1} + \varepsilon_t + \bar{\alpha}_2 \varepsilon_{t-2} \quad t > T/2$

θ	$\bar{\alpha}_2$	T	Wald		LR		LM		PR	
			Sup	Exp	Sup	Exp	Sup	Exp	Sup	Exp
Size properties										
0	0	100	3.5	4.6	2.8	4.7	2.2	3.9	1.5	3.6
		200	3.8	5.2	3.5	4.7	3.3	4.3	2.7	4.5
0.5	0	100	3.5	5.0	3.3	4.6	2.5	4.3	2.5	3.5
		200	4.1	5.0	3.9	5.0	3.3	4.6	2.6	4.4
0.9	0	100	4.8	4.6	4.2	4.1	3.4	3.8	3.2	3.6
		200	4.2	5.4	3.6	4.8	3.4	4.3	3.0	4.5
Power properties										
0	0.5	100	11.7	13.6	11.0	13.3	10.6	11.9	2.3	21.0
		200	13.7	14.9	12.7	14.0	14.1	15.0	61.8	83.2
	−0.5	100	0.9	0.8	1.0	0.9	10.6	0.6	13.8	40.6
		200	1.4	1.7	1.4	1.5	1.3	1.5	81.0	92.4
	−0.9	100	0.5	1.0	0.7	1.1	0.5	0.7	44.7	79.7
		200	0.4	1.1	0.7	1.1	0.6	1.0	98.7	100.0
0.5	0.5	100	8.2	11.6	8.4	11.6	7.8	9.9	1.1	7.2
		200	11.8	16.2	11.8	16.2	12.9	15.7	18.7	39.7
	−0.5	100	6.5	8.7	7.2	8.9	6.7	7.7	18.6	38.6
		200	11.0	14.9	11.6	15.9	10.4	13.8	76.5	88.6
	−0.9	100	7.6	12.7	8.7	14.1	6.4	10.9	47.6	77.4
		200	17.9	25.9	20.1	27.9	15.6	21.8	98.7	99.7
0.9	0.5	100	1.9	3.1	2.4	3.1	1.5	1.3	0.5	1.0
		200	3.5	4.7	3.7	5.0	2.7	2.4	1.6	4.1
	−0.5	100	27.6	30.9	28.3	31.2	27.9	27.6	31.2	35.0
		200	55.9	58.7	56.7	60.0	55.0	54.5	68.1	69.8
	−0.9	100	65.9	72.8	68.0	74.4	63.8	69.2	78.4	86.7
		200	94.9	96.4	95.6	97.0	94.4	95.6	99.9	100.0

Note: In all computation $\pi \in [0.15, 0.85]$.

samples sizes $T = 100$ and $T = 200$. For the autoregressive parameter, we took values $\theta = 0$, 0.5 and 0.9. Size properties of the statistics were simulated by setting $\bar{\alpha}_2 = 0$ in (4.2). The power properties were examined with nonzero values of $\bar{\alpha}_2$. They were set equal to 0.5 as well as -0.5 and -0.9. For the Wald, LR and LM type tests, we considered both the supremum version appearing in (2.6) and the exponential one appearing in (2.8) (replacing in both formula PR by the applicable test statistics). Likewise, for the predictive tests we also reported both versions. The figures reported in Table 2 are based on 1000 simulations with $\pi \in [0.15, 0.85]$.

The top panel of Table 2 reports the size properties in small samples since in all cases $\bar{\alpha}_2 = 0$. There are no important size distortions, sometimes some of the tests are undersized but this seems only to be a minor problem. Let us turn our attention to power properties. They clearly confirm the calculations reported in Table 1 where it was shown that the local asymptotic power of the predictive type tests can be remarkably better. For example, with $\theta = 0$ and $\bar{\alpha}_2 = -0.5$ or -0.9, we notice that the Wald, LR and LM tests have power in the range of 10% to 20% with $T = 200$; whereas for the predictive test, it is between 80% and 100%. With $\theta = 0$ and $\bar{\alpha} = 0.5$, the difference is not so dramatic, yet it is still up to 45%. When the AR coefficient increases the advantage in power of the predictive test reduces, although it remains the most powerful test for this particular setup regardless of the parameter settings and samples sizes.

5. Extensions and concluding remarks

In this paper, we have proposed three versions of the Predictive test which can be used to test for structural stability with an unknown breakpoint in models estimated via GMM. The limiting distributions of the statistics have been derived and tabulated. Our analysis has also enabled us to demonstrate that the Predictive test is asymptotically the sum of the Wald test for parameter variation and the test of the overidentifying restrictions in one of the subsamples. This clarifies the relationship between the Predictive and Wald tests. To conclude, we briefly discuss two interesting extensions of our work: (i) the delineation of conditions under which our statistics are optimal; (ii) the development of separate statistics for testing parameter constancy and the stability of the overidentifying restrictions.

It was mentioned above that the average and exponential versions of the Predictive test are motivated by the recent literature on optimal tests of parameter variation. In fact, these statistics can be shown to be optimal themselves under certain circumstances. Sowell (1996b) develops optimal statistics in the case where (a) under the alternative, the moment condition is invalid before but valid after the breakpoint and so $\eta(s) = \eta 1(\pi_0 \geqslant s)$ in our notation;

(b) the directions of interest under the alternative, η, are weighted by a $N(0, c\Sigma)$ distribution. Sowell's E and L statistics are easily shown to be asymptotically equivalent to versions of the Predictive test calculated using the estimates from $T_2(\pi)$ to evaluate the moment conditions in $T_1(\pi)$; if $c = 0$, the optimal test is AvPR_T and if $c = \infty$, the optimal test is ExpPR_T. However, Hall and Sen (1996a) show that these statistics are not optimal against the alternative in which the moment conditions are valid before, but invalid after, the breakpoint; i.e., $\eta(s) = \eta 1(s \geqslant \pi_0)$. In this case, it can be shown that under (b) above, the optimal statistics for the same values of c are the analogous functionals applied to the Predictive test calculated using the estimates from $T_1(\pi)$ to evaluate the moment conditions in $T_2(\pi)$.

Our asymptotic decomposition of the Predictive test implies this test is sensitive to more general types of instability than just parameter variation. In practice, it may be useful to diagnose whether one or both of these components is the source of a significant statistic. However, this is not possible within our framework because effectively only the sum is calculated. The obvious remedy is to calculate the two components separately and then combine them into a joint test of structural stability. This approach has been proposed independently by Hall and Sen (1996b) and Sowell (1996b). Sowell's (1996b) analysis is motivated by the desire to construct optimal statistics against the alternative described above in which the moment conditions are only invalid before the breakpoint. These optimal statistics, once again, have the average or average exponential form. To test for parameter variation, these functionals are applied to a statistic which is asymptotically equivalent to the Wald statistic. To test the stability of the overidentifying restrictions, these functionals are applied to a statistic which is asymptotically equivalent to the overidentifying restrictions test in $T_1(\pi)$. Hall and Sen (1996b) develop statistics designed to test against an alternative in which the moment condition may be invalid either before or after the breakpoint. Therefore, they propose testing the stability of the overidentifying restrictions by examining two overidentifying restrictions test statistics: one based on $T_1(\pi)$ and the other on $T_2(\pi)$. They similarly propose testing for parameter variation using the Wald statistic. When the breakpoint is known, Hall and Sen's (1996b) statistics involve the same functionals as those used in this paper. It is also possible to adapt the Predictive test framework to obtain statistics which are sensitive to parameter variation alone or the stability of the overidentifying restrictions alone. Guay (1996) considers this extension and develops suitable statistics for both the known and unknown breakpoint case. His approach is based on using transformations to extract the part of the predicted moment condition which goes into the test of parameter variation and the part which tests the stability of the overidentifying restrictions.

Appendix

In this appendix, we describe the set regularity conditions used to derive the asymptotic distribution of the tests. Next we present the proofs of Theorems 2.1 and 3.1. Tables 3–8 present the critical values of the asymptotic distribution.

A.1. Regularity conditions

Assumption A.1. The estimator is based on the argument (2.1) where f is a $\mathbb{R}^q$-valued function of orthogonality conditions.

Table 3
Critical values of SupPR_T for $\pi \in (0.15, 0.85)$

dim $q-p$	dim p							
	1	2	3	4	5	6	7	8
	11.02	13.67	15.97	18.10	19.79	21.71	23.51	25.33
1	12.68	15.41	17.87	20.07	21.93	23.79	25.70	27.69
	16.19	19.53	22.15	24.33	26.10	28.32	30.71	33.87
	13.48	15.63	17.80	19.78	21.65	23.54	25.20	27.06
2	15.31	17.58	19.76	21.97	23.90	25.76	27.43	29.42
	19.47	21.43	24.19	26.34	28.10	30.24	31.96	34.33
	15.35	17.47	19.51	21.53	23.42	24.98	26.93	28.58
3	17.28	19.50	21.61	23.53	25.65	27.22	29.33	30.91
	21.64	23.68	26.36	28.12	30.78	32.09	33.78	36.26
	17.28	19.26	21.26	23.29	25.14	26.66	28.15	29.78
4	19.37	21.30	23.48	25.54	27.34	29.28	30.45	32.38
	23.75	25.86	28.16	30.35	32.39	34.28	35.53	37.91
	19.09	20.97	23.06	24.67	26.32	28.12	29.85	31.33
5	21.26	23.08	25.54	27.00	28.99	30.43	32.27	33.77
	25.94	27.82	30.95	31.88	33.03	35.25	36.85	38.56
	20.78	22.86	24.42	26.27	28.03	29.95	31.34	32.86
6	22.93	25.16	26.68	28.58	30.45	32.46	33.81	35.56
	27.41	30.28	31.46	33.50	35.68	37.78	39.26	40.63
	22.56	24.50	26.26	28.03	29.71	31.29	32.75	34.50
7	24.72	26.80	28.60	30.44	32.15	33.79	35.36	37.23
	29.77	31.41	33.40	35.24	37.64	39.42	40.48	42.70
	24.00	26.00	27.92	29.52	30.81	32.54	34.53	35.81
8	26.32	28.20	30.33	32.09	33.32	35.21	37.16	38.71
	31.22	33.18	35.27	36.90	38.23	40.78	42.21	44.18

Note: First figure is 10%, followed by 5% and 1%.

Table 4
Critical values of SupPR_T for $\pi \in (0.20, 0.80)$

dim $q - p$	dim p							
	1	2	3	4	5	6	7	8
	10.09	11.89	14.13	15.90	18.26	19.97	21.59	23.72
1	11.92	13.59	16.03	18.07	20.42	22.08	24.00	26.15
	15.75	17.50	20.06	22.78	25.16	27.20	28.53	31.65
	11.99	14.81	16.71	18.67	20.18	21.95	23.67	25.25
2	13.94	16.79	18.89	20.94	22.59	24.53	26.13	28.33
	17.87	21.67	23.37	25.75	27.27	29.38	31.28	32.76
	14.94	16.80	18.64	20.54	22.16	23.69	25.58	27.10
3	17.10	19.20	21.00	23.10	24.74	26.21	28.17	29.86
	21.91	23.56	26.03	27.85	30.04	31.30	33.51	35.21
	17.26	18.80	20.60	22.60	24.11	25.88	27.19	28.67
4	19.58	21.32	23.07	25.02	26.75	28.67	29.73	31.41
	24.61	26.55	28.48	29.76	31.94	34.51	35.33	37.47
	19.11	20.80	22.64	24.18	25.84	27.63	29.10	30.48
5	21.58	23.55	25.12	26.84	28.33	30.35	31.79	33.36
	27.51	28.96	31.57	32.86	33.64	36.30	36.70	39.09
	21.14	22.93	24.40	26.11	27.66	29.50	30.78	32.22
6	23.64	25.54	26.98	28.71	30.45	32.24	33.43	35.03
	29.24	31.49	32.68	34.05	36.58	38.80	38.84	41.53
	23.19	24.73	26.55	28.02	29.51	31.16	32.41	34.02
7	25.68	27.51	29.49	30.86	32.34	34.00	35.29	37.28
	31.38	33.10	34.38	36.50	38.29	40.11	41.11	34.26
	24.71	26.56	28.27	29.92	30.95	32.39	34.26	35.61
8	27.47	29.40	31.09	33.09	33.59	35.25	37.49	38.66
	32.98	35.23	37.19	39.23	39.49	41.83	43.31	44.87

Note: First figure is 10%, followed by 5% and 1%.

Assumptions A.2. The true parameter vector θ_0 is an element of the parameter space $\Theta \subset \mathbb{R}^p$.

Assumption A.3. (Θ, σ) is a separable metric space.

Assumption A.4. The function $f(x_t, \theta)$ is Borel measurable for each $\theta \in \Theta$.

We now list a set of regularity conditions to obtain weak convergence of the GMM estimator $\hat{\theta}_T(\pi)$, indexed by π and defined in Section 2, to a function of Brownian motions. The main distributional results will hold under two alternative assumptions regarding the stochastic process x_t. Following Hansen (1982),

Table 5
Critical values of ExpPR_T for $\pi \in (0.20, 0.80)$

dim $q - p$	dim p 1	2	3	4	5	6	7	8
	2.60	3.56	4.50	5.37	6.16	6.98	7.86	8.68
1	3.24	4.25	5.33	6.23	7.09	7.88	8.83	9.71
	4.67	5.93	7.09	8.18	9.13	10.07	10.84	12.38
	3.54	4.49	5.39	6.27	7.02	7.85	8.63	9.42
2	4.23	5.28	6.25	7.15	7.96	8.75	9.61	10.62
	5.95	7.11	8.10	9.08	9.91	10.87	11.70	12.87
	4.50	5.32	6.14	6.99	7.84	8.57	9.44	10.16
3	5.29	6.15	7.06	7.96	8.88	9.58	10.49	11.22
	7.19	7.93	9.10	10.11	11.13	11.76	12.55	13.66
	5.32	6.13	7.00	7.88	8.62	9.43	10.09	10.80
4	6.22	7.04	7.90	8.83	9.66	10.59	11.08	11.94
	8.09	9.07	10.17	10.93	11.94	12.94	13.38	14.56
	6.16	6.87	7.79	8.54	9.24	10.58	10.84	11.58
5	7.08	7.94	8.80	9.48	10.36	11.31	11.97	12.96
	9.39	9.99	11.37	11.90	12.69	13.60	14.19	14.92
	6.90	7.80	8.45	9.28	10.10	10.88	11.58	12.26
6	7.88	8.79	9.54	10.38	11.16	12.11	12.67	13.50
	9.91	11.00	11.67	12.56	13.60	14.56	14.96	16.01
	7.75	8.54	9.37	10.12	10.93	11.59	12.25	13.01
7	8.73	9.59	10.53	11.20	11.91	12.76	13.43	14.27
	110.2	11.82	12.67	13.43	14.74	15.49	15.97	16.85
	8.43	9.23	10.04	10.79	11.45	12.20	13.12	13.69
8	9.48	10.33	11.19	12.00	12.53	13.41	14.33	14.91
	11.78	12.76	13.62	14.37	14.86	15.92	16.83	17.62

Note: First figure is 10%, followed by 5% and 1%.

one can impose stationarity and ergodicity conditions or else, as in Gallant and White (1988) and Andrews (1993), one can also consider a setup with conditions on a triangular array of random variables x_{Tt}. Assumptions A.5–A.13 are taken from Andrews (1993), who provides a complete discussion.

Assumption A.5. The process x_t is stationary and ergodic.

Assumption A.6. The $\{x_{Tt}: \leqslant T, T \geqslant 1\}$ is a triangular array of X-value random vector that is L°-near epoch dependence on a strong mixing base $\{y_{Tt}:$

Table 6
Critical values of ExpPR_T for $\pi \in (0.15, 0.85)$

dim $q-p$	dim p 1	2	3	4	5	6	7	8
	2.79	3.45	4.35	5.15	6.17	6.92	7.69	8.61
1	3.48	4.17	5.18	6.02	7.14	7.89	8.74	9.69
	5.16	5.80	6.99	8.13	9.37	10.16	10.70	12.38
	3.65	4.76	5.56	6.44	7.08	7.86	8.59	9.43
2	4.32	5.69	6.53	7.35	8.16	8.93	9.76	10.66
	6.01	7.71	8.64	9.48	10.32	11.17	12.11	12.96
	4.93	5.70	6.47	7.34	8.05	8.73	9.56	10.23
3	5.85	6.72	7.45	8.41	9.17	9.78	10.75	11.49
	8.07	8.65	9.80	10.66	11.61	12.24	13.14	13.92
	5.94	6.64	7.38	8.24	8.90	9.74	10.37	11.04
4	6.97	7.76	8.50	9.41	10.16	11.05	11.48	12.31
	9.35	10.22	11.06	11.63	12.60	13.65	14.15	15.19
	6.85	7.50	8.31	9.07	9.73	10.58	11.23	11.92
5	7.93	8.69	9.43	10.22	10.96	11.75	12.53	13.20
	10.45	11.32	12.41	13.06	13.47	14.62	14.75	15.84
	7.70	8.51	9.16	9.88	10.66	11.46	12.06	12.75
6	8.88	9.73	10.33	11.14	11.95	12.76	13.23	14.03
	11.43	12.34	13.05	13.67	14.79	15.55	15.74	17.01
	8.64	9.37	10.17	10.84	11.51	12.25	12.83	13.60
7	9.84	10.54	11.51	12.12	12.81	13.60	14.15	15.11
	12.55	13.36	13.87	14.75	15.52	16.36	16.99	18.17
	9.41	10.18	11.00	11.68	12.25	12.90	13.72	14.34
8	10.68	11.49	12.28	13.16	13.46	14.21	15.19	15.86
	13.29	14.27	15.07	16.03	16.06	17.27	18.08	18.94

Note: First figure is 10%, followed by 5% and 1%.

$t = , \ldots, 0, 1, \ldots ; T \geqslant 1\}$ where X is a Borel subset of R^k, and $\{\frac{1}{T}\Sigma_1^T x_{Tt}\colon T \geqslant 1\}$ is tight on X.[5]

Assumption A.7. For some $r \geqslant 2, f(x_{Tt}, \theta)$: $t \leqslant T$, $T \geqslant 1$ is a triangular array of R^p-valued random vectors that is L^2-near epoch dependence of size $-1/2$ on

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

Table 7
Critical values of $AvPR_T$ for $\pi \in (0.20, 0.80)$

dim $q-p$	dim p 1	2	3	4	5	6	7	8
	3.68	4.98	6.31	7.44	8.69	9.92	11.05	12.32
1	4.58	5.88	7.24	8.48	9.87	11.11	12.26	13.58
	6.55	7.89	9.40	10.95	12.11	13.60	15.02	16.51
	5.10	6.51	7.74	8.89	10.02	11.23	12.39	13.72
2	6.09	7.65	8.84	10.09	11.19	12.49	13.64	15.10
	8.37	10.03	11.32	12.82	13.87	15.02	16.35	17.97
	6.70	7.88	9.00	10.25	11.40	12.47	13.75	14.87
3	7.89	9.06	10.16	11.54	12.75	13.85	15.15	16.28
	10.47	11.54	12.95	14.32	15.66	16.86	17.80	19.25
	8.02	9.14	10.37	11.60	12.17	13.88	15.00	16.10
4	9.37	10.47	11.70	13.00	14.22	15.44	16.46	17.60
	12.12	13.50	14.76	15.63	16.95	18.84	19.65	20.86
	9.40	10.46	11.71	12.80	13.78	15.12	16.24	17.38
5	10.81	11.89	13.28	14.23	15.36	17.01	17.78	18.86
	13.94	14.74	16.23	17.66	18.61	20.26	20.87	22.14
	10.67	11.87	12.91	14.09	15.30	16.40	17.37	18.56
6	12.08	13.35	14.37	15.62	16.82	18.14	18.94	20.17
	15.03	16.64	17.54	18.98	20.22	21.35	22.16	23.80
	12.01	13.08	14.33	15.45	16.53	17.64	18.62	19.88
7	13.50	14.58	15.89	17.02	18.11	19.25	20.18	21.58
	16.76	17.91	19.19	20.08	21.68	22.71	24.12	25.12
	13.18	14.33	15.37	16.59	17.61	18.74	20.02	20.87
8	14.73	15.87	17.23	18.35	19.29	20.19	21.72	22.75
	18.32	19.39	20.56	21.71	22.42	23.77	25.19	26.33

Note: First figure is 10%, followed by 5% and 1%.

a strong mixing base $\{y_{Tt}: t = , \ldots, 0, 1, \ldots; T \geqslant 1\}$ or size $-r/(r-2)$ and $\sup_{t \leqslant T, T \geqslant 1} E\|f(x_{Tt}, \theta)\|^r < \infty$.

Now, we defined the following matrices indexed by π:

$$\bar{f}_1(\pi) = \frac{1}{[\pi T]}\, Yx \sum_{T}^{[\pi T]} f(x_t, \hat{\theta}(\pi)) \quad \text{and} \quad \bar{f}_2(\pi)$$

$$= \frac{1}{T - [\pi T]} \sum_{[\pi T]+1}^{T} f(x_t, \hat{\theta}(\pi)), \tag{A.1}$$

Table 8
Critical values of AvPR_T for $\pi \in (0.15, 0.85)$

dim $q - p$	dim p							
	1	2	3	4	5	6	7	8
	4.26	5.28	6.65	7.84	9.44	10.69	11.80	13.16
1	5.38	6.32	7.72	9.07	10.81	12.14	13.27	14.69
	7.76	8.61	10.30	12.21	13.73	15.28	16.64	18.40
	5.68	7.46	8.97	9.87	10.99	12.23	13.31	14.80
2	6.73	8.87	10.08	11.32	12.47	13.79	15.05	16.54
	9.35	11.90	13.28	14.46	15.85	16.98	18.48	20.25
	7.84	9.09	10.23	11.54	12.69	13.78	15.08	16.08
3	9.27	10.55	11.68	13.12	14.30	15.39	16.77	17.99
	12.58	13.73	15.35	16.78	18.23	19.09	20.30	21.69
	9.45	10.56	11.75	13.02	14.21	15.44	16.50	17.60
4	11.09	12.33	13.51	14.79	16.09	17.45	18.36	19.49
	14.17	16.22	17.24	18.39	19.72	21.61	22.44	23.60
	11.06	12.16	13.39	14.56	15.45	17.06	18.02	19.20
5	12.91	13.96	15.16	16.24	17.43	19.20	19.90	21.15
	16.76	18.06	19.33	20.48	21.48	23.13	23.84	25.22
	12.54	13.75	14.78	16.06	17.20	18.43	19.38	20.58
6	14.25	15.76	16.64	17.86	19.24	20.56	21.37	22.69
	18.18	19.53	20.85	22.18	23.56	24.72	25.32	27.17
	14.15	15.26	16.47	17.59	18.68	19.76	20.78	22.16
7	16.09	17.24	18.56	19.61	20.74	21.93	22.87	24.35
	20.25	21.53	22.71	23.74	25.25	26.24	27.63	28.77
	15.55	16.68	17.80	19.07	19.97	21.08	22.45	23.23
8	17.52	18.68	20.09	21.21	22.18	23.23	24.71	25.73
	22.01	23.14	24.32	25.53	25.87	27.96	29.23	30.09

Note: First figure is 10%, followed by 5% and 1%.

$$\hat{F}_1(\pi) = \frac{1}{[\pi T]} \sum_{1}^{[\pi T]} \frac{\delta}{\delta \theta'} f(x_t, \hat{\theta}(\pi)) \quad \text{and} \quad \hat{F}_2(\pi) = \frac{1}{T - [\pi T]} \sum_{[\pi T]+1}^{T} \frac{\delta}{\delta \theta'} f(x_t, \hat{\theta}(\pi)), \tag{A.2}$$

$$\hat{S}_1(\pi) = \frac{1}{[\pi T]} \sum_{k=-\ell(\pi T)}^{\ell(\pi T)} \sum_{1}^{T\pi} \omega(\pi T, k) (f(x_t, \theta(\pi)) - \bar{f}_1(\pi)) (f(x_{t-k}, \theta(\pi)) - \bar{f}_1(\pi))', \tag{A.3}$$

$$\hat{S}_2(\pi) = \frac{1}{T - [\pi T]} \sum_{k=-\ell(\pi T)}^{\ell(\pi T)} \sum_{[T\pi]+1}^{T} \omega(T - \pi T, k)(f(x_t, \theta(\pi)) - \bar{f}_2(\pi)) \times (f(x_{t-k}, \theta(\pi)) - \bar{f}_2(\pi))', \tag{A.4}$$

where $\hat{\theta}(\pi) \equiv \hat{\theta}_T(\pi)$ defined in Section 2; $\ell(\cdot)$ and $w(\cdot)$ are respectively the bandwidth and kernel used in the covariance matrix estimator.[6]

Assumption A.8. $\mathrm{Var}((1/\sqrt{T}) \sum_1^{[\pi T]} f(x_{Tt}, \theta)) \to \pi S, \forall \pi \in [0, 1]$ for some positive $q \times q$ matrix S.

Assumption A.9. $\sup_{\pi \in \Pi} \| \hat{\theta}(\pi) - \theta_0 \| \xrightarrow{p} 0$ for some θ_0 in the interior of Θ.

Assumption A.10. $\sup_{\pi \in \Pi} \| \hat{W}(\pi) - W(\pi) \| \xrightarrow{p} 0$ for some $q \times q$ matrices $W(\pi)$ for which $\sup_{\pi \in \Pi} \| W(\pi) \| < \infty$.

We define

$$\mathrm{F} = \lim_{T \to \infty} \frac{1}{T} \sum_1^T \mathrm{E} \frac{\delta}{\delta \theta'} f(x_{Tt}, \theta_0) \in R^{q \times q}. \tag{A.5}$$

Assumption A.11. $F(\pi)' S^{-1}(\pi) F(\pi)$ is nonsingular $\forall \pi \in \Pi$ and has eigenvalues bounded away from zero.

Assumption A.12. $f(x_{Tt}, \theta)$ is partially differentiable in $\theta \in \Theta_0$ $\forall x \in X$, where Θ_0 is some neighbourhood of θ_0, $(\delta/\delta\theta') f(x_{Tt}, \theta)$ is continuous in (x, θ) on $X \times \Theta_0$, and $\limsup_{T \to \infty} (1/T) \sum_1^T \mathrm{E} \sup_{\theta \in \Theta_0} |(\delta/\delta\theta') f(x_{Tt}, \theta)|^{1+\varepsilon} < \infty$ or some $\varepsilon > 0$.

Assumption A.13. $\lim_{T \to \infty} (1/T) \sum_1^{T\pi} \mathrm{E}\ (\delta/\delta\theta')\ f(x_{Tt}, \theta)$ exists uniformly over $\pi \in \Pi$ equals πF $\forall \pi \in \Pi$.

A.2. Proof of Theorem 2.1

We denote $\hat{\theta}_1(\pi)$ as the estimator of θ_0 for the first πT observations and $\hat{\theta}_2(\pi)$, the estimator for the remaining subsample $(\pi T + 1, \ldots, T)$. From Theorem 1 of Andrews (1993):

$$\sqrt{T}(\hat{\theta}_1(\pi) - \theta_0 \Rightarrow -\frac{1}{\pi} ((F'(\pi) S^{-1}(\pi) F(\pi))^{-1} F(\pi)' S^{-1/2}(\pi) B(\pi),$$

[6] One suitable choice is $\ell(a) = 0(a^{1/4})$, $w(a, k) = 1 - k/[\ell(a) + 1]$ for $a = [\pi T]$ or $T - [\pi T]$; see Newey and West (1987). For other possible choices see the references in the main text.

$$\sqrt{T}(\hat{\theta}_2(\pi) - \theta_0 \Rightarrow -\frac{1}{(1-\pi)}((F'(\pi)S^{-1}(\pi)F(\pi)))^{-1}$$
$$\times F(\pi)'S^{-1/2}(\pi)B(((1) - B(\pi),$$

where $B(\pi)$ is a p-vector of independent Brownian motions on $[0, 1]$ and $\Rightarrow$ denotes weak convergence as defined by Pollard (1984).

Using a first-order Taylor series expansion for $f_t(\theta) = f(x_t, \theta)$ evaluated as $\hat{\theta}_1 = \hat{\theta}_1(\pi)$ around θ_0, we obtain

$$\frac{1}{\sqrt{(T-[\pi T])}} \sum_{[\pi T]+1}^{T} f_t(\hat{\theta}_1)$$
$$= \frac{1}{\sqrt{(T-[\pi T])}} \sum_{[\pi T]+1}^{T} f_t(\theta_0)$$
$$+ \sqrt{1-\pi}\left[\frac{1}{(T-[\pi T])}\sum_{[\pi T]+1}^{T} \frac{\delta}{\delta\theta'} f_t(\theta_0)\right]\sqrt{T}(\hat{\theta}_1 - \theta_0)$$
$$+ o_p(1) \tag{A.6}$$

where all summations are over t.

In premultiplying by $\sqrt{1-\pi}S^{-1/2}$, we obtain that

$$\frac{1}{\sqrt{T}} S^{-1/2} \sum_{[\pi T]+1}^{T} f_t(\hat{\theta}_1) \Rightarrow B(1) - B(\pi) - \left[\frac{(1-\pi)}{\pi}\right]$$
$$\times S^{-1/2}F(F'\, S^{-1/2}F)^{-1}F'S^{-1/2}B(\pi).$$

We now decompose the matrix of the projection spanned by $S^{-1/2}F'$ like Sowell (1996a) namely:

$$S^{-1/2}F(F'\, S^{-1/2}F)^{-1}F'S^{-1/2} = H'\Lambda H, \tag{A.7}$$

where $HH' = I_q$ and

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

It follows that

$$\frac{1}{\sqrt{T}} S^{-1/2} \sum_{[\pi T]+1}^{T} f_t(\hat{\theta}_1) \Rightarrow B(1) - B(\pi) - \left[\frac{(1-\pi)}{\pi}\right] H'\Lambda HB(\pi),$$

and premultiplying by H yields

$$\frac{1}{\sqrt{T}} HS^{-1/2} \sum_{[\pi T]+1}^{T} f_t(\hat{\theta}_1) \Rightarrow \begin{bmatrix} -(BH_1(\pi) - \pi BH_1(1))/\pi \\ BH_2(1) - BH_2(\pi) \end{bmatrix},$$

where BH_1 and BH_2 are vectors of independent Brownian motions of dimensions p and $q-p$ respectively.

Since, H and S are full rank matrix, then $\mathrm{PR}(\pi)$ equals

$$\left[\frac{1}{\sqrt{(T-[\pi T])}} HS^{-1/2} \sum_{[\pi T]+1}^{T} f_t(\hat{\theta}_1)\right]' [HS^{-1/2}\hat{V}S^{-1/2}H']^{-1}$$
$$\times\left[\frac{1}{\sqrt{(T-[\pi T])}} HS^{-1/2} \sum_{[\pi T]+1}^{T} f_t(\hat{\theta}_1)\right]. \tag{A.8}$$

some algebra yield that

$$[HS^{-1/2}\hat{V}S^{-1/2}H']^{-1} = \begin{bmatrix} \pi I_p & 0 \\ 0 & I_{(q-p)} \end{bmatrix}, \tag{A.9}$$

then, we obtain the desired result:

$$\mathrm{PR}(\pi) \Rightarrow \left[\frac{[BH_1(\pi) - \pi BH_1(1)]'[BH_1(\pi) - \pi BH_1(1)]}{\pi(1-\pi)}\right]$$
$$+\left[\frac{[BH_2(1) - BH_2(\pi)]/[BH_2(1) - BH_2(\pi)]}{(1-\pi)}\right].$$

A.3. Proof of Theorem 3.1

From (A.6) it follows that

$$T^{-1/2}HS^{-1/2} \sum_{[\pi T]+1}^{T} f_t(\hat{\theta}_1) \Rightarrow HB(1) - HB(\pi) + HS^{-1/2}\mu_2(\pi)$$
$$- dHH'AH[B(\pi) + S^{-1/2}\mu_1(\pi)].$$

The result then follows directly from (A.7), (A.8), $HH'AH = H_1$ and the symmetry of the distribution of $BH_1(\pi) - \pi BH_1(1)$.

References

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

Andrews, D.W.K., Fair, R., 1988. Inference in econometric models with structural change. Review of Economic Studies 55, 615–640

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

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

Andrews, D.W.K., Ploberger, W., 1995. Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. Annals of Statistics 5, 1609–1629.

Epstein, L., Zin, S., 1991. Substitution, risk aversion and the temporal behavior of asset returns, Journal of Political Economy 99, 263–286.

Gallant, A.R., White, H., 1988. A United Theory of Estimation of Inference for Nonlinear Dynamic Models. Basil Blackwell, New York.

Ghysels, E., Hall, A., 1990a. A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator. International Economic Review 31, 355–364.

Ghysels, E., Hall, A., 1990b. Are consumption-based intertemporal capital asset pricing models structural?, J. Econometrics 45, 121–139.

Guay, A., 1996. Structural change tests based on moment conditions. Unpublished mimeo (Centre for Research on Economic Fluctuations and Employment, UQAM, Montreal, Canada).

Hall, A., 1993. Some aspects of generalized method of moments estimation. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics, Vol. 11, Chap. 15. North-Holland, Amsterdam.

Hall, A., Sen, A., 1996a. On the interpretation of optimality in certain tests of structural stability. Unpublished mimeo (North Carolina State University, Raleigh, NC).

Hall, A., Sen, A., 1996b. Structural stability testing in models estimated via generalized method of moments. Unpublished mimeo (North Carolina State University, Raleigh, NC).

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Kramer, W., Ploberger, W., Alt, R., 1988. Testing for structural change in dynamic models. Econometrica 56, 1355–1369.

Newey, W.K., 1993. Efficient estimation of models with conditional moment restrictions. In: Maddala, G.S., Rao, C.R., Vinod, H.D., (Eds.), Handbook of Statistics 11, Econometrics. North-Holland, Amsterdam, pp. 419–454.

Newey, W.K., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Ogaki, M., 1993. Generalized method of moments: econometric applications. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics 11, Econometrics. North-Holland, Amsterdam, pp. 455–488.

Sowell, F., 1996a. Optimal tests of parameter variation in the generalized method of moments framework. Econometrica, forthcoming.

Sowell, F., 1996b. Tests for violations of moment conditions. Unpublished mimeo (Carnegie Mellon University, Pittsburgh, PA).