Optimally Testing General Breaking Processes in Linear Time Series Models*

Graham Elliott
University of California, San Diego
9500 Gilman Drive
LA JOLLA, CA, 92093-0508

Ulrich K. Müller
University of St. Gallen
Dufourstr. 50
9000 St. Gallen, Switzerland

April 11, 2003

Abstract

There are a large number of tests for instability or breaks in coefficients in regression models designed for different possible departures from a stable regression. We make two contributions to this literature. First, we provide conditions under which optimal tests are asymptotically equivalent. Our conditions allow for models with many or relatively few breaks, clustered breaks, regularly occurring breaks or smooth transitions to changes in the regression coefficients. Thus we show nothing is gained asymptotically by knowing the exact breaking process. Second, we provide a statistic that is simple to compute, avoids any need for searching over high dimensions when there are many breaks, is valid for a wide range of data generating processes and has high power for many alternative models.

*Graham Elliott is grateful to the NSF for financial assistance under grant SES 0111238. Ulrich Müller is currently visiting the Department of Economics, Princeton University, and gratefully acknowledges financial support of the Swiss National Science Foundation. The authors thank Mark Watson, Giorgio Primiceri, Allan Timmermann, Paolo Giordani, Piotr Eliasz and seminar participants at Berkeley, UCSD, Princeton, EUI, Boston University, LSE, University of Aarhus and Yale for useful comments.

1 Introduction

It is reasonable to expect that there is some instability in most econometric relationships, across time or space. In cross sections, there is likely (as is typically found in longitudinal data) some degree of heterogeneity amongst agents. In time series, changing market conditions, rules and regulations etc. will also result in heterogeneity in the relationships. So long as this heterogeneity is not 'too strong', standard regression methods still have reasonable properties with the replacement of 'true' values of the coefficients with averages of the individual or intertemporal true values of coefficients (see White (2001) for precise results for limit theory under heterogeneity). If the heterogeneity is of a stronger form, then inference using standard methods will be misleading.

For this reason there is a large literature on testing for instability, or 'breaks' in parameters in time series regressions (restrictions to time series reduce the dimension of the problem since there is a natural ordering to the data). We consider linear models of the form $y_t = X_t'(\bar{\beta} + \beta_t) + Z_t'\delta + \varepsilon_t$ and consider the possibility of nonconstant β_t. Numerous difficulties have arisen in testing this possibility. First and foremost is the problem that there are many possible ways (i.e. models where) β_t can be nonconstant. Tests developed for one model of a nonconstant β_t may not be useful for other possible alternative models. Compounding this difficulty is that for time varying models, some nuisance parameters fail to be identified under the null hypothesis of constant β_t. This renders the typical intuitions of optimality of general Likelihood Ratio, Lagrange Multiplier, and Wald tests and asymptotic normality inapplicable in general. So arriving at a good (i.e. optimal) test may both be involved (deriving new distributions and requiring potentially novel asymptotic theory) and also may depend quite strongly on the alternative model the researcher has in mind.

Because of these difficulties, research has concentrated on very specific breaking processes. LaMotte and McWorther (1978), Franzini and Harvey (1983) and Shively (1988b), for instance, consider models where β_t is subject to Gaussian breaks of constant variance every period. But even for this apparently simple model the test statistics become such complicated functions of the observables that they are difficult to analyze in an asymptotic framework. Andrews and Ploberger (1994) (denoted AP in the sequel) derive asymptotically optimal tests when β_t is subject to a finite number of breaks by employing an average asymptotic power criterion. Bai and Perron (1998) promote the maximum of a sequence of F-statistics for this type of parameter variation. Even for a very moderate number of breaks, say, five, the AP statistics and the maximal F-statistics become computationally extremely involved, since they require searching over all possible combinations of the break dates. In a nutshell, tests of parameter constancy are mostly based on highly complicated statistics tailor-made for one specific breaking process, and little is known about their efficiency against other processes.

This paper makes two contributions. First, we show that for a very general set of breaking processes and a normality assumption on the disturbances, optimal tests of coefficient stability are asymptotically equivalent. The set of breaking processes includes breaks that occur in a random fashion, serial correlation in the changes of the coefficients, a clustering of break dates and so forth. Our results imply that any specific optimal test will have the same asymptotic power against any other breaking process of the set. We hence show that leaving the exact breaking process unspecified (apart from a scaling parameter) does not result in a loss of power, at least asymptotically.

Second, we suggest an easy-to-compute statistic that is asymptotically optimal for this set of breaking processes. Its computation requires no more than $k+1$ OLS regressions, where k is the dimension of the vector X_t. We investigate the local asymptotic power of the statistic and compare it to other popular statistics for breaking models. Given the first result, one would not expect great differences in performance, and we show that this is true even when one steps out of the set of breaking models for which our statistic is optimal. At the same time, the statistic suggested in this paper does have a somewhat superior performance for many models. We also show that it remains asymptotically valid under very general assumptions on the disturbances.

The following section examines the testing problem and describes the new test statistic. In the third section we establish the asymptotic equivalence of optimal tests for a large class of breaking processes. The construction of feasible tests with the same asymptotic power is taken up in Section 4, and Section 5 evaluates the asymptotic and small sample power of a number of tests for time variation in β_t and brings out the implications of the theory. A final section concludes. Proofs are collected in an appendix.

2 The Model and Tests for Breaks

The model this paper is concerned with is

$$y_t = X_t'(\bar{\beta} + \beta_t) + Z_t'\delta + \varepsilon_t \quad t = 1, \cdots, T \tag{1}$$

where y_t is a scalar, X_t, $\bar{\beta}$ and β_t are $k \times 1$ vectors, Z_t and δ are $d \times 1$, $\{y_t, X_t, Z_t\}$ are observed, $\bar{\beta}$, β_t and δ are unknown and ε_t is a mean zero disturbance. As a normalization we set $\beta_1 = 0$. The hypotheses to be tested are

$$H_0 : \beta_t = 0 \ \forall t \quad \text{against} \quad H_1 : \beta_t \neq 0 \text{ for some } t > 1. \tag{2}$$

so that under the null hypothesis the model reduces to

$$y_t = X_t'\bar{\beta} + Z_t'\delta + \varepsilon_t \quad t = 1, \cdots, T \tag{3}$$

In words, we want to test whether the coefficient vector β_t that links the observables X_t to y_t remains stable over time, while allowing for other stable links between y_t and the observables through Z_t.

The hypothesis test (2) of model (1) has received a great deal of attention in both the statistical and econometrics literature. The major reason why the literature has taken so many different approaches to the problem is that the alternative of a nonconstant β_t is so general. Obviously there exists a very large variety of ways β_t might evolve under the alternative, and any specific assumption leads to a different testing problem. The huge literature on this problem might be organized into two strands: the 'structural break' literature, which views the path of β_t under the alternative as unknown but fixed and described by vector of unknown parameters, and the 'time varying parameter' literature, which views $\{\beta_t\}$ under the alternative as random with some distribution.

In the 'structural break' literature, by far the most attention has been given to the single break model, in which

$$\beta_t = 0 \quad \text{for } t < \tau \qquad (4)$$
$$\beta_t = \bar{\beta}_1 \quad \text{for } t \geq \tau$$

for some $\bar{\beta}_1 \neq 0$. In this literature, $\bar{\beta}_1$ and τ are a fixed but unknown parameters. If τ was known, then one can rely on the usual F-test to distinguish (4) from parameter stability, an idea that goes back to Chow (1960). But in practice, τ is usually unknown, making it a nuisance parameter in the testing problem. Quandt (1958, 1960) suggested using the maximum F-statistic over all values of τ as a means to test the stability of β_t. This search over a set of dependent F-statistics affects the asymptotic distribution of the test, which ceases to be χ^2. Andrews (1993) explores the properties of such tests in a very general setting. Brown, Durbin and Evans (1975) suggested testing the stability of β_t by considering the partial sums of the standardized forecast errors of rolling regressions of (3), leading to the so called CUSUM test. Ploberger, Krämer and Kontrus (1989) and Ploberger and Krämer (1992) propose tests that are functions of the partial sum of OLS residuals of regression (3). These tests are straightforward to compute, but nothing is known about their optimality.

Conceptually it is also straightforward to extend Quandt's idea to the case of multiple breaks, where

$$\begin{aligned}
\beta_t &= 0 \quad \text{for } t < \tau_1 \\
\beta_t &= \bar{\beta}_1 \quad \text{for } \tau_1 \leq t < \tau_2 \\
&\vdots \\
\beta_t &= \bar{\beta}_{N-1} \quad \text{for } \tau_{N-1} \leq t < \tau_N \\
\beta_t &= \bar{\beta}_N \quad \text{for } \tau_N \leq t \leq T
\end{aligned} \qquad (5)$$

Bai and Perron (1998), for instance, examine the maximum of the F-statistic over all combinations of $(\tau_1, \cdots, \tau_N)$.[1] Because the number of break date combinations becomes huge

[1] In an asymptotic set-up, one must exclude breaking dates that are too close to the beginning and end of the sample in order to obtain a stable asymptotic distribution. A similar caveat applies for some circumstances if two break dates are too close to each other. See Bai and Perron (1998) for details.

even for moderate N (with $T = 100$ and $N = 5$, there are $\binom{100}{5} = 75,287,520$ combinations), they require some clever dynamic programming to implement such a test.

Although these maximum F-statistics can be naturally motivated as generalized likelihood ratio tests, this does not necessarily make them desirable tests. The reason is that under the null hypothesis, the break dates τ_j are unidentified, which strips standard testing procedures like the likelihood ratio, Wald or LM-tests of their usual asymptotic optimality properties.[2] AP have devised an optimal method for dealing with testing problems of this kind, which can also be applied to testing structural stability against (4) or (5). Their procedure is (asymptotically) optimal in the sense of maximizing a weighted average power criterion, where the weighting is both over the size of the breaks $(\bar{\beta}_j - \bar{\beta}_{j-1})$ and over the combinations of the break dates under the alternative. Using the same criterion, Sowell (1996) derives asymptotically optimal tests for the set of statistics that are continuous functionals of the partial sums of the sample moment condition. By choosing the weighting of the size of the breaks as a Gaussian distribution function, the expressions for these test statistics become much more compact, but still involve a sum over all combinations of break dates. While not posing any conceptual difficulties, even a moderate N thus leads to computationally very cumbersome test statistics. Andrews, Ploberger and Lee (1996) and Forchini (2002) derive analogous small sample optimal statistics, but in none of these papers optimal statistics are calculated for $N > 1$.

The 'time varying parameter' literature approaches the problem from a seemingly very different angle. There the nonconstant β_t under the alternative is viewed as being random, and contributions to this strand differ in the probability law they pose for β_t. A number of studies investigate models in which β_t deviates only temporarily from zero, so that the 'long-run' value remains $\bar{\beta}$. Watson and Engle (1985) and Shively (1988a), for instance, investigate the case of a stable autoregressive process for β_t. Most Markov switching models with recurring states and threshold autoregressive models are also closely related to this class.

The majority of studies in the time varying parameter literature, however, have con-

[2] Andrews and Ploberger (1995) showed, however, that the maximum F-statistic does possess a weak optimality property.

sidered the model where deviations of β_t from zero are permanent. In these models the alternative hypothesis is that β_t follows a random walk. In the case where $X_t = 1$ this model is the 'unobserved components' model examined in Chernoff and Zacks (1964) and Nyblom and Mäkeläinen (1983). For more general stationary X_t the model has been examined in Garbade (1977), LaMotte and McWhorter (1978), Franzini and Harvey (1983), Nabeya and Tanaka (1988), Shively (1988) and Leybourne and McCabe (1989) — see the annotated bibliography by Hackl and Westlund (1989) for further references. By making distributional assumptions for $\{\varepsilon_t\}$ and $\{\beta_t\}$ the difficulty in this approach consists of analyzing the likelihood of the model under both the null and alternative hypotheses — even for independent Gaussian disturbances $\{\varepsilon_t\}$ and a Gaussian Random Walk of $\{\beta_t\}$ under the alternative the resulting expressions are so complicated functions of the observables that little is known about the asymptotic properties of the tests these authors promote.

At least at first sight, the 'structural break' and the 'time varying parameter' literature seem very distinct. And surely, the typical path of β_t in a time varying parameter model with $\beta_t = \sum_{s=1}^{t} w_s$, w_s independent zero mean Gaussian variates is quite different from a model with N breaks such as (5). But it is, of course, perfectly possible to let w_t have a continuous distribution with probability p and $w_t = 0$ with probability $(1-p)$. The number of breaks N in β_t (i.e. the number $\Delta\beta_t$ which are nonzero) then follows a Poisson distribution with $E[N] = (T-1)p$. The outcome of such a model can hence be cast in terms of model (5), with N and $\{\bar{\beta}_1, \cdots, \bar{\beta}_N\}$ being random variables. By allowing for a suitable dependence in $\{w_t\}$, a model with a fixed number of breaks can be written in the time varying parameter form, too.

The relationship between these models does not stop there. Tests of model (5) that are optimal in the weighted average power sense of AP and Andrews, Ploberger and Lee (1996) will have to specify weight functions on (i) the number of breaks (ii) the distribution of break dates given their number and (iii) the distribution of the breaks given their dates and number. A reinterpretation of these weights as probability measures naturally leads to a particular time varying parameter model. Thinking about the unobserved β_t as fixed and using weights for their outcomes under the alternative or treating them as random hence amounts to the same thing. These relationships are due to the close link between Bayesian methods and optimal classical methods in general — see Berger (1985), for instance, for more details on this point. The time varying parameter literature and the structural break literature hence treat the exact same problem, albeit with a different emphasis on what the typical alternative looks like.

But against what kind of alternatives the hypothesis test (2) should be most concerned with? This obviously depends crucially on why we want to test for parameter constancy in the first place. Three main motivations come to mind:

First, the stability of relationship (3) might be an important question in its own right. A standard example is the stability of the link between monetary aggregates and output, a crucial question for conducting appropriate monetary policy (see Clarida et al. (2000) for a recent example). Also tests for the Lucas critique (which has been difficult to detect in practice) arise directly as tests of parameter instability (Engle et al. (1983), Engle and Hendry (1993), Linde (2001)). Typically, the alternative of interest here is absence of *any* stable relationship, including long-term relations. Relevant alternatives are hence those in which changes in β_t are permanent.

Second, one might want to take advantage of relationship (3) for forecasting y_t. But if (3) turns out to be unstable, then the appropriate forecast will have to be modified in at least two respects: On the one hand, a good forecast of y_t will then be driven more by the recent past than by the distant past, since recent observations of relationship (1) will be closer to the (unknown) future relationship than past observations — see Chernoff and Zacks (1964), Clements and Hendry (1999) and Stock and Watson (1996). On the other hand, the perceived instability will evidently also affect the confidence in the accuracy of the forecast, resulting in wider confidence intervals. From this perspective, the most 'damaging' form of a time varying $\{\beta_t\}$ is again one in which the true relationship has changed permanently compared to the beginning of the sample, since ignoring the time variation will then lead to biased forecasts, even at long horizons.

Third, the hypothesis test (2) is a crucial specification test when $\bar{\beta}$ is to be interpreted as a structural parameter. When $X_t = 1$ and there is no Z_t, then model (1) with $\beta_t = \sum_{s=1}^{t} w_s$

is an unobserved components model where y_t contains a unit root. While it is possible to estimate the sample mean of y_t for any realization, it is impossible to interpret it in any meaningful way as a parameter of the model. More generally, whenever β_t varies in a permanent fashion (as, for instance, in (4)), ignoring its variation and computing averages makes little sense — the computed average value has no interpretation as describing the effect on y_t of a marginal change X_t, since the true marginal effect depends on time t. Note that temporary deviations of β_t from zero do not necessarily lead to the same interpretational difficulties. In the extreme temporary case of β_t being independent and identically distributed, $\beta_t' X_t$ can usefully be thought of as part of a heteroskedastic disturbance. There is no problem in interpreting $\bar{\beta}$ as a meaningful and interesting parameter of the model. The more persistent $\{\beta_t\}$ becomes, however, the more $\bar{\beta}$ becomes an inadequate description of the time homogenous marginal effect on y_t of a marginal change in X_t. These interpretational difficulties are reflected in the behavior of standard inference about $\bar{\beta}$. Up to a certain degree, heterogeneity of β_t will not affect the asymptotic properties of F-tests on $\bar{\beta}$ at all — see White (2001) for a precise statement of such results. If β_t contains a unit root, however, then the $F-$test ceases to follow its standard distribution, even asymptotically.

All three motivations are hence more pervasive the more persistent the changes in β_t. At least when carried out for one of the three reasons discussed above, a useful test of parameter stability should hence maximize its power against permanent changes of β_t. While this suggests a focus on alternatives with a permanently varying β_t, the obvious problem remains that there exist many different persistent breaking processes. Intuitively, it seems that knowledge about the precise form of the variation in β_t under the alternative is required in order to carry out an efficient hypothesis test of parameter instability.

This paper shows that this intuition is largely mistaken. We will show that for a very large class of breaking processes with persistently varying β_t, and an assumption on the distribution of the disturbances, the optimal small sample statistics will be asymptotically equivalent. In other words, the precise form of the breaking process $\{\beta_t\}$ under the alternative is irrelevant for the asymptotic power of the tests. The one parameter that drives the asymptotic power of the optimal statistics is the expected average size of the breaks. This result dramatically simplifies the practice of testing against parameter instability, because it allows the applied researcher to leave the exact form of the alternative unspecified without foregoing (asymptotic) power. Additionally, we derive an easy-to-compute test statistic that shares this asymptotic optimality. The statistic is asymptotically valid under very general assumptions on the disturbances and the regressors — see Section 4 for details.

Specifically, we will show in the next section that under some additional regularity conditions concerning $\{X_t, Z_t\}$ and independent Gaussian disturbances $\{\varepsilon_t\}$, all optimal small sample statistics for testing (2) are asymptotically equivalent as long as the alternative of a time varying parameter satisfies the following Condition.

Condition 1 *Let $\{\Delta\beta_{T,t}\}$ be a double array of $k \times 1$ random vectors $\Delta\beta_{T,t,i}$. Assume*
(i) $\{T\Delta\beta_{T,t}\}$ is uniform mixing with mixing coefficient of size $-r/(2r-2)$ or strong mixing of size $-r/(r-2)$, $r > 2$
(ii) $E[\Delta\beta_{T,t}] = 0$ and there exists $K < \infty$ such that $E[|T\Delta\beta_{T,t,i}|^r] < K$ for all T, t, i
(iii) $\{T\Delta\beta_{T,t}\}$ is globally covariance stationary with nonsingular long-run covariance matrix Ω.

For notational simplicity, we will drop the dependence on T of all elements defined in Condition 1 and subsequent similar conditions. The dependence of the scale of β_t on T is introduced because optimal tests in an asymptotic framework will have power in a local neighborhood of the null hypothesis of parameter constancy. The appropriate neighborhood of nontrivial power of optimal tests is where the global covariance matrix of $\{\Delta\beta_t\}$ is of order T^{-2}. Also recall from the discussion above that optimal tests against a random $\{\beta_t\}$ as described in Condition 1 may equally be interpreted as optimal tests that maximize weighted average power over alternatives with nonstochastic $\{\beta_t\}$, where the weighting is according to a distribution that satisfies Condition 1.

Condition 1 allows for a multitude of diverse breaking models, although scenarios in which the number of breaks remains finite irrespective of the sample size are ruled out. For any finite sample, however, even a model with a single break satisfies Condition 1. The asymptotic thought experiment then entails that a larger sample from the same data generating process will contain more breaks eventually. Furthermore, models which are subject to breaks every period with probability p and arbitrary mean zero distribution with covariance Ω_p in case of a break also satisfy the condition. In this case, $\Omega = p\Omega_p$. Thus Condition 1 spans a wide range of specifications from models with rare large breaks to models with frequent small breaks. This covers the economically interesting case of persistent stochastic shocks that hit the economy infrequently but repeatedly. Autocorrelations in $\Delta\beta_t$ allow the coefficient vector to smoothly adjust to a new level after a break. The effect of an oil price shock, for instance, might take several periods before it is fully felt in the economy. Furthermore, mixing allows for variation in the variance of $\Delta\beta_t$, thus generating periods of fewer or more changes. Similar to the randomly occurring breaks, the Condition covers the case of breaks that occur with a certain regular pattern, say, every sixteen quarters. Such a set-up might be motivated by policy changes following for example presidential elections. In essence, virtually any persistent breaking process is captured by Condition 1.

The possibility of obtaining optimal statistics against a wide class of models has been noted before by Nyblom (1989). He derives small sample locally optimal tests for model (1) when β_t follows a martingale, the disturbances $\{\varepsilon_t\}$ have known density and $E[\Delta\beta_t\Delta\beta_t']$ is known for all t. The locally optimal test (that is the test which maximizes power for very small $E[\Delta\beta_t\Delta\beta_t'])$ is then independent of the exact distribution of $\Delta\beta_t$. Given that for any fixed number of breaks N, $\{\beta_t\}$ is a martingale as long as breaks have (conditional) mean zero, this result implies that the locally optimal test is independent of the number of breaks N. Note that Nyblom's assumption and Condition 1 are truly distinct: Not all martingales satisfy Condition 1 (a counter example being $\{\beta_t\}$ with exactly one break for any sample size), and Condition 1 covers assumptions on $\{\beta_t\}$ that fail to be a martingale.

The result of the next section implies that, at least for testing purposes, the precise form of the time variation is of very little importance. The power of any tailor-made statistic against a certain breaking process approaches the power of any other optimal statistic as the sample size increases, as long as the breaking processes are such that Condition 1 holds. Intuition suggests that this independence of optimal tests of the exact breaking process might, at least qualitatively, extend to optimal tests for processes that are not covered by Condition 1, but which are similar in nature. We will show in Section 5 that this intuition is correct for the statistics suggested by AP and others. This implies that in practice, tests against time variation in the parameters can be carried out with any reasonable statistic.

In practice, we recommend basing inference on the test statistic $\hat{J}$. For the special case of $X_t = 1$ and serially uncorrelated $\{\varepsilon_t\}$, $\hat{J}$ is the Most Powerful Invariant (MPI) test in a Gaussian unobserved component model, as analyzed by Franzini and Harvey (1983) and Shively (1988). For more general assumptions on X_t and ε_t, $\hat{J}$ does not correspond to a test previously suggested in the literature.

$\hat{J}$ can be computed in a few simple steps, involving only matrix manipulations and OLS regressions:

- Step 1: Compute the OLS residuals $\{\hat{\varepsilon}_t\}$ by regressing $\{y_t\}$ on $\{X_t, Z_t\}$

- Step 2: Construct a consistent estimator $\hat{V}_X$ of the $k \times k$ long-run covariance matrix of $\{X_t \varepsilon_t\}$. When ε_t can be assumed uncorrelated, a natural choice is the heteroskedasticity robust estimator $\hat{V}_X = T^{-1} \sum_{t=1}^{T} X_t X_t' \hat{\varepsilon}_t^2$. For the more general case of possibly autocorrelated ε_t, many such estimators have been suggested; see Newey and West (1987) or Andrews (1991) and the discussion in Section 4.

- Step 3: Compute $\{\hat{U}_t\} = \{\hat{V}_X^{-1/2} X_t \hat{\varepsilon}_t\}$ and denote the k elements of $\{\hat{U}_t\}$ by $\{\hat{U}_{t,i}\}$, $i = 1, \cdots, k$.

- Step 4: For each series $\{\hat{U}_{t,i}\}$, compute a new series, $\{\hat{w}_{t,i}\}$ via $\hat{w}_{t,i} = \bar{r} \hat{w}_{t-1,i} + \Delta \hat{U}_{t,i}$ and $\hat{w}_{1,i} = \hat{U}_{1,i}$, where $\bar{r} = 1 - 10/T$.

- Step 5: Compute the squared residuals from OLS regressions of $\{\hat{w}_{t,i}\}$ on $\{\bar{r}^t\}$ individually, and sum all of those for $i = 1, \cdots, k$.

- Step 6: Multiply this sum of sum of squared residuals by $\bar{r}$ and subtract $\sum_{i=1}^{k} \sum_{t=1}^{T} (\hat{U}_{t,i})^2$.

The null hypothesis of parameter stability is rejected for small values of $\hat{J}$ and asymptotic critical values are given in Table 1 below for $k = 1, \cdots, 10$. The critical values are independent of the dimension of Z_t.

3 Asymptotically Optimal Tests for General Breaking Processes

In this section we will show that under certain regularity conditions, any small sample optimal test statistic of the hypothesis test (2) will be asymptotically equivalent if under the alternative the breaking processes is as described in Condition 1. The focus of this enterprise lies on establishing that knowledge about the breaking process over and above what is stated in Condition 1 is not helpful for constructing a more powerful test, at least asymptotically. It turns out, however, that optimal tests do depend on the average magnitude of the breaks, as described by Ω of Condition 1. But with our focus on understanding the impact of breaking processes of different form — rather than differences in breaking processes that simply arise by some unknown scaling — we will treat Ω as known in this section. This will establish the relevant benchmark case, in which additional knowledge about the exact form of the breaking process is entirely without asymptotic value for the testing problem.

In order to be able to write the model in matrix form, define the $T \times k$ matrix $X = (X_1, \cdots, X_T)'$, the $T \times d$ matrix $Z = (Z_1, \cdots, Z_T)'$, the $T \times 1$ vector $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_T)'$, the $(kT) \times 1$ vector $\beta = (\beta_1', \cdots, \beta_T')'$ and the $T \times (kT)$ matrix $\Xi = \begin{pmatrix} X_1' & 0 & \cdots & 0 \\ 0 & X_2' & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & X_T' \end{pmatrix}$, and, for future reference, the $T \times (d+k)$ matrix $Q = (X, Z)$, the $T \times (T-d-k)$ matrix B_Q that satisfies $B_Q' B_Q = I_{T-d-k}$ and $B_Q' Q = 0$, and $M = B_Q B_Q' = I - Q(Q'Q)^{-1}Q'$.

With these definitions, model (1) can be rewritten as

$$y = \Xi\beta + X\bar{\beta} + Z\delta + \varepsilon \tag{6}$$

We will derive optimal statistics that are invariant to transformations of the data of the form

$$(y, Q) \to (y + X\bar{b} + Z\bar{d}, Q) \quad \text{for any } \bar{b} \text{ and } \bar{d} \tag{7}$$

Note that all standard tests against structural breaks satisfy this property. In words the invariance requirement means that the outcome of the test for parameter stability does not depend on how exactly the linear regression is formulated. For an autoregressive process of order one, for instance, invariance to the group of transformations (7) implies that the test of parameter stability comes to the same conclusion independent of whether $\{y_t\}$ is regressed on $\{y_{t-1}\}$ or $\{\Delta y_t\}$ is regressed on $\{y_{t-1}\}$.

From the theory of invariant tests as described in Lehmann (1986), pp. 282–364, any invariant test can be written as a function of a maximal invariant of the group of transformations (7). One maximal invariant is given by

$$(h, Q) = (B_Q' y, Q).$$

The small sample optimal statistic that is invariant to the group of transformations (7) can now be computed by comparing the likelihood of (h, Q) under the null of no breaks with the likelihood of (h, Q) under the alternative of a nonzero β. To be able to write down the likelihood, we must make a distributional assumption for ε, which we will assume to be multivariate Gaussian. Additionally, we impose some regularity conditions concerning $\{Q_t\}$.

Condition 2 *(i)* $y_{T,t} = X_{T,t}'(\bar{\beta} + \beta_{T,t}) + Z_{T,t}'\delta + \varepsilon_t$
(ii) $\{\varepsilon_t\}$ *is i.i.d.* $\mathcal{N}(0, \sigma^2)$ *and* $\{\varepsilon_t\}$ *is independent of* $\{\beta_{T,t}\}$
(iii) $\hat{Q}_{T,t} = Q_{T,t}|\{Q_{T,t-1}, Q_{T,t-2}, \cdots, y_{T,t-1}, y_{T,t-2}, ...\}$ *is independent of* $\{\varepsilon_t\}$*, there exist measures* $\bar{\nu}_{\hat{Q},T,t}$ *such that* $\hat{Q}_{T,t}$ *has density* $f_{\hat{Q},T,t}$ *with respect to* $\bar{\nu}_{\hat{Q},T,t}$*, and* $\{f_{\hat{Q},T,t}, \bar{\nu}_{\hat{Q},T,t}\}$ *do not depend on* $\bar{\beta}$*,* δ *and* $\{\beta_{T,t}\}$*.*
Furthermore, if $\beta_{T,t} = 0 \forall t$*, then additionally*
(iv) $\{Q_{T,t}\}$ *is uniform mixing of size* $-r/(2r-2)$ *or strong mixing of size* $-r/(r-2)$*,* $r > 2$
(v) $E[Q_{T,t}Q_{T,t}'] = \Sigma_Q$*,* $T^{-1}\sum_{t=1}^{[sT]} Q_{T,t}Q_{T,t}' \xrightarrow{p} s\Sigma_Q$ *uniformly in* s*,* Σ_Q *and* $T^{-1}\sum_{t=1}^{T} Q_{T,t}Q_{T,t}'$ *are positive definite for all* T *and there exists* $K < \infty$ *such that* $E[|Q_{T,t,i}|^r] < K$ *for all* T, t, i*.*

The distributional assumption on ε_t is crucial for the development of an optimal statistic, but our test will be valid under much less stringent conditions on ε_t — see Section 4 below. Part (iii) of Condition 2 requires the conditional distribution of Q_t given past values of Q_t and y_t not to depend on $\bar{\beta}$, δ and $\{\beta_t\}$, which is the assumption of weak exogeneity as described in detail by Engle, Hendry and Richard (1983). This assumption will allow a factorization of the likelihood of (y, Q) into two pieces, one capturing the contribution to the likelihood of $\{\varepsilon_t = y_t - X_t'(\bar{\beta} + \beta_t) - Z_t'\delta\}$ and the other the contribution of $\{\hat{Q}_t = Q_t | \{Q_{t-1}, Q_{t-2}, \cdots, y_{t-1}, y_{t-2}, ...\}\}$. The independence of the latter piece of $\{\beta_t\}$ will ensure that it cancels in the ratio of the likelihoods of the null and alternative hypothesis, making the resulting optimal statistic independent of the exact form of either $\{f_{\hat{Q},t}\}$ or $\{\bar{\nu}_{\hat{Q},t}\}$.

Further restrictions on $\{Q_t\}$ in parts (iv) and (v) are only required to hold under the null hypothesis of $\beta_t = 0 \forall t$. The assumptions are rather weak, allowing for stationary as well as non-stationary behavior of the regressors. They do not, however, accommodate deterministic or stochastic trends.

Under Condition 2, we find that for a given β, the density of the data is

$$\begin{aligned} f_{y,Q|\beta}(y, Q) &= \prod_{t=1}^{T} (2\pi)^{-1/2} \sigma^{-1} \exp\left[-\tfrac{1}{2}\varepsilon_t^2/\sigma^2\right] f_{\hat{Q},t} \\ &= (2\pi\sigma^2)^{-T/2} \exp\left[-\tfrac{1}{2}\varepsilon'\varepsilon/\sigma^2\right] \prod_{t=1}^{T} f_{\hat{Q},t} \end{aligned}$$

where $\varepsilon = y - X\bar{\beta} - \Xi\beta - Z\delta$ is to be interpreted as a function of $\bar{\beta}$, δ and $\{\beta_t\}$ and the data (y, Q), whereas $\{\hat{Q}_t\}$ is a function of the data alone. The unconditional density may hence be written as

$$f_{y,Q}^1(y, Q) = (2\pi\sigma^2)^{-T/2} \int \exp\left[-\tfrac{1}{2}\varepsilon'\varepsilon/\sigma^2\right] d\nu_\beta \prod_{t=1}^{T} f_{\hat{Q},t} \tag{8}$$

where ν_β is the measure of β.

Now $h = B_Q' y$, so by standard calculations we find from (8) for the density of (h, Q) under the alternative hypothesis

$$f_{h,Q}^1(h, Q) = \int (2\pi\sigma^2)^{-(T-k-d)/2} \exp\left[-\tfrac{1}{2}\sigma^{-2}(h - B_Q'\Xi\beta)'(h - B_Q'\Xi\beta)\right] d\nu_\beta \prod_{t=1}^{T} f_{\hat{Q},t}$$

and clearly, under the null hypothesis of $h = B_Q'\varepsilon$

$$f_{h,Q}^0(h, Q) = (2\pi\sigma^2)^{-(T-k-d)/2} \exp\left[-\tfrac{1}{2}\sigma^{-2}h'h\right] \prod_{t=1}^{T} f_{\hat{Q},t}.$$

We therefore find the likelihood ratio statistic of the maximal invariant (h, Q) to be

$$LR_T = \int \exp\left[\sigma^{-2}h'B_Q'\Xi\beta - \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta\right] d\nu_\beta. \tag{9}$$

The essential problem for obtaining the optimal test for a particular break process (i.e. a particular choice of ν_β) revolves around the complexity of evaluating LR_T. For any specific choice of ν_β, it is in principle possible to write LR_T as an explicit function of y and Q. But even for moderately complex breaking processes, the resulting function becomes analytically intractable. The usual way of obtaining asymptotic optimality results — writing down the small sample optimal statistic and taking limits — is thus not feasible here.

Rather, we will show that LR_T converges in probability under both the null and alternative hypothesis to another, much more tractable statistic $\widetilde{LR}_T$, that depends on the distribution of β only through Ω. On the one hand, this is will prove the claim that all small sample optimal statistics for any breaking process that satisfies Condition 1 will be asymptotically equivalent. On the other hand, we will choose $\widetilde{LR}_T$ in a way which makes the actual computation of the statistic straightforward, thus making progress towards the goal of deriving a simple statistic with good power for any of these breaking processes.

For the definition of $\widetilde{LR}_T$ and the subsequent proofs, we will need some additional notation and definitions. Let

$$\Omega^* = \sigma^{-2}\Sigma_X^{1/2}\Omega\Sigma_X^{1/2},$$

where $E[X_t X_t'] = \Sigma_X$ is the upper $k \times k$ block of Σ_Q, and note that Ω^* is the long-run variance of $\{T\Delta\beta_t^*\} = \{\sigma^{-1}T\Sigma_X^{1/2}\Delta\beta_t\}$. Ω^* is the average size of the breaks after having normalized the model for the covariance of $\{X_t\}$ and the variance of ε_t, a more appropriate measure for the relative magnitude of the breaking process.

The spectral decomposition of Ω^* will play a major role in the subsequent analysis. Let P^* be the $k \times k$ orthonormal matrix of the eigenvectors of Ω^* and let $\Lambda = \text{diag}(a_1^2, \cdots, a_k^2)$ be the diagonal matrix of the eigenvalues of Ω^* (such that $\Omega^* = P^*\Lambda P^{*\prime}$), where we define a_i, $i = 1, \cdots, k$, to be nonnegative. Furthermore, define the $T \times 1$ vector e of ones, $M_e = I - e(e'e)^{-1}e'$, the $k \times 1$ vector $\iota_{k,i}$ with a one in the ith row and zeros elsewhere, the $T \times T$ matrix $F = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$, the $T \times T$ matrix $G_a = H_a^{-1} - H_a^{-1}e(e'H_a^{-1}e)^{-1}e'H_a^{-1}$, where $H_a =$

$r_a^{-1}FA_aA_a'F'$, $A_a = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -r_a & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -r_a & 1 \end{pmatrix}$ and $r_a = \frac{1}{2}(2+a^2T^{-2}-T^{-1}\sqrt{4a^2+a^4T^{-2}}) = 1 - aT^{-1} + o(T^{-1})$. Further define the following random elements, that are needed for the ensuing arguments: let $T\tilde{\beta} \sim \mathcal{N}(0, FF' \otimes \Omega)$, let $\tilde{\gamma}$ be a $(Tk \times 1)$ vector, and let $\{\Delta\gamma_{T,t}\}$ be a double array of $k \times 1$ random vectors with elements $\Delta\gamma_{T,t,i}$, where (i) $\tilde{\gamma}$ has the same distribution as $\tilde{\beta}$ and $\{\Delta\gamma_{T,t}\}$ has the same distribution as $\{\Delta\beta_{T,t}\}$ of Condition 1 and (ii) $\tilde{\beta}$, $\tilde{\gamma}$ and $\{\Delta\gamma_{T,t}\}$ are mutually independent and independent of $\{\varepsilon_t\}$, $\{Q_{T,t}\}$ and $\{\Delta\beta_{T,t}\}$.

We will show that LR_T is asymptotically equivalent to the the statistic

$$\widetilde{LR}_T = \int \exp\left[\sigma^{-2}h'B_Q'\Xi[M_e \otimes I_k]\tilde{\beta} - \tfrac{1}{2}\sigma^{-2}\tilde{\beta}'[M_e \otimes \Sigma_X]\tilde{\beta}\right] d\nu_{\tilde{\beta}}. \tag{10}$$

Note that $\widetilde{LR}_T$ is not a feasible statistic, since it depends on the generally unknown parameters σ^2 and Σ_X. But Theorem 3 in Section 4 below establishes that it is possible to construct a feasible statistic that does not depend on such knowledge, but that has the same (nondegenerate) asymptotic distribution under both the null and alternative hypothesis.

We begin by considering the asymptotic behavior of $\widetilde{LR}_T$, and will then show $LR_T - \widetilde{LR}_T \xrightarrow{p} 0$. Because $\tilde{\beta}$ is multivariate normal, we can explicitly carry out the integration in (10) by 'completing the square'. By some matrix manipulations detailed in the appendix, we arrive at the following equality.

Lemma 1

$$\widetilde{LR}_T = \prod_{i=1}^{k}\left[\frac{1-r_{a_i}^{2T}}{T(1-r_{a_i}^2)r_{a_i}^{T-1}}\right]^{-1/2} \exp\left[-\tfrac{1}{2}v_i'[G_{a_i} - M_e]v_i\right]$$

where the t^{th} element of v_i is the $((t-1)k+i)th$ element of $[I \otimes P^{*\prime}\sigma^{-1}\Sigma_X^{-1/2}]\Xi My$ or, equivalently, $v_i = [I \otimes \iota_{k,i}'P^{*\prime}\sigma^{-1}\Sigma_X^{-1/2}]\Xi My$.

A test based on the statistic

$$J(\Omega^*) = \sum_{i=1}^{k} v_i'[G_{a_i} - M_e]v_i \tag{11}$$

will hence be exactly equivalent to a test based on $\widetilde{LR}_T$, since $J(\Omega^*)$ is just a monotone transformation of $\widetilde{LR}_T$. Being an explicit function of observables, it is tedious but straightforward to derive the asymptotic distribution of $J(\Omega^*)$ under the null hypothesis, which is an obvious special case of the following Lemma (the greater generality is needed for an argument in the proof of Theorem 1 below). Here and in subsequent derivations, the limits of integration are understood to be zero and one, if not stated otherwise. Further, $\int G$ stands for $\int G(s)ds$ an so forth.

Lemma 2 *Under Condition 2 and the null hypothesis of $h = B_Q'\varepsilon$, for any positive $c_1, \cdots, c_k$*

$$\sum_{i=1}^{k} v_i'[G_{c_i} - M_e]v_i$$
$$\Rightarrow \sum_{i=1}^{k} \left[-c_i J_i(1)^2 - c_i \int J_i^2 - \frac{2c_i}{1-e^{-2c_i}} \left[e^{-c_i} J_i(1) + c_i \int e^{-c_i s} J_i \right]^2 + \left[J_i(1) + c_i \int J_i \right]^2 \right]$$

where $J_i(s) = W_{\varepsilon,i}(s) - \int_0^s e^{-c_i(s-\lambda)} W_{\varepsilon,i}(\lambda)d\lambda$ and $W_{\varepsilon,i}$ and $W_{\beta,i}$ are the ith elements of the independent $k \times 1$ standard Wiener processes W_ε and W_β.

We now turn to the argument that $LR_T - \widetilde{LR}_T \xrightarrow{p} 0$. Given that it is not feasible to compute the integral in the expression for LR_T explicitly, we will take advantage of the similarity of the expressions inside the integral in expressions (9) and (10). The strategy will be to do the asymptotic reasoning 'inside the integration'.

Lemma 3 *Under Conditions 1 and 2 the following weak convergences hold jointly with the convergence in Lemma 2*

(i) $\sigma^{-2}(\varepsilon' M \Xi \beta, \varepsilon' M \Xi \gamma) \Rightarrow (\int \bar{W}_\beta' \Lambda^{1/2} dW_\varepsilon, \int \bar{W}_\gamma' \Lambda^{1/2} dW_\varepsilon)$

(ii) $\sigma^{-2}(\beta' \Xi' M \Xi \beta, \gamma' \Xi' M \Xi \gamma) \Rightarrow (\int \bar{W}_\beta' \Lambda \bar{W}_\beta, \int \bar{W}_\gamma' \Lambda \bar{W}_\gamma)$

(iii) $\sigma^{-2}(\varepsilon' M \Xi [M_e \otimes I_k] \tilde{\beta}, \varepsilon' M \Xi [M_e \otimes I_k] \tilde{\gamma}) \Rightarrow (\int \bar{W}_{\tilde{\beta}}' \Lambda^{1/2} dW_\varepsilon, \int \bar{W}_{\tilde{\gamma}}' \Lambda^{1/2} dW_\varepsilon)$

(iv) $\sigma^{-2}(\tilde{\beta}' [M_e \otimes \Sigma_X] \tilde{\beta}, \tilde{\gamma}' [M_e \otimes \Sigma_X] \tilde{\gamma}) \Rightarrow (\int \bar{W}_{\tilde{\beta}}' \Lambda \bar{W}_{\tilde{\beta}}, \int \bar{W}_{\tilde{\gamma}}' \Lambda \bar{W}_{\tilde{\gamma}})$

where W_β, W_γ, $W_{\tilde{\beta}}$, $W_{\tilde{\gamma}}$ and W_ε are independent $k \times 1$ standard Wiener processes and bars denote demeaned Wiener processes.

Parts (i) to (iv) of Lemma 3 imply that the integrands in expressions (9) and (10) converge weakly to the same limit under the null hypothesis, where β and $\tilde{\beta}$ are interpreted as random vectors with distriubtions ν_β and $\nu_{\tilde{\beta}}$, respectively. While highly suggestive, this result in itself is not enough for the convergence of $LR_T - \widetilde{LR}_T \xrightarrow{p} 0$ because the convergence in probability is a statement of the asymptotic behavior of the integrals (9) and (10).

To tackle this problem, it will be useful to note that LR_T and $\widetilde{LR}_T$ can be alternatively written as integrals with respect to the measures of γ and $\tilde{\gamma}$, respectively, since these measures are identical to those of β and $\tilde{\beta}$

$$
\begin{aligned}
LR_T &= \int \exp\left[\sigma^{-2}h'B_Q'\Xi\gamma - \tfrac{1}{2}\sigma^{-2}\gamma\Xi'M\Xi\gamma\right] d\nu_\gamma \\
\widetilde{LR}_T &= \int \exp\left[\sigma^{-2}h'B_Q'\Xi[M_e\otimes I_k]\tilde{\gamma} - \tfrac{1}{2}\sigma^{-2}\tilde{\gamma}'[M_e\otimes\Sigma_X]\tilde{\gamma}\right] d\nu_{\tilde{\gamma}}
\end{aligned}
$$

Letting

$$
\begin{aligned}
\xi(\beta) &= \exp\left[\sigma^{-2}h'B_Q'\Xi\beta - \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta\right] \\
\tilde{\xi}(\tilde{\beta}) &= \exp\left[\sigma^{-2}h'B_Q'\Xi[M_e\otimes I_k]\tilde{\beta} - \tfrac{1}{2}\sigma^{-2}\tilde{\beta}'[M_e\otimes\Sigma_X]\tilde{\beta}\right]
\end{aligned}
$$

allows us to write

$$
\begin{aligned}
E\left[(LR_T - \widetilde{LR}_T)^2\right] &= E\left[\left(\int\xi(\beta)\nu_\beta - \int\tilde{\xi}(\tilde{\beta})\nu_{\tilde{\beta}}\right)\left(\int\xi(\gamma)d\nu_\gamma - \int\tilde{\xi}(\tilde{\gamma})d\nu_{\tilde{\gamma}}\right)\right] \\
&= E\left[\xi(\beta)\xi(\gamma) - \xi(\beta)\tilde{\xi}(\tilde{\gamma}) - \tilde{\xi}(\tilde{\beta})\xi(\gamma) + \tilde{\xi}(\tilde{\beta})\tilde{\xi}(\tilde{\gamma})\right]. \qquad (12)
\end{aligned}
$$

Note that the interpretation of β, $\tilde{\beta}$, γ and $\tilde{\gamma}$ changes from dummy variables of integration in the first line to random vectors in the second. Now all four terms inside the expectation operator in (12) converge weakly to the same limit by the Continuous Mapping Theorem and Lemma 3. But convergence in distribution implies convergence in expectation for uniformly integrable random variables. So if the respective products of $\xi(\beta)$, $\xi(\gamma)$, $\tilde{\xi}(\tilde{\beta})$ and $\tilde{\xi}(\tilde{\gamma})$ could be shown to be uniformly integrable, we would find that $LR_T - \widetilde{LR}_T \to 0$ in mean square under the null hypothesis, and the convergence in probability follows.

In the appendix we make a bounding argument for $\xi(\cdot)$ and $\tilde{\xi}(\cdot)$ to ensure the uniform integrability, i.e. we show that replacing $\xi(\cdot)$ and $\tilde{\xi}(\cdot)$ by $\xi(\cdot)\mathbf{1}[\xi(\cdot) < K']$ and $\tilde{\xi}(\cdot)\mathbf{1}[\tilde{\xi}(\cdot) < K']$ in the definition of LR_T and $\widetilde{LR}_T$, respectively, with $\mathbf{1}$ being the indicator function, is innocuous when K' is chosen large enough.

Theorem 1 *Under Conditions 1 and 2, as $T \to \infty$,*

$$LR_T - \widetilde{LR}_T \xrightarrow{p} 0$$

under the null hypothesis of $h = B'_Q \varepsilon$.

In order to substantiate the claim of asymptotic equivalence of tests based on LR_T and $\widetilde{LR}_T$ we still lack the crucial additional step of showing that the convergence in probability of Theorem 1 also holds under the alternative hypothesis. A brute force approach of running through the same arguments that led to Theorem 1 also for the alternative hypothesis is extremely cumbersome and barely tractable, since a varying β_t will lead to changes in y_t that in general will feed back to changes in Q_t, given that Condition 2 allows weakly exogenous regressors. Furthermore, the bounding argument referred to above is not easily implemented under the alternative data generating process.

Out of these reasons we rather follow AP in taking the more indirect route of proving that the density of (h, Q) under the alternative hypothesis is *contiguous* to the density of (h, Q) under the null. Contiguity can be thought of as a generalization of the concept of absolute continuity to sequences of densities; if a sequence of densities of a data generating process can be shown to be contiguous to another sequence of densities, then all statements of convergence in probability of the latter automatically also hold under the former data generating process. The reader is referred to the excellent survey of Pollard (2001) for a more detailed introduction to the concept.

Theorem 2 *Under Conditions 1 and 2, the sequence of densities $\{f^1_{h,Q}(h, Q)\}_T$ are contiguous to the densities $\{f^0_{h,Q}(h, Q)\}_T$.*

Corollary 1 *Under Conditions 1 and 2 the convergence in probability of Theorem 1 also holds under the alternative hypothesis of $h = B'_Q(\varepsilon + \Xi\beta)$.*

Since convergence in probability implies convergence in distribution, Theorem 1 and Corollary 1 imply that the small sample optimal statistic LR_T and the statistic $\widetilde{LR}_T$ have the same asymptotic distributions under the null and alternative hypothesis, which in turn implies the same local power. As the sample size gets large, nothing is hence lost by relying on $\widetilde{LR}_T$ rather than the tailor-made LR_T for testing the stability of parameters. Or put differently, the knowledge of the exact breaking process is not helpful for conducting a better test.

Additionally, given that any specific LR_T converges in probability to the same $\widetilde{LR}_T$, any given pair of small sample optimal statistics for Condition 1 breaking processes also converge in probability under the null hypothesis. Furthermore, the densities described by these two breaking processes are both contiguous to the null density by Theorem 2, hence the convergence in probability continues to hold under both these alternatives. Theorems 1 and 2 thus also imply that one can rely on any one specific small sample optimal statistic for a breaking process that satisfies Condition 1 to obtain the same asymptotic power against *any* breaking process that is covered in Condition 1.[3] Each optimal test has asymptotically the same ability to distinguish each possible alternative in our class of models.

4 Feasible asymptotically optimal test statistics

As shown in the last section, the statistic $J(\Omega^*)$ is asymptotically optimal for the testing problem (2) under Conditions 1 and 2. $J(\Omega^*)$ is not a feasible statistic, however, since it requires $\sigma^2 = E[\varepsilon_t^2]$ and $\Sigma_X = E[X_t X_t']$ to be known. This section is concerned with the derivation of feasible statistics that share the asymptotic optimality property of $J(\Omega^*)$, but that remain asymptotically valid under much wider assumptions concerning the disturbance and its relationship to the regressors. Additionally, we motivate our implicit choice of $\Omega^* = \sigma^{-2}\Sigma_X^{1/2}\Omega\Sigma_X^{1/2}$ for the statistic $\hat{J}$ described at the end of Section 2.

We consider data generating processes for $\{\varepsilon_t\}$ and $\{Q_t = (X_t', Z_t')'\}$ of the following form.

Condition 3 *Let* $\{Q_{T,t}\}$ *and* $\{\varepsilon_{T,t}\}$ *be double arrays of* $(d+k)\times 1$ *and* 1×1 *random with elements* $Q_{T,t,i}$ *and* $\varepsilon_{T,t}$*, respectively. With some* $K<\infty$*, assume that under the null*

[3] This statement is not true for aribtrary sequences of processes that satisfy Conditon 1, though, since the convergence statements are not shown to hold uniformly over all processes that satisfy Condition 1. In fact, such a uniform convergence result does not hold.

hypothesis $h = B_Q'\varepsilon$

(i) $E[Q_{T,t}\varepsilon_{T,t}] = 0$ *for all* T, t

(ii) $\{Q_{T,t}\}$ *and* $\{\varepsilon_{T,t}\}$ *are either uniform mixing sequences of size* $-r/(r-1)$ *or strong mixing sequences of size* $-2r/(r-2)$, $r > 2$

(iii) $E[Q_{T,t}Q_{T,t}'] = \Sigma_Q$, $E[|Q_{T,t,i}\varepsilon_{T,t}|^{2r}] < K$, $T^{-1}\sum_{t=1}^{[sT]} Q_{T,t}Q_{T,t}' \xrightarrow{p} s\Sigma_Q$ *uniformly in* s *and* Σ_Q *and* $T^{-1}\sum_{t=1}^{T} Q_{T,t}Q_{T,t}'$ *are positive definite for all* T

(iv) $\{Q_{T,t}\varepsilon_{T,t}\}$ *is globally covariance stationary with nonsingular long-run covariance matrix* V_Q.

In comparison to Condition 2 of Section 3, the assumptions on the disturbances of Condition 3 are much weaker. Among the many possibilities are nonstationary, heteroskedastic and autocorrelated $\{\varepsilon_t\}$, which are allowed to be correlated with lagged values of $\{Q_t\}$. The assumptions on the regressors $\{Q_t\}$ are similar to those of Condition 2, the moment and memory conditions are strengthened to allow for a consistent estimator the long-run covariance matrix V_X of $\{X_t\varepsilon_t\}$. $\{Q_t\}$ is not required to be stationary, although only relatively mild heterogeneity of $\{Q_t\}$ is allowed under Condition 3. See Hansen (2000) for a possible approach to relaxing this assumption.

To obtain a valid test statistic under Condition 3, we will substitute the unknown quantity $\sigma^{-1}\Sigma_X^{-1/2}$ in the definition (11) of $J(\Omega^*)$ by a consistent estimator $\hat{V}_X^{-1/2}$ of $V_X^{-1/2}$, where V_X is long-run covariance matrix of $\{X_t\varepsilon_t\}$. If it is known that $\{\varepsilon_t\}$ is not autocorrelated, a natural estimator of V_X is given by the heteroskedasticity robust estimator $\hat{V}_X = T^{-1}\sum_{t=1}^{T} X_tX_t'\hat{\varepsilon}_t^2$. In the more general case of possibly autocorrelated $\{\varepsilon_t\}$, one might employ estimators of the form

$$\hat{V}_X = T^{-1}\sum_{t=1}^{T} X_tX_t'\hat{\varepsilon}_t^2 + \sum_{l=1}^{b_T} w_{T,l}T^{-1}\sum_{t=1+l}^{T}(X_tX_{t-l}' + X_{t-l}X_t')\hat{\varepsilon}_t\hat{\varepsilon}_{t-l}. \tag{13}$$

Theorem 6.21 of White (2001) establishes the consistency of $\hat{V}_X$ in (13) under Condition 3 as long as $b_T \to \infty$ as $T \to \infty$ such that $b_T = o(T^{1/4})$, and $1 \geq w_{T,l} \to 1$ for all l as $T \to \infty$.

The feasible estimator $\hat{J}(\Omega^*)$ is hence defined as

$$\begin{aligned}\hat{J}(\Omega^*) &= \sum_{i=1}^{k} \hat{v}_i'[G_{a_i} - M_e]\hat{v}_i \\ \hat{v}_i &= [I \otimes \iota_{k,i}'P^{*\prime}\hat{V}_X^{-1/2}]\Xi My\end{aligned}$$

and its asymptotic properties are investigated in the following Theorem.

Theorem 3 *If $\hat{V}_X \xrightarrow{p} V_X$ and Condition 3 holds, then*
(i) the asymptotic distribution of $\hat{J}(\Omega^*)$ under the null hypothesis of $h = B_Q'\varepsilon$ is the same as the asymptotic distribution of $J(\Omega^*)$
(ii) if in addition Conditions 1 and 2 hold, then $\hat{J}(\Omega^*) - J(\Omega^*) \xrightarrow{p} 0$ under both the null and alternative hypothesis.

The asymptotic distribution of $\hat{J}(\Omega^*)$ under the null hypothesis of parameter stability in (1) is identical for any data generating process covered by Condition 3. Tests based on this statistic are hence asymptotically valid. In addition, part (ii) of Theorem 3 implies $\hat{J}(\Omega^*)$ shares the asymptotic optimality of $J(\Omega^*)$ under Conditions 1 and 2, as derived in the last section. Asymptotically, nothing is lost by not knowing Σ_X and σ^2.

The discussion so far has concentrated on assessing the additional informational value of knowing the exact form of a Condition 1 breaking process, and in conjunction Theorem 1, Corollary 1 and Theorem 3 show that there is no asymptotic gain of such information for the hypothesis test (2). As a by-product of our analysis, we have found a feasible statistic $\hat{J}(\Omega^*)$. Theorem 3 shows this statistic to be asymptotically optimal under Conditions 1 and 2, and valid under a wide range of error distributions. These properties make $\hat{J}(\Omega^*)$ a potentially attractive choice for applied work.

The statistic $\hat{J}(\Omega^*)$ maximizes power against breaking processes of relative magnitude Ω^*. But Ω^* is typically unknown in practice. One possibility is to draw on the ideas of King (1988) and simply compute $\hat{J}(\bar{\Omega}^*)$ for some constant $\bar{\Omega}^*$, while being aware that $\bar{\Omega}^*$ and the true Ω^* will typically differ. Note that such a discrepancy does not affect the size properties of the statistic, but it will affect its power under the alternative of a time-varying $\{\beta_t\}$.

The asymptotic power of tests based on $\hat{J}(\bar{\Omega}^*)$ will be a continuous function of the true value Ω^*. Small deviations between $\bar{\Omega}^*$ and Ω^* will hence result in small losses of power only. A good choice for $\bar{\Omega}^*$ should have the property of having close to optimal power over a wide range of true values Ω^*. When X_t, β_t and Ω^* are scalars, then a choice of $\bar{\Omega}^* = \bar{a}^2$ for $\bar{a} = 10$ will have this property as we will show in the next section. As long as the eigenvalues of Ω^* are of similar magnitude, the same holds for a choice of $\bar{\Omega}^* = \bar{a}^2 I_k$ for $k > 1$. The eigenvalues of Ω^* describe the average magnitude of the normalized breaking process $\{\beta_t^* = \sigma^{-1}\Sigma_X^{1/2}\beta_t\}$ in the direction of the corresponding eigenvectors. But if, for instance, in truth there is only one moderately sized break (so that only one eigenvalue of Ω^* is positive), then a choice of $\bar{\Omega}^* = \bar{a}^2 I_k$ for $k > 1$ will have less power than a test with $\bar{\Omega}^* = \Omega^*$. This simply reflects that looking for breaks in all directions if in truth the breaks occur only in a specific direction is suboptimal. But usually neither the number nor the direction of the breaks in $\{\beta_t^*\}$ are known, so that setting $\bar{\Omega}^* = \bar{a}^2 I_k$ remains a reasonable choice.

There are alternative motivations for choosing $\bar{\Omega}^* = \bar{a}^2 I_k$, some of them outlined in Nyblom (1989). Note that only with a choice of $\bar{\Omega}^*$ proportional to I_k the outcome of tests based on $\hat{J}(\bar{\Omega}^*)$ will become invariant to the group of transformations

$$X \to X\tilde{P} \text{ for any nonsingular } k \times k \text{ matrix } \tilde{P} \tag{14}$$

A desire to be invariant to the transformations (14) may be reinterpreted as meaning that the direction of breaks under the alternative should not affect the outcome of the test. The supF test for any number of breaks, Nyblom's (1989) statistic and both tests suggested by AP are invariant to the transformations (14). And in fact, the power of these statistic and of $\hat{J}(\bar{a}^2 I_k)$ under a Condition 1 alternative depends only on the eigenvalues of Ω^*, but not on the eigenvectors P^*.

With $\bar{\Omega}^* = \bar{a}^2 I_k$, $\hat{J}(\bar{\Omega}^*)$ simplifies to

$$\begin{aligned} \hat{J}(\bar{a}^2 I_k) &= \sum_{i=1}^{k} \hat{v}_i'[G_{\bar{a}} - M_e]\hat{v}_i \\ \hat{v}_i &= [I \otimes \iota_{k,i}' \hat{V}_X^{-1/2}]\Xi M y \end{aligned}$$

and the equivalence to the statistic described at the end of Section 2 follows with $\bar{a} = 10$ after straightforward manipulations from the definition of $G_{\bar{a}}$. Table 1 contains asymptotic critical values of $\hat{J}(\bar{a}^2 I_k)$ for $\bar{a} = 10$ for $k = 1, \cdots, 10$.

5 Asymptotic and Small Sample Power

We now present evidence on the asymptotic and small sample power of the test derived in this paper as well as other tests of parameter stability. This exercise serves several purposes:

Table 1: Asymptotic Critical Values of $\hat{J}$ (reject for small values)

k	1	2	3	4	5	6	7	8	9	10
1%	-11.05	-17.57	-23.42	-29.18	-35.09	-40.24	-45.85	-51.18	-56.46	-61.77
5%	-8.36	-14.32	-19.84	-25.28	-30.60	-35.74	-40.80	-46.18	-51.10	-56.14
10%	-7.14	-12.80	-18.07	-23.37	-28.55	-33.45	-38.49	-43.59	-48.48	-53.38

Note: Percentiles reported are calculated from 40000 draws from distributions of the random variable reported in Lemma 2 with $c_i = 10$ for all i using 2000 standard normal steps to approximate Wiener Processes.

First, the results justify our choice of Ω^* for $\hat{J}$ described in Section 4. Second, we show that even for breaking processes not covered by Condition 1, the asymptotic power of $\hat{J}$ remains comparable to tailor-made optimal statistics. At least qualitatively, the analytical result of Section 3 that the exact specification of the breaking process does not matter for optimal testing hence carries over to an even larger class of breaking processes than that desribed in Condition 1. Third, we assess the relative asymptotic efficiency of $\hat{J}$ with other popular tests for parameter constancy against a variety of alternatives. Finally, we demonstrate that these asymptotic results are a very good guide to small sample behavior, at least for some standard data generating processes.

In order to compute the local asymptotic power of $J(\Omega^*)$ and $\hat{J}$ under Conditions 1 and 2, note that these statistics can be written (up to an $o_p(1)$ term) as a functional from the $k \times 1$ vector partial sum process

$$m_T(s) = T^{-1/2}\sigma^{-1/2}\Sigma_X^{-1/2}\sum_{t=1}^{[sT]} X_t\hat{\varepsilon}_t.$$

The exact form of the functional is rather complicated; Lemma 6 in the appendix implies that

$$J(\Omega^*) = \sum_{i=1}^{k}\left[-a_i\tilde{J}_i(1)^2 - a_i\int\tilde{J}_i^2 - \frac{2a_i}{1-e^{-2a_i}}\left[e^{-a_i}\tilde{J}_i(1) + a_i\int e^{-a_i s}\tilde{J}_i\right]^2 + [\tilde{J}_i(1) + a_i\int\tilde{J}_i]^2\right] + o_p(1)$$

where $\tilde{J}_i(s) = m_{T,i}(s) - \int_0^s e^{-a_i(s-\lambda)}m_{T,i}(\lambda)d\lambda$ and $m_{T,i}$ is the i^{th} element of $P^{*\prime}m_T(s)$. The expression for $\hat{J}$ is a special case, because $\hat{J} = J(\bar{a}^2 I_k) + o_p(1)$.[4]

[4] In principle, these equalities could be used as an alternative way of computing a statistic that is asymp-

As long as the regressors (X_t, Z_t) are strictly exogenous, under local alternatives $m_T(s)$ will satisfy

$$m_T(s) \Rightarrow W(s) - sW(1) + P^* \Lambda^{1/2} \int_0^s (W_\beta(\lambda) - \lambda W_\beta(1))\, d\lambda$$

where W and W_β are independent $k \times 1$ vector Wiener processes. By simulating the resulting asymptotic distributions of $J(\Omega^*)$ and $\hat{J}$, we obtained Figure 1. It depicts the asymptotic local power of $\hat{J}$ and the power envelope under Conditions 1 and 2 for strictly exogenous regressors. Here and in the following figures we consider the power of 5 percent level tests. The first panel is for $k = 1$, where $\hat{J}$ is asymptotically optimal only for the local alternative $(\Omega^*)^{1/2} = \bar{a} = 10$. But the asymptotic power of $\hat{J}$ remains extremely close to the power envelope for all other alternatives, too. The second and third panel of Figure 1 examine the $k = 2$ case. In the second panel, breaks of the same magnitude occur in both dimensions: the eigenvalues of Ω^* are equal and equal to the square of the value reported on the x-axis. The power of $\hat{J}$ is again very close to the envelope for all alternatives. This substantiates the claim of Section 4 that relying on $\hat{J} = \hat{J}(\bar{a}^2 I_k)$ with $\bar{a} = 10$ leads to only very small losses compared to the unfeasible optimal statistic $\hat{J}(\Omega^*)$, at least when the eigenvalues of Ω^* are the same. In the third panel, one of the eigenvalues of Ω^* is set to zero, such that only the other one varies. This corresponds to the case where only one component of the 2×1 vector β_t breaks, but this is not known. Then $\hat{J}$ does lose power compared to the envelope, which is the same as the envelope for the $k = 1$ case.

As second step in our analysis, we want to investigate the relative inefficiency of $\hat{J}$ when the true breaking process does not satisfy Condition 1, i.e. when the asymptotic optimality of $\hat{J}$ does not hold. Condition 1 entails that, as $T \to \infty$, there are an infinite number of breaks. One might hence expect that $\hat{J}$ does poorest when the true alternative is such that there is only a single break, even asymptotically. In order to assess the relative efficiency of $\hat{J}$ in this case, we compare it to the asymptotic power of tests that have been specifically constructed for a single break at an unknown date: the Quandt SupF statistic (Andrews, 1993) and the Andrews and Ploberger (1996) Average LM (APav) and Exponential LM (APexp) statistics

totically equivalent to $\hat{J}$. In practice, however, following the steps outlined at the end of Section 2 seems much more straightforward.

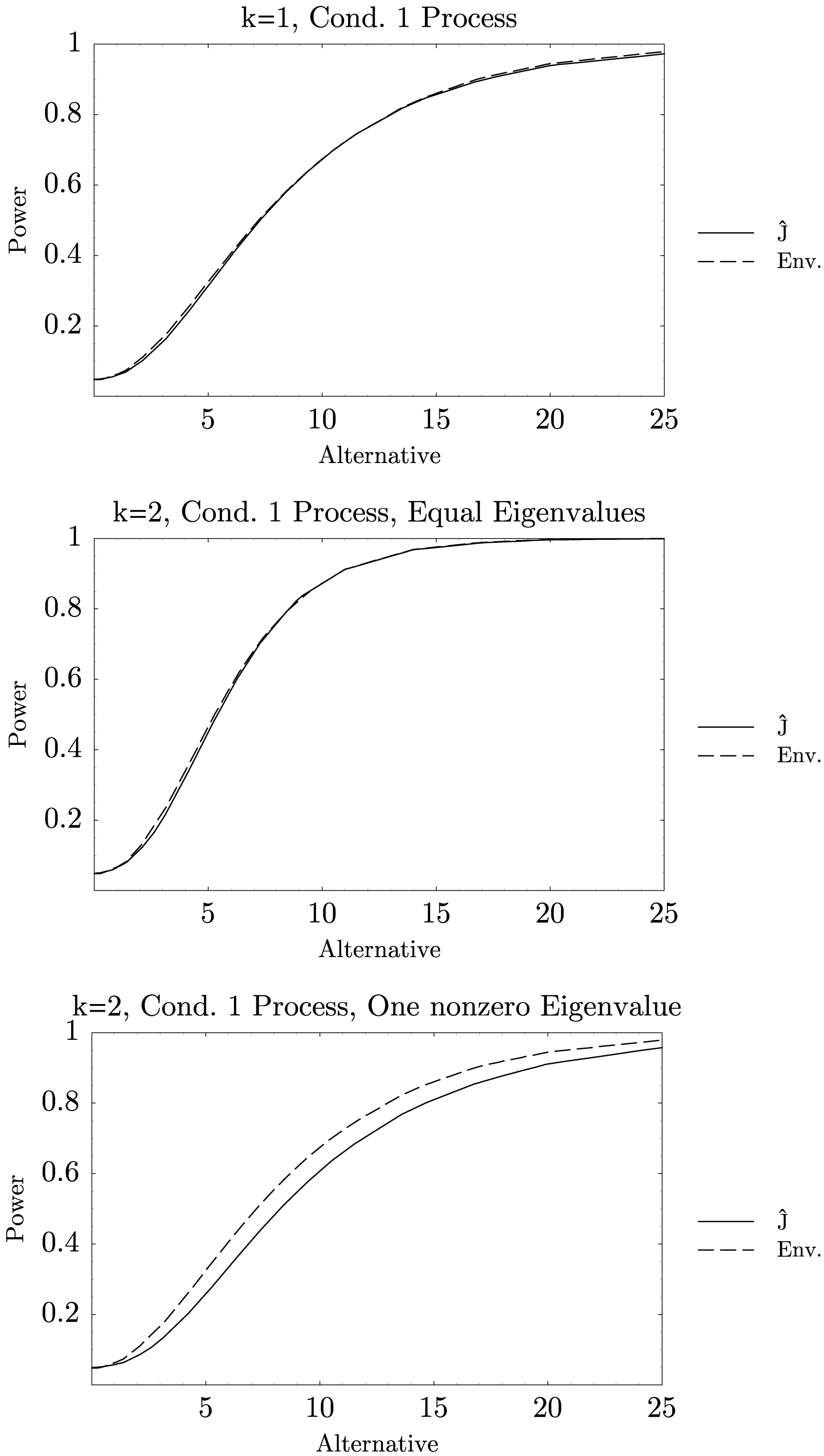


Figure 1: Asymptotic Local Power

Table 2: Test Statistics for Structural Breaks

Test Statistic	Functional
Nyblom	$\sum \int m_{T,i}(s)^2$
APav	$\int_{\lambda_0}^{1-\lambda_0} \sum \frac{m_{T,i}(s)^2}{s(1-s)} ds$
APexp	$\int_{\lambda_0}^{1-\lambda_0} \exp \left[\frac{1}{2} \sum \frac{m_{T,i}(s)^2}{s(1-s)} \right] ds$
SupF	$\sup_{\lambda_0 < s < 1-\lambda_0} \sum \frac{m_{T,i}(s)^2}{s(1-s)}$

Sums are over $i = 1, ..., k$ and λ_0 is a 'trimming' parameter we have chosen to be 0.02 for the APexp and APav, as suggested by Andrews and Ploberger (1996), and 0.05 for the supF statistic, as suggested by Bai and Perron (1998).

for independent normal disturbances. The APav and APexp statistics maximize a weighted average asymptotic power criterion against such alternatives, and also the supF statistic satisfies an asymptotic optimality property (Andrews and Ploberger, 1995). For comparison purposes, we also include the Nyblom (1989) statistic. Just like $\hat{J}$, also these four statistics can be written (up to $o_p(1)$ terms) as continuous functionals of $m_T(s)$; see Table 2.

Figure 2 compares the local asymptotic power of these statistics when under the alternative, $\beta_t = cT^{-1/2}\mathbf{1}[t \geq \pi_0 T]$, i.e. there is a single break after $100\pi_0$ percent of the sample. In panels one to three, π_0 is set to 50%, 70% and 90%, respectively, whereas in panel four of Figure 2 π_0 is uniformly drawn from $[0, 1]$.[5] Despite the fact that the supF and AP statistics have been specifically constructed for this alternative, the power of these tests is very much comparable to $\hat{J}$. The supF statistic does relatively better for a break close to the end of the sample, while the Nyblom statistic does relatively better for $\pi_0 = 50\%$. Overall, one might say that the APexp statistic does best, but the differences in power compared to $\hat{J}$ are very modest.

We can hence conclude that at least qualitatively, the result of Section 3 remains valid: using an optimal statistic for a breaking process different from the true one results in very little loss in terms of power. Because a process with a single break asymptotically is ar-

[5] Power is symmetric around $\pi_0 = 50\%$, i.e. power for $\pi_0 = 30\%$ is identical to power for $\pi_0 = 70\%$.

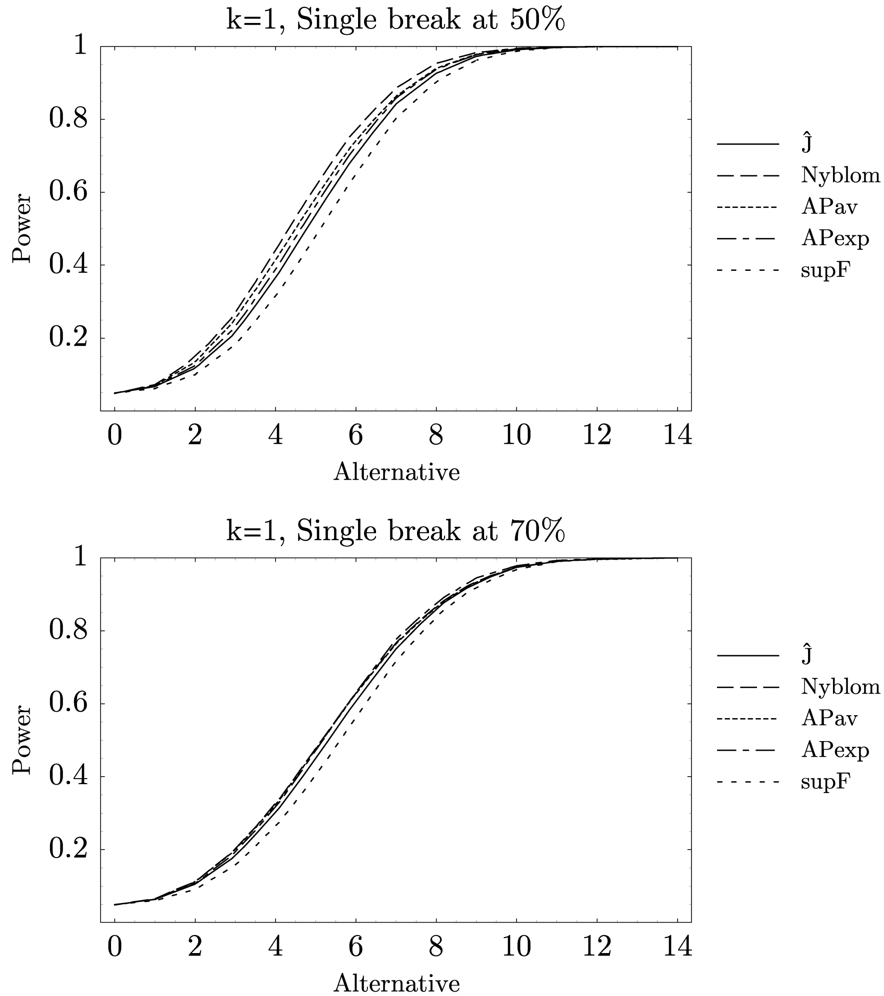
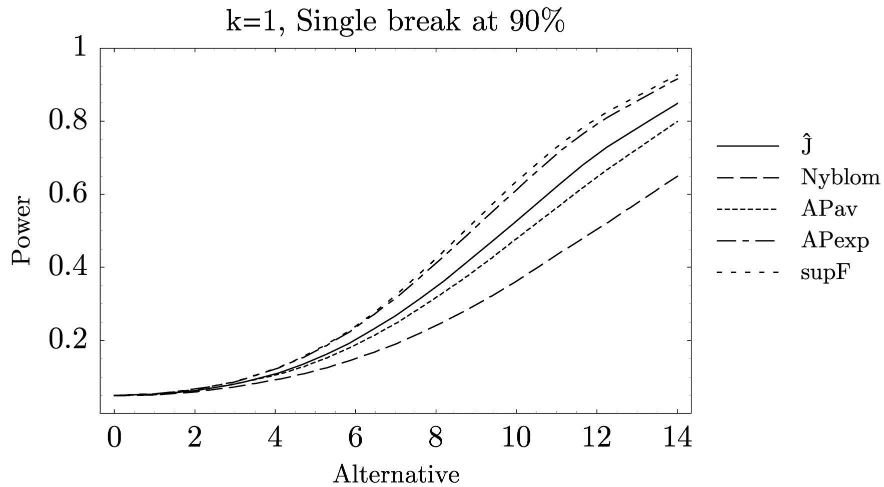


Figure 2: Asymptotic Local Power



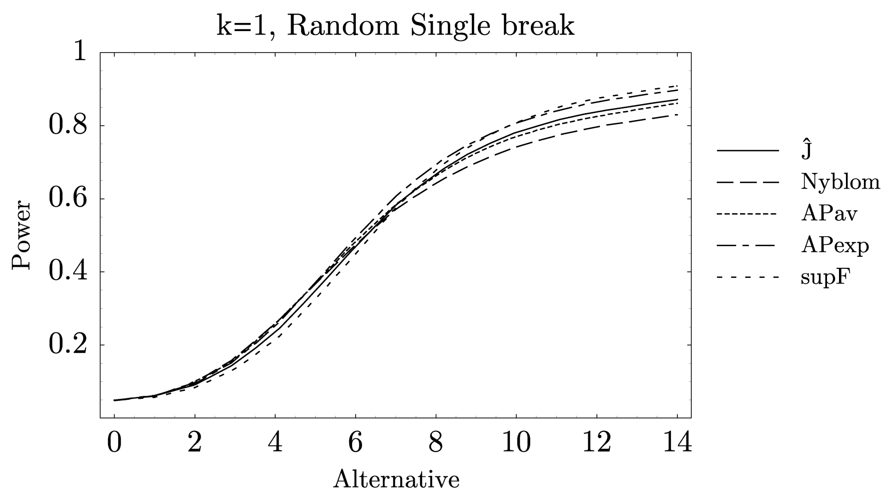


Figure 3: Asymptotic Local Power

Table 3: Overview of Small Sample Power

Fig.	Breaking Process			X_t	Z_t
	1st	2nd	3rd		
5	Break each Period	5 Expected Breaks	1 Random Break	1	–
6	1 Break at 50%	1 Break at 70%	1 Break at 90%	1	–
7	Break each Period	5 Expected Breaks	1 Random Break	$\tilde{X}_t$	1
8	1 Break at 50%	1 Break at 70%	1 Break at 90%	$\tilde{X}_t$	1
9	2-dim. Brk each Period	1-dim. Brk each Period		$(1, \tilde{X}_t)'$	–

Notes: 1st, 2nd and 3rd refers to the first, second and third panel in each figure. $\tilde{X}_t$ is independent Gaussian.

guably the most extreme deviation from a Condition 1 process in the class of all persistent processes, this suggests this qualitative result holds over a very wide range of possible breaking processes.

Turning this argument on its head implies that the statistics of Table 2 also have power against processes that satisfy Condition 1. This is examined in Figure 3 for $k = 1$ and $k = 2$. Again, we do not find large differences in performance. Given its optimality, it is not surprising that $\hat{J}$ does somewhat better than the other statistics, especially at more distant alternatives. Indeed, this property along with the way in which the power curves flatten out means that for alternatives that are not very close to the null hypothesis the Pitman efficiencies of $\hat{J}$ over the other tests are not negligible ($\hat{J}$ is around 15–30% more efficient, meaning that one needs a 15–30% larger sample size for the other tests to achieve the same power as tests based on $\hat{J}$). Of the other tests, APexp seem to be the best performer. For $k = 2$, we again consider both the case that both eigenvalues of Ω^* are equal and the case that one eigenvalue is set to zero. While the former case does lead to considerable more power, the rankings of the test are very much comparable to the results for $k = 1$.

In Figures 5–9 we examine whether the asymptotic theory presented so far can serve as a reasonable guide for small samples. To this end we investigate the properties of the same set of statistics for a sample size of $T = 100$ and iid Gaussian disturbances. We consider

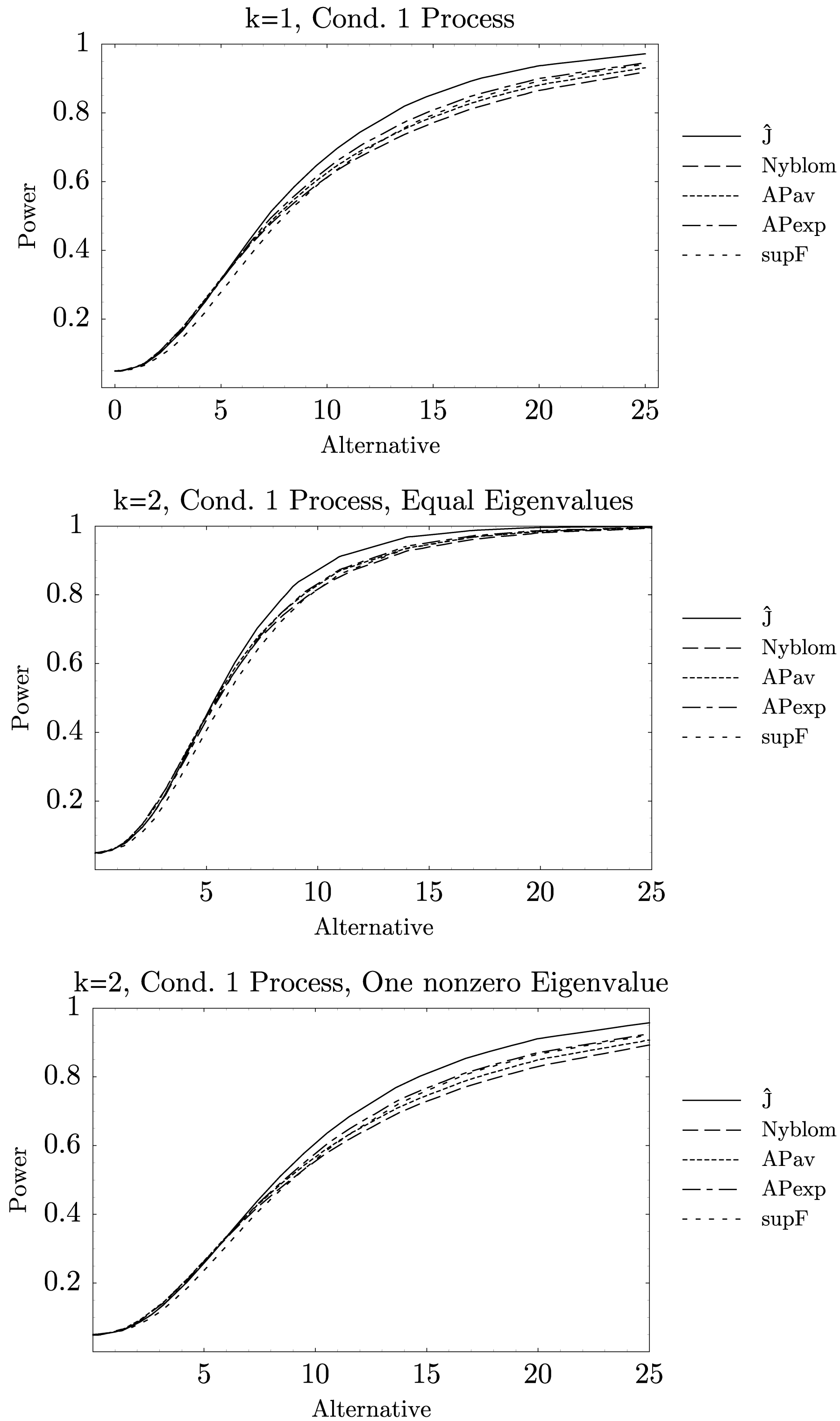


Figure 4: Asymptotic Local Power

breaking processes of a Gaussian break every period ($\beta_t = \sum_{t=1}^{T} \Delta\beta_t$, $\Delta\beta_t \sim iid\mathcal{N}(0, \Omega^*)$), of a break that may occur at any period with probability $p = 5\%$, such that the expected number of breaks in 100 observations is $pT = 5$ ($\beta_t = \sum_{t=1}^{T} \Delta\beta_t$, $\Delta\beta_t$ is iid and with probability p, $\Delta\beta_t \sim \mathcal{N}(0, \Omega_p)$ and $\Delta\beta_t = 0$ otherwise, where $\Omega^* = p\Omega_p$), and of a single break ($\beta_t = cT^{-1/2}\mathbf{1}[t \geq \pi_0 T]$). Table 3 shows the Monte Carlo designs for Figures 5–9. In the first panel of Figure 9, $(\Omega^*)^{1/2} = aT^{-1}I_2$, such that the coefficients on both regressors are subject to independent Gaussian breaks every period of the same magnitude. In the second panel of Figure 9, $(\Omega^*)^{1/2} = aT^{-1}\,\mathrm{diag}(1, 0)$, such that only the coefficient on the constant is subject to breaks every period. In all scenarios, we depict power using asymptotic critical values. Size control of all tests is very reasonable: $\hat{J}$ has size of 5.5 to 6.5 percent, the APexp statistic has size of 6 to 7 percent and the other statistics are within one percentage point of the nominal level of 5 percent. Figures of size–adjusted power look almost identical and are omitted. All scenarios map qualitatively and to a large extent even quantitatively very closely to the predictions of the asymptotic theory.

6 Conclusions

Parameter instability of a permanent nature is interesting economically, causes problems for forecasting and can invalidate inference in linear regression models. This has led researchers to construct many different tests for the stability of regression parameters, almost all specific to a particular breaking process under the alternative. Intuition suggests that reasonable tests for a specific breaking process should have some power also against other breaking processes. An optimal test for a break every other period, for instance, will have power also against an alternative with a break every period. We show not only that this intuition is correct, but a much stronger claim: The optimal test for a break every other period will do just as well as the optimal test for breaks every period when in fact there is a break every period, at least for a large enough sample size. This (asymptotic) equivalence extends over a very large class of breaking processes.

The result has three implications. First, the exact breaking process under the alternative is usually unknown to the applied researcher. But since power is similar over a wide range of

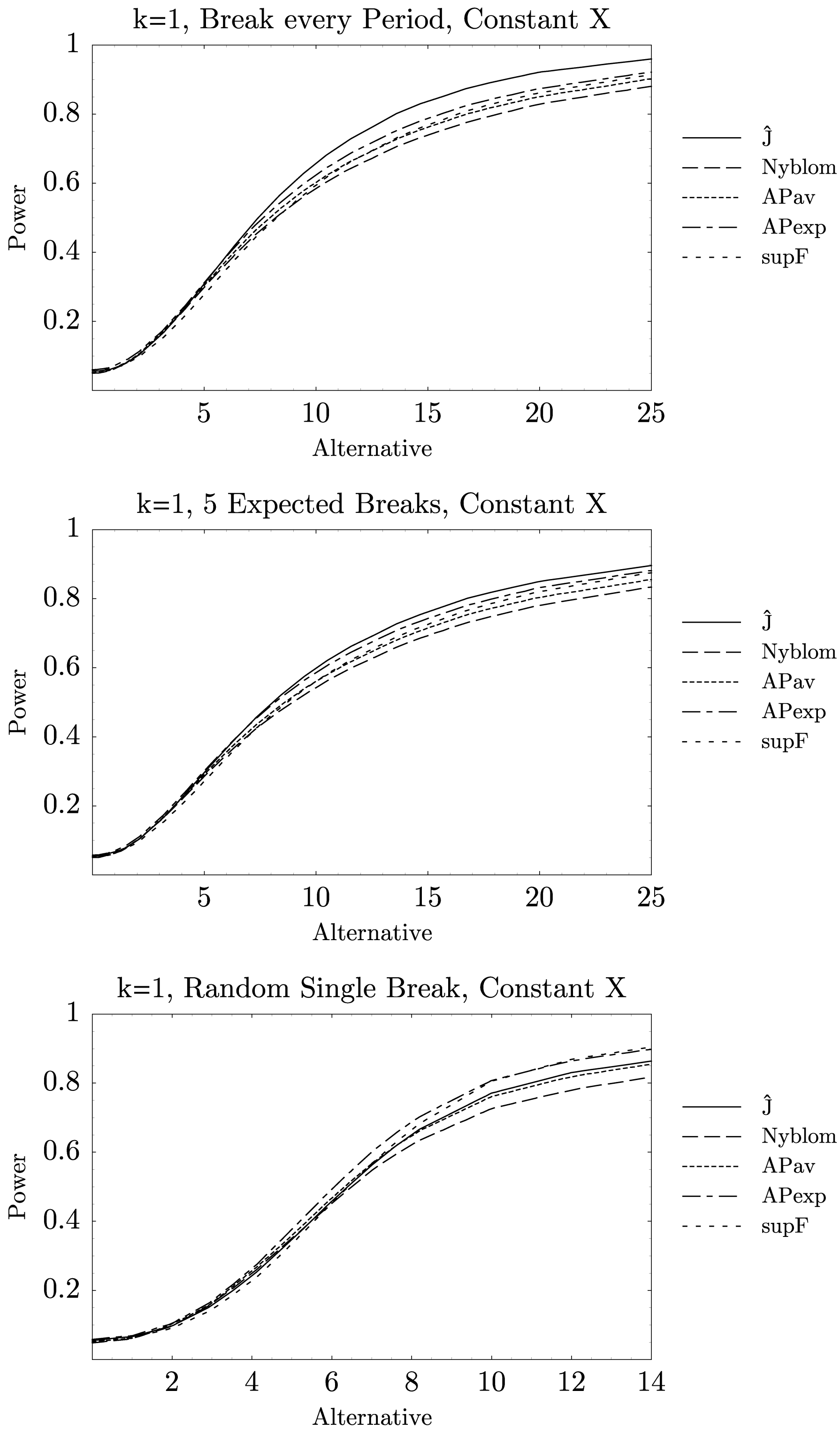


Figure 5: Small Sample Power

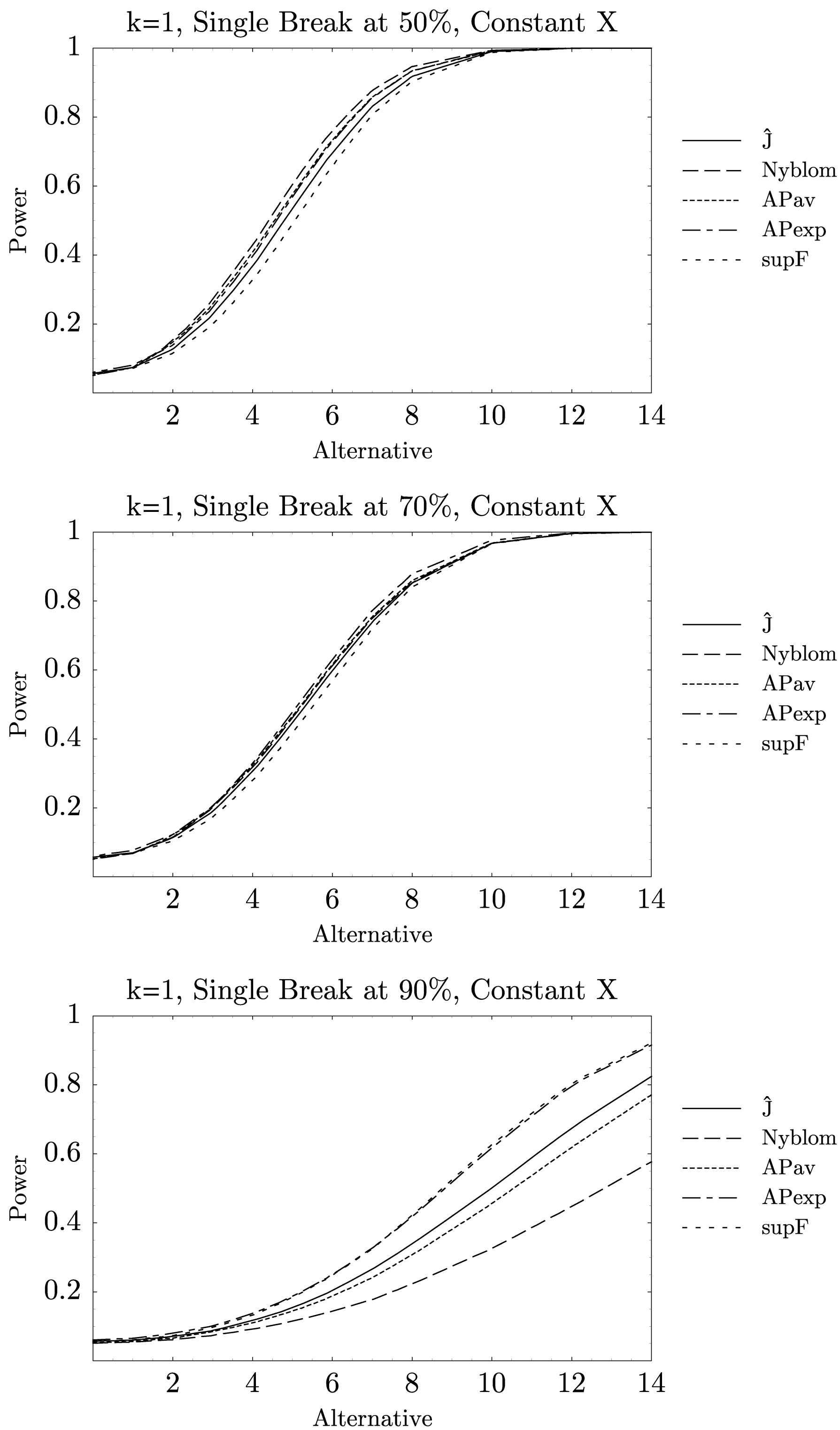


Figure 6: Small Sample Power

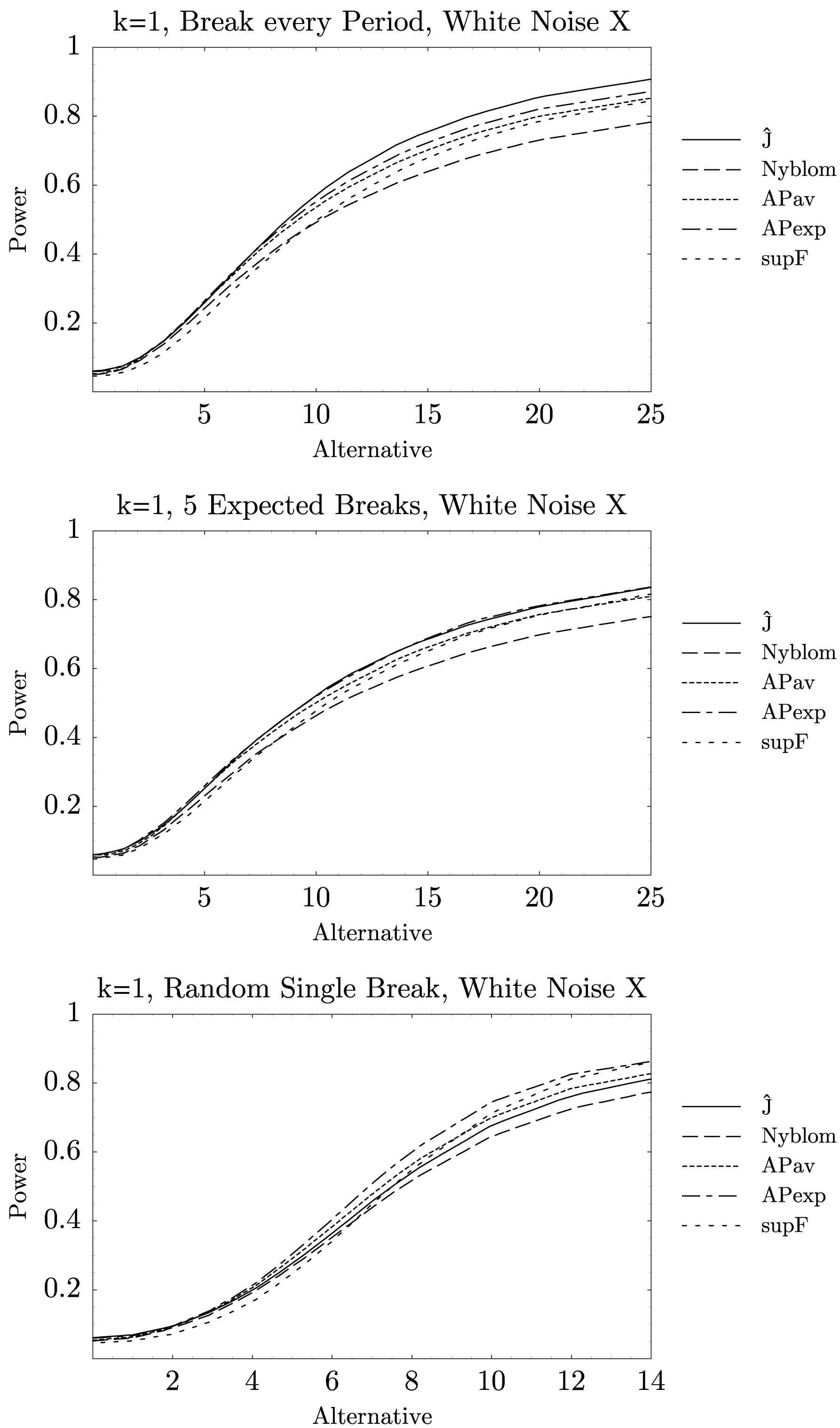


Figure 7: Small Sample Power

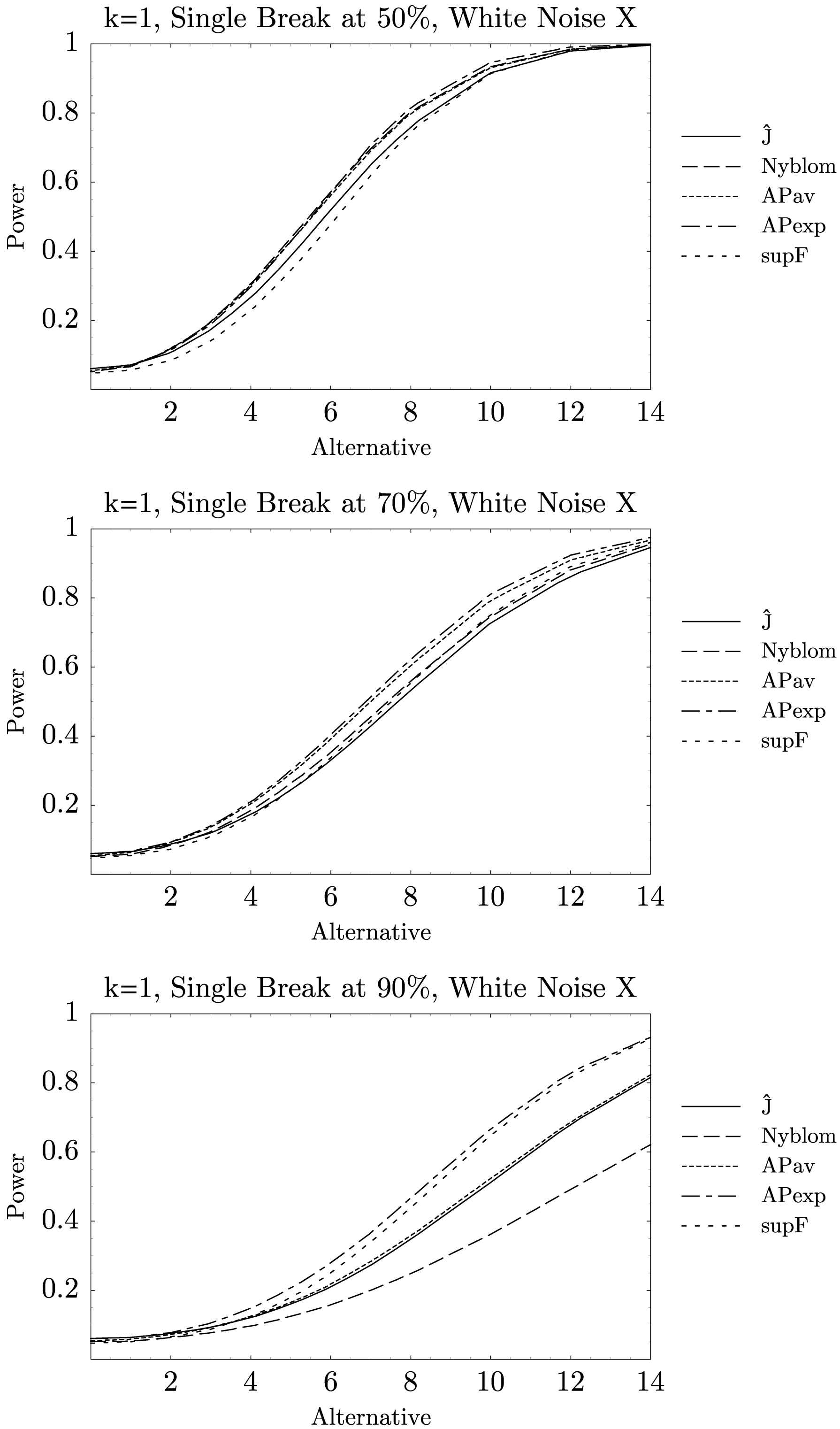
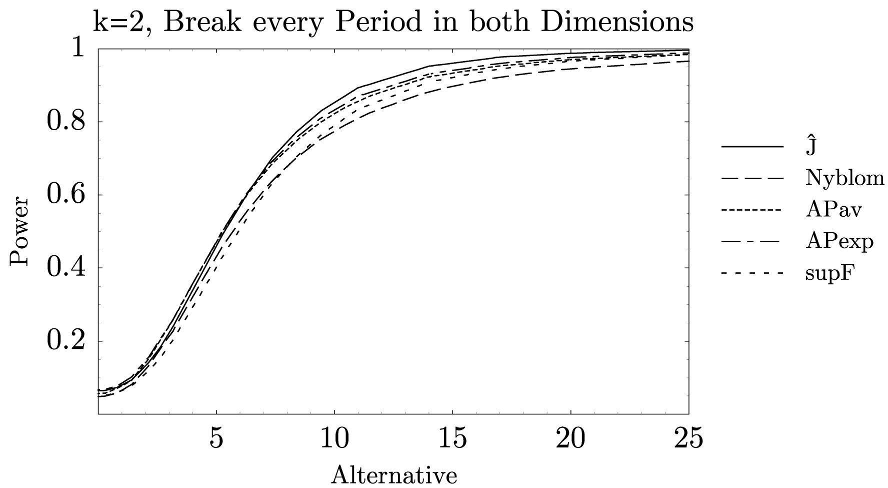


Figure 8: Small Sample Power



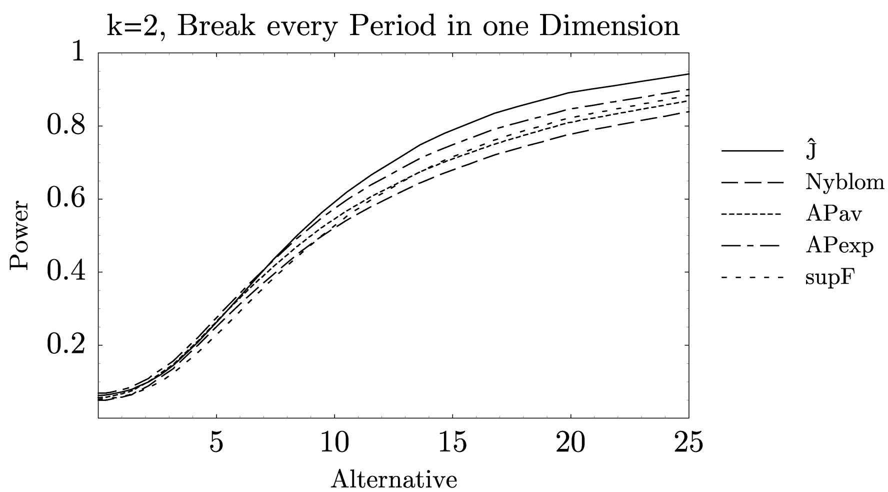


Figure 9: Small Sample Power

breaking processes for any reasonable test statistic, this ignorance does not matter for being able to conduct a powerful test. The applied researcher is hence relieved of having to make stark choices about the assumed breaking process under the alternative.

Second, under local alternatives standard tests for parameter instability contain very little information about the exact form of the breaking process. This is simply the flip side of all tests behaving roughly the same, no matter how the breaking process precisely looks. If a test that has been designed against the alternative of a single break rejects, say, then this does by no means imply that the true breaking process consists in fact of a single break. While for nonlocal alternatives, i.e. for breaks that are large asymptotically, methods have been developed to discern the number and location of breaks (Bai and Perron (1998)), distinguishing local breaking processes requires a different approach.

Third, complicated tailor-made tests will not result in significant gains in power over any other reasonable statistic. This considerably simplifies the practice of testing parameter stability, because tailor-made tests have nonstandard distributions (so that one needs a set of critical values for each special case) and many of them are very difficult to compute. Our results suggest that one can choose any specific breaking process for which the optimal statistic has a simple form. Very little power will be foregone by basing inference on this simple statistic even if it is known that the true breaking process under the alternative is not of the form the simple statistic has been constructed for.

We suggest such an easy-to-compute statistic that has an asymptotic optimality property for the class of breaking processes we focus on. Unsurprisingly, given its optimality, the statistic has superior asymptotic power against alternatives that fall into this class. But its relative efficiency extends to other natural breaking processes not in the class, making the statistic an appealing choice for applied work.

7 Appendix

Many subsequent results are easier to obtain by working with regressors having identity covariance matrix. To this end, let C be the $(k+d)\times(k+d)$ matrix with $(\Sigma_X^{-1/2}, 0_{k\times d})$ in its upper $k\times(d+k)$ block that satisfies $C\Sigma_Q C' = I_{k+d}$. Denote $Q^* = QC'$, $X^* = X\Sigma_X^{-1/2}$, let Ξ^* be defined just as Ξ with $X_t^* = \Sigma_X^{-1/2}X_t$ replacing X_t, and let $\varepsilon_t^* = \sigma^{-1}\varepsilon_t$, $\beta^* = [I\otimes\sigma^{-1}\Sigma_X^{1/2}]\beta$ and $\tilde{\beta}^* = [I\otimes\sigma^{-1}\Sigma_X^{1/2}]\tilde{\beta}$.

Note that the long-run variance of $\{T\Delta\beta_t^*\}$ is given by $\sigma^{-2}\Sigma_X^{1/2}\Omega\Sigma_X^{1/2} = \Omega^* = P^*\Lambda P^{*\prime}$.

Further define B_e to be the $(T-1)\times T$ matrix that satisfies $B_e' B_e = I_{T-1}$ and $B_e' e = 0$, so that $B_e B_e' = M_e$. Also let $L = F^{-1}$.

We proceed by establishing several Lemmas that are needed in preparation for the proofs of the Lemmas and Theorems in the main text.

Lemma 4 *(i)*

$$B_e'(a^2T^{-2}FF' + I)B_e = B_e' H_a B_e$$

(ii)

$$B_e(B_e' H_a B_e)^{-1} B_e' = G_a$$

(iii)

$$|B_e' H_a B_e| = \frac{1 - r_a^{2T}}{T(1 - r_a^2) r_a^{T-1}}$$

Proof. (i)

$$\begin{aligned} B_e'(a^2T^{-2}FF' + I)B_e &= B_e'(a^2T^{-2}FF' + I + (1 - r_a)ee')B_e \\ &= B_e' F(LL' + a^2T^{-2}I + (1 - r_a)\iota_{T,1}\iota_{T,1}')F' B_e \end{aligned}$$

Now from a direct calculation

$$\begin{aligned} & LL' + a^2T^{-2}I + (1 - r_a)\iota_{T,1}\iota_{T,1}' \\ = & \begin{pmatrix} a^2T^{-2} + 2 - r_a & -1 & 0 & \cdots & 0 & 0 \\ -1 & a^2T^{-2} + 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & a^2T^{-2} + 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a^2T^{-2} + 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & a^2T^{-2} + 2 \end{pmatrix} \end{aligned}$$

and

$$r_a^{-1} A_a A_a' = \begin{pmatrix} r_a^{-1} & -1 & 0 & \cdots & 0 & 0 \\ -1 & r_a^{-1} + r_a & -1 & \cdots & 0 & 0 \\ 0 & -1 & r_a^{-1} + r_a & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & r_a^{-1} + r_a & -1 \\ 0 & 0 & 0 & \cdots & -1 & r_a^{-1} + r_a \end{pmatrix}$$

so that

$$\begin{aligned} B_e'(a^2T^{-2}FF' + I)B_e &= B_e' F(r_a^{-1} A_a A_a')F' B_e \\ &= B_e' H_a B_e \end{aligned}$$

(ii) see Rao (1973), p. 77.

(iii) From (ii)

$$\begin{aligned} B_e(B_e' H_a B_e)^{-1} B_e' &= G_a \\ (B_e' H_a B_e)^{-1} &= B_e' G_a B_e \end{aligned}$$

yielding $|B_e' H_a B_e| = |B_e' G_a B_e|^{-1}$. Now note that $(T^{-1/2}e, B_e)'(T^{-1/2}e, B_e) = I$, so that

$$
\begin{aligned}
|H_a^{-1}| &= \left| \begin{pmatrix} T^{-1/2}e & B_e \end{pmatrix}' H_a^{-1} \begin{pmatrix} T^{-1/2}e & B_e \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} T^{-1}e'H_a^{-1}e & T^{-1/2}e'H_a^{-1}B_e \\ T^{-1/2}B_e'H_a^{-1}e & B_e'H_a^{-1}B_e \end{pmatrix} \right| \\
&= |T^{-1}e'H_a^{-1}e||B_e'H_a^{-1}B_e - B_e'H_a^{-1}e(e'H_a^{-1}e)^{-1}e'H_a^{-1}B_e| \\
&= |T^{-1}e'H_a^{-1}e||B_e'G_aB_e| \\
&= |T^{-1}e'H_a^{-1}e||B_e'H_aB_e|^{-1}
\end{aligned}
$$

But $|H_a| = |r_a^{-1}FA_aA_a'F'| = r_a^{-T}$ and

$$
\begin{aligned}
|e'H_a^{-1}e| &= r_a \iota_{T,1}' A_a'^{-1} A_a^{-1} \iota_{T,1} \\
&= r_a \sum_{j=0}^{T-1} r_a^{2j} = r_a \frac{1-r_a^{2T}}{1-r_a^2}
\end{aligned}
$$

and we find

$$
|B_e'H_aB_e| = \frac{1-r_a^{2T}}{T(1-r_a^2)r_a^{T-1}}
$$

■

Proof of Lemma 1:

Let $\tilde{\beta}_e = [B_e' \otimes I_k]\tilde{\beta}$ and $\nu_{\tilde{\beta}_e}$ its measure, and let $K_a = T^{-2}a^2B_e'FF'B_e$, $K_\Omega = T^{-2}B_e'FF'B_e \otimes \Omega$ and $K_\Lambda = T^{-2}B_e'FF'B_e \otimes \Lambda$. Recall that $\Omega = \sigma^2\Sigma_X^{-1/2}P^*\Lambda P^{*\prime}\Sigma_X^{-1/2}$. We compute

$$
\begin{aligned}
\widetilde{LR}_T &= \int \exp\left[\sigma^{-2}h'B_Q'\Xi[M_e \otimes I_k]\tilde{\beta} - \tfrac{1}{2}\sigma^{-2}\tilde{\beta}'[M_e \otimes \Sigma_X]\tilde{\beta}\right] d\nu_{\tilde{\beta}} \\
&= \int \exp\left[\sigma^{-2}h'B_Q'\Xi[B_e \otimes I_k]\tilde{\beta}_e - \tfrac{1}{2}\tilde{\beta}_e'[I_{T-1} \otimes \sigma^{-2}\Sigma_X]\tilde{\beta}_e\right] d\nu_{\tilde{\beta}_e} \\
&= \int (2\pi)^{-k(T-1)/2}|K_\Omega|^{-1/2} \exp\left[\sigma^{-2}h'B_Q'\Xi[B_e \otimes I_k]\tilde{\beta}_e - \tfrac{1}{2}\tilde{\beta}_e'[K_\Omega^{-1} + I_{T-1} \otimes \sigma^{-2}\Sigma_X]\tilde{\beta}_e\right] d\tilde{\beta}_e \\
&= |K_\Omega|^{-1/2}|K_\Omega^{-1} + I_{T-1} \otimes \sigma^{-2}\Sigma_X|^{-1/2} \\
&\quad \exp\left[\tfrac{1}{2}\sigma^{-4}h'B_Q'\Xi[B_e \otimes I_k][K_\Omega^{-1} + I_{T-1} \otimes \sigma^{-2}\Sigma_X]^{-1}[B_e' \otimes I_k]\Xi'B_Qh\right] \\
&= |I_{T-1} \otimes I_k + K_\Lambda|^{-1/2} \\
&\quad \exp\left[\tfrac{1}{2}\sigma^{-2}h'B_Q'\Xi[B_e \otimes \Sigma_X^{-1/2}P^*][K_\Lambda^{-1} + I_{T-1} \otimes I_k]^{-1}[B_e' \otimes P^{*\prime}\Sigma_X^{-1/2}]\Xi'B_Qh\right]
\end{aligned}
$$

Now

$$
\begin{aligned}
[K_\Lambda^{-1} + I_{T-1} \otimes I_k]^{-1} &= K_\Lambda \left[K_\Lambda + I_{T-1} \otimes I_k\right]^{-1} \\
&= I_{T-1} \otimes I_k - \left[K_\Lambda + I_{T-1} \otimes I_k\right]^{-1}
\end{aligned}
$$

and

$$
\begin{aligned}
[K_\Lambda + I_{T-1} \otimes I_k]^{-1} &= \left[\sum_{i=1}^{k} K_{a_i} \otimes (\iota_{k,i}\iota_{k,i}') + I_{T-1} \otimes I_k \right]^{-1} \\
&= \left[\sum_{i=1}^{k} (K_{a_i} + I_{T-1}) \otimes (\iota_{k,i}\iota_{k,i}') \right]^{-1} \\
&= \sum_{i=1}^{k} [K_{a_i} + I_{T-1}]^{-1} \otimes (\iota_{k,i}\iota_{k,i}')
\end{aligned}
$$

so that

$$
\begin{aligned}
& [B_e \otimes I_k][K_\Lambda^{-1} + I_{T-1} \otimes I_k]^{-1}[B_e' \otimes I_k] \\
= & \sum_{i=1}^{k} B_e(I_{T-1} - [K_{a_i} + I_{T-1}]^{-1})B_e' \otimes (\iota_{k,i}\iota_{k,i}') \\
= & \sum_{i=1}^{k} (M_e - G_{a_i}) \otimes (\iota_{k,i}\iota_{k,i}')
\end{aligned}
$$

where the last line relies on Lemma 4 above. Furthermore, again relying on Lemma 4, we find

$$
\begin{aligned}
|K_\Lambda + I_{T-1} \otimes I_k| &= \prod_{i=1}^{k} |a_i^2 T^{-2} B_e' F F' B_e + I_{T-1}| \\
&= \prod_{i=1}^{k} |B_e'(a_i^2 T^{-2} F F' + I) B_e| = \prod_{i=1}^{k} \left[\frac{1 - r_{a_i}^{2T}}{T(1 - r_{a_i}^2) r_{a_i}^{T-1}} \right]
\end{aligned}
$$

Therefore

$$
\widetilde{LR}_T = \prod_{i=1}^{k} \left[\frac{1 - r_{a_i}^{2T}}{T(1 - r_{a_i}^2) r_{a_i}^{T-1}} \right]^{-1/2} \exp\left[-\tfrac{1}{2} v_i'[G_{a_i} - M_e] v_i \right]
$$

with $v_i = [I \otimes \iota_{k,i}' P^{*\prime} \sigma^{-1} \Sigma_X^{-1/2}] \Xi M y$.

Lemma 5 *Let $\{Q_t\}$ satisfy Condition 2 and assume that $\{\upsilon_t\}$ is independent of $\{Q_t\}$ and satisfies $\upsilon_{[Ts]} \Rightarrow A_\upsilon W_\upsilon(s)$, where A_υ is a $(d+k) \times (d+k)$ nonstochastic, possibly singular matrix and W_υ is a $(d+k) \times 1$ Wiener process. Then*
(i) $T^{-1} \sum_{t=1}^{T} (Q_t^* Q_t^{*\prime} - I_{k+d}) \upsilon_t \xrightarrow{p} 0$
(ii) $T^{-1} \sum_{t=1}^{T} (Q_t^* Q_t^{*\prime} - I_{k+d}) \upsilon_t \upsilon_t' \xrightarrow{p} 0$

Proof. (i) We will show convergence in probability of

$$
T^{-1} \sum (Q_{t,i}^* Q_{t,j}^* - \delta_{i,j}) \upsilon_{t,j}
$$

for any $i, j \in \{1, \cdots, d+k\}$, where $\delta_{i,j} = 1$ if $i = j$ and zero otherwise and here and in the following computations sums are taken from $t = 1$ to T if not stated otherwise. The proof relies on a truncation argument with respect to $\upsilon_{t,j}$. For all t and T, define $\tilde{\upsilon}_{t,j} = \upsilon_{t,j}$ if $|\upsilon_{t,j}| < K_\upsilon$ and $\tilde{\upsilon}_{t,j} = 0$ otherwise. Then

$$\begin{aligned} P[\exists t &: \tilde{\upsilon}_{t,j} \neq \upsilon_{t,j}] = P[\max_t |\upsilon_{t,j}| > K_\upsilon] \\ &\to P[\sup_s |A_{\upsilon,j} W_\upsilon(s)| > K_\upsilon] \end{aligned} \tag{15}$$

where $A_{\upsilon,j}$ is the jth row of A_υ and the last line follows from the CMT and the definition of weak convergence.

We will first show that $(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}$ is a L_1 adapted mixingale with respect to the σ-field $\mathfrak{F}^*_t$ generated by $\{Q_t, Q_{t-1}, \cdots, \upsilon_T, \upsilon_{T-1}, \cdots\}$. Apart from the presence of $\tilde{\upsilon}_{t,j}$, the reasoning is similar to Example 16.4 of Davidson (1994), p. 249. From $E[Q^*_{T,t}Q^{*\prime}_{T,t}] = I_{k+d}$ and the independence of $\{Q^*_t\}$ from $\{\upsilon_t\}$, $E[(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}] = 0$. Since $\{Q^*_{t,i}\}$ and $\{Q^*_{t,j}\}$ are L_r-bounded and $|\tilde{\upsilon}_{t,j}| \leq K_\upsilon$, $\{(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}\}$ is $L_{r/2}$-bounded

$$\begin{aligned} \|(Q^*_{t,i}Q^*_{t,j} - I_{k+d})\tilde{\upsilon}_{t,j}\|_{r/2} &\leq K_\upsilon(\|Q^*_{t,i}\|_r \|Q^*_{t,j}\|_r + \delta_{i,j}) \\ &\leq K'_\upsilon + \delta_{i,j} \end{aligned}$$

for some $K'_\upsilon < \infty$ for all t and T, where the first inequality follows from the Cauchy-Schwarz inequality. Furthermore, because $r > 2$, this implies that $\{(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}\}$ is uniformly integrable. For the L_1 mixingale property, we need to bound $|E[(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}|\mathfrak{F}^*_{t-m}]|$. Note that the mixing properties of $\{Q_t\}$ extend to $\{(CQ_tQ'_tC - \delta_{i,j})\}$ with C defined at the beginning of the appendix (cf. Theorem 3.49 of White, 2001).

Now under strong mixing, Theorem 14.2 of Davidson (1994) is applicable and we find

$$\begin{aligned} |E[(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}|\mathfrak{F}^*_{t-m}]| &\leq K_\upsilon |E[(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})|\mathfrak{F}^*_{t-m}]| \\ &\leq 6K_\upsilon \alpha_m^{1-2/r} \|(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\|_{r/2} \\ &\leq 6K_\upsilon \alpha_m^{1-2/r}(K'_\upsilon + \delta_{i,j}) \end{aligned}$$

with α_m the m^{th} strong mixing coefficient. Since $\alpha_m = O(m^{-r/(r-2)-\epsilon})$ for some $\epsilon > 0$, we find that $\alpha_m^{1-2/r} = O(m^{-1-\epsilon'})$ for some $\epsilon' > 0$, so that under strong mixing, $\{(Q^*_tQ^{*\prime}_t - I_{k+d})\tilde{\upsilon}_t, \mathfrak{F}^*_t\}$ is a L_1 mixingale of size -1 (with constants that do not depend on t).

Under uniform mixing, we can apply Theorem 14.4 of Davidson (1994) to find

$$|E[(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}|\mathfrak{F}^*_{t-m}]| \leq 2K_\upsilon \phi_m^{1-2/r}(K'_\upsilon + \delta_{i,j})$$

with ϕ_m the m^{th} uniform mixing coefficient. Since $\phi_m = O(m^{-r/(2r-2)-\epsilon})$ for some $\epsilon > 0$, we find $\phi_m^{1-2/r} = O(m^{-1/2-\epsilon'})$ for some $\epsilon' > 0$, so that $\{(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}, \mathfrak{F}^*_t\}$ becomes a L_1 mixingale of size $-1/2$ with constants that do not depend on t when $\{Q_t\}$ is uniform mixing.

But Theorem 19.11 of Davidson (1994), p. 302, shows that the mean of a uniformly integrable L_1 mixingale of any size with respect to constants that do not depend on t converges to zero in the L_1-norm. Since convergence in L_1 implies convergence in probability, for any $\epsilon, \eta > 0$ there exists a T^* such that for all $T > T^*$

$$P\left[|T^{-1}\sum(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}| > \epsilon\right] < \eta$$

Furthermore, from (15) there exists a T^{**} such that for all $T > T^{**}$

$$P[\exists t : \tilde{\upsilon}_{t,j} \neq \upsilon_{t,j}] \leq P[\sup_s |A_{\upsilon,j}W_\upsilon(s)| > K_\upsilon] + \eta$$

Therefore, for all $T > T^* \vee T^{**}$, we find

$$\begin{aligned}
& P\left[|T^{-1}\sum(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\upsilon_{t,j}| > \epsilon\right] \\
= \; & P\left[|T^{-1}\sum(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}| > \epsilon|\tilde{\upsilon}_t = \upsilon_t \forall t\right](1 - P[\exists t : \tilde{\upsilon}_{t,j} \neq \upsilon_{t,j}]) \\
& \quad + P\left[|T^{-1}\sum(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\upsilon_{t,j}| > \epsilon|\exists t : \tilde{\upsilon}_{t,j} \neq \upsilon_{t,j}\right] P[\exists t : \tilde{\upsilon}_{t,j} \neq \upsilon_{t,j}] \\
\leq \; & P\left[|T^{-1}\sum(Q^*_{t,i}Q^*_{t,j} - \delta_{i,j})\tilde{\upsilon}_{t,j}| > \epsilon\right] + P[\exists t : \tilde{\upsilon}_{t,j} \neq \upsilon_{t,j}] \\
\leq \; & 2\eta + P[\sup_s |A_{\upsilon,j}W_\upsilon(s)| > K_\upsilon]
\end{aligned}$$

By choosing K_υ large, $P[\sup_s |A_{\upsilon,j}W_\upsilon(s)| > K_\upsilon]$ can be made arbitrarily small, and η was arbitrary, which concludes the proof.

(ii) The proof is analogous to part (i), the only difference is that now the j, lth element of $\upsilon_t\upsilon_t'$ is truncated. The probability of such of a truncation taking place is then

$$P[\max_t |[\upsilon_t\upsilon_t']_{j,l}| > K_\upsilon] \to P[\sup_s |[A_\upsilon W_\upsilon(s)W_\upsilon(s)'A_\upsilon']_{j,l}| > K_\upsilon]$$

which can also be made arbitrarily small by choosing K_υ large. ■

Lemma 6 *Let $B_u(s)$ be a stochastic process on the unit interval that has bounded uniformly continuous sample paths with probability one. Let $u = (u_1, \cdots, u_T)'$ be such that $T^{-1/2}\sum_{t=1}^{[T\cdot]} u_t \Rightarrow B_u(\cdot)$. Then*

$$u'[G_c - M_e]u = g_J(T^{-1/2}\sum_{t=1}^{[T\cdot]} u_t) + o_p(1)$$

where g_J is

$$g_J\left(T^{-1/2}\sum_{t=1}^{[T\cdot]} u_t\right) = -cJ_u(1)^2 - c^2\int J_u^2 - \frac{2c}{1-e^{-2c}}\left[e^{-c}J_u(1) + c\int e^{-cs}J_u\right]^2 + [J_u(1) + c\int J_u]^2$$

and $J_u(s) = T^{-1/2}\sum_{t=1}^{[Ts]} u_t - \int_0^s e^{-c(s-\lambda)}\left(T^{-1/2}\sum_{t=1}^{[T\lambda]} u_t\right)d\lambda$. Furthermore, $g_J(a(s) + \kappa s) = g_J(a(s))$ for any κ.

Proof. Write $u'[G_c - M_e]u = u'(H_c^{-1} - I)u - u'H_c^{-1}e(e'H_c^{-1}e)^{-1}e'H_c^{-1}u + (T^{-1/2}e'u)^2$. Define $B = A_c^{-1}u$, so that the t^{th} element of B satisfies $B_t = \sum_{s=1}^t r_c^{t-s}u_s$, and let $B_{-1} = (0, B_1, \cdots, B_{T-1})'$. Also note that $A_c^{-1}LA_c = L$. Then

$$\begin{aligned}
u'(H_c^{-1} - I)u &= u'(r_cL'A_c^{-1\prime}A_c^{-1}L - I)u \\
&= r_cB'L'LB - u'u \\
&= r_c(u + (r_c - 1)B_{-1})'(u + (r_c - 1)B_{-1}) - u'u \\
&= (r_c - 1)u'u + r_c(r_c - 1)^2B_{-1}'B_{-1} + 2r_c(r_c - 1)B_{-1}'u
\end{aligned}$$

Now from $u + r_cB_{-1} = B$, we find $u'u + 2r_cB_{-1}'u + r_c^2B_{-1}'B_{-1} = B'B$, yielding

$$B_{-1}'u = (2r_c)^{-1}\left[B_T^2 + (1 - r_c^2)B_{-1}'B_{-1} - u'u\right]$$

so after rearranging we have

$$u'(H_c^{-1} - I)u = (r_c - 1)B_T^2 - (1 - r_c)^2 B_{-1}' B_{-1}.$$

By direct calculation $T^{-1}e'H_c^{-1}e = r_c \frac{1-r_c^{2T}}{T(1-r_c^2)} = \frac{1-e^{-2c}}{2c} + o(1)$. Also

$$\begin{aligned} T^{-1/2}e'H_c^{-1}u &= r_c T^{-1/2} e' L' A_c^{-1\prime} A_c^{-1} L A_c A_c^{-1} u \\ &= r_c T^{-1/2} \iota_{T,1}' A_c^{-1\prime} L B \\ &= T(1-r_c)T^{-3/2} \sum_{t=1}^{T-1} r_c^t B_t + r_c^T T^{-1/2} B_T \end{aligned}$$

For the final term $T^{-1/2}e'u = T^{-1/2}e'A_c B = T^{-1/2}B_T + T(1-r_c)T^{-3/2}e'B_{-1}$. The first claim of the Lemma now follows after noting that $T^{-1/2}B_T = J_u(1) + o_p(1)$, $T^{-2}\sum_{t=1}^T B_{t-1}^2 = \int J_u^2 + o_p(1)$, $T^{-1}\sum_{t=1}^T r_c^{t-1} B_{t-1} = \int e^{-cs} J_u + o_p(1)$, $T^{-1}\sum_{t=1}^T B_{t-1} = \int J_u + o_p(1)$, and $r_c^T = e^{-c} + o(1)$.

For the second part of the Lemma, simply note that $G_c - M_e = M_e(G_c - M_e)M_e$, such that $u'[G_c - M_e]u$ is invariant to transformations of u of the form $u \to u + \kappa e$, which implies the claimed invariance of g_J. ■

Proof of Lemma 2:

We have

$$\sum_{i=1}^k v_i'[G_{c_i} - M_e]v_i = \sigma^{-2}\varepsilon' M \Xi [I \otimes \Sigma_X^{-1/2} P^*] \left[\sum_{i=1}^k (G_{c_i} - M_e) \otimes (\iota_{k,i}\iota_{k,i}') \right] [I \otimes P^{*\prime}\Sigma_X^{-1/2}] \Xi' M \varepsilon$$

as defined in the main text. Noting that $\{Q_t^* \varepsilon_t^*\}$ is a mixing sequence with the same mixing coefficient as $\{Q_t\}$, we find that the sum of the first $[sT]$ $k \times 1$ vectors of $\sigma^{-1}[M_e \otimes P^{*\prime}\Sigma_X^{-1/2}]\Xi M \varepsilon$ satisfies

$$\begin{aligned} & T^{-1/2}\left[(e_{[sT]}', 0_{T-[sT]}') I \otimes I_k\right] [I \otimes P^{*\prime}\Sigma_X^{-1/2}] \Xi' M \varepsilon^* \\ = \; & T^{-1/2} P^{*\prime} \left(\sum_{t=1}^{[sT]} X_t^* \varepsilon_t^* \right) - P^{*\prime} \left(T^{-1} \sum_{t=1}^{[sT]} X_t^* Q_t^{*\prime} \right) (T^{-1} Q^{*\prime} Q^*)^{-1} T^{-1/2} \sum_{t=1}^T Q_t^* \varepsilon_t^* \\ \Rightarrow \; & P^{*\prime} P^* W_\varepsilon(s) - s P^{*\prime} P^* W_\varepsilon(1) = W_\varepsilon(s) - s W_\varepsilon(1) \end{aligned}$$

from a FCLT for mixing sequences as in White (2001), p. 189 and $\left(T^{-1}\sum_{t=1}^{[sT]} X_t^* Q_t^{*\prime} - sT^{-1}\sum_{t=1}^T X_t^* Q_t^{*\prime} \right) \xrightarrow{p} 0$ from the uniform convergence of $T^{-1}\sum_{t=1}^{[sT]} Q_t Q_t' \xrightarrow{p} s\Sigma_Q$ in s. We hence have

$$\begin{aligned} T^{-1/2}(e_{[sT]}', 0_{T-[sT]}') v_i &= T^{-1/2}(e_{[sT]}', 0_{T-[sT]}')[I \otimes \iota_{k,i}' P^{*\prime} \Sigma_X^{-1/2}] \Xi' M \varepsilon \\ &= \iota_{k,i}' T^{-1/2}(e_{[sT]}', 0_{T-[sT]}')[I \otimes P^{*\prime}\Sigma_X^{-1/2}] \Xi' M \varepsilon \\ &\Rightarrow W_{\varepsilon,i}(s) - s W_{\varepsilon,i}(1) \end{aligned}$$

An application of Lemma 6 to each v_i (with $B_u = W_{\varepsilon,i}$) now yields the result.

Proof of Lemma 3:

We rely on a weak convergence result for mixing sequences as described in Theorem 7.45 of White (2001), p. 201 for the following computations concerning the weak convergence of

$\{Q_t\varepsilon_t, T\Delta\beta_t, T\Delta\gamma_t, T\Delta\tilde{\beta}_t, T\Delta\tilde{\gamma}_t\}$. Furthermore, we make repeated use of parts (i) and (ii) of Lemma 5 above.

(i) We treat β only, since the identical distribution of γ obviously leads to the same result. Now

$$\begin{aligned}\sigma^{-2}\varepsilon' M\Xi\beta &= \varepsilon^{*\prime} M\Xi^*\beta^* \\ &= \varepsilon^{*\prime}\Xi^*\beta^* - \varepsilon^{*\prime}Q^*(Q^{*\prime}Q^*)^{-1}Q^{*\prime}\Xi^*\beta^*\end{aligned}$$

The long-run variance of $\{Q_t^*\varepsilon_t^*\}$ is given by $E[Q_t^*Q_t^{*\prime}(\varepsilon_t^*)^2] = \sigma^{-2}E[CQ_tQ_t'C'\varepsilon_t^2] = I_{k+d}$, and $E[|Q_t^*\varepsilon_t^*|] = E[|Q_t^*|]E[|\varepsilon_t^*|] < \infty$ uniformly in T from the moment restriction on Q_t. Hence

$$\begin{aligned}\sum \beta_t^{*\prime}X_t^*\varepsilon_t^* &= tr\left[P^*\Lambda^{1/2}\left(\sum \Lambda^{-1/2}P^{*\prime}\beta_t^*X_t^{*\prime}\varepsilon_t^*\right)\right] \\ &\Rightarrow tr\left[P^*\Lambda^{1/2}\int W_\beta d\tilde{W}_\varepsilon'\right] \\ &= \int W_\beta'\Lambda^{1/2}dW_\varepsilon\end{aligned}$$

where $\tilde{W}_\varepsilon$ is a $k\times 1$ Wiener process and $W_\varepsilon = P^*\tilde{W}_\varepsilon$. Note that since P^* is orthonormal, W_ε is a Wiener process, too. Furthermore

$$T^{-1/2}\sum Q_t^*\varepsilon_t^* \Rightarrow \begin{pmatrix} P^*W_\varepsilon(1) \\ W_{Z\varepsilon}(1)\end{pmatrix}$$

where $W_{Z\varepsilon}$ is a $d\times 1$ Wiener process independent of W_ε, and from Lemma 5 above

$$T^{-1/2}\sum Q_t^*X_t^{*\prime}\beta_t^* = T^{-1/2}\sum Q_t^*Q_t^{*\prime}\begin{pmatrix}\beta_t^* \\ 0\end{pmatrix} \Rightarrow \begin{pmatrix} P^*\Lambda^{1/2}\int W_\beta \\ 0\end{pmatrix}$$

so that

$$\begin{aligned}T^{-1/2}\left(\sum Q_t^*\varepsilon_t^*\right)'(T^{-1}Q^{*\prime}Q^*)^{-1}T^{-1/2}\sum Q_t^*X_t^{*\prime}\beta_t^* &\Rightarrow \begin{pmatrix} P^*W_\varepsilon(1) \\ W_{Z\varepsilon}(1)\end{pmatrix}'\begin{pmatrix} P^*\Lambda^{1/2}\int W_\beta \\ 0\end{pmatrix} \\ &= (\textstyle\int W_\beta)'\Lambda^{1/2}W_\varepsilon(1)\end{aligned}$$

We hence find

$$\sigma^{-2}\varepsilon' M\Xi\beta \Rightarrow \int W_\beta'\Lambda^{1/2}dW_\varepsilon - (\textstyle\int W_\beta)'\Lambda^{1/2}W_\varepsilon(1) = \int \bar{W}_\beta'\Lambda^{1/2}dW_\varepsilon.$$

(ii)

$$\begin{aligned}\sigma^{-2}\beta'\Xi' M\Xi\beta &= \beta^{*\prime}\Xi^{*\prime}M\Xi^*\beta^* \\ &= \beta^{*\prime}\Xi^{*\prime}\Xi^*\beta^* - \beta^{*\prime}\Xi^{*\prime}Q^*(Q^{*\prime}Q^*)^{-1}Q^{*\prime}\Xi^*\beta^*\end{aligned}$$

Now

$$\begin{aligned}\beta^{*\prime}\Xi^{*\prime}\Xi^*\beta^* &= \sum \beta_t^{*\prime}X_t^*X_t^{*\prime}\beta_t^* \\ &= tr\left[\sum X_t^*X_t^{*\prime}\beta_t^*\beta_t^{*\prime}\right] \\ &\Rightarrow tr\left[P^*\Lambda^{1/2}\left(\textstyle\int W_\beta W_\beta'\right)\Lambda^{1/2}P^{*\prime}\right] = \int W_\beta'\Lambda W_\beta\end{aligned}$$

using part (ii) of Lemma 5, and

$$T^{-1/2}Q^{*\prime}\Xi^*\beta^* \Rightarrow \begin{pmatrix} P^*\Lambda^{1/2}\int W_\beta \\ 0\end{pmatrix}$$

as in part (i), so that

$$\sigma^{-2}\beta'\Xi' M\Xi\beta \Rightarrow \int W_\beta'\Lambda W_\beta - (\int W_\beta)'\Lambda(\int W_\beta) = \int \bar{W}_\beta'\Lambda\bar{W}_\beta$$

(iii) We treat $\tilde{\beta}$, which implies the result for $\tilde{\gamma}$ since $\tilde{\beta}$ and $\tilde{\gamma}$ have the same distribution.

$$\begin{aligned}
\sigma^{-2}\varepsilon' M\Xi[M_e\otimes I_k]\tilde{\beta} &= \varepsilon^{*\prime}M\Xi^*[M_e\otimes I_k]\tilde{\beta}^* \\
&= \varepsilon^{*\prime}M\Xi^*\tilde{\beta}^* - \varepsilon^{*\prime}M\Xi^*[e(e'e)^{-1}e'\otimes I_k]\tilde{\beta}^* \\
&= \varepsilon^{*\prime}M\Xi^*\tilde{\beta}^* - \varepsilon^{*\prime}\Xi^*[e(e'e)^{-1}e'\otimes I_k]\tilde{\beta}^* + \varepsilon^{*\prime}Q^*(Q^{*\prime}Q^*)^{-1}Q^{*\prime}\Xi^*[e(e'e)^{-1}e'\otimes I_k]\tilde{\beta}^*
\end{aligned}$$

The first term will converge to $\varepsilon^{*\prime}M\Xi^*\tilde{\beta}^* \Rightarrow \int \bar{W}_{\tilde{\beta}}'\Lambda^{1/2}dW_\varepsilon$ from the same reasoning as in part (i) of the proof. For the remaining terms, we find

$$\begin{aligned}
\varepsilon^{*\prime}\Xi^*[e(e'e)^{-1}e'\otimes I_k]\tilde{\beta}^* &= \left(T^{-1/2}\sum X_t^*\varepsilon_t^*\right)'\left(T^{-1/2}\sum \tilde{\beta}_t^*\right) \\
&\Rightarrow W_\varepsilon(1)'P^*P^{*\prime}\Lambda^{1/2}\int W_{\tilde{\beta}} \\
T^{-1/2}Q^{*\prime}\Xi^*[e(e'e)^{-1}e'\otimes I_k]\tilde{\beta}^* &= \left(T^{-1}\sum Q_t^*X_t^{*\prime}\right)\left(T^{-1/2}\sum \tilde{\beta}_t^*\right) \Rightarrow \begin{pmatrix} P^*\Lambda^{1/2}\int W_{\tilde{\beta}} \\ 0 \end{pmatrix} \\
T^{1/2}\varepsilon^{*\prime}Q^*(Q^{*\prime}Q^*)^{-1} &= \left(T^{-1/2}\sum Q_t^*\varepsilon_t^*\right)'(T^{-1}Q^{*\prime}Q^*)^{-1} \Rightarrow \begin{pmatrix} P^*W_\varepsilon(1) \\ W_{Z\varepsilon}(1) \end{pmatrix}'
\end{aligned}$$

yielding the result.

(iv)

$$\begin{aligned}
\sigma^{-2}\tilde{\beta}'[M_e\otimes\Sigma_X]\tilde{\beta} &= \tilde{\beta}^{*\prime}[M_e\otimes I_k]\tilde{\beta}^* \\
&= \sum \tilde{\beta}_t^{*\prime}\tilde{\beta}_t^* - \tilde{\beta}^{*\prime}[e(e'e)^{-1}e'\otimes I_k]\tilde{\beta}^*
\end{aligned}$$

where

$$\sum \tilde{\beta}_t^{*\prime}\tilde{\beta}_t^* = tr\left[\sum \tilde{\beta}_t^*\tilde{\beta}_t^{*\prime}\right] \Rightarrow \int W_{\tilde{\beta}}'\Lambda W_{\tilde{\beta}}$$

and

$$\begin{aligned}
\tilde{\beta}^{*\prime}[e(e'e)^{-1}e'\otimes I_k]\tilde{\beta}^* &= T^{-1}\tilde{\beta}^{*\prime}[(e\otimes I_k)(e'\otimes I_k)]\tilde{\beta}^* \\
&= \left(T^{-1/2}\sum \tilde{\beta}_t^{*\prime}\right)\left(T^{-1/2}\sum \tilde{\beta}_t^*\right) \\
&\Rightarrow (\int W_{\tilde{\beta}})'\Lambda(\int W_{\tilde{\beta}})
\end{aligned}$$

so that

$$\sigma^{-2}\tilde{\beta}'[M_e\otimes\Sigma_X]\tilde{\beta} \Rightarrow \int \bar{W}_{\tilde{\beta}}'\Lambda\bar{W}_{\tilde{\beta}}$$

The joint convergence is an immediate consequence of the independence of β, $\tilde{\beta}$, γ and $\tilde{\gamma}$.

Proof of Theorem 1:

All computations in the proof are made under H_0, i.e. under the assumption of $h = B_Q'\varepsilon$. Specifically, ν_h will denote the measure associated with $h = B_Q'\varepsilon$. Furthermore, let $d\nu_{\hat{Q}}$ denote integrations with respect to the measure of the conditional distributions $\hat{Q}_t = Q_t|\{Q_{t-1}, Q_{t-2}, \cdots, y_{t-1}, y_{t-2}, ...\}$, i.e. a shortcut for $\prod_{t=1}^T f_{\hat{Q},t}d\bar{\nu}_{\hat{Q},t}$.

Let $\phi = \sum_{i=1}^k v_i'[G_{\sqrt{2}a_i} - M_e]v_i$ and recall that $LR_T = \nu_\beta \xi(\beta)$ and $\widetilde{LR}_T = \nu_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta})$, where $\nu_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta})$ is short-hand notation for $\int \tilde{\xi}(\tilde{\beta})d\nu_{\tilde{\beta}}$ and so forth. Further define

$$\begin{aligned}
\widetilde{LR}_T^0 &= \mathbf{1}[\phi \le K_0]\nu_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta}) \\
\widetilde{LR}_T^0(K') &= \mathbf{1}[\phi \le K_0]\nu_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta})\mathbf{1}[\tilde{\xi}(\tilde{\beta}) \le K'] \\
LR_T^0 &= \mathbf{1}[\phi \le K_0]\nu_\beta \xi(\beta) \\
LR_T^0(K') &= \mathbf{1}[\phi \le K_0]\nu_\beta \xi(\beta)\mathbf{1}[\xi(\beta) \le K']
\end{aligned}$$

Note that

$$\begin{aligned}
P(|LR_T - \widetilde{LR}_T| > 5\epsilon) &\le P(|LR_T - LR_T^0| > \epsilon) + P(|LR_T^0 - LR_T^0(K')| > \epsilon) \\
&+ P(|LR_T^0(K') - \widetilde{LR}_T^0(K')| > \epsilon) + P(|\widetilde{LR}_T^0 - \widetilde{LR}_T^0(K')| > \epsilon) + P(|\widetilde{LR}_T - \widetilde{LR}_T^0| > \epsilon)
\end{aligned}$$

We hence need to show: (i) for any $\eta > 0$ there exists T^*, K' and K_0 such that for all $T \ge T^*$, $P(|LR_T - LR_T^0| > \epsilon) < \eta$, $P(|LR_T^0 - LR_T^0(K')| > \epsilon) < \eta$, $P(|\widetilde{LR}_T^0 - \widetilde{LR}_T^0(K')| > \epsilon) < \eta$ and $P(|\widetilde{LR}_T - \widetilde{LR}_T^0| > \epsilon) < \eta$ and (ii) for all K' and K_0, $LR_T^0(K') - \widetilde{LR}_T^0(K') \xrightarrow{p} 0$.

We show (i) first. Under H_0

$$\begin{aligned}
P(|\widetilde{LR}_T - \widetilde{LR}_T^0| &> \epsilon) \le P(\phi > K_0) \\
P(|LR_T - LR_T^0| &> \epsilon) \le P(\phi > K_0)
\end{aligned}$$

But by Lemma 2

$$\phi \Rightarrow \sum_{i=1}^k \left[-c_i J_i(1)^2 - c_i^2 \int J_i^2 - \frac{2c_i}{1-e^{-2c_i}} \left[e^{-c_i} J_i(1) + c_i \int e^{-c_i s} J_i \right]^2 + [J_i(1) + c_i \int J_i]^2 \right]$$

where $c_i = \sqrt{2}a_i$, so that by choosing K_0 large enough, $P(\phi > K_0)$ can be made arbitrarily small for sufficiently large T.

To show that $P(|\widetilde{LR}_T^0 - \widetilde{LR}_T^0(K')| > \epsilon)$ can be made arbitrarily small by choosing K' large, we will show that $\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta})$ is uniformly integrable. By the definition of uniform integrability, this implies that $E[\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta})\mathbf{1}[\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta}) > K']]$ can be made arbitrarily small uniformly over T by choosing K' large. Since

$$\begin{aligned}
P(|\widetilde{LR}_T^0 - \widetilde{LR}_T^0(K')| &> \epsilon) \le \epsilon^{-1} E|\widetilde{LR}_T^0 - \widetilde{LR}_T^0(K')| \\
&= \epsilon^{-1} E[\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta})\mathbf{1}[\tilde{\xi}(\tilde{\beta}) > K']] \\
&= \epsilon^{-1} E[\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta})\mathbf{1}[\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta}) > K']]
\end{aligned}$$

this is sufficient for the claim. For the uniform integrability of $\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta})$, we will show that the second moment of $\mathbf{1}[\phi \le K_0]\tilde{\xi}(\tilde{\beta})$ is bounded uniformly over T (the 'Crystal Ball' condition for uniform integrability). Letting $\tilde{\beta}_e$, H_Ω and H_Λ as in the proof of Lemma 1, we find

$$\begin{aligned}
E[(\mathbf{1}[\phi \leq K_0]\tilde{\xi}(\tilde{\beta}))^2] &= E[\mathbf{1}[\phi \leq K_0](\tilde{\xi}(\tilde{\beta}))^2] \\
&= \int\int\int \mathbf{1}[\phi \leq K_0] \exp\left[2\sigma^{-2}h'B_Q'\Xi[M_e \otimes I_k]\tilde{\beta} - \sigma^{-2}\tilde{\beta}'[M_e \otimes \Sigma_X]\tilde{\beta}\right] d\nu_{\tilde{\beta}} d\nu_h d\nu_{\hat{Q}} \\
&= \int\int \mathbf{1}[\phi \leq K_0] \int \exp\left[2\sigma^{-2}h'B_Q'\Xi[B_e \otimes I_k]\tilde{\beta}_e - \tilde{\beta}_e'[I_{T-1} \otimes \sigma^{-2}\Sigma_X]\tilde{\beta}_e\right] d\nu_{\tilde{\beta}_e} d\nu_h d\nu_{\hat{Q}} \\
&= \int\int \mathbf{1}[\phi \leq K_0] \int (2\pi)^{-k(T-1)/2}|K_\Omega|^{-1/2} \\
&\quad \exp\left[2\sigma^{-2}h'B_Q'\Xi[B_e \otimes I_k]\tilde{\beta}_e - \tfrac{1}{2}\tilde{\beta}_e'[K_\Omega^{-1} + 2I_{T-1} \otimes \sigma^{-2}\Sigma_X]\tilde{\beta}_e\right] d\tilde{\beta}_e d\nu_h d\nu_{\hat{Q}} \\
&= \int\int \mathbf{1}[\phi \leq K_0]|K_\Omega|^{-1/2}|K_\Omega^{-1} + 2I_{T-1} \otimes \sigma^{-2}\Sigma_X|^{-1/2} \\
&\quad \exp\left[2\sigma^{-4}h'B_Q'\Xi[B_e \otimes I_k][K_\Omega^{-1} + 2I_{T-1} \otimes \sigma^{-2}\Sigma_X]^{-1}[B_e' \otimes I_k]\Xi'B_Q h\right] d\nu_h d\nu_{\hat{Q}} \\
&= \int\int \mathbf{1}[\phi \leq K_0]|I_{T-1} \otimes I_k + K_{2\Lambda}|^{-1/2} \\
&\quad \exp\left[\sigma^{-2}h'B_Q'\Xi[B_e \otimes \Sigma_X^{-1/2}P^*][K_{2\Lambda}^{-1} + I_{T-1} \otimes I_k]^{-1}[B_e' \otimes P^{*\prime}\Sigma_X^{-1/2}]\Xi'B_Q h\right] d\nu_h d\nu_{\hat{Q}} \\
&= \int\int \mathbf{1}[\phi \leq K_0]\,[\exp\phi]\, d\nu_h d\nu_{\hat{Q}} \prod_{i=1}^{k}\left[\frac{1 - r_{\sqrt{2}a_i}^{2T}}{T(1 - r_{\sqrt{2}a_i}^2)r_{\sqrt{2}a_i}^{T-1}}\right]^{-1/2} \\
&\leq \exp K_0 \prod_{i=1}^{k}\left[\frac{1 - r_{\sqrt{2}a_i}^{2T}}{T(1 - r_{\sqrt{2}a_i}^2)r_{\sqrt{2}a_i}^{T-1}}\right]^{-1/2}
\end{aligned}$$

the computation of the integral follows closely the computations in Lemma 1. From the last line, we conclude $\sup_T E[(\mathbf{1}[\phi \leq K_0]\tilde{\xi}(\tilde{\beta}))^2] < \infty$, so that the 'Crystal Ball' condition holds.

The final piece for (i) is the term $P(|LR_T^0 - LR_T^0(K')| > \epsilon)$. We have

$$
\begin{aligned}
P(|LR_T^0 - LR_T^0(K')| &> \epsilon) \leq \epsilon^{-1} E|LR_T^0 - LR_T^0(K')| \\
&= \epsilon^{-1} \int \int |LR_T^0 - LR_T^0(K')| d\nu_h d\nu_{\hat{Q}} \\
&= \epsilon^{-1} \int \int \int \mathbf{1}[\phi \leq K_0] \mathbf{1}[\xi(\beta) > K'] \xi(\beta) d\nu_\beta d\nu_h d\nu_{\hat{Q}} \\
&\leq \epsilon^{-1} \int \int \int \mathbf{1}[\xi(\beta) > K'] \xi(\beta) d\nu_h d\nu_\beta d\nu_{\hat{Q}} \\
&= \epsilon^{-1} \int \int \int (2\pi\sigma^2)^{-(T-k-d)/2} \mathbf{1}[\xi(\beta) > K'] \\
&\quad \exp\left[-\tfrac{1}{2}\sigma^{-2}(h'h - 2h'B_Q'\Xi\beta + \beta'\Xi'M\Xi\beta)\right] dh d\nu_\beta d\nu_{\hat{Q}} \\
&= \epsilon^{-1} \int \int \int (2\pi\sigma^2)^{-(T-k-d)/2} \mathbf{1}[\sigma^{-2}h'B_Q'\Xi\beta - \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta > \ln K'] \\
&\quad \exp\left[-\tfrac{1}{2}\sigma^{-2}(h - B_Q'\Xi\beta)'(h - B_Q'\Xi\beta)\right] dh d\nu_\beta d\nu_{\hat{Q}} \\
&= \epsilon^{-1} \int \int \int (2\pi\sigma^2)^{-(T-k-d)/2} \mathbf{1}[\sigma^{-2}h'B_Q'\Xi\beta + \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta > \ln K'] \\
&\quad \exp\left[-\tfrac{1}{2}\sigma^{-2}h'h\right] dh d\nu_\beta d\nu_{\hat{Q}} \\
&= \epsilon^{-1} \int \int \int \mathbf{1}[\sigma^{-2}h'B_Q'\Xi\beta + \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta > \ln K'] d\nu_h d\nu_\beta d\nu_{\hat{Q}} \\
&= \epsilon^{-1} P(\sigma^{-2}h'B_Q'\Xi\beta + \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta > \ln K') \\
&= \epsilon^{-1} P(\sigma^{-2}\varepsilon'M\Xi\beta + \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta > \ln K')
\end{aligned}
$$

But

$$
\sigma^{-2}\varepsilon'M\Xi\beta + \tfrac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta \Rightarrow \int \bar{W}_\beta'\Lambda^{1/2} dW_\varepsilon + \tfrac{1}{2} \int \bar{W}_\beta'\Lambda\bar{W}_\beta
$$

from Lemma 3, so that by making K' sufficiently large, $P(\sigma^{-2}\varepsilon'M\Xi\beta + \frac{1}{2}\sigma^{-2}\beta'\Xi'M\Xi\beta > \ln K')$ can be made arbitrarily small for sufficiently large T.

We are hence left to show (ii), i.e. that $LR_T^0(K') - \widetilde{LR}_T^0(K') \xrightarrow{p} 0$ for all $0 < K' < \infty$ and $0 < K_0 < \infty$. Introducing the notation $\tilde{\psi}(\tilde{\beta}) = \tilde{\xi}(\tilde{\beta})\mathbf{1}[\tilde{\xi}(\tilde{\beta}) \leq K']$ and $\psi(\beta) = \xi(\beta)\mathbf{1}[\xi(\beta) \leq K']$ we find

$$
\begin{aligned}
LR_T^0(K') &= \mathbf{1}[\phi \leq K_0] \nu_\beta \psi(\beta) \\
&= \mathbf{1}[\phi \leq K_0] \nu_\gamma \psi(\gamma)
\end{aligned}
$$

and

$$
\begin{aligned}
\widetilde{LR}_T^0(K') &= \mathbf{1}[\phi \leq K_0] \nu_{\tilde{\beta}} \tilde{\psi}(\tilde{\beta}) \\
&= \mathbf{1}[\phi \leq K_0] \nu_{\tilde{\gamma}} \tilde{\psi}(\tilde{\gamma})
\end{aligned}
$$

so that we can write

$$
\begin{aligned}
E\left[\left(LR_T^0(K') - \widetilde{LR}_T^0(K')\right)^2\right] &= \nu_{\hat{Q}}\nu_h \left(LR_T^0(K') - \widetilde{LR}_T^0(K')\right)^2 \\
&= \nu_{\hat{Q}}\nu_h[\mathbf{1}[\phi \le K_0][\nu_\beta\psi(\beta)\nu_\gamma\psi(\gamma) - \nu_\beta\psi(\beta)\nu_{\tilde{\gamma}}\tilde{\psi}(\tilde{\gamma}) \\
&\quad - \nu_{\tilde{\beta}}\tilde{\psi}(\tilde{\beta})\nu_\gamma\psi(\gamma) + \nu_{\tilde{\beta}}\tilde{\psi}(\tilde{\beta})\nu_{\tilde{\gamma}}\tilde{\psi}(\tilde{\gamma})]] \\
&= E[\mathbf{1}[\phi \le K_0]\psi(\beta)\psi(\gamma)] - E[\mathbf{1}[\phi \le K_0]\psi(\beta)\tilde{\psi}(\tilde{\gamma})] \\
&\quad -E[\mathbf{1}[\phi \le K_0]\tilde{\psi}(\tilde{\beta})\psi(\gamma)] + E[\mathbf{1}[\phi \le K_0]\tilde{\psi}(\tilde{\beta})\tilde{\psi}(\tilde{\gamma})]
\end{aligned}
$$

Now Lemmas 2 and 3 and the CMT imply that $\mathbf{1}[\phi \le K_0]\psi(\beta)\psi(\gamma)$, $\mathbf{1}[\phi \le K_0]\psi(\beta)\tilde{\psi}(\tilde{\gamma})$, $\mathbf{1}[\phi \le K_0]\tilde{\psi}(\tilde{\beta})\psi(\gamma)$ and $\mathbf{1}[\phi \le K_0]\tilde{\psi}(\tilde{\beta})\tilde{\psi}(\tilde{\gamma})$ have the same asymptotic distribution, which is given by

$$
\begin{aligned}
&\mathbf{1}[\sum_{i=1}^k \left[-c_i J_i(1)^2 - c_i^2 \int J_i^2 - \frac{2c_i}{1-e^{-2c_i}} \left[e^{-c_i} J_i(1) + c_i \int e^{-c_i s} J_i\right]^2 + [J_i(1) + c_i \int J_i]^2\right] \le K_0]\times \\
&\quad \mathbf{1}[\exp\left[\int \bar{W}_0'\Lambda^{1/2} dW_\varepsilon - \tfrac{1}{2}\int \bar{W}_0'\Lambda\bar{W}_0\right] \le K'] \exp\left[\int \bar{W}_0'\Lambda^{1/2} dW_\varepsilon - \tfrac{1}{2}\int \bar{W}_0'\Lambda\bar{W}_0\right] \times \\
&\quad \mathbf{1}[\exp\left[\int \bar{W}_1'\Lambda^{1/2} dW_\varepsilon - \tfrac{1}{2}\int \bar{W}_1'\Lambda\bar{W}_1\right] \le K'] \exp\left[\int \bar{W}_1'\Lambda^{1/2} dW_\varepsilon - \tfrac{1}{2}\int \bar{W}_1'\Lambda\bar{W}_1\right]
\end{aligned}
$$

where $\bar{W}_0$ and $\bar{W}_1$ are mutually independent $k\times 1$ demeaned standard Wiener processes independent of W_ε, $c_i = \sqrt{2}a_i$ and J_i (which are continuous functionals of $W_\varepsilon(\cdot)$) are defined as in Lemma 2. But weak convergence together with the boundedness of $\tilde{\psi}(\cdot)$ and $\psi(\cdot)$ by K' implies convergence in expectation, so that

$$
E\left[\left(LR_T(K') - \widetilde{LR}_T(K')\right)^2\right] \to 0
$$

$LR_T(K') - \widetilde{LR}_T(K')$ hence converge in mean square, which implies convergence in probability.

Proof of Theorem 2:

In order to establish contiguity, we need to show that (i) LR_T converges weakly to some random variable $\widetilde{LR}$ under the null hypothesis of $h = B_Q'\varepsilon$ and (ii) $E[\widetilde{LR}] = 1$.

For (i), first note that by Theorem 1, $LR_T - \widetilde{LR}_T \xrightarrow{p} 0$ under the null hypothesis. But convergence in probality implies convergence in distribution, after noting that

$$
\left[\frac{1-r_a^{2T}}{T(1-r_a^2)r_a^{T-1}}\right]^{-1} \to \frac{2ae^{-a}}{1-e^{-2a}}
$$

as $T \to \infty$ the result is immediate from the CMT and and Lemma 2, with $\widetilde{LR} = \prod_{i=1}^k \widetilde{LR}^i$ where

$$
\begin{aligned}
\widetilde{LR}^i &= \left[\frac{2a_i e^{-a_i}}{1-e^{-2a_i}}\right]^{1/2} \\
&\quad \exp\left[-\tfrac{1}{2}\left[-a_i J_i(1)^2 - a_i^2 \int J_i^2 - \frac{2a_i}{1-e^{-2a_i}}\left[e^{-a_i} J_i(1) + a_i \int e^{-a_i s} J_i\right]^2 + [J_i(1) + a_i \int J_i]^2\right]\right]
\end{aligned}
$$

and J_i is defined in Lemma 2.

Turning to (ii), from the independence of the processes $J_i(\cdot)$ we find

$$E[\widetilde{LR}] = \prod_{i=1}^{k} E[\widetilde{LR}^i]$$

so that it is clearly sufficient to show that

$$E[\widetilde{LR}^i] = 1.$$

Now from Girsanov's (1960) Theorem as described in Tanaka (1996), p. 109, a change of measure yields

$$E[\widetilde{LR}^i] = \left[\frac{2a_i e^{-a_i}}{1-e^{-2a_i}}\right]^{1/2}$$
$$E\left[\exp\left[-\tfrac{1}{2}\left[-a_i - \frac{2a_i}{1-e^{-2a_i}}\left[e^{-a_i}W_i(1) + a_i \int e^{-a_i s}W_i(s)ds\right]^2 + [W_i(1) + a_i \int W_i(s)ds]^2\right]\right]\right]$$

where W_i is a Wiener process. Define

$$Z_W = \begin{pmatrix} W_i(1) + a_i \int W_i(s)ds \\ W_i(1) + a_i \int e^{a_i(1-s)}W_i(s)ds \end{pmatrix}$$

and

$$\Lambda_W = \begin{pmatrix} 1 & 0 \\ 0 & -2a_i e^{-2a_i}/(1-e^{-2a_i}) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2a_i/(1-e^{2a_i}) \end{pmatrix}$$

so that

$$E[\widetilde{LR}^i] = \left[\frac{2a_i}{1-e^{-2a_i}}\right]^{1/2} E\left[\exp[-\tfrac{1}{2}Z_W' \Lambda_W Z_W]\right]$$

With

$$Z_W = \int \begin{pmatrix} 1 + a_i(1-s) \\ e^{a_i(1-s)} \end{pmatrix} dW_i(s)$$

we find $Z_W \sim \mathcal{N}(0, V_W)$, where

$$V_W = E[Z_W Z_W'] = \begin{pmatrix} 1 + a_i + a_i^2/3 & e^{a_i} \\ e^{a_i} & (e^{2a_i}-1)/(2a_i) \end{pmatrix}.$$

By completing the square we compute

$$\begin{aligned} E[\widetilde{LR}^i] &= \left[\frac{2a_i}{1-e^{-2a_i}}\right]^{1/2} \int (2\pi)^{-1}|V_W|^{-1/2} \exp\left[-\tfrac{1}{2}Z_W'\left[\Lambda_W + V_W^{-1}\right]Z_W\right] dZ_W \\ &= \left[\frac{2a_i}{1-e^{-2a_i}}\right]^{1/2} |(\Lambda_W + V_W^{-1})V_W|^{-1/2} \\ &= \left[\frac{2a_i}{1-e^{-2a_i}}|\Lambda_W V_W + I_2|^{-1}\right]^{1/2} \end{aligned}$$

and a direct calculation shows

$$\Lambda_W V_W + I_2 = \begin{pmatrix} 2 + a_i + a_i^2/3 & e^{a_i} \\ -2a_i e^{a_i}/(e^{2a_i}-1) & 0 \end{pmatrix}$$

so that $|\Lambda_W V_W + I_2| = 2a_i e^{2a_i}/(e^{2a_i}-1)$, yielding the desired result.

Proof of Theorem 3:

(i) Noting that $\{X_t^*\varepsilon_t^*\} = \{\sigma^{-1}\Sigma_X^{-1/2}X_t\varepsilon_t\}$, Condition 3 implies that the long-run covariance of $\{X_t^*\varepsilon_t^*\}$ is given by $\sigma^{-2}\Sigma_X^{-1/2}V_X\Sigma_X^{-1/2}$, so that

$$\begin{aligned}
& T^{-1/2}\left[(e'_{[sT]}, 0'_{T-[sT]}) \otimes I_k\right][I \otimes P^{*\prime}\hat{V}_X^{-1/2}]\Xi' M\varepsilon \\
= \; & T^{-1/2}P^{*\prime}\hat{V}_X^{-1/2}\sigma\Sigma_X^{1/2}\sum_{t=1}^{[sT]}X_t^*\varepsilon_t^* \\
& -P^{*\prime}\hat{V}_X^{-1/2}\sigma\Sigma_X^{1/2}\left(T^{-1}\sum_{t=1}^{[sT]}X_t^*Q_t^{*\prime}\right)(T^{-1}Q^{*\prime}Q^*)^{-1}T^{-1/2}\sum_{t=1}^{T}Q_t^*\varepsilon_t^* \\
\Rightarrow \; & P^{*\prime}V_X^{-1/2}V_X^{1/2}P^*W_\varepsilon(s) - sP^{*\prime}V_X^{-1/2}V_X^{1/2}P^*W_\varepsilon(1) \\
= \; & W_\varepsilon(s) - sW_\varepsilon(1)
\end{aligned}$$

where the weak convergence follows from the uniform convergence of $T^{-1}\sum_{t=1}^{[sT]}Q_t^*Q_t^{*\prime} \xrightarrow{p} sI_{k+d}$, the consistency of $\hat{V}_X$, the CMT and the FCLT for mixing series as in the proof of Lemma 3.

(ii) We first prove the convergence in probability under the null hypothesis. By definition

$$\begin{aligned}
\hat{J}(\Omega^*) &= \sum_{i=1}^{k}\hat{v}_i'[G_{a_i} - M_e]\hat{v}_i \\
J(\Omega^*) &= \sum_{i=1}^{k}v_i'[G_{a_i} - M_e]v_i \\
\hat{v}_i &= [I \otimes \iota'_{k,i}P^{*\prime}\hat{V}_X^{-1/2}]\Xi M y \\
v_i &= [I \otimes \iota'_{k,i}P^{*\prime}\sigma^{-1}\Sigma_X^{-1/2}]\Xi M y
\end{aligned}$$

From an application of Lemma 6,

$$v_i'[G_{a_i} - M_e]v_i - \hat{v}_i'[G_{a_i} - M_e]\hat{v}_i = g_J\left(T^{-1/2}\sum_{t=1}^{[\cdot T]}v_{i,t}\right) - g_J\left(T^{-1/2}\sum_{t=1}^{[\cdot T]}\hat{v}_{i,t}\right) + o_p(1)$$

But with $\{\hat{\varepsilon}_t\}$ the residuals of a OLS regression of $\{\varepsilon_t\}$ on $\{Q_t\}$, we have

$$\begin{aligned}
\sup_s T^{-1/2}\left|\sum_{t=1}^{[sT]}\hat{v}_{i,t} - \sum_{t=1}^{[sT]}v_{i,t}\right| &\leq |\iota'_{k,i}P^{*\prime}\hat{V}_X^{-1/2} - \iota'_{k,i}P^{*\prime}\sigma^{-1}\Sigma_X^{-1/2}|\sup_s\left|T^{-1/2}\sum_{t=1}^{[sT]}X_t\hat{\varepsilon}_t\right| \\
&\xrightarrow{p} 0
\end{aligned}$$

since $\sup_s|T^{-1/2}\sum_{t=1}^{[sT]}X_t\hat{\varepsilon}_t| = O_p(1)$, and by the continuity of g_J in the sup-norm the result is established. Convergence under the alternative follows immediately from Theorem 2, since $\hat{V}_X$ and $\{\hat{\varepsilon}_t\}$ are functions of (h, Q).

References

ANDREWS, D. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

ANDREWS, D., I. LEE, AND W. PLOBERGER (1996): "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics*, 70, 9–38.

ANDREWS, D., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

——— (1995): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative," *Annals of Statistics*, 23, 1609–1629.

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

BERGER, J. (1985): *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, second edn.

BROWN, R., J. DURBIN, AND J. EVANS (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society Series B*, 37, 149–163.

CHERNOFF, H., AND S. ZACKS (1964): "Estimating the Current Mean of a Normal Distribution Which is Subject to Changes in Time," *Annals of Mathematical Statistics*, 35, 999–1028.

CHOW, G. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605.

CLARIDA, R., J. GALI, AND M. GERTLER (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–180.

CLEMENTS, M., AND D. HENDRY (1999): *Forecasting Non-Stationary Economic Time Series*. MIT Press, Cambridge, Mass.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, New York.

Engle, R., and D. Hendry (1993): “Testing Super Exogeneity and Invariance in Regression Models,” *Journal of Econometrics*, 56, 119–139.

Engle, R., D. Hendry, and J.-F. Richard (1983): “Exogeneity,” *Econometrica*, 51, 277–304.

Forchini, G. (2002): “Optimal Similar Tests for Structural Change for the Linear Regression Model,” *Econometric Theory*, 18, 853–867.

Franzini, L., and A. Harvey (1983): “Testing for Deterministic Trend and Seasonal Components in Time Series Models,” *Biometrika*, 70, 673–682.

Garbade, K. (1977): “Two Methods for Examining the Stability of Regression Coefficients,” *Journal of the American Statistical Association*, 72, 54–63.

Girsanov, I. (1960): “On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution of Measure,” *Theory of Probability and Its Applications*, 5, 285–301.

Hackl, P., and A. Westlund (1989): “Statistical Analysis of "Structural Change": An Annotated Bibliography,” *Empirical Economics*, 143, 167–192.

Hansen, B. (2000): “Testing for Structural Change in Conditional Models,” *Journal of Econometrics*, 97, 93–115.

King, M. (1988): “Towards a Theory of Point Optimal Testing,” *Econometric Reviews*, 6, 169–218.

LaMotte, L., and J. A. McWorther (1978): “An Exact Test for the Presence of Random Walk Coefficients in a Linear Regression Model,” *Journal of the American Statistical Association*, 73, 816–820.

Lehmann, E. (1986): *Testing Statistical Hypotheses*. Wiley, New York, second edn.

Leybourne, S., and B. McCabe (1989): “On the Distribution of some Test Statistics for Coefficient Constancy,” *Biometrika*, 76, 169–177.

Linde, J. (2001): “Testing for the Lucas-Critique: A Quantitative Investigation,” *American Economic Review*, 91, 986–1005.

Nabeya, S., and K. Tanaka (1988): "Asymptotic Theory of a Test for Constancy of Regression Coefficients Against the Random Walk Alternative," *Annals of Statistics*, 16, 218–235.

Newey, W., and K. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Nyblom, J., and T. Mäkeläinen (1983): "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association*, 78, 856–864.

Ploberger, W., and W. Krämer (1992): "The CUSUM Test with OLS Residuals," *Econometrica*, 60, 271–285.

Ploberger, W., W. Krämer, and K. Kontrus (1989): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

Pollard, D. (2001): "Contiguity," unpublished draft, http://www.stat.yale.edu/~pollard/Paris2001/Lectures/Contiguity.pdf.

Quandt, R. (1958): "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53, 873–880.

——— (1960): "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–330.

Rao, C. (1973): *Linear Statistical Inference and Its Applications*. Wiley, New York.

Shively, T. (1988a): "An Analysis of Tests for Regression Coefficient Stability," *Journal of Econometrics*, 39, 367–386.

——— (1988b): "An Exact Test for a Stochastic Coefficient in a Time Series Regression Model," *Journal of Time Series Analysis*, 9, 81–88.

Sowell, F. (1996): "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework," *Econometrica*, 64, 1085–1107.

Stock, J., and M. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

Tanaka, K. (1996): *Time Series Analysis — Nonstationary and Noninvertible Distribution Theory*. Wiley, New York.

Watson, W., and R. Engle (1985): "Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative," *Review of Economics and Statistics*, 67, 341–346.

White, H. (2001): *Asymptotic Theory for Econometricians*. Academic Press, San Diego, revised edn.