Efficient Tests for General Persistent Time Variation in Regression Coefficients*

Graham Elliott
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093
gelliott@weber.ucsd.edu

Ulrich K. Müller†
Princeton University
Economics Department
Princeton, NJ, 08544
umueller@princeton.edu

Abstract

There are a large number of tests for instability or breaks in coefficients in regression models designed for different possible departures from the stable model. We make two contributions to this literature. First, we consider a large class of persistent breaking processes that lead to asymptotically equivalent efficient tests. Our class allows for many or relatively few breaks, clustered breaks, regularly occurring breaks or smooth transitions to changes in the regression coefficients. Thus asymptotically nothing is gained by knowing the exact breaking process of the class. Second, we provide a test 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 good power and size properties even in heteroskedastic models.

JEL Classification: C12, C22

Keywords: Structural Breaks, Nonstandard Testing, Time Varying Parameters, Contiguity

*This paper is a revision of an earlier working paper circulated under the title "Optimally Testing General Breaking Processes in Linear Time Series Models". Graham Elliott is grateful to the NSF for financial assistance under grant SES 0111238, and Ulrich Müller gratefully acknowledges financial support of the Swiss National Science Foundation. The authors thank two anonymous referees, the editor, Bernard Salanié, and Mark Watson, Giorgio Primiceri, Allan Timmermann, Paolo Giordani, Piotr Eliasz and seminar participants at EUI, Berkeley, UCSD, Princeton, Boston University, LSE, Aarhus, Yale, Harvard and at the ReStud Tour 2003 for helpful comments.

†Corresponding author.

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 tests of the null hypothesis of a stable linear model $y_t = X_t'\bar{\beta} + Z_t'\delta + \varepsilon_t$ against the alternative of a partially unstable model $y_t = X_t'\beta_t + Z_t'\delta + \varepsilon_t$, where the variation in β_t is of the strong form.

The literature on this problem is huge as numerous difficulties have arisen in testing this possibility. The diversity of testing approaches in this literature follows primarily from the diversity of possible ways $\{\beta_t\}$ can be nonconstant. Tests have been derived for many different posited models of the breaking process $\{\beta_t\}$, despite theory giving relatively little guide as to what alternatives to expect in practice. Tests developed for one model of a nonconstant $\{\beta_t\}$ may not be useful for other possible models. A secondary effect is that many of the tests for each of these models are justified on the grounds that they provide consistent tests rather than an appeal to being optimal for some particular model. Progress has been made in examining optimality for particular special models of the alternative. But for many plausible models of the breaking process optimal tests have not been derived and to do so requires overcoming many difficulties. Even when optimal tests have been derived, they often become computationally very involved, since they require searching over all possible combinations of the break dates. Alternatively, for other cases the test statistic is simple to compute but asymptotic distributions are not available for construction of valid critical values under realistic assumptions.

This paper analyzes tests for parameter stability in a single unified framework. We point out that when restricting attention to efficient tests, the seemingly different approaches of 'structural breaks' and 'random coefficients' are in fact equivalent. Thus approaches that describe the breaking process with a number of nonrandom parameters are unified with tests that specify stochastic processes for $\{\beta_t\}$. The crucial determinant of any efficient test for structural stability is the assumption it makes for the evolution of $\{\beta_t\}$. We argue that for most applications, it is reasonable to focus on testing against the alternative of persistent time variation in $\{\beta_t\}$, although that clearly leaves a myriad of possibilities for the exact evolution of $\{\beta_t\}$. Making further progress hence requires a systematic investigation of the impact of the specific choice of persistent process under the alternative on efficient tests of parameter stability.

This paper carries out such an investigation in a novel analytical framework. We consider general mean-zero and persistent breaking processes, such that the scaled parameter coefficients converge weakly to a Wiener process. The processes we study include breaks that occur in a random fashion, serial correlation in the changes of the coefficients, a clustering of break dates and so forth. The main result is that under a normality assumption on the disturbances, small sample efficient tests in this broad set are asymptotically equivalent. Optimal tests for any specific breaking process that satisfies our assumptions are interchangeable with an optimal test for another breaking process that satisfies the assumptions—the tests are asymptotically equally capable of distinguishing the null hypothesis from any of the breaking processes we examine. We hence show that leaving the exact breaking process unspecified (apart from a scaling parameter) does not result in a loss of power in large samples. In a simulation section, we show this asymptotic result to be an accurate prediction for some simple small sample data generating processes: For 100 observations and a martingale assumption on $\{\beta_t\}$ with five known break dates, the gain in power of the small sample efficient invariant tests over asymptotically equivalent tests that focus on Gaussian random walk variation in $\{\beta_t\}$ never exceeds five percentage points.

An important precursor to this work is Nyblom's (1989) result that the small sample locally best test is unique as long as $\{\beta_t\}$ follows a martingale. Locally best tests maximize the slope of the power function at the null hypothesis of a stable model, where power and the level of a test coincide. In contrast, our asymptotic equivalence result concerns the behavior of tests at distinguishable alternatives, that is at alternatives where tests have power between level and unity. At the same time, our set of breaking processes neither contains nor is contained in the class of all martingales.

Our equivalence result has a number of positive implications for testing for breaks, both theoretically and empirically. From a theoretical perspective, the equivalence of power over many models means that there is little point in deriving further optimal tests for particular processes in our set. Doing so will not lead to any substantive power gains over tests already in the literature or the one developed here. From a practical perspective it means that the researcher does not have to specify the exact path of the breaking process in order to be able to carry out (almost) efficient inference. This is fortunate for applied work—for many of the breaking processes optimal tests are highly cumbersome to derive and in most cases theory provides no guidance as to what form of time variation to expect.

Finally, we suggest a new, easy-to-compute statistic that is asymptotically point-optimal for our broad set of breaking processes. It remains valid under very general specifications of the error term, including heteroskedasticity as well as general assumptions on the covariates. The proposed test has a number of advantages over previously suggested tests. Computation of the test statistic requires no more than $k+1$ OLS regressions, where k is the dimension of the vector X_t. This is a significant simplification over tests that require computations for each possible combination of break dates. Our test statistic requires no trimming of the data, often a feature of other tests for breaks. This simplifies testing not only in regards to programming the test statistic, but also avoids the awkward and empirically relevant dependence of the test outcome on the trimming parameter. Finally, we find that our statistic has superior size control in small samples than other popular tests, particularly when the disturbances are heteroskedastic. Since the implications of our theoretical results are that amongst reasonable tests power will be very comparable, it would seem that simplicity of construction and good size control are strong reasons to choose between available tests.

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 the recommended point-optimal statistic is taken up in Section 4, and Section 5 evaluates the small sample size and power of a number of tests for time variation in β_t. A final section concludes. Proofs are collected in an appendix.

2 The Model and Tests for Breaks

This paper is concerned with distinguishing the null hypothesis of a stable regression model

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

from the alternative hypothesis of the unstable model

$$y_t = X_t'\beta_t + Z_t'\delta + \varepsilon_t \quad t = 1, \cdots, T \tag{2}$$

with nonconstant $\{\beta_t\}$, where y_t is a scalar, X_t, β_t are $k \times 1$ vectors, Z_t and δ are $d \times 1$, $\{y_t, X_t, Z_t\}$ are observed, $\bar{\beta}$, $\{\beta_t\}$ and δ are unknown and ε_t is a mean zero disturbance. In words, we want to test whether the coefficient vector 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. We focus on a situation where there is little or no reliable information on the form of potential instabilities.

Hypothesis tests that distinguish between models (1) and (2) have received a great deal of attention in both the statistical and econometrics literature. It might usefully be organized into two strands: the 'structural break' literature, which views the path of $\{\beta_t\}$ 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.

The 'structural break' literature posits a model with a fixed number N of breaks at fixed points in time τ_i,

$$\begin{aligned} \beta_t &= \bar{\beta}_0 \text{ for } t < \tau_1 \\ \beta_t &= \bar{\beta}_1 \text{ for } \tau_1 \leq t < \tau_2 \\ &\vdots \\ \beta_t &= \bar{\beta}_{N-1} \text{ for } \tau_{N-1} \leq t < \tau_N \\ \beta_t &= \bar{\beta}_N \text{ for } \tau_N \leq t \leq T \end{aligned} \tag{3}$$

where $\bar{\beta}_i$ are nonzero for $i = 1, ..., N$.

By far the most attention has been given to the single break model, in which $N = 1$. In this literature, $\bar{\beta}_0$, $\bar{\beta}_1$ and τ_1 are fixed but unknown parameters. With τ_1 unknown, Quandt (1958, 1960) suggested considering the maximum of the usual Chow (1960) F-tests computed over all possible τ_1, denoted here by supF. This search over a set of dependent F-statistics results in the asymptotic distribution of the test ceasing to be χ^2. Andrews (1993) derives asymptotic properties of such tests. Many other tests have been suggested (e.g. Brown, Durbin and Evans (1975), Ploberger, Krämer and Kontrus (1989) and Ploberger and Krämer (1992)). Fewer results are available when $N > 1$. Bai and Perron (1998) extend the Quandt approach and examine the maximum of the F-statistics over all combinations of $(\tau_1, \cdots, \tau_N)$. Because the number of break date combinations becomes huge even for moderate N (with $T = 100$ and $N = 5$, there are over 75 million combinations), they employ some clever dynamic programming and additional assumptions on the breaking process to implement such a test.

Most tests are motivated on consistency grounds, which often provides no reason to distinguish between them or think that they provide 'best' tests. And though the supF statistics can be naturally motivated as generalized likelihood ratio tests, this does not necessarily make them desirable tests. 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.[1] Andrews and Ploberger (1994) have devised an optimal method for dealing with testing problems of this kind, which can also be applied to testing structural stability against (3). Their procedure is (asymptotically) optimal in the sense of maximizing a weighted average power criterion: For each fixed set of break dates and magnitudes $(\bar{\beta}_j - \bar{\beta}_{j-1})$, the power of a test is potentially different. In choosing among possible tests, Andrews and Ploberger (1994) identify the test that maximizes the weighted average of these powers, where the nonnegative weighting is over the magnitude of the breaks and over 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 over the magnitude of the breaks as a Gaussian density 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 calculations are only made for $N = 1$.

The 'time varying parameter' literature approaches the problem from a seemingly very different angle. There the nonconstant $\{\beta_t\}$ is viewed as being random, and contributions to this strand differ in the probability law they pose for $\{\beta_t\}$. Whilst some studies investigate models in which $\{\beta_t\}$ deviates only temporarily from zero (e.g. Watson and Engle (1985) and Shively (1988a)), the majority of studies have considered the model where deviations of $\{\beta_t\}$ from zero are permanent. In these models the alternative hypothesis is that $\{\beta_t\}$ 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 (1988b), Leybourne and McCabe (1989), Nyblom (1989) and Saikkonen and Luukonen (1993)—see the annotated

[1] Andrews and Ploberger (1995) showed, however, that the supF statistic does possess an optimality property.

bibliography by Hackl and Westlund (1989) for further references. More recently, there has been a renewed interest in a fully Bayesian analysis of such models. Chib (1998), Koop and Potter (2004) and Giordani, Kohn and van Dijk (2005) employ Markov Chain Monte Carlo methods to overcome computational challenges. Also see Pesaran, Pettenuzzo, and Timmermann (2004) for an application to forecasting.

As with the deterministic approach, optimal tests have been derived in closed form only for the simplest cases. With distributional assumptions for $\{\varepsilon_t\}$ and $\{\beta_t\}$, efficient tests are given by the likelihood ratio statistic. The difficulty consists of analyzing the likelihood of the model under both the null and alternative hypothesis—even for independent Gaussian disturbances $\{\varepsilon_t\}$ and a Gaussian Random Walk of $\{\beta_t\}$ the resulting expressions are such complicated functions of the observables that the asymptotic distributions have not been derived for nonconstant covariates. As a by-product of our derivations in Section 3, we derive this analytical result, enabling the computation of asymptotic critical values. For more complicated processes, this is even more difficult, and in general depends on the specific alternative.

As noted in the introduction, a focus on the slope of the power function greatly simplifies testing for time variation: Nyblom (1989) establishes the remarkable result that for β_0 and δ known, the small sample locally best test of parameter constancy is unique as long as $\{\beta_t\}$ follows a martingale. The generality and implications of this result are not quite clear, however. Not all economically interesting processes for $\{\beta_t\}$ are martingales. Furthermore, in nonstandard testing problems local optimality does not necessarily imply good power relative to other tests even for alternatives very close to the null hypothesis. In the case of testing for a unit root, for instance, the locally best test has significantly lower power against local alternatives than nearly all other tests—see Stock (1994).

Despite their different rationales, we would suspect that tests against a time varying parameter have power against the alternative of structural breaks and vice versa. Ploberger et al. (1989) show the consistency of their approach against a wide range of alternatives, and Stock and Watson (1998) derive the asymptotic local power of the supF statistic and the Andrews and Ploberger (1994) tests in a time varying parameter model. But the relationship between tests for these two models runs deeper than this insight.

Consider the typical path of $\{\beta_t\}$ in a time varying parameter model with $\beta_t = \sum_{s=1}^{t} w_s$, w_s independent zero mean Gaussian variates. This is quite different from a model with N breaks such as (3). But, as noted by Nyblom (1989), we could 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 of $\Delta\beta_t$ 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 (3), with N and $\{\bar{\beta}_0, \bar{\beta}_1, \cdots, \bar{\beta}_N\}$ 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.

Similarly, tests of model (3) that are optimal in the weighted average power sense of Andrews and Ploberger (1994) 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 $\{\beta_t\}$ as fixed and using weights for their outcomes under the alternative or treating them as random hence amounts to the same thing. This equivalence between efficient frequentist tests in the weighted average power sense and Bayesian decision rules has long been understood—see, for instance, Ferguson (1967) or Berger (1985) for general treatments and Andrews and Ploberger (1994) for the application to tests for parameter stability.

With this insight, the one essential determinant of efficient tests—derived from either the structural break perspective or the time varying parameter perspective—is hence its weighting function, or, equivalently, the probability distribution it posits for $\{\beta_t\}$ under the alternative. As argued, the time varying parameter literature and the structural break literature have emphasized quite different processes for $\{\beta_t\}$. This raises the important question how this weighting function should be chosen. The answer 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 a relationship can be an important question in its own right. The stability of the link between monetary aggregates and output, for instance, is a crucial question for monetary policy makers (see Clarida et al. (2000) for a recent example). Also, tests for the Lucas critique arise directly as tests of parameter instability (Engle et al. (1983), Engle and Hendry (1993), Oliner, Rudebusch, and Sichel (1996), 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 $\{\beta_t\}$ are permanent. Second, if a model turns out to be unstable, then this must be taken into account in the construction of appropriate forecasts, since recent observations of the relationship (2) will be closer to the (unknown) future relationship than past observations—see Chernoff and

Zacks (1964), Cohen and Kushary (1994) and Clements and Hendry (1999), among others. Temporary unforecastable breaks are important for the width of confidence intervals but less so for computing the point forecasts, since they can be thought of as an extra source of stationary noise. Third, parameter stability tests are a crucial specification test for standard inference on $\bar{\beta}$. Consider model (2) with $X_t = 1$ and no Z_t. If $\beta_t = \sum_{s=1}^{t} w_s$, then model (2) 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 difficult to interpret it in a meaningful way. More generally, whenever $\{\beta_t\}$ varies in a permanent fashion (as, for instance, in (3)), 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 $\{\beta_t\}$ from zero do not necessarily lead to the same interpretational difficulties. In the extreme temporary case of $\{\beta_t\}$ being independent and identically distributed with mean $\bar{\beta}$, $X_t'(\beta_t - \bar{\beta})$ 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 and the more misleading standard inference ignoring parameter variation will be.

All three motivations are more pervasive the more persistent the changes in $\{\beta_t\}$. When carried out for one the reasons discussed, a useful test of parameter stability should hence maximize its power against persistent changes of $\{\beta_t\}$. While this suggests a focus on alternatives with a persistently varying $\{\beta_t\}$, the obvious problem remains that there exist many different persistent breaking processes. What is being called for, then, is a systematic investigation of the impact of alternative assumptions for persistently varying $\{\beta_t\}$ on the properties of efficient tests. Given that the stochastic properties of the process $\{\beta_t\}$ (or, equivalently, the weighting function employed over various fixed paths of $\{\beta_t\}$) are the one determinant of efficient tests for parameter instability, the answer to this question is of great theoretical and practical interest. At least intuitively, it seems that knowledge about the precise form of the (persistent) variation in $\{\beta_t\}$ under the alternative is required in order to carry out an efficient hypothesis test of parameter stability.

This paper shows that this intuition is largely mistaken. We develop a new analytical framework to show that for a large class of breaking processes with persistently varying $\{\beta_t\}$, and an assumption on the distribution of the disturbances, the optimal small sample statistics are asymptotically equivalent. In other words, the precise form of the breaking process $\{\beta_t\}$ 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. The set of breaking processes we consider is as follows.

Condition 1 *Let $\{\Delta\beta_{T,t}\}$ be a double array of $k\times 1$ random vectors $\Delta\beta_{T,t} = (\Delta\beta_{T,t,1}, \cdots, \Delta\beta_{T,t,k})'$. Assume that*
(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, i.e. $\lim_{T\to\infty} T^{-1}E[(\sum_{t=1}^{[sT]} T\Delta\beta_{T,t})(\sum_{t=1}^{[sT]} T\Delta\beta'_{T,t})] = s\Omega$ for all s.

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 $\{\Delta\beta_t\}$ 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 Ω of $\{\Delta\beta_t\}$ is of order T^{-2}. We stress 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 enables the application of Theorem 7.30 in White (2001), ensuring that suitably scaled, the breaking process $\{\beta_t - \beta_0\}$ is asymptotically well approximated by a $k \times 1$ Wiener process. This allows for a multitude of diverse breaking models, from relatively rare to very frequent breaks. For any finite sample 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. The alternative thought experiment of having a finite number of breaks independent of the sample size—as employed by Andrews and Ploberger (1994), for instance—is not covered by Condition 1. Also stationary processes $\{\beta_t\}$, such as those typically arising from Markov switching models, are ruled out by Condition 1. Note that Nyblom's (1989) martingale assumption is different from Condition 1: not all Condition 1 processes are martingales and not all martingales are Condition 1 processes.

Examples of Condition 1 processes include 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. 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. Such breaking processes are, of course, not martingales. 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, Condition 1 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 presidential elections.

In the next section, we derive small sample efficient tests of parameter stability for any selected alternative process $\{\beta_t\}$ of the set described by Condition 1, which by the Neyman-Pearson Lemma might be based on the likelihood ratio statistic LR_T. We also consider an approximate statistic $\widetilde{LR}_T$ that depends on the selected process only through Ω. Additional regularity conditions concerning $\{X_t, Z_t\}$ and the assumption of Gaussian disturbances $\{\varepsilon_t\}$ are summarized in Condition 2 below. The following Theorem, which implies the asymptotic equivalence of all small sample efficient tests of parameter stability against Condition 1 processes with a common Ω, is the main result of this paper.

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

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

under both the null and alternative hypotheses.

The equivalence of efficient tests in this class is not only of theoretical interest, but also dramatically simplifies the practice of testing against parameter instability: it allows the applied researcher to leave the exact form of the alternative unspecified without foregoing (asymptotic) power. 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. This insight allows us to suggest an easy-to-compute test statistic $\widehat{\mathrm{qLL}}$, based on a "quasi Local Level" model, that is asymptotically point-optimal for Condition 1 processes.

For the special case of $X_t = 1$ and serially uncorrelated, homoskedastic $\{\varepsilon_t\}$, $\widehat{\mathrm{qLL}}$ is the Most Powerful Invariant (MPI) test in a Gaussian unobserved component model, as analyzed by Franzini and Harvey (1983) and Shively (1988b). For more general assumptions on $\{X_t\}$

and $\{\varepsilon_t\}$, $\widehat{\text{qLL}}$ does not correspond to a test previously suggested in the literature. The statistic is asymptotically valid under very general assumptions on the disturbances and the regressors—see Section 4 for details. The test requires no trimming at the end points, and the estimation of only $k+1$ regressions. This contrasts with considerable computational complexity of test statistics against, say, four breaks. In addition, we find $\widehat{\text{qLL}}$ to have very attractive small sample properties in our simulations in Section 5.

$\widehat{\text{qLL}}$ is computed in the following simple steps:

- 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 over $i = 1, \cdots, k$.
- 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 $\widehat{\text{qLL}}$ and asymptotic critical values are given in Table 1 for $k = 1, \cdots, 10$. The critical values are independent of the dimension of Z_t.

The intuition for these computations is most easily developed in the 'local-level model' (see Harvey (1989)), where X_t is constant and there is no Z_t. In this model, an efficient test for stability can be based on the difference in the log-likelihood between the stable model $y_t = \bar{\beta} + \varepsilon_t$ and the time varying model $y_t = \beta_t + \varepsilon_t$, where $T^{-1}\sum_{t=1}^{T}\beta_t = \bar{\beta}$. The requirement that the average value of the parameter path in the unstable model equals that of the stable model ensures that power is directed entirely at detecting parameter instability, rather than different average parameter values between the null and alternative hypothesis. Theorem 1

Table 1: Asymptotic Critical Values of $\widehat{\mathrm{qLL}}$ (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	-48.78	-53.38

Note: Percentiles reported are calculated from 40000 draws from distributions of the random variable reported in Lemma 2, p. 18, with $c_i = 10$ for all i using 2000 standard normal steps to approximate Wiener Processes.

shows the equivalence of the small sample efficient test statistics for alternative Condition 1 assumptions, so that we might conveniently derive the optimal test when $\beta_t - \beta_0$ follows a Gaussian Random walk. Under this alternative, the first differences of y_t, $\Delta y_t = \Delta\beta_t + \Delta\varepsilon_t$, follow a Gaussian MA(1), i.e. $\Delta y_t \sim \eta_t + r_\eta \eta_{t-1}$ for $\eta_t \sim iidN(0, \sigma_\eta^2)$ and constant $r_\eta < 1$, where r_η and σ_η^2 are functions of σ_ε^2, T and Ω. If η_1 was known, one could easily solve for η_t recursively via $\eta_t = \Delta y_t - r_\eta \eta_{t-1}$, mirroring Step 4, and evaluate the log-likelihood (apart from constants) under the alternative as $-\frac{1}{2}\sigma_\eta^{-2}\sum \eta_t^2$, interpreted as a function of $\{\Delta y_t\}$. Under the null hypothesis, the log-likelihood is $-\frac{1}{2}\sigma_\varepsilon^{-2}\sum(\Delta y_t)^2$, such that an efficient test statistic would be given by $\frac{\sigma_\varepsilon^2}{\sigma_\eta^2}\sum \eta_t^2 - \sum(\Delta y_t)^2$ (Step 6). Now η_1 is not known, but with the restriction $T^{-1}\sum_{t=1}^T \beta_t = \bar{\beta}$, it turns out after some matrix algebra (cf. Lemma 4 of the appendix) that the appropriate modifications can be written in terms of regressions (Steps 1 and 5). In practice, also Ω is of course unknown, but we argue in Section 4 below for the specific choice that underlies $\widehat{\mathrm{qLL}}$ as defined above.

The intuition for general regressors X_t is that for purposes of testing the stability of β_t, the unstable model $y_t = X_t'\beta_t + \varepsilon_t$ is asymptotically well approximated by the $k \times 1$ vector local level model $X_t y_t = \hat{\Sigma}_X \beta_t + X_t \varepsilon_t$, where $\hat{\Sigma}_X = T^{-1}\sum X_t X_t'$. Our results show that treating $X_t y_t$ as if they were Gaussian observations with time varying mean $\hat{\Sigma}_X \beta_t$ under the alternative yields an asymptotically efficient test. See Eliasz, Stock, and Watson (2004) for a similar heuristic argument concerning the equivalence of an unstable linear regression to a vector local level model.

3 Asymptotic Equivalence of Optimal Tests for General Breaking Processes

We now turn to the proof of the asymptotic equivalence result in Theorem 1. It turns out that optimal tests depend on the average magnitude of the breaks, as described by Ω of Condition 1, even asymptotically. With our focus on understanding the impact of assuming breaking processes of different forms—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 without asymptotic value for the testing problem.

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 $\Xi = \text{diag}(X_1', \cdots, X_T')$, the $T \times 1$ vector e of ones, $M_e = I_T - e(e'e)^{-1}e'$ and, for future reference, the $T \times (d+k)$ matrix $Q = (X, Z)$ and $M = I_T - Q(Q'Q)^{-1}Q'$. The symbol '$\otimes$' denotes the Kronecker product.

By the Neyman-Pearson Lemma, optimal tests to distinguish the unstable model (2) from the stable model (1) can be based on the likelihood ratio statistic. We are interested in deriving efficient tests of stability of the coefficient of X_t. For this purpose, it makes sense to ensure that under the alternative, the average value of the random parameter path is always the same as that under the stable model, i.e. $T^{-1}\sum_{t=1}^{T}\beta_t = \bar{\beta}$. Under Condition 1, this can be achieved by letting $\beta_0 = \bar{\beta} - T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t}\Delta\beta_s$. This normalization ensures that the likelihood ratio statistic efficiently detects variation in the coefficient of X_t, rather than differences between the average value of the parameter. In matrix form, the stable and unstable model can then be rewritten as

$$\begin{aligned} y &= X\bar{\beta} + Z\delta + \varepsilon \\ y &= \Xi[M_e \otimes I_k]\beta + X\bar{\beta} + Z\delta + \varepsilon. \end{aligned} \tag{4}$$

To be able to write down the likelihood, we must make distributional assumptions on ε and Q. Let $\mathfrak{F}_{T,t}$ be the sigma field generated by $\{Q_{T,t}, y_{T,t}, Q_{T,t-1}, y_{T,t-1}, \cdots, Q_{T,1}, y_{T,1}\}$, and $\mathfrak{F}_{T,0}$ the trivial sigma field. (We assume all random elements introduced here and below to be defined on the same probability space.)

Condition 2 *In the stable model (1) and the unstable model (2)*
(i) $Q_{T,t}$ and $\varepsilon_{T,t}$ are conditionally independent given $\mathfrak{F}_{T,t-1}$, and the conditional distribution of $\varepsilon_{T,t}$ given $\mathfrak{F}_{T,t-1}$ is $\mathcal{N}(0, \sigma^2)$, for $t = 1, \cdots, T$ and all T.

(ii) $Q_{T,t}$ given $\mathfrak{F}_{T,t-1}$ has density $f_{Q,T,t}$ with respect to the sigma-finite measures $\bar{\nu}_{Q,T,t}$, and $\{f_{Q,T,t}, \bar{\nu}_{Q,T,t}\}$ do not depend on $\bar{\beta}$, $\{\beta_t\}$ and δ for all $t = 1, \cdots, T$ and all T.
Furthermore, the stable model (1) satisfies
(iii) $\{Q_{T,t}, \varepsilon_{T,t}\}$ is either uniform mixing of size $-r/(2r-2)$ or strong mixing of size $-r/(r-2)$, $r > 2$.
(iv) $E[Q_{T,t}Q'_{T,t}] = \Sigma_Q$, $T^{-1}\sum_{t=1}^{[sT]} Q_{T,t}Q'_{T,t} \xrightarrow{p} s\Sigma_Q$ 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 the elements $Q_{T,t,i}$ of $Q_{T,t}$ satisfy $E[|Q_{T,t,i}|^r] < K$ 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 (ii) of Condition 2 requires the conditional distribution of Q_t given past values of Q_t and y_t not to depend on $\bar{\beta}$, $\{\beta_t\}$ and δ, which is the assumption of weak exogeneity as described in detail by Engle, Hendry and Richard (1983). This will allows the factorization of the likelihood of (y, Q) under the alternative into two pieces, one capturing the contribution to the likelihood of $\{\varepsilon_t = y_t - X'_t\beta_t - Z'_t\delta\}$ and the other the contribution of Q_t given $\mathfrak{F}_{t-1}$. 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 $\{f_{Q,t}\}$ or $\{\bar{\nu}_{Q,t}\}$.

Further restrictions on $\{Q_t\}$ in parts (iii) and (iv) are only required to hold under the null hypothesis of a stable model. 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 parameter path $\beta = b$, the conditional density of the data is

$$\begin{aligned} f_{y,Q|\beta=b}(y, Q) &= \prod_{t=1}^{T}(2\pi)^{-1/2}\sigma^{-1}\exp\left[-\tfrac{1}{2}(y_t - X'_t(\bar{\beta} + b_t - T^{-1}\sum_{s=1}^{T} b_s) - Z'_t\delta)^2/\sigma^2\right] f_{Q,t}(Q_t) \\ &= (2\pi\sigma^2)^{-T/2}\exp\left[-\tfrac{1}{2}(h - \Xi[M_e \otimes I_k]b)'(h - \Xi[M_e \otimes I_k]b)/\sigma^2\right]\prod_{t=1}^{T} f_{Q,t}(Q_t) \end{aligned}$$

where $h = y - X\bar{\beta} - Z\delta$. The unconditional density under the alternative may hence be written as

$$f^1_{y,Q}(y, Q) = (2\pi\sigma^2)^{-T/2}\int \exp\left[-\tfrac{1}{2}(h - \Xi[M_e \otimes I_k]b)'(h - \Xi[M_e \otimes I_k]b)/\sigma^2\right] d\nu_\beta(b)\prod_{t=1}^{T} f_{Q,t}(Q_t) \tag{5}$$

where ν_β is the measure of β, whereas under the null hypothesis, clearly

$$f^0_{y,Q}(y,Q) = (2\pi\sigma^2)^{-T/2} \exp\left[-\tfrac{1}{2}\sigma^{-2}h'h\right] \prod_{t=1}^{T} f_{Q,t}(Q_t). \tag{6}$$

We therefore find the likelihood ratio statistic to be

$$LR_T = \int \exp\left[\sigma^{-2}h'\Xi[M_e \otimes I_k]b - \tfrac{1}{2}\sigma^{-2}b'[M_e \otimes I_k]\Xi'\Xi[M_e \otimes I_k]b\right] d\nu_\beta(b). \tag{7}$$

Note that computation of LR_T, a function of h, requires knowledge of $\bar{\beta}$ and δ. One way to resolve this difficulty is to derive the efficient test that is invariant to transformations of the form

$$(y, Q) \to (y + X\bar{b} + Z\bar{d}, Q) \;\; \text{for any } \bar{b} \text{ and } \bar{d}. \tag{8}$$

All standard tests for structural breaks satisfy this invariance requirement. 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 (8). One maximal invariant is given by (My, Q), which can be computed without knowing $\bar{\beta}$ or δ. By the Neyman-Pearson Lemma, the optimal invariant test can be based on the likelihood ratio statistic LR^I_T of (My, Q). When the regressors $\{Q_t\}$ are independent of $\{\varepsilon_t\}$, it follows from standard calculations that

$$LR^I_T = \int \exp\left[\sigma^{-2}y'M\Xi b - \tfrac{1}{2}\sigma^{-2}b'\Xi'M\Xi b\right] d\nu_\beta(b). \tag{9}$$

In the weakly exogenous case, though, the density of the maximal invariant (My, Q) becomes a complicated function of $\{f_{Q,t}\}$, since $\{f_{Q,t}\}$ depends on $\{y_t\}$. Note that invariance is an additional restriction on the class of tests under consideration, so that the power of LR^I_T can be at most as large as that of LR_T. At the same time, $\widetilde{LR}_T$ defined below in (10) is a function of the maximal invariant (My, Q), so that a test based on LR^I_T has at least as much power as a test based on $\widetilde{LR}_T$. But Theorem 1 shows that the asymptotic power of LR_T and $\widetilde{LR}_T$ is identical, so that, by this sandwich argument, this also holds for LR^I_T. The following derivations therefore focus exclusively on the analysis of LR_T.

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 the much more tractable statistic $\widetilde{LR}_T$, that depends on the distribution of β only through Ω. On the one hand, this proves 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 against any Condition 1 process.

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 left $k \times k$ block of Σ_Q, and note that Ω^* is the long-run variance of $\{T\Delta\beta_t^*\} = \{T\sigma^{-1}\Sigma_X^{1/2}\Delta\beta_t\}$. 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 $k \times 1$ 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 $(Tk \times 1)$ random 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 $\{\Delta\beta_{T,t}\}$ in the stable model.

We will show that LR_T is asymptotically equivalent to the statistic

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

Note that $\widetilde{LR}_T$ depends on the generally unknown parameters σ^2 and Σ_X. Let $\overline{LR}_T$ be defined just like $\widetilde{LR}_T$, with σ^2 and Σ_X replaced by the estimators $\hat{\sigma}^2 = T^{-1} y' M y$ and $\hat{\Sigma}_X = T^{-1} X' X$, respectively. We show in Theorem 2 and Corollary 1 below that $\overline{LR}_T$ and $\widetilde{LR}_T$ have the same asymptotic distributions under both the null and alternative hypothesis. The lack of knowledge of σ^2 and Σ_X hence has no cost in terms of asymptotic power.

We begin by considering the asymptotic behavior of $\widetilde{LR}_T$, and will then show $LR_T - \widetilde{LR}_T \xrightarrow{p} 0$. Because $\nu_{\tilde{\beta}}$ is the distribution of a 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 $v_{i,t}$ of v_i is the $((t-1)k + i)$th element of $[I_T \otimes P^{*\prime} \sigma^{-1} \Sigma_X^{-1/2}] \Xi' M y$ or, equivalently, $v_i = [I_T \otimes \iota_{k,i}' P^{*\prime} \sigma^{-1} \Sigma_X^{-1/2}] \Xi' M y$.

A test based on the statistic

$$\text{qLL} = \sum_{i=1}^{k} v_i' [G_{a_i} - M_e] v_i \tag{11}$$

will hence be exactly equivalent to a test based on $\widetilde{LR}_T$, since qLL 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 qLL 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 2 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 $\int G(s) ds$ and so forth.

Lemma 2 *Under Condition 2 and the null hypothesis (1), for any positive $c_1, \cdots, c_k$*

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

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

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 (7) and (10). The strategy will be to do the asymptotic reasoning 'inside the integration'. Recall that γ and β are independent and identically distributed Condition 1 processes, and $\tilde{\gamma}$ and $\tilde{\beta}$ are independent and distributed $\mathcal{N}(0, FF' \otimes T^{-2}\Omega)$, all independent of ε and $Q = (X, Z)$ in the stable model (1).

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

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

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

(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 (7) and (10) converge weakly to the same limit under the null hypothesis, where b is replaced by the random vectors β and $\tilde{\beta}$ with distributions 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 (7) 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 ν_γ and $\nu_{\tilde{\gamma}}$ of γ and $\tilde{\gamma}$, respectively, since these measures are identical to those of β and $\tilde{\beta}$

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

Define

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

so that by the equivalence of the distribution of β and γ and of the distribution of $\tilde{\beta}$ and $\tilde{\gamma}$, $LR_T = \int \xi(b)\nu_\beta(b) = \int \xi(b)d\nu_\gamma(b)$ and $\widetilde{LR}_T = \int \tilde{\xi}(b)\nu_{\tilde{\beta}}(b) = \int \tilde{\xi}(b)d\nu_{\tilde{\gamma}}(b)$. Therefore

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

where β, $\tilde{\beta}$, γ and $\tilde{\gamma}$ in the second line are random vectors with distribution ν_β, $\nu_{\tilde{\beta}}$, ν_γ and $\nu_{\tilde{\gamma}}$, respectively. 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 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.

By exploiting the conditionally Gaussian likelihood structure of Condition 2 and the Gaussianity of $\tilde{\beta}$ and $\tilde{\gamma}$, we show that these products can be approximated arbitrarily accurately by uniformly integrable random variables.

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

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

under the null hypothesis (1).

In order to substantiate the claim of asymptotic equivalence of tests based on LR_T, $\widetilde{LR}_T$ and $\overline{LR}_T$, we still lack the crucial additional step of showing that the convergence in probability of Theorem 2 also holds under the alternative hypothesis. A brute force approach of running through the same arguments that led to Theorem 2 also for the alternative hypothesis is extremely cumbersome and barely tractable, since a nonconstant $\{\beta_t\}$ 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.

We therefore rather follow Andrews and Ploberger (1994) in taking the more indirect route of proving that the density of (y, Q) under the alternative hypothesis is *contiguous* to the density of (y, 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 describing 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 3 *Under Conditions 1 and 2, the sequence of densities* $\{f^1_{y,Q}(y,Q)\}_T$ *is contiguous to the densities* $\{f^0_{y,Q}(y,Q)\}_T$.

Corollary 1 *Under Conditions 1 and 2 the convergences in probability of Theorem 2 also hold under the alternative hypothesis (2).*

Theorem 2 and Corollary 1 imply Theorem 1 of Section 2 above. Since convergence in probability implies convergence in distribution, Theorem 1 implies 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 Condition 1 breaking process is not helpful for conducting a better test.

Additionally, given that any specific LR_T satisfies $LR_T - \widetilde{LR}_T \xrightarrow{p} 0$, the difference of any given pair of small sample optimal statistics for Condition 1 breaking processes also converge in probability to zero under the null hypothesis. The densities implied by these two breaking processes are both contiguous to the null density by Theorem 3, hence the convergence in probability continues to hold under both these alternatives. Theorems 2 and 3 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.[2] Each optimal test has asymptotically the same ability to distinguish each possible alternative in our class of models.

4 An Asymptotically Point-Optimal Test Statistic

The main result of this paper is the asymptotic equivalence of small sample efficient tests against alternatives that are covered by Condition 1. As a by-product of our analysis, we found the statistic qLL to be an asymptotically optimal test statistic against the processes of Condition 1. This makes qLL an attractive choice for applied work, and this section deals

[2]This statement is not true for arbitrary sequences of processes that satisfy Condition 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.

with the issues that led to our recommendation of basing inference on $\widehat{\text{qLL}}$ at the end of section 2.

As noted, qLL depends on the scaling parameter Ω that describes the average magnitude of the breaking process under the alternative. Clearly, the effect of Ω depends on the scale of $\{X_t\}$ and $\{\varepsilon_t\}$—the larger $\{X_t\}$ and the smaller $\{\varepsilon_t\}$, the larger the 'signal-to-noise' ratio of the time varying $\{\beta_t\}$ is going to be. In many regressions, the scale of $\{X_t\}$ is arbitrary (think of units of measurement), so that it makes sense to think of the 'signal-to-noise ratio' as measured by $\Omega^* = \sigma^{-2}\Sigma_X^{1/2}\Omega\Sigma_X^{1/2}$ in the rotated problem where regressors have identity covariance matrix and the disturbances have unit variance.

In fact, when one faces the problem of choosing a specific Ω for the construction of a test, a natural requirement is to demand that the outcome of a test of structural stability to be independent of the rotation the original regression is written in. A violation of such a requirement would lead to the counterintuitive possibility that the test outcome depends on the units of measurement of X_t. This is the same reasoning that led Wald (1943) to the construction of the usual F-statistic, and was employed in the construction of structural break tests by Nyblom (1989) and Andrews and Ploberger (1994). Rotational invariance is achieved by letting $\Omega^* = a^2 I_k$. As in Andrews and Ploberger (1994), the problem thus reduces to choosing a single parameter a that describes the distance between the alternative and the null hypothesis.

Not knowing this exact distance between the null and the alternative hypothesis will lead to losses in asymptotic power compared to the benchmark case. But drawing on the ideas of King (1988) regarding point optimal testing, it might be possible to find a certain choice of $a = \bar{a}$ in the construction of the test that makes these losses small. Figure 1 shows the asymptotic local power envelope for $k = 1$ and $k = 2$ along with the power of our recommended statistic $\widehat{\text{qLL}}$, which is asymptotically point-optimal for $\bar{a} = 10$. As can be seen, $\widehat{\text{qLL}}$ has asymptotic power very close to the power envelope for any distance from the null hypothesis. This also holds for larger k, the largest difference in power for $k = 5$ is smaller than four percentage points. If not theoretically, at least in practice the lack of knowledge of the distance of the alternative is hence a minor issue.

In applications, in addition to being powerful, the validity of a test statistic over a wide range of data generating processes is of major importance. 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 vectors with elements $Q_{T,t,i}$ and $\varepsilon_{T,t}$, respectively. With some $K < \infty$, assume that under the null*

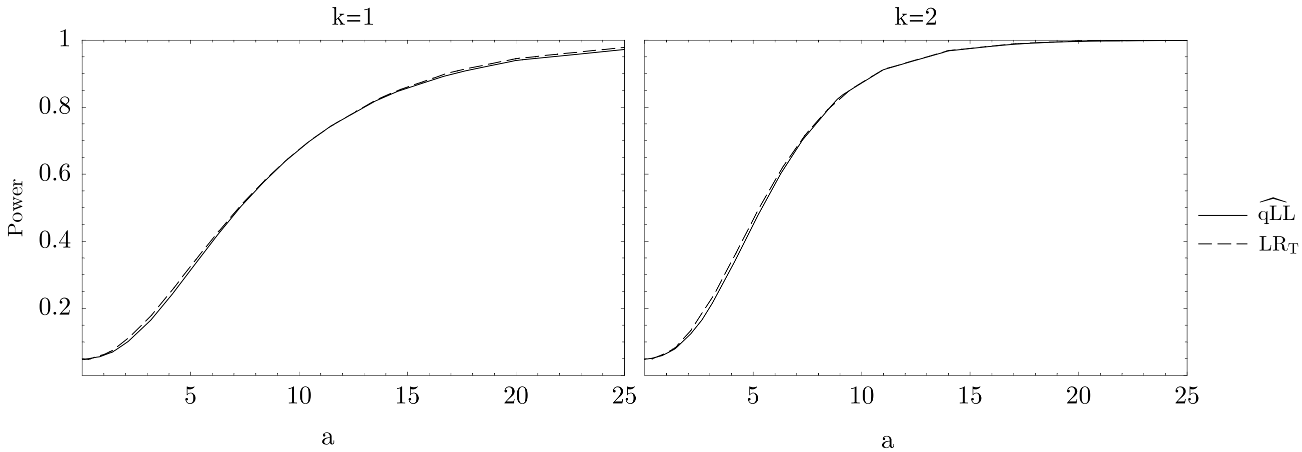

Figure 1: Local Asymptotic Power against Condition 1 Processes with $\Omega^* = a^2 I_k$

hypothesis (1)
(i) $E[Q_{T,t}\varepsilon_{T,t}] = 0$ *for all* T, t
(ii) $\{Q_{T,t}, \varepsilon_{T,t}\}$ *is either a uniform mixing sequence of size* $-r/(r-1)$ *or strong mixing sequence 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* $T^{-1}\sum_{t=1}^{T} Q_{T,t}Q'_{T,t}$ *is positive definite almost surely for all large enough* 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 qLL (which depends on $\sigma^{-1}\Sigma_X^{-1/2}$ through v_i as defined in Lemma 1) 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 $\hat{V}_X = T^{-1}\sum_{t=1}^{T} X_t X_t' \hat{\varepsilon}_t^2$. In the more general case of possibly autocorrelated $\{\varepsilon_t\}$, one might employ estimators of the form

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

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 $T \to \infty$. Alternatively, the long-run covariance matrix estimators studied by Andrews (1991) may be employed.

The feasible estimator $\widehat{\mathrm{qLL}}$ is hence defined as

$$\begin{aligned} \widehat{\mathrm{qLL}} &= \sum_{i=1}^{k} \hat{v}_i' [G_{\bar{a}} - M_e] \hat{v}_i \\ \hat{v}_i &= [I_T \otimes \iota_{k,i}' \hat{V}_X^{-1/2}] \Xi' M y \end{aligned}$$

with $\bar{a} = 10$ and the equivalence to the statistic described at the end of Section 2 follows after manipulations from the definition of $G_{\bar{a}}$. The asymptotic properties of $\widehat{\mathrm{qLL}}$ are investigated in the following Theorem.

Theorem 4 *Under Condition 3, the asymptotic null distribution of* $\widehat{\mathrm{qLL}}$ *is given in Lemma 2 with* $c_i = 10$, $i = 1, \cdots, k$.

In addition to being a point-optimal test statistic against Condition 1 breaking processes, $\widehat{\mathrm{qLL}}$ is hence asymptotically valid over a wide range of data-generating processes. Table 1 above contains asymptotic critical values of $\widehat{\mathrm{qLL}}$ for $k = 1, \cdots, 10$.

5 Monte Carlo Evidence

The main analytical result of this paper—the equivalence of a large class of efficient structural break tests—is asymptotic in nature. The question hence arises whether this asymptotic insight may serve as guide to small samples the econometrician faces in practice. We address this issue in this section for some simple designs, and find the asymptotic predictions to be quite accurate. In addition, we compare the small sample size and power properties of $\widehat{\mathrm{qLL}}$ with other popular structural break tests. Tests based on $\widehat{\mathrm{qLL}}$ turn out to have considerably superior size control, especially when disturbances are potentially heteroskedastic.

The small sample data generating process we consider consists of a constant and a stationary, mean-zero AR(1) process $\{\zeta_t\}$ with coefficient 0.5 and unit unconditional variance, and

$$y_t = \mu + \alpha\zeta_t + \varepsilon_t, \tag{14}$$

where the disturbances $\{\varepsilon_t\}$ are independent standard normal and independent of $\{\zeta_t\}$. In all experiments, we consider tests of 5% nominal level and a sample size of $T = 100$.

We investigate power against two types of parameter instability: On the one hand, a single zero mean Gaussian break whose date is uniformly distributed on $\{2, \cdots, T\}$, and on the other hand five persistent Gaussian breaks occurring at dates $\varsigma = \{11, 31, 51, 71, 91\}$, i.e. $\beta_t = \sum_{s=1}^{t} \mathbf{1}[s \in \varsigma]\varrho_s$ for independent Gaussian $\{\varrho_s\}$. In both cases, the variances of the breaks are normalized such that under the alternative denoted by c, $(\beta_T - \beta_0)$ has a variance of $c^2/T^2 I_k$. We consider the three cases where only μ, only α or both μ and α are time varying under the alternative.

These small sample assumptions on $\{\beta_t\}$ can be embedded in different asymptotic thought experiments, which determines whether Condition 1 is satisfied. In the thought experiment with a single single break under the alternative for all T, Condition 1 does not hold, and the same is true for a breaking process with five breaks independent of T. But a breaking process $\{\beta_t\}$ that is subject to a mean zero Gaussian break every 20 periods does satisfy Condition 1, as well as a process $\{\beta_t\}$ that is subject to mean zero Gaussian increments that are randomly distributed over all 100 observation segments. The relevance of the asymptotic equivalence results based on Condition 1 for small samples hinges on whether an embedding of such small sample breaking processes in Condition 1 asymptotics yields useful approximations.

This issue is investigated by a comparison of the small sample power of three infeasible tests: The small sample efficient test based on LR_T as defined in (7), the small sample efficient invariant test based on LR_T^I as defined in (9), and the test based on $\widetilde{LR}_T$ as defined in (10). These tests are infeasible, since they rely on knowledge of $\sigma^2 = E[\varepsilon_t^2]$, $\Omega = c^2 I_k$ and, in addition, on knowledge of μ and α in the case of LR_T, and on knowledge of Σ_X in the case of $\widetilde{LR}_T$. Note that LR_T and LR_T^I efficiently exploit knowledge of the number of breaks under the alternative, and the break dates ς in the case of the five break process. The statistic $\widetilde{LR}_T$, in contrast, can be thought of as efficiently testing for the presence of a Gaussian random walk of known variance in the mean of $\{y_t X_t\}$. Theorem 1 and the discussion of LR_T^I in Section 3 imply that for large enough T and breaks every 20 periods, the difference in power of these three tests converges to zero. The small sample results of

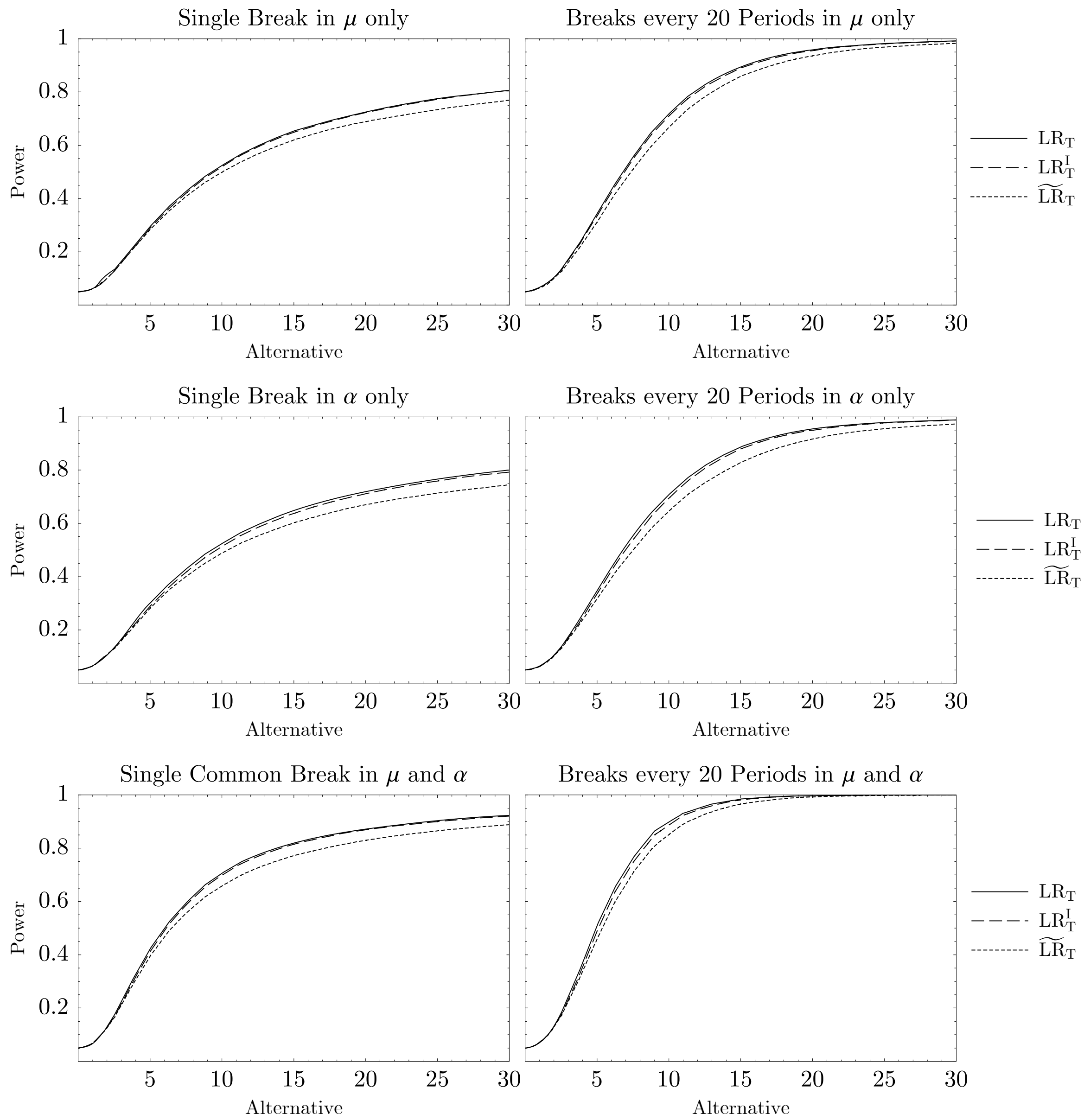

Figure 2: Small Sample Power of Infeasible Tests based on LR_T, LR_T^I and $\widetilde{LR}_T$

Figure 2, based on 20,000 replications,[3] shows this to be a useful prediction. In fact, if one considers only the two invariant statistics LR_T^I and $\widetilde{LR}_T$, whose computation do not require knowledge of α and μ, the difference in power never exceeds five percentage points. A similar result holds for the single break case, even though a single break is arguably the most extreme deviation from the break every period case $\widetilde{LR}_T$ is constructed against. These results show that *for testing purposes*, the two considered breaking processes can be well approximated by a Gaussian Random walk.

We now turn to the size and power performance of feasible tests. Specifically, we compare $\widehat{\text{qLL}}$ to tests that have been especially constructed for a single break at an unknown date: the supF statistic (Andrews, 1993), and the Andrews and Ploberger (1996) exponentially weighted F-statistics (AP∞) for independent normal disturbances. In addition, we include the Nyblom (1989) statistic in our experiments, denoted Ny.

In most applications, heteroskedasticity cannot be ruled out a priori, so that we follow Stock and Watson (1996) and consider both heteroskedasticity robust and non-robust versions of all statistics. For the non-heteroskedasticity robust version of AP∞ and supF, the sequence of Chow F-statistics is computed without a heteroskedasticity correction and $\hat{V}_X$ in $\widehat{\text{qLL}}$ and Ny is given by $\hat{V}_X = (T-2)^{-1}\left(\sum \hat{\varepsilon}_t^2\right) T^{-1} \sum X_t X_t'$, where $\{\hat{\varepsilon}_t\}$ are the OLS residuals in (14). The heteroskedasticity robust versions employ the White (1980) correction in the construction of the sequence of F-statistics and use $\hat{V}_X = (T-2)^{-1} \sum X_t X_t' \hat{\varepsilon}_t^2$ for $\widehat{\text{qLL}}$ and Ny. Following Andrews (1993) and Stock and Watson (1996), we chose a 15% trimming in the construction of AP∞ and supF. We experimented with less trimming, and found comparable power but substantially worse size control.

Table 2 shows the empirical small sample size of these statistics, using 20,000 repetitions. We consider model (14) with homoskedastic disturbances (HOMO), as well as with heteroskedastic disturbances (HET), where $\{\varepsilon_t\}$ in (14) is given by $\{\varepsilon_t\} = \{|\zeta_t|\tilde{\varepsilon}_t\}$ for i.i.d. standard normal $\{\tilde{\varepsilon}_t\}$. When the disturbances are homoskedastic (HOMO), size control is very good for all statistics that are constructed without the heteroskedasticity correction. The heteroskedasticity robust versions of AP∞ and supF, however, are substantially oversized even when the disturbances are homoskedastic, especially in the case where the stability of both μ and α are tested. With heteroskedastic disturbances (HET), non-heteroskedasticity robust tests perform very poorly when the stability of α is examined. Unsurprisingly, the

[3] The statistics LR_T and LR_T^I are efficient test statistics conditional on the realization of $\{\zeta_t\}$. Accordingly, Figures 2–4 show the average power of conditionally size adjusted tests. The Figures are based on 500 repetitions for 40 independent draws of $\{\zeta_t\}$. In general, power is not sensitive to the realization of $\{\zeta_t\}$.

Table 2: Empirical Small Sample Size in Percent

	heteroskedasticity non-robust				heteroskedasticity robust			
DGP	$\widehat{\text{qLL}}$	Ny	AP∞	supF	$\widehat{\text{qLL}}$	Ny	AP∞	supF
$X_t = 1, Z_t = \zeta_t$,HOMO	4.4	4.4	5.5	4.7	4.4	4.4	6.6	6.4
$X_t = \zeta_t, Z_t = 1$,HOMO	5.3	4.2	5.5	4.1	4.5	4.2	10.0	10.3
$X_t = (1, \zeta_t)'$,HOMO	5.1	3.9	5.9	4.6	4.6	3.8	14.2	15.3
$X_t = 1, Z_t = \zeta_t$,HET	4.0	4.0	4.9	3.9	4.0	4.0	5.6	4.3
$X_t = \zeta_t, Z_t = 1$,HET	69.4	30.3	42.5	42.4	4.3	3.9	13.9	14.0
$X_t = (1, \zeta_t)'$,HET	53.0	22.3	35.1	33.9	4.8	3.5	17.4	17.7

heteroskedasticity robust tests control size much better, but $\widehat{\text{qLL}}$ and Ny stand out as being by far the best performers in this regard.

Figures 3 and 4 show size adjusted power of the feasible statistics, along with the power envelope based on the infeasible optimal invariant statistic LR_T^I. The size adjusted power of the feasible test statistics is close throughout. Compared to Figure 2, the loss in power of $\widehat{\text{qLL}}$ compared to the benchmark LR_T^I is somewhat greater, especially for the heteroskedasticity robust version of $\widehat{\text{qLL}}$, but the absolute loss is still quite moderate.

Overall, these results underline the relevance of the asymptotic result derived in Section 3. At the same time, we find considerably better size control properties and moderately higher power of $\widehat{\text{qLL}}$ compared to AP∞ and supF in many scenarios, making $\widehat{\text{qLL}}$ an attractive choice for applied work.

6 Conclusions

Permanent parameter instability at unknown dates is interesting economically, causes problems for forecasting and typically invalidates 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 permanent break every other period, for instance, will have power also against an alternative with a permanent break every period. We show not only that this intuition is correct, but a much stronger claim: Conditional on the average magnitude of the breaks being the same, 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

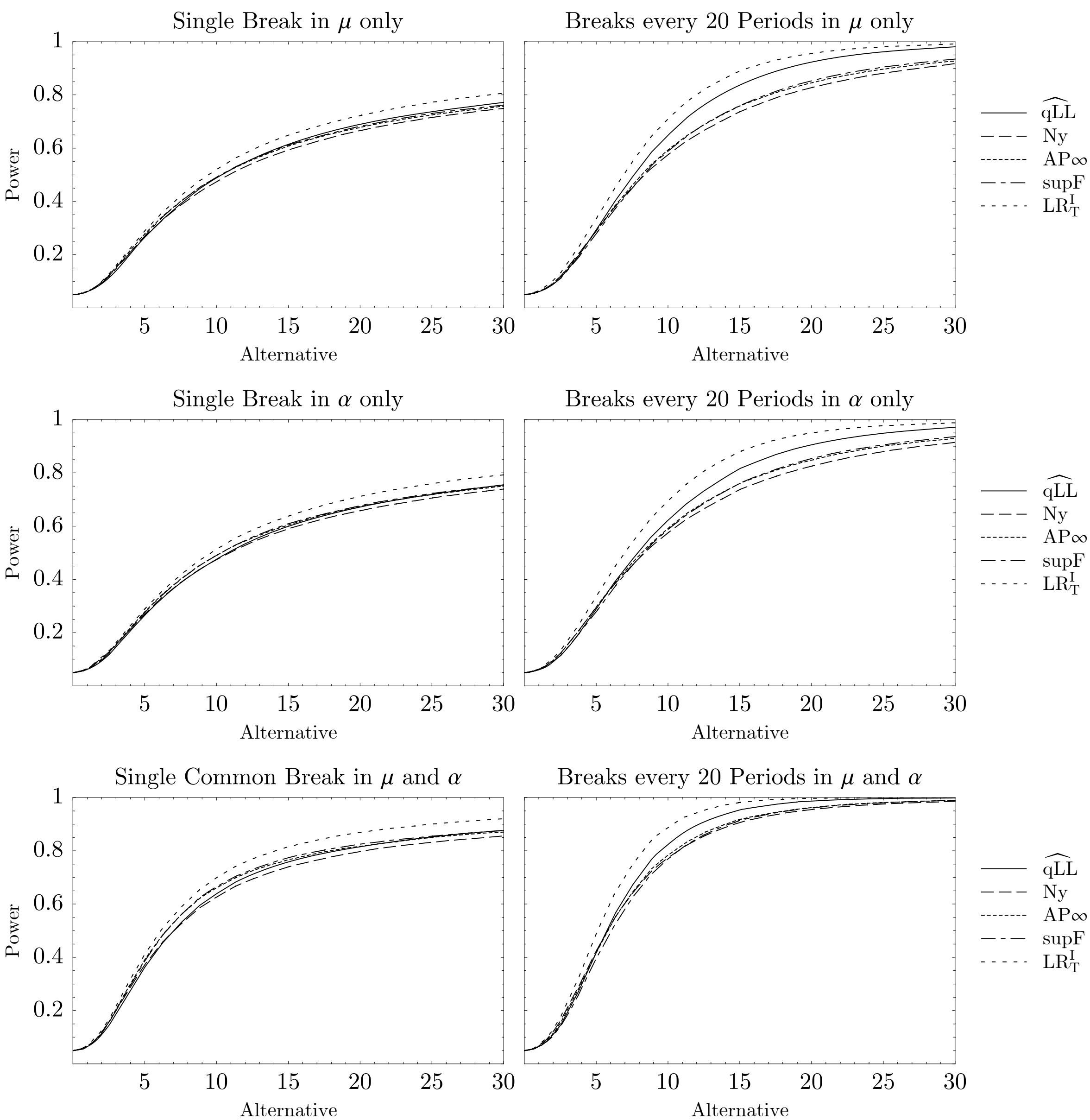

Figure 3: Small Sample Size-Adjusted Power of Non-Heteroskedasticity Robust Tests

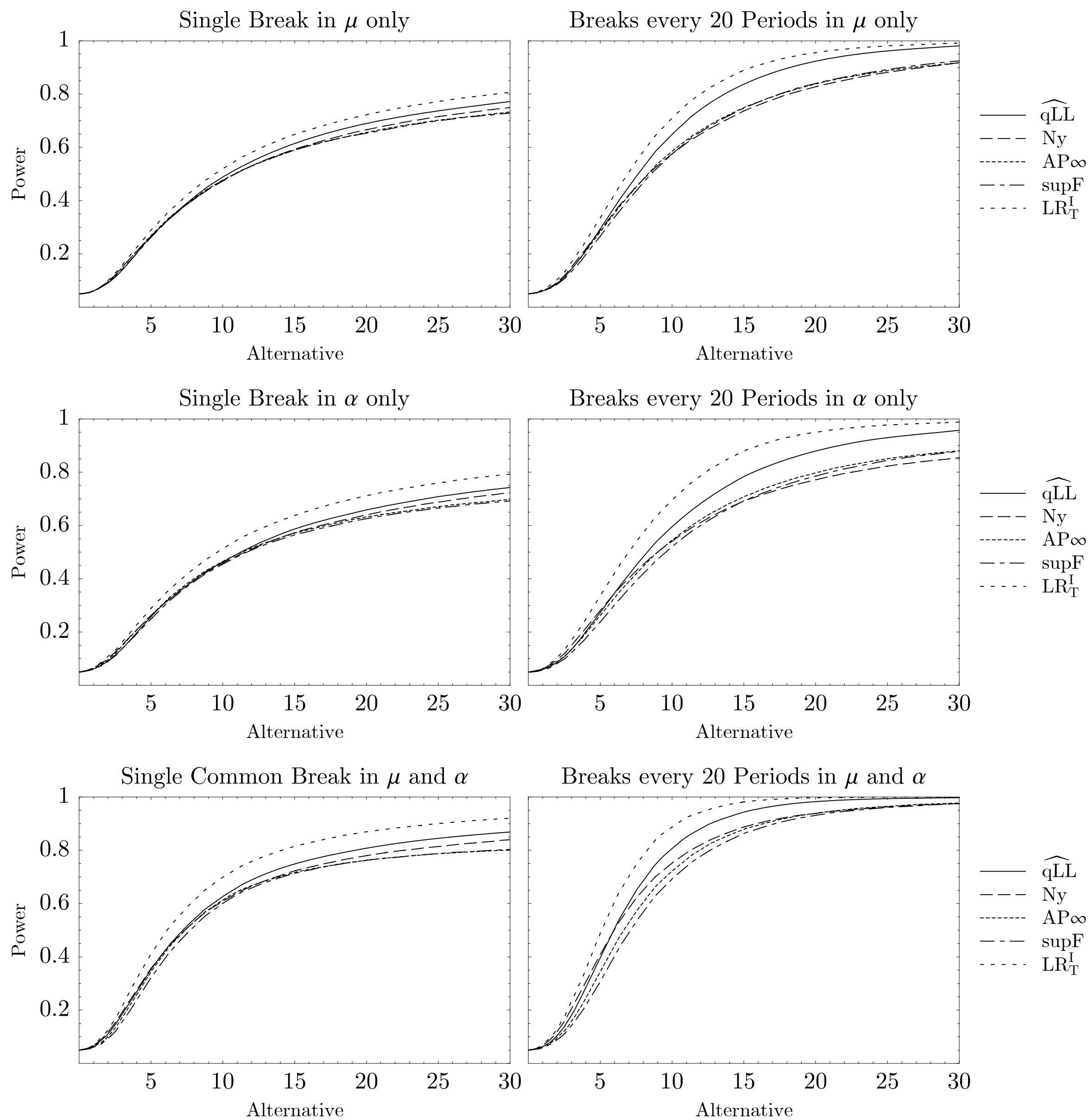

Figure 4: Small Sample Size-Adjusted Power of Heteroskedasticity Robust Tests

a break every period, at least for a large enough sample size. This (asymptotic) equivalence extends over a large class of persistent 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 close over a wide range of 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 five breaks rejects, say, then this does by no means imply that the true breaking process consists in fact of five breaks. While for non-local 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 is more difficult (see Elliott and Müller (2004)).

Third, for a large class of mean zero, persistent breaking processes, 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 is asymptotically point-optimal for the class of breaking processes we focus on. We find tests based on this statistic to have very good small-sample size and power properties, 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, $\varepsilon_t^* = \sigma^{-1}\varepsilon_t$ and $\beta^* = [I_T\otimes\sigma^{-1}\Sigma_X^{1/2}](\beta - [e\otimes I_k]\beta_1)$. Note that the long-run variance of $\{\Delta\beta_t^*\}$ 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 $T\times(T-1)$ matrix that satisfies $B_e'B_e = I_{T-1}$ and $B_e'e = 0$, so that $B_eB_e' = M_e$. 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)*

$$B_e'(a^2T^{-2}FF' + I_T)B_e = B_e'H_aB_e$$

(ii)

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

(iii)

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

Proof. (i) We have

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

From a direct calculation

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

and

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

so that

$$\begin{aligned} B_e'(a^2T^{-2}FF' + I_T)B_e &= B_e'F(r_a^{-1}A_aA_a')F'B_e \\ &= B_e'H_aB_e. \end{aligned}$$

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

(iii) From (ii)

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

yielding $|B_e'H_aB_e| = |B_e'G_aB_e|^{-1}$. Now note that $(T^{-1/2}e, B_e)'(T^{-1/2}e, B_e) = I_T$, 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

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

Now

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

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

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

Therefore

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

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

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 $(d+k) \times 1$ standard Wiener process. Then*
(i) $T^{-1} \sum_{t=1}^{T} (Q_t^* Q_t^{*\prime} - I_{k+d}) \upsilon_t \xrightarrow{p} 0$ and
(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

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

for any $i, j \in \{1, \cdots, d+k\}$, where $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. 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

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

where $A_{\upsilon,j}$ is the jth row of A_υ and the last line follows from the Continuous Mapping Theorem (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 $E[Q^*_t Q^{*\prime}_t] = I_{k+d}$ and the independence of $\{Q^*_t\}$ and $\{\upsilon_t\}$, $E[(Q^*_{t,i}Q^*_{t,j}-\delta_{i,j})\tilde{\upsilon}_{t,j}] = 0$. Since $\{Q^*_{t,i}\}$ and $\{Q^*_{t,j}\}$ are L_r-bounded and $\sup_{t\leq T}|\tilde{\upsilon}_{t,j}| \leq K_\upsilon$ a.s., $\{(Q^*_{t,i}Q^*_{t,j}-\delta_{i,j})\tilde{\upsilon}_{t,j}\}$ is $L_{r/2}$-bounded

$$\begin{aligned} E[|(Q^*_{t,i}Q^*_{t,j}-\delta_{i,j})\tilde{\upsilon}_{t,j}|^{r/2}]^{2/r} &\leq K_\upsilon E[|Q^*_{t,i}Q^*_{t,j}-\delta_{i,j}|^{r/2}]^{2/r} \\ &\leq K_\upsilon(E[|Q^*_{t,i}|^r]^{1/r}E[|Q^*_{t,j}|^r]^{1/r}+\delta_{i,j}) \leq K_\upsilon K_Q \end{aligned}$$

for some $K_Q < \infty$ for all t and T, where the second inequality follows from the triangle and Cauchy-Schwarz inequalities. 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 $|E[(Q^*_{t,i}Q^*_{t,j}-\delta_{i,j})\tilde{\upsilon}_{t,j}|\mathfrak{F}^*_{t-m}]|$.

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

$$\begin{aligned} |E[(Q^*_{t,i}Q^*_{t,j}-\delta_{i,j})\tilde{\upsilon}_{t,j}|\mathfrak{F}^*_{t-m}]| &\leq K_\upsilon|E[(Q^*_{t,i}Q^*_{t,j}-\delta_{i,j})|\mathfrak{F}^*_{t-m}]| \\ &\leq 6K_\upsilon\alpha_m^{1-2/r}E[|Q^*_{t,i}Q^*_{t,j}-\delta_{i,j}|^{r/2}]^{2/r} \\ &\leq 6K_\upsilon\alpha_m^{1-2/r}K_Q \end{aligned}$$

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, $\{(Q^*_{t,i}Q^{*\prime}_{t,j}-\delta_{i,j})\tilde{\upsilon}_{t,j}, \mathfrak{F}^*_t\}$ 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

$$|E[(Q^*_{t,i}Q^*_{t,j}-\delta_{i,j})\tilde{\upsilon}_{t,j}|\mathfrak{F}^*_{t-m}]| \leq 2K_\upsilon\phi_m^{1-2/r}K_Q$$

with ϕ_m the m^{th} uniform mixing coefficient. Since $\phi_m = O(m^{-r/(2r-2)-\epsilon})$ for some $\epsilon > 0$, we find $\phi_m^{1-2/r} = O(m^{-(r-2)/(2r-2)-\epsilon'})$ 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 $-(r-2)/(2r-2)$ 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, and hence in probability. Since the probability of the truncation (15) can be made arbitrarily small by choosing K_υ large for large enough T, the result follows.

(ii) The proof is analogous to part (i), the only difference is that now the j,lth element $[\upsilon_t\upsilon'_t]_{j,l}$ 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 for large enough T. ■

Lemma 6 *Let the $T\times 1$ vector $u = (u_1, \cdots, u_T)'$ be such that $T^{-1/2}\sum_{t=1}^{[T\cdot]}u_t \Rightarrow W_u(\cdot)$, where W_u is a standard scalar Wiener process. Then*

$$u'[G_c - M_e]u \Rightarrow -cJ_u(1)^2 - c^2\int J_u^2 - \frac{2c}{1-e^{-2c}}\left[e^{-c}J_u(1) + c\int e^{-cs}J_u\right]^2 + \left[J_u(1)+c\int J_u\right]^2$$

where $J_u(s) = W_u(s) - c\int_0^s e^{-c(s-\lambda)}W_u(\lambda)d\lambda$.

Proof. Write $u'[G_c - M_e]u = u'(H_c^{-1} - I_T)u - u'H_c^{-1}e(e'H_c^{-1}e)^{-1}e'H_c^{-1}u + (T^{-1/2}e'u)^2$. 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 $A_c^{-1}LA_c = L$. For the first term, we compute

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

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

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

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

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

For the final term $T^{-1/2}e'u = T^{-1/2}e'A_cB = T^{-1/2}B_T + T(1-r_c)T^{-3/2}e'B_{-1}$. The Lemma now follows from the joint convergence of $T^{-1/2}B_T \Rightarrow J_u(1)$, $T^{-2}\sum_{t=1}^{T}B_{t-1}^2 \Rightarrow \int J_u^2$, $T^{-1}\sum_{t=1}^{T-1}r_c^tB_t \Rightarrow \int e^{-cs}J_u(s)ds$, $T^{-1}\sum_{t=1}^{T}B_{t-1} \Rightarrow \int J_u$, $r_c^T \to e^{-c}$ and the CMT. ■

Proof of Lemma 2:

Since under Condition 2, $\{Q_t^*\varepsilon_t^*\}$ is a mixing sequence, $E[|Q_{t,j}^*\varepsilon_t^*|] = E[|Q_{t,j}^*|]E[|\varepsilon_t^*|]$ is uniformly bounded in T for $j = 1, \cdots, k+d$ and the long-run variance of $\{Q_t^*\varepsilon_t^*\}$ is given by $E[Q_t^*Q_t^{*\prime}(\varepsilon_t^*)^2] = \sigma^{-2}E[CQ_tQ_t'C'\varepsilon_t^2] = I_{k+d}$ by the law of iterated expectations, we find that the sum of the first $[sT]$ $k \times 1$ vectors of $\sigma^{-1}[I_T \otimes P^{*\prime}\Sigma_X^{-1/2}]\Xi'M\varepsilon$ satisfies

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

where $\tilde{W}_\varepsilon$ is a $k \times 1$ Wiener process and $W_\varepsilon = P^{*\prime}\tilde{W}_\varepsilon$ from a Functional Central Limit Theorem (FCLT) for mixing sequences as in White (2001), p. 189 and $\left(T^{-1}\sum_{t=1}^{[sT]} X_t^*Q_t^{*\prime} - sT^{-1}\sum_{t=1}^{T} X_t^*Q_t^{*\prime}\right) \xrightarrow{p} 0$ by the uniform convergence of $T^{-1}\sum_{t=1}^{[sT]} Q_tQ_t' \xrightarrow{p} s\Sigma_Q$ in s. Note that since P^* is orthonormal, W_ε is a standard Wiener process, too.

Since $v_i'(G_{a_i} - M_e)v_i = (v_i + q_i e)'(G_{a_i} - M_e)(v_i + q_i e)$ for any choice of scalar q_i, we find with $q_i = T^{-1}\iota_{k,i}' P^{*\prime} X^{*\prime}\varepsilon^*$ (so that $T^{1/2}q_i \Rightarrow W_{\varepsilon,i}(1)$)

$$\begin{aligned} T^{-1/2}(e_{[sT]}', 0_{T-[sT]}')(v_i + q_i e) &= T^{-1/2}(e_{[sT]}', 0_{T-[sT]}')[I_T \otimes \iota_{k,i}' P^{*\prime}]\Xi^{*\prime} M^* \varepsilon^* + [sT]T^{-3/2}\iota_{k,i}' P^{*\prime} X^{*\prime}\varepsilon^* \\ &\Rightarrow W_{\varepsilon,i}(s) - sW_{\varepsilon,i}(1) + sW_{\varepsilon,i}(1) = W_{\varepsilon,i}(s). \end{aligned}$$

An application of Lemma 6 with $u = v_i + q_i e$, $i = 1, \cdots, k$ and the CMT now yield the result.

Proof of Lemma 3:

As in the proof of Lemma 2, we rely on a FCLT 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. We explicitly consider terms involving β and $\tilde{\beta}$ only, the identical distributions of γ and $\hat{\gamma}$ obviously lead to the analogous results. Let $\sum$ stand for summation over $t = 1, \cdots, T$.

(i)

$$\begin{aligned} \sigma^{-2}\varepsilon'\Xi[M_e \otimes I_k]\beta &= \varepsilon^{*\prime}\Xi^*[M_e \otimes I_k]\beta^* \\ &= \varepsilon^{*\prime}\Xi^*\beta^* - T^{-1}\varepsilon^{*\prime}X^*[e' \otimes I_k]\beta^* \end{aligned}$$

But

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

where $W_\varepsilon = P^{*\prime}\tilde{W}_\varepsilon$, and

$$\begin{aligned} T^{-1/2}X^{*\prime}\varepsilon^* &= T^{-1/2}\sum X_t^*\varepsilon_t^* \Rightarrow P^* W_\varepsilon(1) \\ T^{-1/2}[e' \otimes I_k]\beta^* &= T^{-1/2}\sum \beta_t^* \Rightarrow P^*\Lambda^{1/2}\int W_\beta \end{aligned}$$

so that by the CMT,

$$\sigma^{-2}\varepsilon'\Xi[M_e \otimes I_k]\beta \Rightarrow \int W_\beta'\Lambda^{1/2}dW_\varepsilon - (\int W_\beta')\Lambda^{1/2}W_\varepsilon(1) = \int \bar{W}_\beta'\Lambda^{1/2}dW_\varepsilon.$$

(ii)

$$\begin{aligned} \sigma^{-2}\beta'[M_e \otimes I_k]\Xi'\Xi[M_e \otimes I_k]\beta = \beta^{*\prime}\Xi^{*\prime}\Xi^*\beta^* - T^{-1}2\beta^{*\prime}(e \otimes I_k)X^{*\prime}\Xi^*\beta^* \\ + T^{-2}\beta^{*\prime}(e \otimes I_k)X^{*\prime}X^*(e \otimes I_k)'\beta^* \end{aligned}$$

Now

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

using part (ii) of Lemma 5, and similarly

$$T^{-1/2}X^{*\prime}\Xi^*\beta^* = T^{-1/2}\sum X_t^*X_t^*\beta_t^* \Rightarrow P^*\Lambda^{1/2}\int W_\beta$$

so that, with the results established in part (i) above, $T^{-1}X^{*\prime}X^{*} \xrightarrow{p} I_k$ and the CMT,

$$\sigma^{-2}\beta'[M_e \otimes I_k]\Xi'\Xi[M_e \otimes I_k]\beta \Rightarrow \int W_\beta'\Lambda W_\beta - (\int W_\beta)'\Lambda(\int W_\beta) = \int \bar{W}_\beta'\Lambda\bar{W}_\beta.$$

(iii) Define $\tilde{\beta}^*$ in analogy to β^*.

$$\sigma^{-2}\varepsilon' M\Xi[M_e \otimes I_k]\tilde{\beta} = \varepsilon^{*\prime}\Xi^*[M_e \otimes I_k]\tilde{\beta}^* - \varepsilon^{*\prime}Q^*(Q^{*\prime}Q^*)^{-1}Q^{*\prime}\Xi^*[M_e \otimes I_k]\tilde{\beta}^*.$$

From the same reasoning as in part (i), $\varepsilon^{*\prime}\Xi^*[M_e \otimes I_k]\tilde{\beta}^* \Rightarrow \int \bar{W}_{\tilde{\beta}}'\Lambda^{1/2}dW_\varepsilon$. Furthermore, with $W_{Z\varepsilon}$ a $d \times 1$ standard Wiener process independent of W_ε,

$$\begin{aligned}
T^{1/2}\varepsilon^{*\prime}Q^*(Q^{*\prime}Q^*)^{-1} &= \left(T^{-1/2}\sum Q_t^*\varepsilon_t^*\right)'(T^{-1}Q^{*\prime}Q^*)^{-1} \Rightarrow \begin{pmatrix} P^*W_\varepsilon(1) \\ W_{Z\varepsilon}(1) \end{pmatrix}' \\
T^{-3/2}Q^{*\prime}\Xi^*[ee' \otimes I_k]\tilde{\beta}^* &= \left(T^{-1}\sum Q_t^*X_t^{*\prime}\right)\left(T^{-1/2}\sum \tilde{\beta}_t^*\right) \Rightarrow \begin{pmatrix} P^*\Lambda^{1/2}\int W_{\tilde{\beta}} \\ 0 \end{pmatrix} \\
T^{-1/2}Q^{*\prime}\Xi^*\tilde{\beta}^* &= T^{-1/2}\sum Q_t^*X_t^{*\prime}\tilde{\beta}_t^* \Rightarrow \begin{pmatrix} P^*\Lambda^{1/2}\int W_{\tilde{\beta}} \\ 0 \end{pmatrix}
\end{aligned}$$

yielding the result.

(iv) From the same reasoning as in part (i) and the CMT,

$$\begin{aligned}
\sigma^{-2}\tilde{\beta}'[M_e \otimes \Sigma_X]\tilde{\beta} &= \operatorname{tr}\left[\sum \tilde{\beta}_t^*\tilde{\beta}_t^{*\prime}\right] - \left(T^{-1/2}\sum \tilde{\beta}_t^{*\prime}\right)\left(T^{-1/2}\sum \tilde{\beta}_t^*\right) \\
&\Rightarrow \int W_{\tilde{\beta}}'\Lambda W_{\tilde{\beta}} - (\int W_{\tilde{\beta}})'\Lambda(\int W_{\tilde{\beta}}) = \int \bar{W}_{\tilde{\beta}}'\Lambda\bar{W}_{\tilde{\beta}}.
\end{aligned}$$

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

Proof of Theorem 2:

All computations in the proof are made in the stable model (1), i.e. under the assumption that $h = \varepsilon$. Denote by E_β, E_γ, $E_{\tilde{\beta}}$ and $E_{\tilde{\gamma}}$ integration with respect to ν_β, ν_γ, $\nu_{\tilde{\beta}}$ and $\nu_{\tilde{\gamma}}$, so that $LR_T = E_\beta\xi(\beta)$ and $\widetilde{LR}_T = E_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta})$. Let $\phi = \sum_{i=1}^k v_i'[G_{2a_i} - M_e]v_i$ and $\hat{\Sigma}_{X^*} = T^{-1}\sum_{t=1}^T X_t^*X_t^{*\prime}$, and for real constants K_G and K_S, define the two indicator functions

$$\begin{aligned}
G_T &= \mathbf{1}[\phi < K_G]\mathbf{1}[\operatorname{tr}\hat{\Sigma}_{X^*} < k+1] \\
S_T(b) &= \mathbf{1}[\sup_{t \le T} Tb_t'b_t < K_S].
\end{aligned}$$

Further define

$$\begin{aligned}
\widetilde{LR}_{G,T} &= G_T E_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta}) \\
\widetilde{LR}_{S,T} &= G_T E_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta})S_T(\tilde{\beta}^*) \\
LR_{G,T} &= G_T E_\beta\xi(\beta) \\
LR_{S,T} &= G_T E_\beta\xi(\beta)S_T(\beta^*).
\end{aligned}$$

Note that

$$\begin{aligned}
P(|LR_T - \widetilde{LR}_T| > 5\epsilon) \le{}& P(|LR_T - LR_{G,T}| > \epsilon) + P(|LR_{G,T} - LR_{S,T}| > \epsilon) \\
&+ P(|LR_{S,T} - \widetilde{LR}_{S,T}| > \epsilon) + P(|\widetilde{LR}_{S,T} - \widetilde{LR}_{G,T}| > \epsilon) + P(|\widetilde{LR}_{G,T} - \widetilde{LR}_T| > \epsilon).
\end{aligned}$$

We hence need to show: (i) for any $\epsilon, \eta > 0$ there exists T^*, K_G and K_S such that for all $T \geq T^*$, $P(|LR_T - LR_{G,T}| > \epsilon) < \eta$, $P(|LR_{G,T} - LR_{S,T}| > \epsilon) < \eta$, $P(|\widetilde{LR}_{S,T} - \widetilde{LR}_{G,T}| > \epsilon) < \eta$ and $P(|\widetilde{LR}_{G,T} - \widetilde{LR}_T| > \epsilon) < \eta$ and (ii) for all K_G and K_S, $LR_{S,T} - \widetilde{LR}_{S,T} \xrightarrow{p} 0$.

We show (i) first. Now

$$\begin{aligned} P(|\widetilde{LR}_{G,T} - \widetilde{LR}_T| &> \epsilon) \leq P(\phi \geq K_G) + P(\operatorname{tr}\hat{\Sigma}_{X^*} \geq k+1) \\ P(|LR_T - LR_{G,T}| &> \epsilon) \leq P(\phi \geq K_G) + P(\operatorname{tr}\hat{\Sigma}_{X^*} \geq k+1). \end{aligned}$$

Since $\hat{\Sigma}_Q \xrightarrow{p} \Sigma_Q$, $P(\operatorname{tr}\hat{\Sigma}_{X^*} \geq k+1) \to 0$. Also, 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 $c_i = 2a_i$, $i = 1, \cdots, k$, so that by choosing K_G large enough, $P(\phi \geq K_G)$ can be made smaller than η for sufficiently large T.

Before proceeding further, note that by computations close to those in the proof of Lemma 1,

$$\begin{aligned} E[G_T \tilde{\xi}(\tilde{\beta})^4] &= EG_T E_{\tilde{\beta}} \exp\left[4\sigma^{-2}\varepsilon' M \Xi [M_e \otimes I_k]\tilde{\beta} - 2\sigma^{-2}\tilde{\beta}'[M_e \otimes \Sigma_X]\tilde{\beta} \right] \\ &= EG_T \int (2\pi)^{-k(T-1)/2} |K_\Omega|^{-1/2} \\ &\quad \exp\left[4\sigma^{-2}\varepsilon' M \Xi [B_e \otimes I_k] b_e - \tfrac{1}{2} b_e' [K_\Omega^{-1} + 4I_{T-1} \otimes \sigma^{-2}\Sigma_X] b_e \right] db_e \\ &= EG_T |K_\Omega|^{-1/2} |K_\Omega^{-1} + 4I_{T-1} \otimes \sigma^{-2}\Sigma_X|^{-1/2} \\ &\quad \exp\left[8\sigma^{-4}\varepsilon' M \Xi [B_e \otimes I_k][K_\Omega^{-1} + 4I_{T-1} \otimes \sigma^{-2}\Sigma_X]^{-1}[B_e' \otimes I_k]\Xi' M \varepsilon \right] \\ &= EG_T |I_{T-1} \otimes I_k + K_{4\Lambda}|^{-1/2} \\ &\quad \exp\left[8\sigma^{-2}\varepsilon' M \Xi [B_e \otimes \Sigma_X^{-1/2} P^*][K_{4\Lambda}^{-1} + I_{T-1} \otimes I_k]^{-1}[B_e' \otimes P^{*\prime}\Sigma_X^{-1/2}]\Xi' M \varepsilon \right] \\ &= EG_T \exp[8\phi] \prod_{i=1}^{k} \left[\frac{1 - r_{2a_i}^{2T}}{T(1 - r_{2a_i}^2) r_{2a_i}^{T-1}} \right]^{-1/2} \\ &\leq \exp[8K_G] \prod_{i=1}^{k} \left[\frac{1 - r_{2a_i}^{2T}}{T(1 - r_{2a_i}^2) r_{2a_i}^{T-1}} \right]^{-1/2} \end{aligned}$$

so that there exists a constant K' (that depends on K_G) such that $\sup_T E[G_T \tilde{\xi}(\tilde{\beta})^4] < K'$.

With this result, by applying Markov's and Hölder's inequalities

$$\begin{aligned} P(|\widetilde{LR}_{S,T} - \widetilde{LR}_{G,T}| &> \epsilon) \leq \epsilon^{-1} E\tilde{\xi}(\tilde{\beta}) G_T (1 - S_T(\tilde{\beta}^*)) \\ &\leq \epsilon^{-1} [E(1 - S_T(\tilde{\beta}^*))^{4/3}]^{3/4} [E\tilde{\xi}(\tilde{\beta})^4 G_T]^{1/4} \\ &\leq \epsilon^{-1} P(\sup_t T \tilde{\beta}_t^{*\prime} \tilde{\beta}_t^* \geq K_S)^{3/4} (K')^{1/4}. \end{aligned}$$

Also, applying again Markov's inequality

$$
\begin{aligned}
P(|LR_{G,T} - LR_{S,T}| &> \epsilon) \leq \epsilon^{-1} E\xi(\beta) G_T (1 - S_T(\beta^*)) \\
&\leq \epsilon^{-1} E\xi(\beta)(1 - S_T(\beta^*)) \\
&= \epsilon^{-1} \int (1 - S_T((I_T \otimes \sigma^{-1}\Sigma_X^{1/2})(b - (e \otimes I_k) b_1)) E[\frac{f_{y,Q|\beta=b}(y,Q)}{f^0_{y,Q}(y,Q)}] d\nu_\beta(b) \\
&= \epsilon^{-1} E(1 - S_T(\beta^*)) \\
&= \epsilon^{-1} P(\sup_t T\beta_t^{*\prime}\beta_t^* \geq K_S)
\end{aligned}
$$

where the interchange of the order of integration in the third line is allowed by Fubini's Theorem, and the second equality follows from the fact that the density $f_{y,Q|\beta=b}(y,Q)$ integrates to unity for all b. Since $P(\sup_t T\tilde{\beta}_t^{*\prime}\tilde{\beta}_t^* \geq K_S) \to P(\sup_s W_{\tilde{\beta}}(s)'\Lambda W_{\tilde{\beta}}(s) \geq K_S)$ and $P(\sup_t T\beta_t^{*\prime}\beta_t^* \geq K_S) \to P(\sup_s W_\beta(s)'\Lambda W_\beta(s) \geq K_S)$, by choosing K_S large enough, $P(|\widetilde{LR}_{S,T} - \widetilde{LR}_{G,T}| > \epsilon)$ and $P(|LR_{G,T} - LR_{S,T}| > \epsilon)$ can hence be made smaller than η for large enough T.

We are left to show (ii), i.e. that $LR_{S,T} - \widetilde{LR}_{S,T} \xrightarrow{p} 0$ for all $0 < K_G < \infty$ and $0 < K_S < \infty$. We will show below that $\sup_T E[(LR_{S,T})^4] < \infty$ and $\sup_T E[(\widetilde{LR}_{S,T})^4] < \infty$, which implies that $(LR_{S,T} - \widetilde{LR}_{S,T})^2$ is uniformly integrable. Let $\psi(\beta) = \xi(\beta)S_T(\beta^*)$ and $\tilde{\psi}(\tilde{\beta}) = \xi(\tilde{\beta})S_T(\tilde{\beta}^*)$. Then, as in (12),

$$
E[(LR_{S,T} - \widetilde{LR}_{S,T})^2] = EG_T\psi(\beta)\psi(\gamma) - EG_T\psi(\beta)\tilde{\psi}(\tilde{\gamma}) - EG_T\tilde{\psi}(\tilde{\beta})\psi(\gamma) + EG_T\tilde{\psi}(\tilde{\beta})\tilde{\psi}(\tilde{\gamma}).
$$

Now Lemmas 2 and 3 and the CMT imply that $G_T\psi(\beta)\psi(\gamma)$, $G_T\psi(\beta)\tilde{\psi}(\tilde{\gamma})$, $G_T\tilde{\psi}(\tilde{\beta})\psi(\gamma)$ and $G_T\tilde{\psi}(\tilde{\beta})\tilde{\psi}(\tilde{\gamma})$ have the same asymptotic distribution, which is given by

$$
\begin{aligned}
\mathbf{1}[\sum_{i=1}^k \left[-c_i J_i(1)^2 - c_i^2 \textstyle\int J_i^2 - \frac{2c_i}{1 - e^{-2c_i}} \left[e^{-c_i} J_i(1) + c_i \int e^{-c_i s} J_i \right]^2 + \left[J_i(1) + c_i \int J_i \right]^2 \right] < K_G] \\
\cdot \mathbf{1}[\sup_s W_0(s)'\Lambda W_0(s) < K_S] \mathbf{1}[\sup_s W_1(s)'\Lambda W_1(s) < K_S] \\
\cdot \exp\left[\textstyle\int \bar{W}_0'\Lambda^{1/2} dW_\varepsilon - \frac{1}{2}\int \bar{W}_0'\Lambda\bar{W}_0 \right] \exp\left[\int \bar{W}_1'\Lambda^{1/2} dW_\varepsilon - \frac{1}{2}\int \bar{W}_1'\Lambda\bar{W}_1 \right]
\end{aligned}
$$

where W_0 and W_1 are mutually independent $k \times 1$ standard Wiener processes independent of W_ε, $c_i = 2a_i$ and J_i (which are continuous functionals of $W_\varepsilon(\cdot)$) are defined as in Lemma 2. The expectation of the asymptotic distribution of $(LR_{S,T} - \widetilde{LR}_{S,T})^2$ is hence zero, and by the uniform integrability of $(LR_{S,T} - \widetilde{LR}_{S,T})^2$, this implies that

$$
E[(LR_{S,T} - \widetilde{LR}_{S,T})^2] \to 0
$$

so that $LR_{S,T} - \widetilde{LR}_{S,T} \xrightarrow{p} 0$.

Now in order to show $\sup_T E[(\widetilde{LR}_{S,T})^4] < \infty$, note that by Jensen's inequality

$$
\begin{aligned}
E[(\widetilde{LR}_{S,T})^4] &= E[(G_T E_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta})S_T(\tilde{\beta}^*))^4] \\
&\leq E[G_T E_{\tilde{\beta}}\tilde{\xi}(\tilde{\beta})^4] < K'
\end{aligned}
$$

uniformly in T from the result above, and for $\sup_T E[(LR_{S,T})^4]$, again by Jensen's inequality,

$$\begin{aligned} E[(LR_{S,T})^4] &= E[(G_T E_\beta \xi(\beta) S_T(\beta^*))^4] \\ &\leq EG_T \xi(\beta)^4 S_T(\beta^*) \\ &= EG_T \exp\left[4\sigma^{-2}\varepsilon'\Xi\beta - 2\sigma^{-2}\beta'[M_e \otimes I_k]\Xi'\Xi[M_e \otimes I_k]\beta\right] S_T(\beta^*) \\ &\leq EG_T \exp\left[4\varepsilon^{*\prime}\Xi^*\beta^*\right] S_T(\beta^*). \end{aligned}$$

From a repeated application of the Law of Iterated Expectations and the conditional Gaussianity of $\{\varepsilon_t\}$ of Condition 2

$$\begin{aligned} EG_T \exp[4\varepsilon^{*\prime}\Xi^*\beta^*]S_T(\beta^*) &\leq E\mathbf{1}[\mathrm{tr}\,\hat{\Sigma}_{X^*} < k+1] \exp[4\varepsilon^{*\prime}\Xi^*\beta^*]S_T(\beta^*) \\ &= E[\mathbf{1}[\mathrm{tr}\,\hat{\Sigma}_{X^*} < k+1] \exp[8\sum_{t=1}^{T}(\beta_t^{*\prime}X_t^*)^2]S_T(\beta^*)] \\ &\leq E[\mathbf{1}[\mathrm{tr}\,\hat{\Sigma}_{X^*} < k+1] \exp[8K_S T^{-1}\sum_{t=1}^{T} X_t^{*\prime}X_t^*]] \\ &\leq \exp[8K_S(k+1)] \end{aligned}$$

uniformly in T.

For the convergence $\widetilde{LR}_T - \overline{LR}_T \xrightarrow{p} 0$, let $\tilde{v}_i = [I_T \otimes \iota_{k,i}' P^{*\prime}\hat{\sigma}^{-1}\hat{\Sigma}_X^{-1/2}]\Xi' M\varepsilon$. Note that $\hat{\sigma}^2$ and $\hat{\Sigma}_X$ are consistent by standard arguments. Therefore,

$$(T^{-1/2}\sum_{t=1}^{[sT]} v_{i,t}, T^{-1/2}\sum_{t=1}^{[sT]} \tilde{v}_{i,t}) \Rightarrow (W_{\varepsilon,i}(s) - sW_{\varepsilon,i}(1), W_{\varepsilon,i}(s) - sW_{\varepsilon,i}(1))$$

and proceeding as in the proof of Lemma 2, the CMT yields

$$v_i'[G_{a_i} - M_e]v_i - \tilde{v}_i'[G_{a_i} - M_e]\tilde{v}_i \Rightarrow 0.$$

But weak convergence to a constant is equivalent to convergence in probability, and the result follows by Slutsky's Theorem.

Proof of Theorem 3:

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 $h = \varepsilon$ and (ii) $E[\widetilde{LR}] = 1$.

For (i), first note that by Theorem 2, $LR_T - \widetilde{LR}_T \xrightarrow{p} 0$ under the null hypothesis. But convergence in probability implies convergence in distribution, and 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 Lemma 2, with $\widetilde{LR} = \prod_{i=1}^{k} \widetilde{LR}^i$ where

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

and J_i is defined in Lemma 2.

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

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

so that it is sufficient to show that $E[\widetilde{LR}^i] = 1$. 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 4:

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

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

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. Proceeding as in the proof of Lemma 2 now yields the result.

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.

Chib, S. (1998): "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 75, 221–241.

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.

Cohen, A., and D. Kushary (1994): "Adaptive and Unbiased Predictors in a Change Point Regression Model," *Statistical and Probability Letters*, 20, 131–138.

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

Eliasz, P., J. Stock, and M. Watson (2004): "Optimal Tests for Reduced Rank Time Variation in Regression Coefficients and for Level Variation in the Multivariate Local Level Model," *mimeo, Princeton University*.

Elliott, G., and U. Müller (2004): "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *UCSD Working Paper 2004-10*.

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.

Ferguson, T. (1967): *Mathematical Statistics — A Decision Theoretic Approach*. Academic Press, New York and London.

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.

Giordani, P., R. Kohn, and D. V. Dijk (2005): "A Unified Approach to Nonlinearity, Structural Change and Outliers," *Econometric Insitute Report EI 2005-09*.

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.

Harvey, A. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

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

Koop, G., and S. Potter (2004): “Forecasting and Estimating Multiple Change-Point Models with an Unknown Number of Change-Points,” *Department of Economics at University of Leicester Working Paper 04/31.*

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.

Oliner, S., G. Rudebusch, and D. Sichel (1996): "The Lucas Critique Revisited — Assesssing the Stability of Empirical Euler Equations for Investment," *Journal of Econometrics*, 70, 291–316.

Pesaran, M., D. Pettenuzzo, and A. Timmermann (2004): "Forecasting Time Series Subject to Multiple Structural Breaks," *mimeo*.

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," mimeo, Yale University.

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.

Saikkonen, P., and R. Luukonen (1993): "Point Optimal Tests for Testing the Order of Differencing in ARIMA Models," *Econometric Theory*, 9, 343–362.

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.

——— (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.

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

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

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. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–830.

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